

## Geographically and temporally weighted neural network for winter wheat yield prediction



Luwei Feng<sup>a,b,1</sup>, Yumiao Wang<sup>a,1</sup>, Zhou Zhang<sup>b,\*</sup>, Qingyun Du<sup>a,2</sup>

<sup>a</sup> School of Resources and Environmental Science, Wuhan University, Wuhan, 430079, China

<sup>b</sup> Department of Biological Systems Engineering, University of Wisconsin-Madison, Madison, WI 53706, USA

### ARTICLE INFO

Edited by: Marie Weiss

**Keywords:**

Winter wheat  
Yield prediction  
Remote sensing  
Spatiotemporal non-stationarity  
Geographically and temporally weighted neural network

### ABSTRACT

Accurate prediction of crop yield is essential for agricultural trading, market risk management and food security. Although various statistical models and machine learning models have been developed to enhance prediction accuracy, spatial and temporal non-stationarity, an intrinsic attribute of many geographical processes, is still rarely considered in crop yield modeling. From a statistical point of view, this study respectively provided evidence for the existence of spatial non-stationarity and temporal non-stationarity in winter wheat yield prediction based on geographically weighted regression (GWR) and temporally weighted regression (TWR). Then, a geographically and temporally weighted neural network (GTWNN) model was proposed by integrating artificial neural network (ANN) into geographically and temporally weighted regression (GTWR) using publicly available data sources, including satellite imagery and climate data. For a more credible evaluation, the leave-one-year-out strategy was adopted to make out-of-sample prediction resulting in a total of 12 test years from 2008 to 2019. The experiment results showed that the proposed GTWNN outperformed ANN, GTWR and support vector regression (SVR) achieving the average coefficient of determination ( $R^2$ ) values of 0.766, 0.759 and 0.720 at the three prediction times of end of July, end of June and end of May. Moreover, an extended Moran's I was adopted to assess the degree of spatiotemporal autocorrelation of the prediction errors. The error aggregation of GTWNN was lower than other models, indicating that GTWNN is applicable to addressing spatial non-stationarity in modeling the relationship between predictors and yield response. The methodology proposed in this paper can be extended to handle spatiotemporal non-stationarity in other crop yield predictions and even other environmental phenomena.

### 1. Introduction

Wheat is one of the earliest cultivated crop in the history of agriculture (Bell, 1987), and it continues to be vital in providing calories and nutrients for human and livestock today (Shewry and Hey, 2015). In addition to diet and health, wheat is even associated with general economic and social conditions (Curtis and Halford, 2014). As a major wheat producer and explorer, the United States (US) accounted for 7% of the world's wheat production and 14% of the world's wheat exports in 2019 (USDA, 2020a). According to the growing season, wheat is primarily classified into winter wheat and spring wheat. Winter wheat, planted in fall and harvested in the next summer, represents around 70% of total wheat US production with 35 million metric tons of harvest from

24 million acres in 2019 (USDA, 2020b). Timely, accurate and spatially specific winter wheat yield estimation therefore is important for food security monitoring and marketing planning on the local, national, and international levels (Cai et al., 2019).

The availability of growth conditions during the growing season has made it possible for researchers to perform accurate yield estimates. The low cost and wide spatial coverage of satellite imagery has also made up for the lack of ground observations, so that the scale of yield estimation has been expanded from field to regional, national and even global scales. Several studies have combined satellite imagery and statistical approaches to perform crop yield estimation (Balaghi et al., 2008; Bolton and Friedl, 2013; Kern et al., 2018; Ma et al., 2021). For example, Maselli and Rembold (2001) built linear regression models with

\* Corresponding author.

E-mail addresses: [lwfeng@whu.edu.cn](mailto:lwfeng@whu.edu.cn) (L. Feng), [wymfrank@whu.edu.cn](mailto:wymfrank@whu.edu.cn) (Y. Wang), [zhang347@wisc.edu](mailto:zhang347@wisc.edu) (Z. Zhang), [qydu@whu.edu.cn](mailto:qydu@whu.edu.cn) (Q. Du).

<sup>1</sup> These two authors contributed equally to this work.

<sup>2</sup> Corresponding author.

monthly normalized difference vegetation index (NDVI) derived from the Advanced Very High Resolution Radiometer (AVHRR) images, to forecast cereal crop yield in four North African Countries. Becker-Reshef et al. (2010) developed a winter wheat yield prediction approach with time series NDVI derived from the Moderate resolution Imaging Spectroradiometer (MODIS) data, and the regression model achieved satisfactory performance with an error less than 10%. Similarly, NDVI derived from MODIS was also adopted to predict winter wheat yield in Shandong, China using a linear model (Ren et al., 2008). Despite moderate success, the models established in these studies are global regression models, which means that the relationship between yield and explanatory variables is assumed to be spatially constant. However, in fact the relationship can be highly variable over space especially in large areas, because it is always affected by natural environment such as topography and soil type, and human factors such as farming habits and management system.

The phenomenon that an explanatory variable has different effects on model output over space is known as spatial non-stationarity or spatial heterogeneity (Stewart Fotheringham et al., 1996). Non-stationarity has been recognized by many researchers and has been taken into account in different areas such as environmental analysis and epidemiological spreading (Brunton et al., 2017; Lin and Wen, 2011; Song et al., 2014; Zhai et al., 2018). Moreover, the assumption of non-stationarity has been demonstrated to be of great significance to the improvement of model accuracy in the research of agricultural yield estimation (Shen et al., 2018). Therefore, various local modeling approaches have been developed to deal with the spatial non-stationarity issue (Huang et al., 2013; Imran et al., 2013; Shiu and Chuang, 2019). A commonly used approach in large-scale yield estimation is partition modeling, which assumes that in each partition, the relationship between dependent and independent variables is spatially stationary, and does not follow a location-specific regression function. Therefore, in partition modeling, a number of global models are developed independently for the geographic partitions. For example, to achieve better modeling accuracy, Manjunath et al. (2002) developed a linear yield prediction model for each of the 15 wheat-growing regions in Rajasthan, India, by integrating spectral and meteorological data. Mishra et al. (2008) established multiple district-level linear regression models to make in season yield prediction of sorghum in Burkina Faso, a West African country. However, the separation of limited datasets brings a tradeoff between achieving generalization performance and modeling spatial non-stationarity. Another model designed to address the issue of spatial non-stationarity is geographically weighted regression (GWR), which has been widely applied in many fields, such as geography and ecology, and has been demonstrated to be effective for crop yield prediction in agricultural research. For instance, Imran et al. (2015) compared GWR with conditional autoregression, a global spatial regression, when simulating the yields of sorghum, millet and cotton in West Africa, and found that the average accuracy improvement of GWR in semiarid zone and subhumid zone was 0.1 and 0.3, respectively. Shiu and Chuang (2019) who compared GWR with least squares ordinary in paddy rice yield estimation also observed similar results. Despite success, GWR has limitations in constructing spatial weight matrix due to the heavy reliance on the predefined spatial kernel functions (e.g. Gaussian, Bi-square, Exponential functions, etc). Essentially, spatial kernel functions are predefined distance decay functions in which larger weights are assigned to closer observations based on the Tobler's First Law of Geography (Tobler, 1970), and they are the simplified and idealized expressions of spatial proximity, making it possible to quantify the spatial relationships in modeling process. However, these kernels are relatively simple in addressing complicated interactions inherent in geographical relationships (Du et al., 2020). Additionally, although the selection of optimal bandwidth requires data-driven approaches such as Akaike Information Criterion (AIC), the predefined kernels expressed in formulas may hinder the precise quantification of spatial relationships using the inherent information in the data (Hagenauer and Helbich,

2021).

In addition to space, time is also an important dimension that affects the relationship between predictors and response (Fotheringham et al., 2015). Extensive efforts have been made to incorporate temporal effects into spatial regression. Among these, geographically and temporally weighted regression (GTWR) is one of the successful attempts for modeling non-stationary spatiotemporal relationships. It was originally achieved by redefining the distances between observations in the construction of the weight matrix. For example, Huang et al. (2010) formed a spatiotemporal distance through the linear combination of spatial and temporal distance, which was then improved by adding the interaction term of time and space dimensions on the basis of the original distance formula (Wu et al., 2014). Instead of improving the expression of spatiotemporal distance, Fotheringham et al. (2015) realized the expansion of GWR to GTWR by constructing a spatiotemporal weighting function which was expressed by the product of a temporal kernel function with a unique temporal bandwidth, and a spatial kernel function with a set of segregated spatial bandwidths over time. The various structures of GTWR adapt the interpretability of GWR, that is, they can generate a set of coefficients of the independent variables corresponding to a specific spatiotemporal coordinates, and output bandwidth which can be interpreted as spatiotemporal effects. In agriculture, factors including soil fertility changes and farming technology innovations are all possible causes of temporal non-stationarity. However, although GTWR has been successfully applied in various fields (Guo et al., 2017; Huang et al., 2010), it has never been used for crop yield estimation.

In contrast to conventional statistical methods, machine learning allows the appliance of complex mathematical calculations to large dataset, and it has demonstrated its superior ability in numerous fields such as natural language processing and image recognition (Goldberg, 2016; Nadkarni et al., 2011; Svyrydov et al., 2018). Increasingly, machine learning has been used in agriculture, such as crop yield estimation. For instance, Matsumura et al. (2015) compared the performances between artificial neural network (ANN) and multiple linear regression in predicting maize yield in Jilin province, China, and the results showed that ANN significantly outperformed the linear one. Sakamoto (2020) built a random forest (RF) regression algorithm with time-series MODIS satellite data and environmental records for soybean and corn yield prediction in the US, and the approach accurately predicted the yields of corn and soybean with root mean square error (RMSE) of 0.539 t/ha and 0.206 t/ha, respectively. In addition, several researchers also have tried to establish machine learning based models considering spatiotemporal non-stationarity. For example, Crane-Droesch (2018) developed a semiparametric ANN using weather, soil, geographic coordinates and year data to predict corn yield across multiple geographically contiguous states in the US. Shook et al. (2018) built a Long Short Term Memory based framework with weather and genotype variables for soybean yield estimation in North America, and the model achieved improved performance by incorporating GPS coordinates and predicting year. Similar ideas can be found from these studies that spatial and temporal information are added as independent and parallel factors with other variables into the machine learning method. However, in fact, each location in spatiotemporal dimension is characterized by a series of physical and human properties, and it is these properties, not the location, that directly affect the growth and yield of crops. Moreover, the machine learning model performance can potentially be further enhanced by incorporating prior knowledge of the crop.

In this study, we proposed a geographically and temporally weighted neural network (GTWNN) considering spatial and temporal non-stationarity for county-level winter wheat yield estimation in the US. Before modeling, we first performed statistical tests to provide evidence for the existence of spatial and temporal non-stationarity in winter wheat yield estimation based on GWR and temporally weighted regression (TWR) respectively. A GTWNN model was then developed with the smoothed sequential features extracted from satellite images and meteorological datasets from 2008 to 2019. Furthermore, the

proposed model was compared with GTWR, ANN and support vector regression (SVR) for performance evaluation. Finally, the spatiotemporal patterns of the prediction errors derived from the four models were analyzed.

## 2. Materials

### 2.1. Study area and yield data

The modeling work was performed to estimate winter wheat yield in the US. County-level yield records over years 2008–2019 were collected from the United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) database. Overall, the national annual yield of winter wheat exhibited an upward trend, increasing from 3.167 t/ha in 2008 to 3.605 t/ha in 2019 (Fig. 1b). However, there were also obvious fluctuations in the inter-annual variation of yield. Specifically, the yield in 2016 and 2019 increased significantly, while other years remained stable or declined to varying degrees compared with the previous year. The multi-year average yield of each county was also calculated and mapped in Fig. 1a. The figure indicates the existence of spatial autocorrelation in the distribution of winter wheat yield, which means that neighboring counties tend to have similar yields. Based on the geographical location and yield distribution characteristics, we divided Contiguous United States (CONUS) into six regions including northcentral (NC), northwest (NW), northeast (NE), southcentral (SC), southwest (SW), and southeast (SE) as shown in Fig. 1c. Except for SW region, winter wheat is widely grown throughout the CONUS. Among these regions, NW has the highest yield, followed by NE. Winter wheat is heavily concentrated in the central region, but the corresponding yield is relatively low, especially in the SC region.

### 2.2. Satellite data

The vegetation indices (VIs) are designed to improve the sensitivity to vegetation characteristics while reducing the influence of confounding factors. Due to the effectiveness in quantitative evaluation of vegetation growth, VIs have been widely and successfully adopted in agricultural applications, such as crop yield estimation (Johnson, 2014; Schwalbert et al., 2020). Four VIs, namely NDVI, enhanced vegetation index (EVI), normalized difference water index (NDWI) and green chlorophyll index (GCI), were considered in this study with the formula shown in Eqs. (1)–(4) and they were extracted from the daily MODIS MCD43A3 product which includes both the visible and near-infrared (NIR) spectral ranges at 500-m spatial resolution. As the most commonly used VI, NDVI quantifies vegetation by measuring the difference between red and NIR reflectance (Son et al., 2014). EVI is a modified NDVI that enhances sensitivity to high biomass regions while minimizing soil and atmosphere influences by incorporating blue spectral wavelengths (Pan et al., 2012). NDWI is a numerical indicator of liquid water content of vegetation canopies by combining reflectance of red and shortwave infrared (SWIR) (Gu et al., 2007). GCI provides an accurate estimate of leaf chlorophyll content and therefore can be used as a measurement of physiological stages and health of crops (Kang et al., 2020).

$$NDVI = \frac{NIR - Red}{NIR + Red} \quad (1)$$

$$EVI = 2.5 \times \frac{(NIR - Red)}{NIR + 6 \times Red - 7.5 \times Blue + 1} \quad (2)$$

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

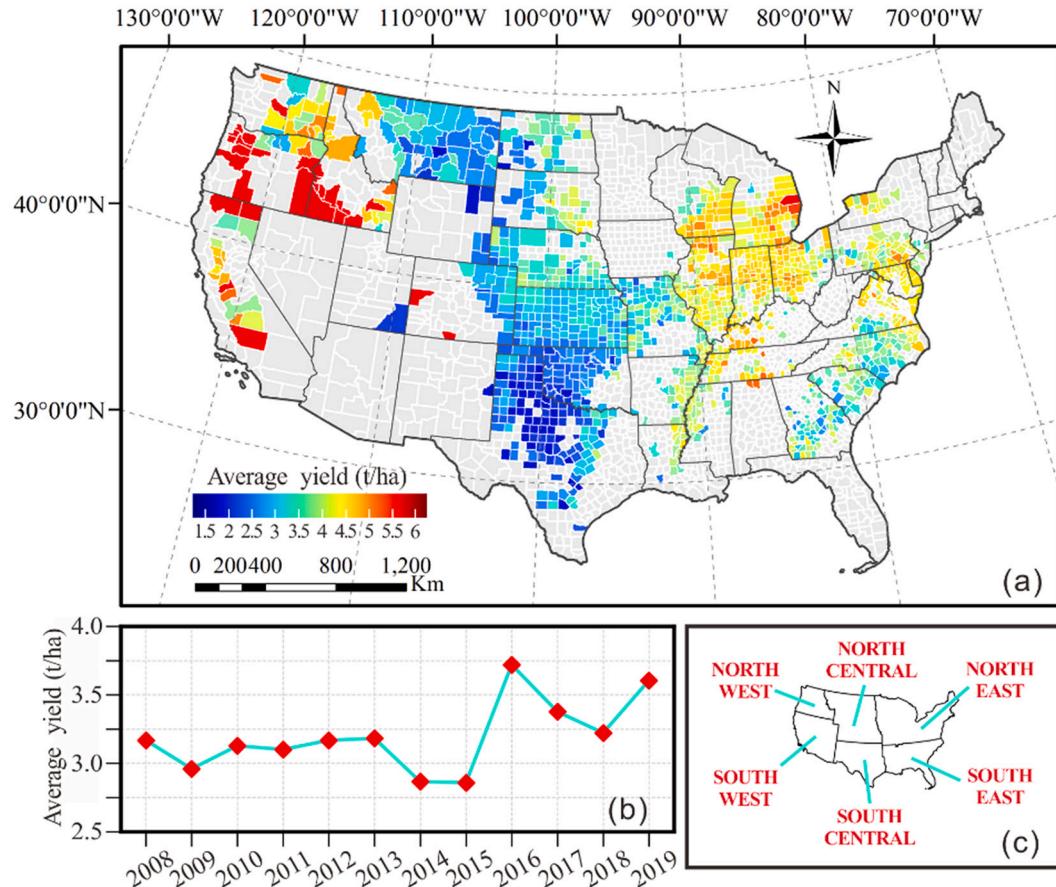


Fig. 1. Distribution and changes of winter wheat yield in the US.

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

where blue, green, red, NIR and SWIR stand for the spectral reflectance in blue, green, red, NIR and SWIR channels, respectively.

Another satellite data source is MODIS MYD11A2 product with a spatial resolution of 1 km, from which the daytime and nighttime land surface temperature (LST\_D and LST\_N) were extracted. These two variables were chosen because of their ability to capture heat stress and water stress especially during the extremely dry growing seasons (Pede et al., 2019).

### 2.3. Climate data

Since agriculture is considered as the most weather-dependent human activity, many studies have included meteorological factors for better crop yield estimation (Oettli et al., 2011; Schwalbert et al., 2020). Climate features involved in the study were extracted from Parameter-elevation Regressions on Independent Slopes Model (PRISM) Dataset with a 4 km spatial resolution. Specifically, these meteorological variables include daily mean dew point temperature (DPT), daily maximum, mean and minimum (Tmax, Tmean and Tmin), daily maximum and minimum vapor pressure deficit (VPDmax and VPDmin), and daily total precipitation (PPT).

### 2.4. Data preprocessing

These datasets from 2008 to 2019 were first aggregated spatially to the county level using the Google Earth Engine platform. The Cropland Data Layer with a 30 m resolution provided by NASS was applied as the mask, and the counties with winter wheat planting areas less than 100 ha were filtered out. The sequential variables including the VIs and climate observations were collected and aggregated to a 16-day interval from October 1st of the previous year to July 30th (or 31th) of the current year, resulting in 19 temporal observations with details shown in Table (A1). Then, Savitzky–Golay filtering algorithm was used to smooth these sequential features by fitting successive seven adjacent data points with a quadratic polynomial (Chen et al., 2004; Kim et al., 2014). Finally, a total of 247 features were obtained for model development, including four VIs and nine climate features with 19 temporal observations for each.

## 3. Methods

### 3.1. Foundations for GTWNN

#### 3.1.1. OLS regression

Ordinary least squares (OLS) regression is a best known statistical method in identifying the relationships between the dependent variable  $y_i$  and the independent variables  $x_{i1}, x_{i2}, \dots, x_{ip}$ . The OLS regression is expressed as:

$$y_i = \beta_0 + \sum_{k=1}^p \beta_k x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (5)$$

where  $\beta_0$  is the intercept term and  $\beta_1, \dots, \beta_p$  are the regressive coefficients of the corresponding independent variables;  $n$  and  $p$  represent the size of the samples and the number of independent variables respectively; and  $\varepsilon$  denotes the error term that is normally distributed with a constant zero mean and variance  $\sigma^2$ . Using matrix notation, OLS regression can be expressed as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon} \quad (6)$$

where

$$\begin{aligned} \mathbf{y} &= \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix}, \\ \mathbf{X} &= \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \\ \boldsymbol{\beta} &= \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_n \end{bmatrix}, \\ \boldsymbol{\varepsilon} &= \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix} \end{aligned}$$

and the estimation of parameters  $\boldsymbol{\beta}$  is given by Eq. (7):

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}^T \mathbf{X})^{-1} \mathbf{X}^T \mathbf{y} \quad (7)$$

#### 3.1.2. GWR and TWR

As a global model, OLS regression provides fixed relationship over space without considering the effects of geographic locations. To address the spatial non-stationarity issue, the GWR model extends the OLS regression by allowing local estimates (Fotheringham et al., 1998), and the model is presented as follows:

$$y_i = \beta_0(u_i, v_i) + \sum_{k=1}^p \beta_k(u_i, v_i)x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (8)$$

where  $(u_i, v_i)$  denotes the coordinates of sample  $i$ ;  $\beta_0(u_i, v_i)$  is the constant coefficient and  $\beta_k(u_i, v_i)$  represents the coefficient of  $k$ th independent variable at location  $(u_i, v_i)$ . Using a weighted least squares, the estimated coefficients can be expressed by Eq. (9):

$$\hat{\boldsymbol{\beta}}(u_i, v_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y} \quad (9)$$

where  $\mathbf{W}(u_i, v_i)$  is a  $n \times n$  space weight matrix, whose diagonal elements  $w_{i1}, w_{i2}, \dots, w_{in}$  are geographical weights of  $(u_i, v_i)$ , and the off-diagonal elements are zero. The predictive value  $\hat{y}_i$  of sample  $i$  can be represented by Eq. (10):

$$\hat{y}_i = \mathbf{x}_i^T \hat{\boldsymbol{\beta}}(u_i, v_i) = \mathbf{x}_i^T (\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{y} \quad (10)$$

where  $\mathbf{x}_i$  is the vector of the  $i$ th row of matrix  $\mathbf{X}$ . Therefore, using matrix notation, the predictive values can be expressed by Eq. (11):

$$\hat{\mathbf{y}} = \begin{bmatrix} \mathbf{x}_1^T [\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \\ \mathbf{x}_2^T [\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \\ \vdots \\ \mathbf{x}_n^T [\mathbf{X}^T \mathbf{W}(u_i, v_i) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i) \end{bmatrix} \mathbf{y} = \mathbf{Ly} \quad (11)$$

where  $\mathbf{L}$  is called the hat matrix through which the observed  $\mathbf{y}$  is transformed into the predicted  $\hat{\mathbf{y}}$ .

The calculation of geographical weights relies on a certain weight kernel. Gaussian, Bi-square, and Exponential functions are commonly used kernels which are defined in Eqs. (12–14).

$$w_{ij}^{\text{Gaussian}} = \exp \left[ - \left( \frac{d_{ij}}{h^{\text{Gaussian}}} \right)^2 \right] \quad (12)$$

$$w_{ij}^{\text{Bi-square}} = \begin{cases} \left[ 1 - \left( \frac{d_{ij}}{h^{\text{Bi-square}}} \right)^2 \right]^2 & d_{ij} \leq h^{\text{Bi-square}} \\ 0 & d_{ij} > h^{\text{Bi-square}} \end{cases} \quad (13)$$

$$w_{ij}^{\text{Exponential}} = \exp \left[ -\frac{d_{ij}}{h^{\text{Exponential}}} \right] \quad (14)$$

where  $d_{ij}$  measures the spatial distance between sample  $i$  and  $j$ ;  $h^{\text{Gaussian}}$  and  $h^{\text{Bi-square}}$  stand for the bandwidths that produce a distance decay effect;  $h^{\text{Bi-square}}$  denotes the bandwidth that determines the distance threshold, within which the kernel function will be performed.

To improve goodness-of-fit, the corrected AIC ( $\text{AIC}_c$ ) defined in Eq. (15) is first used to select the appropriate value of the bandwidth for each kernel, and then the kernel is determined according to the residual sum of squares (RSS) of the model.

$$\text{AIC}_c = n \log_e(\hat{\sigma}^2) + n \log_e(2\pi) + n \left( \frac{n + \text{tr}(\mathbf{S})}{n - 2 - \text{tr}(\mathbf{S})} \right) \quad (15)$$

where  $\hat{\sigma}^2$  is the estimated variance of the error term; and  $\text{tr}(\cdot)$  is a trace operator.

TWR extends OLS regression by allowing local estimates over time, and the model can be expressed by Eq. (16).

$$y_i = \beta_0(t_i) + \sum_{k=1}^p \beta_k(t_i)x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (16)$$

where  $(t_i)$  denotes the time coordinate of sample  $i$ . And the estimated coefficients can be represented by Eq. (17).

$$\hat{\beta}(t_i) = (\mathbf{X}^T \mathbf{W}(t_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(t_i) \mathbf{y} \quad (17)$$

where  $\mathbf{W}(t_i)$  is a  $n \times n$  time weight matrix, whose diagonal elements are calculated based on a kernel function. The kernels used in TWR are similar to those used in GWR, except for the modification from spatial distance to time distance. Additionally, the strategy of determining the kernel and corresponding bandwidth in GWR is applicable in TWR (Huang et al., 2010).

### 3.1.3. GTWR

GTWR was developed to simultaneously consider spatial and temporal non-stationarity in parameter estimation (Huang et al., 2010). The regression equation of GTWR is expressed as:

$$y_i = \beta_0(u_i, v_i, t_i) + \sum_{k=1}^p \beta_k(u_i, v_i, t_i)x_{ik} + \varepsilon_i, \quad i = 1, 2, \dots, n \quad (18)$$

where  $(u_i, v_i, t_i)$  represents the space-time coordinates of sample  $i$ ; and  $\beta(u_i, v_i, t_i)$  is the regression coefficient for sample  $i$ . The matrix expression for the estimated coefficients is presented as:

$$\hat{\beta}(u_i, v_i, t_i) = (\mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{X})^{-1} \mathbf{X}^T \mathbf{W}(u_i, v_i, t_i) \mathbf{y} \quad (19)$$

where  $\mathbf{W}(u_i, v_i, t_i)$  is the space-time weight matrix at position  $(u_i, v_i, t_i)$ , with spatiotemporal weights  $w_{ij}$  on its main diagonal and zeros on off-diagonal parts. To calculate  $w_{ij}$ , a kernel function is adopted as described in 3.1.2. Here,  $d_{ij}^{ST}$  is the spatiotemporal distance between sample  $i$  and sample  $j$ , and it can be calculated using Eq. (20).

$$d_{ij}^{ST} = \sqrt{\left[ (u_i - u_j)^2 + (v_i - v_j)^2 \right] + \lambda (t_i - t_j)^2} \quad (20)$$

where  $\lambda$  is a scale factor employed to balance the weight of temporal and spatial distance.

### 3.2. Design of GTWNN

Although GTWR has demonstrated its superiority to address spatio-temporal non-stationarity, it still has limitations in modeling complex geographical processes and environmental phenomena. First, GTWR estimates the relationship between response and explanatory variables based on local linear assumptions, but in fact, the relationship is usually more complicated. Second, GTWR relies on kernel functions to generate the space-time weight matrix, but the simple and predefined kernel structure is difficult to describe complex nonlinear interactions. Third, the composition of spatiotemporal distance probably ignores the multi-scale difference between space and time. ANN uses the processing of the brain as a basis to develop algorithms, and it consists of a large number of highly interconnected processing and computing artificial neurons that can work together to solve problems. Activation functions are crucial components in ANN which convert the linear input signals of the neurons into non-linear output signals and thus facilitate the learning of more complex relationships of the real world (Sharma, 2017). Considering the advantages of ANN in effectively solving nonlinear problems, we proposed a GTWNN model by integrating ANN into GTWR to better modeling the nonlinear relationships while considering the spatiotemporal effects. Extending from GTWR, GTWNN can be expressed as:

$$y_i = \beta_0(u_i, v_i, t_i) \otimes \beta_1(u_i, v_i, t_i)x_{i1} \otimes \dots \otimes \beta_p(u_i, v_i, t_i)x_{ip} \otimes \varepsilon_i, \quad i = 1, 2, \dots, n \quad (21)$$

where  $\beta_k(u_i, v_i, t_i)$  denotes the influence of space-time position  $(u_i, v_i, t_i)$  on the  $k$ th independent variable; and  $\otimes$  represents an operation that replaces addition operation in GTWR.

The architecture of the proposed GTWNN is presented in Fig. 2. GTWNN consists of two fully connected neural networks, through which the prediction is made in consideration of spatiotemporal non-stationarity. Specifically, the proposed GTWNN starts with a weight estimation neural network ANN1 (blue dashed line), which aims to learn spatiotemporal weight  $\beta_{ik}$  (same as  $\beta_k(u_i, v_i, t_i)$  in Eq. (21)) of each independent variable  $x_{ik}$  from the input spatial and temporal coordinates (longitude  $x_i^{Lon}$ , latitude  $x_i^{Lat}$  and year  $x_i^T$  in this study) of the estimated point  $x_i$ . ANN1 starts with an input layer followed by two hidden layers, and the spatiotemporal weight  $\beta_{ik}$  derived from ANN1 is multiplied by the corresponding independent variable  $x_{ik}$  to form the input of ANN2. Then, two hidden layers in ANN2 (red dashed line) perform nonlinear transformations of the inputs and obtain the output as the predictive value  $\hat{y}_i$ .

Through experimental analysis, Rectified Linear Unit (ReLU) was selected as the activation function. The entire network was trained using the Adam optimization algorithm with a learning rate of 0.0001, and the parameters were updated using the backward propagation of errors calculated by the mean squared error, a commonly used loss function, after each epoch.

### 3.3. Experimental setup

#### 3.3.1. Strategy for feature selection

Regression models, such as OLS and GWR, are susceptible to multicollinearity which may induce parameter estimation instability, counterintuitive parameter signs, and large parameter standard errors (Wheeler and Tiefelsdorf, 2005). To ensure the reliability of the model itself and the subsequent non-stationarity test, we used variance inflation factor (VIF) as the basis for feature selection. As a commonly used method to uncover potential multicollinearity, VIF qualifies the extent of correlation between one and the other predictors using Eq. (22).

$$\text{VIF}_i = \frac{1}{1 - R_i^2} \quad (22)$$

where  $R_i^2$  is the  $R^2$  statistic obtained from the regression, with the  $i$ th

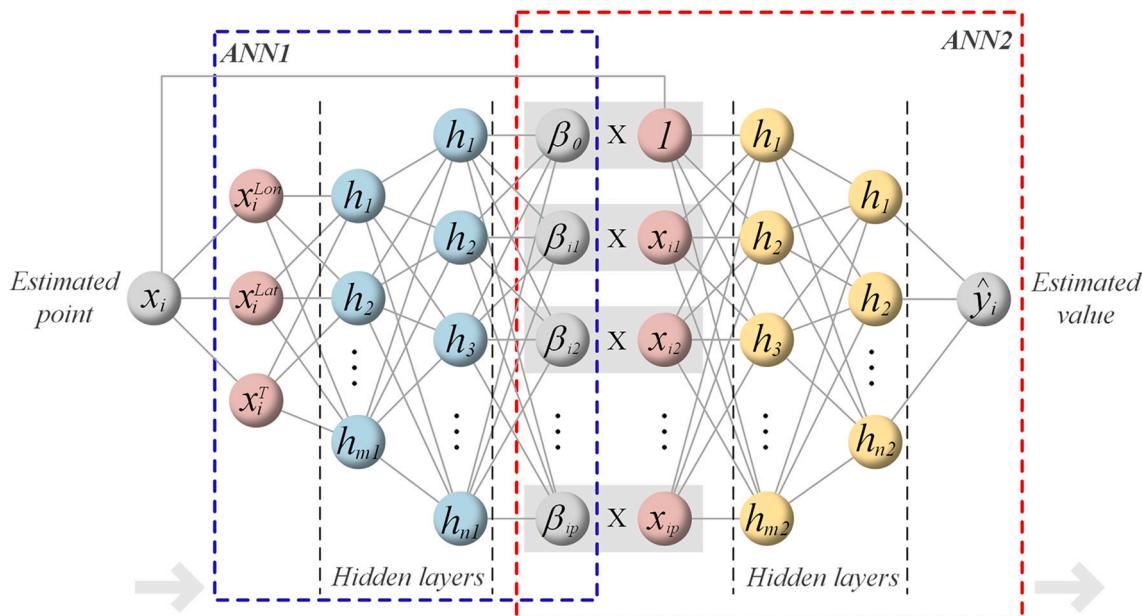


Fig. 2. The architecture of GTWNN.

independent variable as a response variable and the other independent variables as predictors.

All the data from 2008 to 2019 were considered, and 13 sequential variables (NDVI, EVI, NDWI, GCI, LST\_D, LST\_N, Tmax, Tmean, Tmin, DPT, PPT, VPDmax and VPDmin) with 19 temporal observations were considered in feature selection. To conduct feature selection, we first averaged the values of each sequential feature during the growing season and used the mean value for feature selection to eliminate the influence of the seasonal cycle. Then, the VIF values of the initial features were determined and the feature with the highest VIF was removed. The procedure was recursively repeated until the VIF values of all remaining features were less than 10 (Niazian et al., 2018; O'Brien, 2007).

### 3.3.2. Statistical test for spatial and temporal non-stationarity

Over the years, non-stationarity of space and time has been widely recognized by researchers, and several approaches have been proposed by incorporating spatial variation into the modeling (Brunton et al., 2017; Lin and Wen, 2011; Zhai et al., 2018). In agriculture, methods considering spatial non-stationarity such as GWR have been adopted for crop yield prediction, and have achieved better performance than global models (Imran et al., 2015; Shiu and Chuang, 2019). However, few studies statistically discussed whether spatial and temporal non-stationarity exists in the establishment of the relationship between crop yield and explanatory factors. Therefore, we tried to answer the question in this study based on GWR and TWR models from which not only the variation of the parameters but also the significance of the variation can be explored. From a statistical perspective, performing the statistical tests of the following two sub-questions is the key to answering the aforementioned question.

- (1) Does the data description ability of the GWR/TWR model is significantly improved than that of OLS. That is, overall, do the parameters in GWR/TWR exhibit significant variation over the study area?
- (2) Does each set of parameters  $\beta_{ik}$  ( $i = 1, 2, \dots, n$ ) varies significantly over the study area?

To answer the first sub-question, the null hypothesis  $H_0$  and alternative hypothesis  $H_1$  are proposed as:

$H_0$ : no significant difference exists between GWR/TWR and OLS;

$H_1$ : GWR/TWR is significantly better than OLS.

With RSS, a test statistic  $F_1$  is designed as Eq. (23) and its distribution can be approximated by an F-distribution with  $v_1^2/v_2$  and  $(n - p - 1)$  degrees of freedom.

$$F_1 = \frac{(RSS_{OLS} - RSS_{GWR/TWR})/v_1}{RSS_{OLS}/(n - p - 1)} \quad (23)$$

where  $RSS_{OLS}$  and  $RSS_{GWR/TWR}$  represent the RSS of OLS regression and GWR/TWR model;  $n$  and  $p$  are the sample size and the number of independent variables respectively; and  $v_1$  and  $v_2$  can be calculated using the following formulas:

$$v_1 = n - p - 1 - \delta_1 \quad (24)$$

$$v_2 = n - p - 1 - 2\delta_1 + \delta_2 \quad (25)$$

where  $\delta_i = \text{tr}[(\mathbf{I} - \mathbf{L})^T(\mathbf{I} - \mathbf{L})]^i$ ,  $i = 1, 2$ .  $\mathbf{I}$  is an identity matrix of order  $n$ , and  $\mathbf{L}$  is the hat matrix expressed in Eq. (26).

$$\mathbf{L} = \begin{bmatrix} \mathbf{x}_1^T [\mathbf{X}^T \mathbf{W}(1) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(1) \\ \mathbf{x}_2^T [\mathbf{X}^T \mathbf{W}(2) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(2) \\ \vdots \\ \mathbf{x}_n^T [\mathbf{X}^T \mathbf{W}(n) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(n) \end{bmatrix} \quad (26)$$

where  $\mathbf{X}$  is defined as in Eq. (6);  $\mathbf{x}_i^T$  is the  $i$ th row of  $\mathbf{X}$ ; and  $\mathbf{W}(i)$  is the weighting matrix at location  $i$ .

The  $p$ -value  $p_{F1}$  can be generated based on the calculated  $F_1$  value and two degrees of freedom. Given a significance level  $\alpha$  (0.05 in this study), we will reject the null hypothesis and receive the alternative hypothesis if  $p_{F1} < \alpha$ , otherwise, we will receive the null hypothesis.

To answer the second sub-question, the null hypothesis  $H_0$  and alternative hypothesis  $H_1$  are established for each set of parameters  $\beta_{ik}$  ( $i = 1, 2, \dots, n$ ) of  $x_k$ , the value of the  $k$ th independent variable.

$H_0 : \beta_{1k} = \beta_{2k} = \dots = \beta_{nk}$  for a given  $k$ ,

$H_1 : \text{not all } \beta_{ik} (i = 1, 2, \dots, n) \text{ are equal.}$

An appropriate statistic  $F_2(k)$  is then constructed (Eq. (27)) and the distribution can be approximated by an F-distribution with  $v_1^2/v_2$  and  $\delta_1^2/$

$\delta_2$  degrees of freedom.

$$F_2(k) = \frac{V_k^2 / \gamma_1}{\hat{\sigma}^2} \quad (27)$$

where  $V_k^2 = \frac{1}{n} \sum_{i=1}^n \left( \hat{\beta}_{ik} - \frac{1}{n} \sum_{i=1}^n \hat{\beta}_{ik} \right)^2$ ;  $\hat{\sigma}^2 = \frac{RSS_{GWR/TWR}}{\delta_1}$ ;  $\delta_1$  and  $\delta_2$  are defined in Eq. (25); and  $\gamma_1$  and  $\gamma_2$  can be calculated using the following formula:

$$\gamma_i = \text{tr} \left[ \frac{1}{n} \mathbf{B}^T \left( \mathbf{I} - \frac{1}{n} \mathbf{J} \right) \mathbf{B} \right]^i, i = 1, 2 \quad (28)$$

And

$$\mathbf{B} = \begin{bmatrix} \mathbf{e}_k^T [\mathbf{X}^T \mathbf{W}(1) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(1) \\ \mathbf{e}_k^T [\mathbf{X}^T \mathbf{W}(2) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(2) \\ \vdots \\ \mathbf{e}_k^T [\mathbf{X}^T \mathbf{W}(n) \mathbf{X}]^{-1} \mathbf{X}^T \mathbf{W}(n) \end{bmatrix} \quad (29)$$

where  $\mathbf{X}$  and  $\mathbf{W}(i)$  are defined in Eq. (26);  $\mathbf{J}$  is an  $n \times n$  matrix of ones; and  $\mathbf{e}_k$  is a column vector whose  $(k+1)$ th element is one and all other elements are zeros.

Similarly, the p-value  $p_{F_2(k)}$  can be generated in the significance test for the spatial and temporal non-stationarity of univariate regression coefficient based on the calculated  $F_2(k)$  value and two degrees of freedom. Given a significance level  $\alpha$  (0.05 in this study), we will reject the null hypothesis and receive the alternative hypothesis if  $p_{F_2(k)} < \alpha$ , otherwise, we will receive the null hypothesis. More details about the construction theory and process of the two test statistics  $F_1$  and  $F_2(k)$  can be found in (Leung et al., 2000).

### 3.3.3. Model evaluation

To validate the proposed GTWNN model, we implemented three models for comparison: GTWR, ANN and SVR. The choice of GTWR and ANN was to verify whether the GTWNN model outperformed its basic models, and the choice of SVR was due to its wide application and successful performance in crop yield estimation (Cai et al., 2019; Han et al., 2020; Joshi et al., 2020). In addition, leave-one-year-out strategy was applied to all the models. Specifically, data from each year were used for testing and data from the rest of the years were divided into training set and validation set according to the ratio of 7:3, resulting in a total of 12 experiments (Fig. 3). Three metrics including coefficient of determination ( $R^2$ ), RMSE, and mean absolute percentage error (MAPE) were selected to evaluate the models and they were calculated using Eqs. (30)–(32) respectively.

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (30)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (31)$$

$$\text{MAPE} = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (32)$$

where  $n$  denotes the sample size,  $y_i$  and  $\hat{y}_i$  are the observed and the predicted yield of sample  $i$ , and  $\bar{y}$  represents the mean value of the observed yield. A paired sample  $t$ -test was applied to test whether there was a significant accuracy improvement of the proposed model, and the significance level was set to 0.05 in this study.

To further evaluate the model's capability of addressing

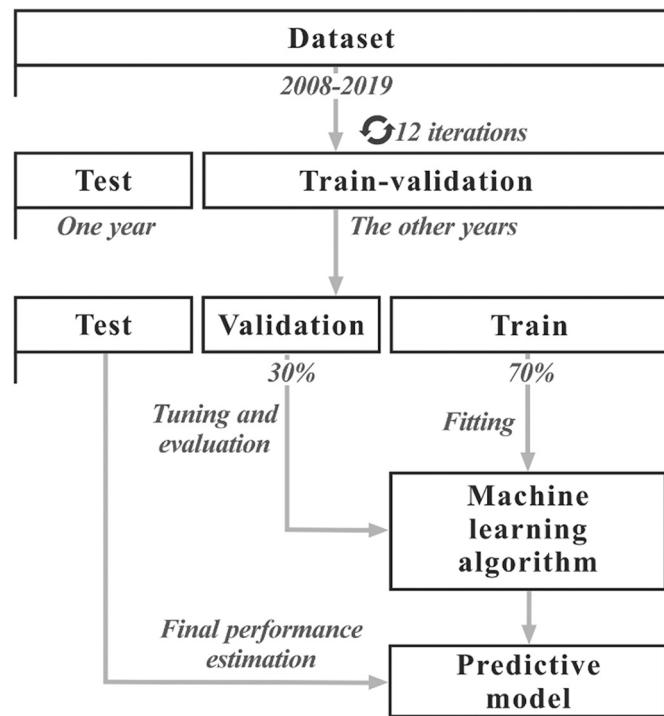


Fig. 3. The experimental workflow of model development.

spatiotemporal non-stationarity, the error distribution pattern from each model was analyzed using an extended Moran's I (Gao et al., 2019), which is a spatiotemporal autocorrelation measurement modified from the classical Moran's I (Moran, 1950). A positive value represents a clustered pattern, a negative value represents a dispersed pattern, and zero value represents a random pattern. In this study, if the value of Moran's I is closer to 0, it means that the errors are less affected by the spatial location, that is, the corresponding model can better consider the influence of spatial variation when simulating the relationship.

## 4. Results

### 4.1. Important feature selection

The input features were determined based on the strategy described in Section 3.3.1, and the correlation between each selected feature and yield was also calculated using Pearson Correlation Coefficient (PCC).

**Table 1**  
Descriptions of the selected features.

Feature	Description	Unit	Source	VIF	Correlation with yield
EVI	Enhanced vegetation index	–	MCD43A3	7.910	0.215
NDWI	Normalized difference water index	–	MCD43A3	3.344	0.513
GCI	Green chlorophyll index	–	MCD43A3	8.978	0.173
LST_D	Daytime land surface temperature	Kelvin	MYD11A2	5.976	-0.553
DPT	Daily mean dew point temperature	°C	PRISM	5.313	-0.030
VPDmin	Daily minimum vapor pressure deficit	hPa	PRISM	2.192	-0.397
PPT	Daily total precipitation	mm	PRISM	3.576	0.253

**Table 1** provides the descriptions of the selected features. Three VIs including EVI, NDVI and GCI were selected in this study, and higher values indicate healthier growing status and lower levels of environmental stress. Therefore, all the three VIs were found to be positively correlated with the yield. On the contrary, a negative correlation was observed between temperature and yield, which means that even considering the benefits of reduced cold exposure, the overall impact of warming on yield is negative (Tack et al., 2015). Beyond temperature, precipitation is also an important meteorological factor affecting crop yield, and it has a weak positive correlation with the yield. As an index related to plant transpiration rate, VPD showed a negative correlation with yield. This is mainly because that low VPD reduces the driving force for water transport, thereby decreasing water loss rate and moderating plant water stress (Yuan et al., 2019).

#### 4.2. Spatial and temporal non-stationarity in yield prediction

To test the existence of spatial and temporal non-stationarity in yield modeling, we performed statistical tests as described in Section 3.3.2 based on GWR and TWR. The mean value of each sequential variable during the growing period was calculated and all the mean values were combined as independent variables. Additionally, all data over years 2008–2019 were used for GWR/TWR modeling to describe the relationship between yield and predictors, thereby improving the reliability of the statistical test results of spatial and temporal non-stationarity. The non-stationarity tests included (1) significance test for model improvement compared with OLS using  $F_1$ , and (2) significance test for spatial/temporal variation of each set of parameters using  $F_2$ . The performance indicators of OLS, GWR and TWR models are listed in Table 2, from which we could notice that both GWR and TWR considerably outperformed OLS. The improvement of GWR to OLS was much more obvious than that of TWR, suggesting that spatial non-stationarity was more essential than temporal non-stationarity in yield prediction. The results of the  $F_1$  test also demonstrated that the relationship between yield and explanatory variables varied significantly over space and time.

To further explore the existence of significant variation in each set of parameters, the  $F_2$  test was then performed on the coefficients of each parameter of GWR and TWR, with the results shown in Table 3. It was notable that all the  $p$ -values were less than the given significance level 0.05, indicating that each parameter was characterized with significant spatial and temporal non-stationarity.

#### 4.3. Model performance and comparison

The developed model GTWNN was compared to three other state-of-the-art approaches, including GTWR, ANN and SVR. The comparison was conducted at end of May, end of June and end of July when the winter wheat was roughly in heading and flowering, ripening, and harvesting stage respectively (USDA, 2020c). Leave-one-year-out strategy was adopted and the evaluation metrics for each of the past 12 years are shown in Table 4. It was obvious that three machine learning models outperformed GTWR, demonstrating their capability in modeling potential complex relationship between multi-source data and the yield response. Among the machine learning models, the proposed GTWNN had the most satisfactory performance with the highest average  $R^2$  and lowest average RMSE of multiple test years. It worth noting that GTWNN performed well in 2012 and 2019, two abnormal years, characterized by anomalous drought and wet respectively (Climate Central, 2020;

**Table 2**  
Results of the  $F_1$  significance test for model improvement.

Model	$R^2$	RMSE (t/ha)	MAPE (%)	RSS	$F_1$	$P$ -value
OLS	0.479	0.937	23.779	6285.391	–	–
GWR	0.876	0.458	12.083	1498.254	2.872	0.000
TWR	0.559	0.862	21.430	5320.281	12.474	0.000

**Table 3**

Results of the  $F_2$  significance test for spatial/temporal variation of parameters.

Parameter	GWR		TWR	
	$F_2$	$P$ -value	$F_2$	$P$ -value
$\beta_{EVI}$	4.891	0.000	4.633	0.000
$\beta_{NDWI}$	3.588	0.000	5.315	0.000
$\beta_{GCI}$	5.601	0.000	12.388	0.000
$\beta_{ST_D}$	3.303	0.000	15.096	0.000
$\beta_{DPT}$	4.213	0.000	4.886	0.000
$\beta_{VPDmin}$	3.345	0.000	8.357	0.000
$\beta_{PPT}$	3.610	0.000	3.906	0.000

Johnson, 2014). ANN was inferior to GTWNN in terms of model stability and accuracy, which illustrated the importance of including reasonable prior knowledge even for models like ANN that have ability to handle complex data and detect hidden patterns. A paired sample  $t$ -test was conducted between GTWNN and each of the three compared models to further evaluate whether GTWNN was statistically better than the other models on the reported  $R^2$ . Each  $p$ -value of the results shown in Table 5 is less than 0.05, indicating that the accuracy improvement of GTWNN was statistically significant.

Winter wheat is widely grown throughout the CONUS, and the growth stages of winter wheat in different regions usually occur at different times (USDA, 2020c). Generally, the three prediction times are in the heading and flowering, ripening, and harvesting stages. The emergence of wheat head marks the beginning of the heading and flowering stage, during which pollination and fertilization occur (Sciencing, 2017). The ripening stage follows fertilization, and this stage witnesses the growth and maturity of the kernels (Knott, 2016). After that, winter wheat can be harvested. Comparing the three prediction times, the accuracy for almost all models increased along with the growth of crops as more information became available. Also, except for GTWR, all other models performed best at the end of July, indicating that although many regions have entered the harvesting stage in late July, the features in this period generally brought more valuable information than noise. Therefore, we adopted the predictions at the end of July for the following analysis.

The scatter plots shown in Fig. 4 present the agreement between the observed and predicted yields for the four models in each test year. Overall, the data points in each scatter plot follow a positive and linear pattern, but the strength of the pattern varies among the four models. The best agreement was again observed in GTWNN for most of the test years and similar agreement was observed in ANN except for a few more outliers. In contrast, the data points in GTWR were more scattered around the diagonal line compared with those in other models, further indicating that GTWR model had limitations in capturing the complex relationships between explanatory variables and yield. It was notable that due to the limited high-yield and low-yield samples in training dataset, the yield of high-yield counties was underestimated, while the yield of low-yield counties was overestimated. Besides, the underestimation and overestimation phenomenon tended to be more pronounced in SVR, which demonstrated that SVR was more susceptible to imbalanced data distribution.

#### 4.4. The spatial pattern of predicted yield

We explored the spatial patterns of the predicted yield by mapping the results in Fig. 5. Overall, the spatial pattern of the predicted yield was consistent with that of the observed yield. Specifically, NW had the highest yield, followed by NE, while SC was the region with the lowest yield, and then NC. Despite fluctuations, the change in yield showed an upward trend from 2008 to 2019, and 2016 was obviously the year with the highest yield among the years. Comparing the predicted yield from the four models, some similarities were easily observed. For example, the models tended to overestimate the yield in SC, a low-yield region,

**Table 4**

Test accuracies of GTWNN, GTWR, ANN, and SVR in winter wheat yield prediction.

Time	Year	GTwNN			GTWR			ANN			SVR		
		R <sup>2</sup>	RMSE	MAPE									
End of July	2008	<b>0.744</b>	<b>0.606</b>	<b>15.356</b>	0.629	0.730	20.411	0.723	0.631	16.140	0.717	0.637	16.379
	2009	<b>0.705</b>	<b>0.655</b>	<b>19.417</b>	0.396	0.937	29.997	0.699	0.662	19.739	0.636	0.727	23.943
	2010	<b>0.724</b>	<b>0.547</b>	<b>14.360</b>	0.563	0.688	18.282	0.682	0.587	15.713	0.686	0.583	15.335
	2011	<b>0.806</b>	<b>0.599</b>	<b>15.970</b>	0.746	0.685	20.454	0.786	0.630	16.808	0.692	0.755	25.719
	2012	<b>0.770</b>	<b>0.574</b>	<b>12.656</b>	0.483	0.860	19.830	0.731	0.621	14.108	0.700	0.655	15.924
	2013	<b>0.801</b>	<b>0.600</b>	<b>16.598</b>	0.727	0.702	20.850	0.717	0.715	23.589	0.764	0.653	18.528
	2014	<b>0.761</b>	<b>0.672</b>	<b>19.815</b>	0.683	0.774	20.094	0.747	0.691	23.074	0.750	0.688	21.217
	2015	<b>0.761</b>	<b>0.639</b>	<b>17.881</b>	0.708	0.706	18.418	0.737	0.670	19.248	0.742	0.663	19.130
	2016	<b>0.773</b>	<b>0.679</b>	<b>15.901</b>	0.704	0.775	17.604	0.749	0.715	16.203	0.734	0.736	16.879
	2017	<b>0.791</b>	<b>0.634</b>	<b>16.231</b>	0.766	0.670	16.722	0.776	0.656	<b>15.016</b>	0.753	0.688	17.696
	2018	<b>0.846</b>	<b>0.548</b>	<b>12.081</b>	0.751	0.695	15.790	0.802	0.621	13.601	0.822	0.588	13.604
	2019	<b>0.707</b>	<b>0.687</b>	<b>15.956</b>	0.643	0.758	16.704	0.684	0.713	16.156	0.672	0.726	19.197
	Ave	<b>0.766</b>	<b>0.620</b>	<b>16.018</b>	0.650	0.748	19.596	0.736	0.659	17.450	0.722	0.675	18.629
	Std	0.040	0.047	2.233	0.111	0.077	3.531	<b>0.037</b>	<b>0.041</b>	3.137	0.047	0.053	3.394
End of June	2008	<b>0.719</b>	<b>0.635</b>	<b>16.482</b>	0.636	0.722	19.713	0.678	0.680	17.496	0.712	0.643	16.695
	2009	<b>0.679</b>	<b>0.683</b>	<b>20.744</b>	0.479	0.870	26.982	0.659	0.704	21.598	0.667	0.695	22.544
	2010	<b>0.716</b>	<b>0.555</b>	<b>14.931</b>	0.547	0.700	18.595	0.665	0.603	15.982	0.660	0.607	15.745
	2011	<b>0.794</b>	<b>0.618</b>	<b>18.481</b>	0.746	0.685	20.728	0.743	0.689	18.953	0.724	0.714	22.426
	2012	<b>0.758</b>	<b>0.589</b>	<b>13.135</b>	0.553	0.800	18.318	0.720	0.633	13.921	0.689	0.667	16.128
	2013	<b>0.799</b>	<b>0.603</b>	<b>16.924</b>	0.737	0.689	20.746	0.756	0.665	19.504	0.757	0.663	18.060
	2014	<b>0.741</b>	<b>0.699</b>	<b>21.361</b>	0.688	0.768	20.048	0.727	0.719	22.842	0.726	0.719	21.633
	2015	<b>0.774</b>	<b>0.620</b>	<b>16.794</b>	0.697	0.720	18.443	0.721	0.690	19.554	0.721	0.690	20.061
	2016	<b>0.772</b>	<b>0.680</b>	<b>16.403</b>	0.725	0.748	16.868	0.731	0.739	16.911	0.745	0.720	16.994
	2017	<b>0.796</b>	<b>0.626</b>	<b>15.820</b>	0.776	0.655	15.717	0.730	0.720	16.022	0.746	0.698	17.868
	2018	<b>0.842</b>	<b>0.555</b>	<b>12.176</b>	0.762	0.681	15.759	0.794	0.633	12.939	0.825	0.583	13.518
	2019	<b>0.715</b>	<b>0.677</b>	<b>16.441</b>	0.632	0.770	16.881	0.585	0.817	19.146	0.677	0.720	18.293
	Ave	<b>0.759</b>	<b>0.628</b>	<b>16.641</b>	0.665	0.734	19.066	0.709	0.691	17.906	0.721	0.677	18.330
	Std	<b>0.044</b>	0.047	<b>2.558</b>	0.092	0.058	2.923	0.052	0.054	2.811	0.044	<b>0.044</b>	2.706
End of May	2008	<b>0.689</b>	<b>0.668</b>	18.003	0.472	0.871	24.951	0.603	0.755	20.015	0.672	0.687	<b>17.301</b>
	2009	0.620	0.743	22.960	0.513	0.841	27.050	0.630	0.733	<b>22.142</b>	<b>0.644</b>	<b>0.719</b>	22.912
	2010	<b>0.658</b>	<b>0.609</b>	<b>16.510</b>	0.398	0.807	22.316	0.591	0.666	17.881	0.607	0.653	17.441
	2011	<b>0.767</b>	<b>0.656</b>	<b>19.683</b>	0.610	0.849	28.919	0.732	0.704	20.620	0.700	0.746	21.499
	2012	<b>0.743</b>	<b>0.606</b>	<b>13.492</b>	0.607	0.750	17.339	0.681	0.676	14.949	0.662	0.695	16.470
	2013	<b>0.754</b>	<b>0.667</b>	<b>17.716</b>	0.707	0.729	23.909	0.703	0.733	22.230	0.740	0.686	18.881
	2014	<b>0.734</b>	<b>0.710</b>	<b>20.875</b>	0.635	0.831	21.160	0.671	0.789	21.593	0.674	0.785	21.082
	2015	<b>0.740</b>	<b>0.666</b>	<b>17.843</b>	0.670	0.750	18.672	0.687	0.731	19.608	0.685	0.733	21.002
	2016	<b>0.755</b>	<b>0.705</b>	<b>16.777</b>	0.678	0.809	18.058	0.703	0.777	17.686	0.711	0.766	17.991
	2017	<b>0.748</b>	<b>0.696</b>	<b>17.233</b>	0.669	0.797	17.053	0.674	0.792	18.202	0.693	0.768	19.562
	2018	0.784	0.648	14.184	<b>0.789</b>	<b>0.641</b>	<b>13.987</b>	0.732	0.722	14.563	0.788	0.642	14.247
	2019	<b>0.650</b>	<b>0.750</b>	<b>18.376</b>	0.556	0.845	20.388	0.493	0.903	22.889	0.617	0.785	19.442
	Ave	<b>0.720</b>	<b>0.677</b>	<b>17.804</b>	0.609	0.793	21.150	0.658	0.748	19.365	0.683	0.722	18.986
	Std	0.050	<b>0.044</b>	2.495	0.104	0.063	4.257	0.066	0.060	2.650	<b>0.048</b>	0.048	<b>2.341</b>

The units of RMSE and MAPE are t/ha and %, respectively.

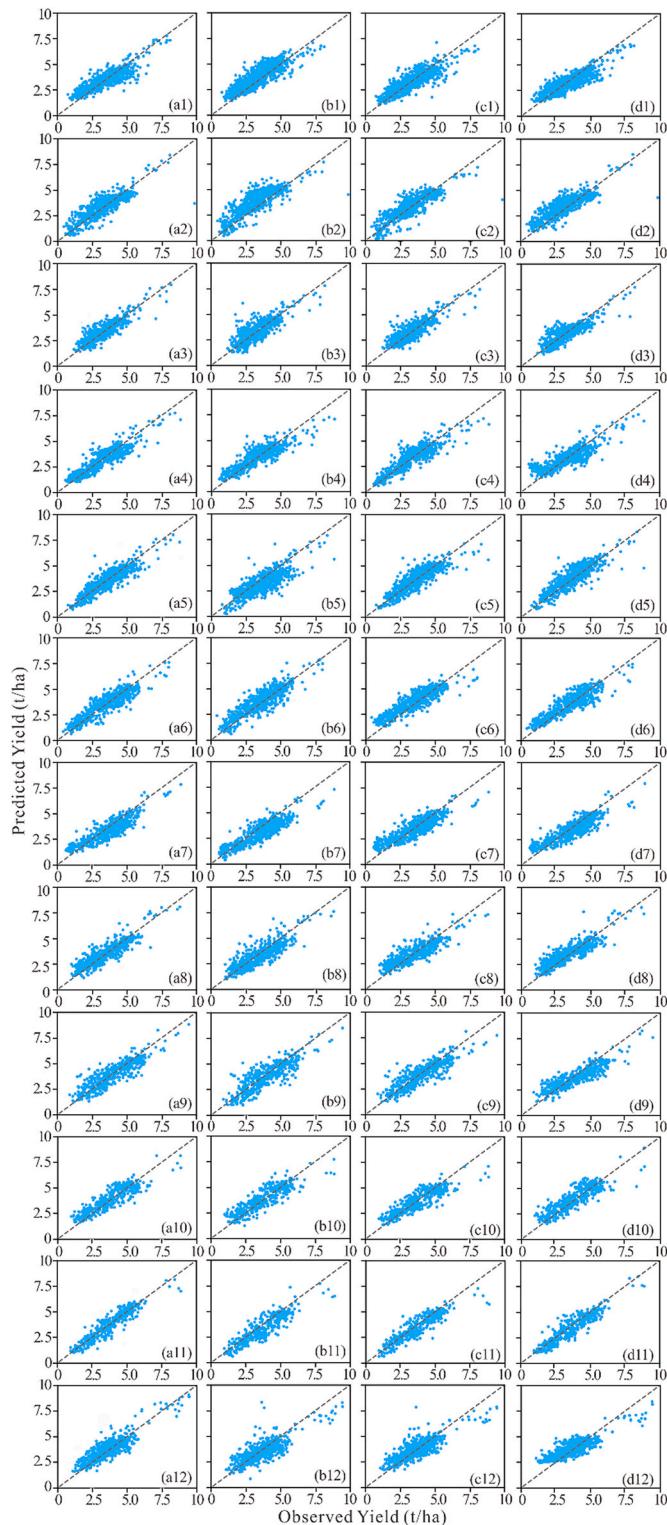
**Table 5**Results from the paired sample t-test between R<sup>2</sup> of GTWNN and that of other models including GTWR, ANN and SVR.

Prediction time	Model	T	P-value
End of July	GTwNN vs. GTWR	4.365	0.001
	GTwNN vs. ANN	5.003	0.000
	GTwNN vs. SVR	5.263	0.000
End of June	GTwNN vs. GTWR	5.239	0.000
	GTwNN vs. ANN	5.982	0.000
End of May	GTwNN vs. SVR	5.932	0.000
	GTwNN vs. GTWR	5.326	0.000
	GTwNN vs. ANN	5.607	0.000
	GTwNN vs. SVR	4.174	0.000

and underestimate the yield in NE, a high-yield region, which was in agreement with the results in section 4.3. Among the models, the proposed GTWNN performed best, and its predictions were closest to the reality. In contrast, GTWR showed the worst predictive ability especially in the transition area between high and low yield. This was not only due to the failure to simulate complex relationship based on local linear assumptions, but also due to its excessive reliance on space and time proximity for yield estimation. For example, the region enclosed by the red circle in 2012 and 2013 was obviously overestimated by GTWR (Fig. 5 (c5), (c6)), and the corresponding predicted yield tended to be

close to the observed yield of the surrounding area in time and space dimensions. Again, the phenomenon of overestimation and underestimation was more obvious in SVR especially for the area circled by green lines (Fig. 5 (e4), (e5) and (e12)), indicating that the uneven data distribution brought more challenges to SVR compared with other models.

We also presented the error maps in Fig. 6 to further illustrate the model adaptability over the study area. Comparing the four models, GTWNN outperformed other models with smaller errors obtained in most test years, while GTWR performed worst especially in 2009. In the ideal case where the spatial non-stationarity is fully expressed, errors should be randomly distributed in the study area. However, the errors of the four models seemed to be spatially clustered to varying degrees in this study. To compare the degree of spatiotemporal aggregation, an autocorrelation analysis described in 3.3.3 was conducted on the errors of each model, with the results shown in Table 6. The Moran's I values were positive and the p-values were lower than 0.05, indicating that the prediction errors of all models in all test years were spatiotemporally clustered to a certain extent. The degree of error aggregation of GTWNN was obviously weaker than that of other models, demonstrating that GTWNN was better at capturing spatiotemporal non-stationarity.



**Fig. 4.** Scatter plots of observed yield vs. predicted yield of (a) GTWNN, (b) GTWR, (c) ANN and (d) SVR in 12 test years from (1) 2008 to (12) 2019.

## 5. Discussion

### 5.1. Predictive factor analysis

To perform accurate winter wheat yield estimation, we selected seven time-series variables based on the strategy described in Section 3.3.1, including three VIs characterizing crop properties and four

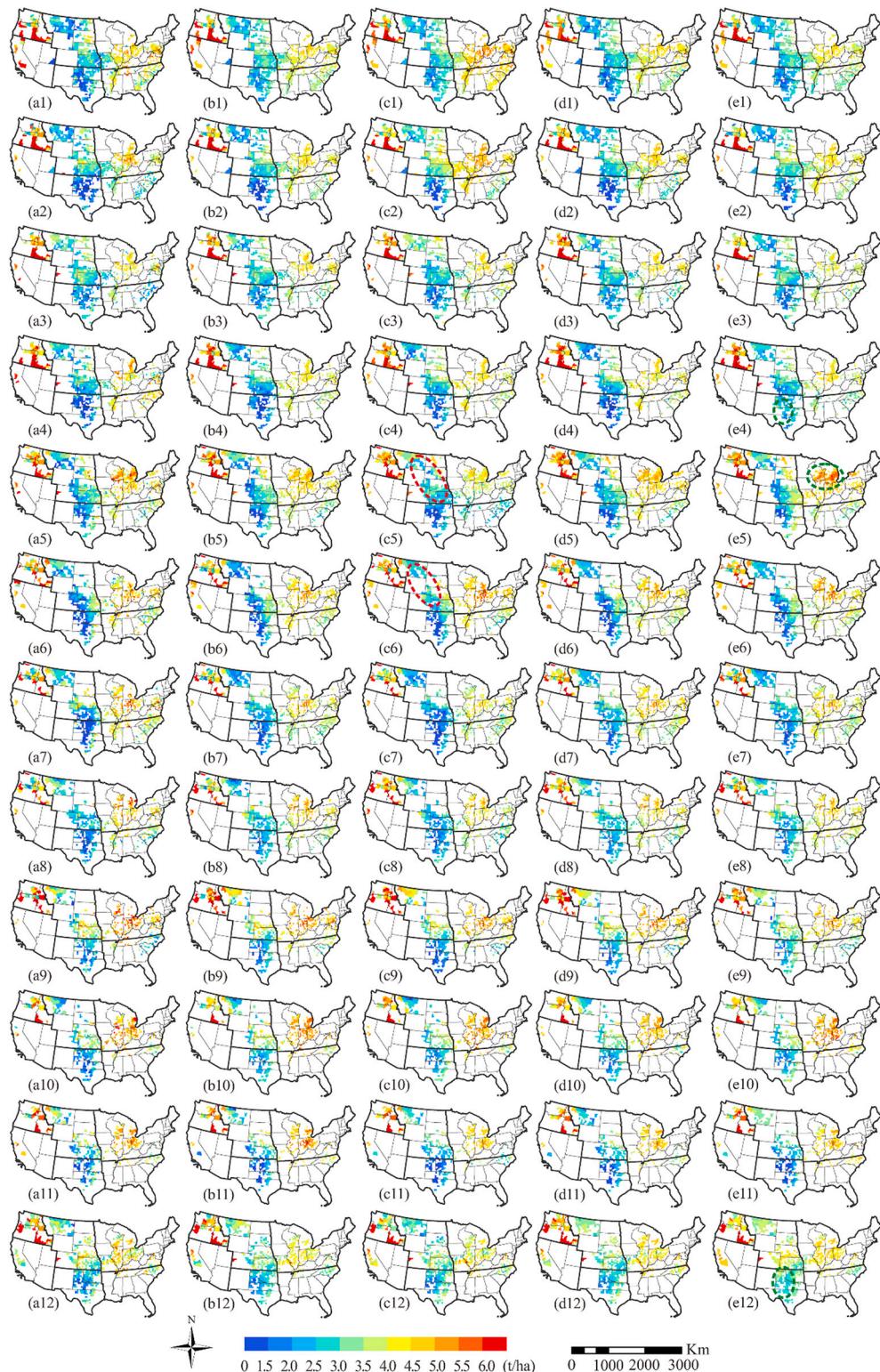
climate factors simulating the growth environment. In this section, the relationship between each variable and yield was further analyzed, and the difference of the variables between regions was also compared.

VIs are widely used to quantify the vegetation properties and indicate yield potential (Kern et al., 2018; Vicente-Serrano et al., 2016), and the effectiveness of EVI, NDWI and GCI in crop yield estimation have been demonstrated in various studies (Chandel et al., 2019; Kang et al., 2020; Kuwata and Shibasaki, 2016). Generally, winter wheat germinates and begins to develop tillers before going dormant in freezing winter (Herbek and Lee, 2009). During this period, NDWI is usually at a low level due to the effects of bare soil background. However, fields in the northern US are usually covered with snow during the long winter, which results in high NDWI values in NW, NC and NE (You, 2020). After surviving the winter months, winter wheat is greening up and becomes active in early spring when NDWI begins to increase. NDWI reaches its peak around the heading stage and decreases with the water content loss during maturity and leaf senescence (Yi et al., 2007). Similar trend was also observed in EVI and GCI. By comparing the VI profiles (Fig. 7) of north and south, we found that the growth process of winter wheat in the south tends to end earlier than in the north. Additionally, as suitable cover crops help retain water and organic matter, reduce soil erosion and suppress pests, the planting of cover crops is becoming a popular option to instead long fallow period after winter wheat harvest (Büchi et al., 2018; Ghimire et al., 2019). The phenomenon also provides explanation for the abnormal decrease of EVI, NDWI and GCI before November and increase after June.

Temperature plays a critical role in the development of winter wheat, and many stages of crop growth require warm weather (Herbek and Lee, 2009). However, the negative correlation between LST and yield indicates that heat stress is a main driver of yield loss in CONUS even considering the benefits of reduced cold exposure (Tack et al., 2015). Excessive accumulated temperature in fall often leads to overgrown seedlings that lack cold resistance (Holman et al., 2011). Also, the increased temperature shortens the grain filling stage, resulting in a decrease in dry matter accumulation (Xiao et al., 2010). These may be the reason why the yield is lower in the south than in the north. Adequate precipitation is also important for the growth of winter wheat by improving stand establishment and increasing kernels, and regions with more precipitation tend to have higher yields (Holman et al., 2011). However, irrigation can make up for the lack of rainfall. For example, despite low precipitation, NW has the highest yield among the regions, partly because of its highest irrigation coverage (Vocke and Ali, 2013). VPD is increasingly recognized as an important driver of atmospheric water demand for vegetation, and increased VPD is found to decline canopy photosynthetic rates and contribute to plant water stress (Basso and Ritchie, 2018; Yuan et al., 2019). Similar conclusion was also observed in this study, namely that higher yield is associated with lower VPD. DPT is the temperature below which the water vapor will condense, causing dew to form (Baghban et al., 2016). Under certain temperature and precipitation conditions, dew on the leaves will trigger plant diseases such as leaf spot (Ben-Asher et al., 2010; Sentelhas et al., 2008).

### 5.2. Advantage analysis of GTWNN

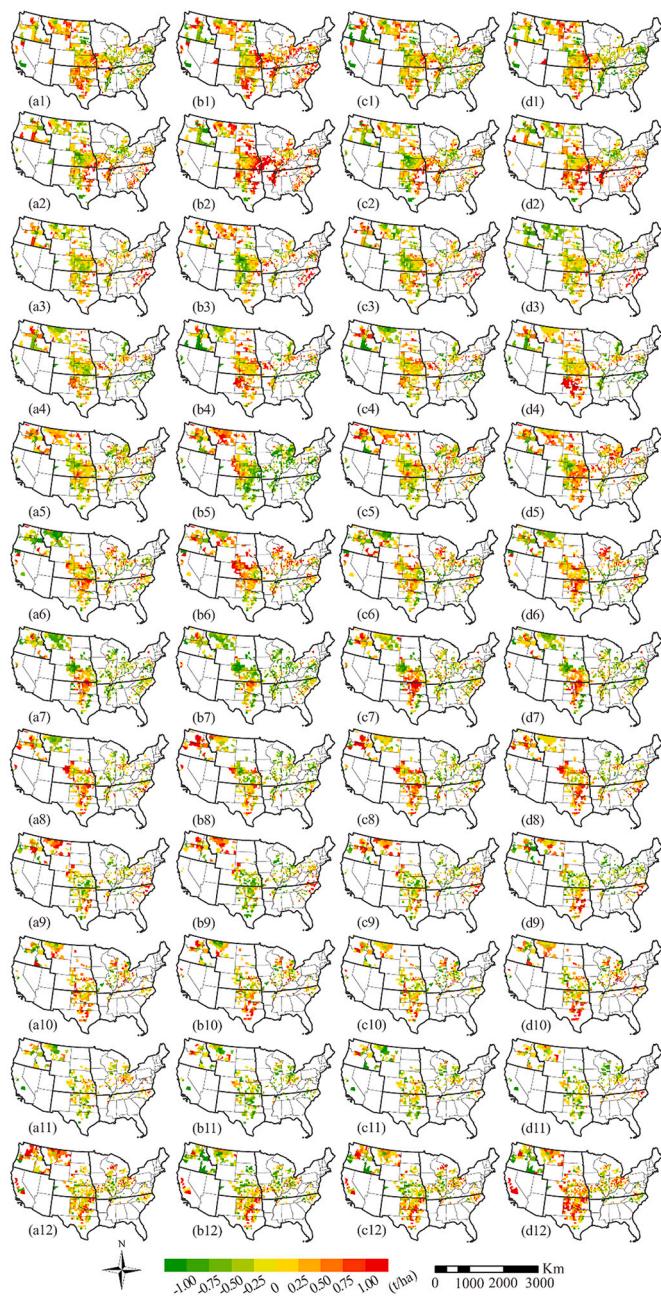
To the best of our knowledge, no previous studies have statistically provided evidence for the existence of non-stationarity in yield modeling, but many researchers have considered spatial non-stationarity in yield prediction to improve model performance (Manjunath et al., 2002; Shiu and Chuang, 2019; Wang et al., 2020). Three kinds of strategies have been often adopted in previous studies. Partition modeling is the simplest strategy, which can be achieved by dividing the study area into small partitions and then establishing multiple independent global models (Manjunath et al., 2002; Mishra et al., 2008). Its success is based on the assumption that a universal relationship exists within each divided region. However, the assumption of spatial



**Fig. 5.** Spatial patterns of the (a) observed yield and predicted yield of (b) GTWNN, (c) GTWR, (d) ANN and (e) SVR in 12 test years from (1) 2008 to (12) 2019.

stationarity is generally unrealistic even in a small area. The separation of limited datasets also brings challenges to the establishment of models with good generalization ability. Another approach is GWR, a well-known statistical method designed to identify spatial variations in relationships. It deals with spatial non-stationarity by allowing the local estimates of the coefficients in the model (Brunsdon et al., 1998). The construction of spatial weight matrix is a key step in the GWR solution

process, and this step depends on the choice of the right spatial kernel function. However, simple and predefined kernel functions limit the expression of complex spatial proximity relation (Du et al., 2020). Also, as an ensemble of local linear models, GWR is inferior to nonlinear models in addressing complicated interactions (Li et al., 2020). The third strategy is more commonly used in yield estimation studies using machine learning models. It includes geographic coordinates as parallel



**Fig. 6.** Spatial patterns of the prediction errors of (a) GTWNN, (b) GTWR, (c) ANN and (d) SVR in 12 test years from (1) 2008 to (12) 2019.

**Table 6**  
Moran's I values of the prediction errors of GTWNN, GTWR, ANN and SVR.

Model	Moran's I	P-value
GTwNN	0.191	0.000
GTWR	0.214	0.000
ANN	0.352	0.000
SVR	0.329	0.000

factors, and relies on the autonomous learning ability of machine learning models to take into account the effect of spatial location (Crane-Droesch, 2018; Shook et al., 2018). This approach has been proved to improve the predictive ability of machine learning models, but a more reasonable structure also needs to be explored. The effect of temporal non-stationarity also exists in crop yield estimation, but few studies

consider it in modeling process.

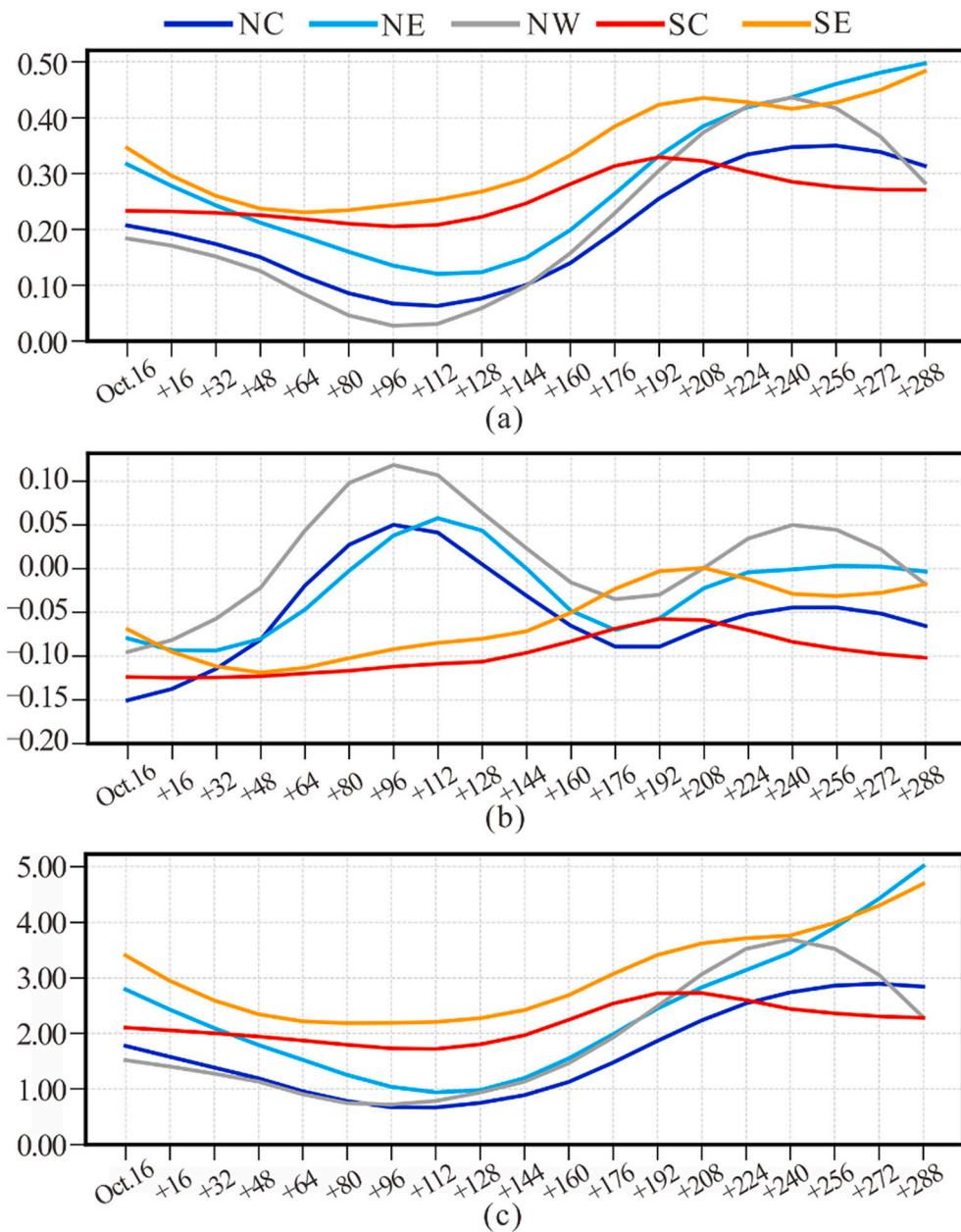
To consider the spatiotemporal non-stationarity in yield prediction, we developed a machine learning approach named GTWNN. It is designed based on the idea of GTWR, that is, allowing the coefficients of the independent variables to change with space and time position, rather than being fixed globally. It also incorporates two ANNs to generate spatiotemporal weights of different independent variables and to make nonlinear and accurate predictions of target values, respectively. The structure of GTWNN makes it superior to GTWR and ordinary ANN in modeling non-stationarity. Compared with GTWR, GTWNN overcomes the dependence on the preset kernel function, avoid the balance of scale effects of space and time when constructing spatiotemporal distance, and replaces the weighted least squares by backpropagation to solve the parameters. Compared with the ANN that incorporates space position and time position as parallel factors, GTWNN modifies the combination method of space-time position factor and other factors based on the fact that spatiotemporal information does not directly affect yield but through other factors attached to spatiotemporal position, such as temperature and precipitation.

To evaluate the effectiveness of GTWNN, it was compared with ANN, GTWR and SVR in winter wheat yield estimation. Geographic coordinates and year were added as parallel factors in ANN and SVR to simulate the solutions to non-stationarity in other studies (Crane-Droesch, 2018; Kang, 2018). Leave-one-year-out strategy was adopted to make out-of-sample predictions, and the four models were evaluated at three prediction times including end of July, end of June and end of May. The results showed that GTWNN had the best prediction performance among the models. The spatiotemporal clustering analysis of errors was further conducted based on a modified Moran's I which extended the classic Moran's I by incorporating temporal autocorrelation. It was found that the proposed GTWNN can better capture the spatiotemporal non-stationarity in yield modeling than other approaches.

### 5.3. Uncertainties and future work

The proposed GTWNN showed satisfactory accuracy in the prediction of winter wheat yield at the county level in the US. In addition, by analyzing the error distribution patterns, it was found that GTWNN could better address the spatiotemporal non-stationarity in winter wheat prediction. However, even the prediction error of GTWNN still had a certain degree of spatiotemporal aggregation. Considering the actual situation of yield modeling, there are two reasons that might lead to the emergence of aggregation of errors. First, there are a large number of factors related to crop yield in actual agricultural production, of which limited ones were included in the modeling of this study. Some unselected factors such as soil type and fertilizer rate are inherent in the specific locations, and the distribution of these factors conforms to the first law of geography, that is, near things are more related. Lack of these factors may lead to the aggregation of errors. Another possible reason is that environmental stress such as drought and heat may make yield response abnormal, and crops in adjacent areas are more likely to experience similar stresses.

To further improve the model, more features such as soil properties can be added to provide additional information. Also, a sensitivity analysis can be performed to determine the appropriate interval to better balance the preservation of vegetation growth details and increase in data dimensions. Moreover, the structure of the model can be further optimized by incorporating advanced architectures, such as LSTM, which is good at capturing the time dependencies in sequential data. To further verify the applicability of the GTWNN model, more experiments should be conducted for various crops at different research scales.



**Fig. 7.** The variation of (a) EVI, (b) NDWI, and (c) GCI during the growing season of winter wheat in different regions.

## 6. Conclusions

Accurate prediction of winter wheat yield is important for marketing, transportation and storage decisions, and helps manage the risks associated with them. In this study, we proposed a GTWNN model for county-level winter wheat yield prediction model. The developed model could address the spatiotemporal non-stationarity in identifying the relationship between predictors and the yield. Specifically, several sequential predictors were extracted from satellite imagery and climate data, and then smoothed using the Savitzky–Golay filtering algorithm. To evaluate the proposed GTWNN model, it was compared to a well-developed local regression model GTWR and two other widely used machine learning models, including ANN and SVR. Through the leave-one-year-out experiments, it was found that GTWNN performed best among the models, with the average  $R^2$  values of 0.766, 0.759 and 0.720 at the three prediction times of end of July, end of June and end of May. Moreover, the error pattern analysis further showed that GTWNN can effectively address spatiotemporal non-stationarity and has stronger

spatiotemporal adaptability.

## Funding

This work was supported by the USDA National Institute of Food and Agriculture, United States Department of Agriculture, Hatch project WIS03026; the University of Wisconsin–Madison, Office of the Vice Chancellor for Research and Graduate Education with funding from the Wisconsin Alumni Research Foundation; the China Scholarship Council (NO.201906270096); and the Wuhan University Graduates International Exchange Program (NO.201905).

## 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.

The authors declare the following financial interests/personal

relationships which may be considered as potential competing interests:

## Appendix A. Appendix

Table A1. The start dates and end dates of the 19 time intervals.

Number of the interval	Start date		End date		Number of the interval	Start date		End date	
	Normal year	Leap year	Normal year	Leap year		Normal year	Leap year	Normal year	Leap year
1	Oct.1	Oct.1	Oct.16	Oct.16	11	Mar.10	Mar.9	Mar.25	Mar.24
2	Oct.17	Oct.17	Nov.1	Nov.1	12	Mar.26	Mar.25	Apr.10	Apr.9
3	Nov.2	Nov.2	Nov.17	Nov.17	13	Apr.11	Apr.10	Apr.26	Apr.25
4	Nov.18	Nov.18	Dec.3	Dec.3	14	Apr.27	Apr.26	May.12	May.11
5	Dec.4	Dec.4	Dec.19	Dec.19	15	May.13	May.12	May.28	May.27
6	Dec.20	Dec.20	Jan.4	Jan.4	16	May.29	May.28	Jun.13	Jun.12
7	Jan.5	Jan.5	Jan.20	Jan.20	17	Jun.14	Jun.13	Jun.29	Jun.28
8	Jan.21	Jan.21	Feb.5	Feb.5	18	Jun.30	Jun.29	Jul.15	Jul.14
9	Feb.6	Feb.6	Feb.21	Feb.21	19	Jul.16	Jul.15	Jul.31	Jul.30
10	Feb.22	Feb.22	Mar.9	Mar.8					

## References

- Baghban, A., Bahadori, M., Rozyn, J., Lee, M., Abbas, A., Bahadori, A., Rahimali, A., 2016. Estimation of air dew point temperature using computational intelligence schemes. *Appl. Therm. Eng.* 93, 1043–1052.
- Balaghi, R., Tychon, B., Eerens, H., Jlibene, M., 2008. Empirical regression models using NDVI, rainfall and temperature data for the early prediction of wheat grain yields in Morocco. *Int. J. Appl. Earth Obs. Geoinf.* 10, 438–452.
- Basso, B., Ritchie, J.T., 2018. Evapotranspiration in high-yielding maize and under increased vapor pressure deficit in the US Midwest. *Agric. Environ. Lett.* 3, 1–6.
- Becker-Reshef, I., Vermote, E., Lindeman, M., Justice, C., 2010. A generalized regression-based model for forecasting winter wheat yields in Kansas and Ukraine using MODIS data. *Remote Sens. Environ.* 114, 1312–1323.
- Bell, G.D.H., 1987. The History of Wheat Cultivation, Wheat Breeding. Springer, pp. 31–49.
- Ben-Asher, J., Alpert, P., Ben-Zvi, A., 2010. Dew is a major factor affecting vegetation water use efficiency rather than a source of water in the eastern Mediterranean area. *Water Resour. Res.* 46.
- Bolton, D.K., Friedl, M.A., 2013. Forecasting crop yield using remotely sensed vegetation indices and crop phenology metrics. *Agric. For. Meteorol.* 173, 74–84.
- Brunsdon, C., Fotheringham, S., Charlton, M., 1998. Geographically weighted regression. *J. Royal Statist. Soc.: Ser. D (Statist.)* 47, 431–443.
- Brunton, L.A., Alexander, N., Wint, W., Ashton, A., Broughan, J.M., 2017. Using geographically weighted regression to explore the spatially heterogeneous spread of bovine tuberculosis in England and Wales. *Stoch. Env. Res. Risk A.* 31, 339–352.
- Büchi, L., Wendling, M., Amossé, C., Necpalova, M., Charles, R., 2018. Importance of cover crops in alleviating negative effects of reduced soil tillage and promoting soil fertility in a winter wheat cropping system. *Agric. Ecosyst. Environ.* 256, 92–104.
- Cai, Y., Guan, K., Lobell, D., Potgieter, A.B., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* 274, 144–159.
- Chandel, N.S., Tiwari, P.S., Singh, K.P., Jat, D., Gaikwad, B.B., Tripathi, H., Golhani, K., 2019. Yield prediction in wheat (*Triticum aestivum* L.) using spectral reflectance indices. *Curr. Sci.* 116, 272.
- Chen, J., Jönsson, P., Tamura, M., Gu, Z., Matsushita, B., Eklundh, L., 2004. A simple method for reconstructing a high-quality NDVI time-series data set based on the Savitzky–Golay filter. *Remote Sens. Environ.* 91, 332–344.
- Climate Central, 2020. It's Official: 2019 was the Nation's Second Wettest Year on Record. Available online: <https://www.climatecentral.org/gallery/maps/its-official-2019-was-the-nations-second-wettest-year-on-record>.
- Crane-Droesch, A., 2018. Machine learning methods for crop yield prediction and climate change impact assessment in agriculture. *Environ. Res. Lett.* 13, 114003.
- Curtis, T., Halford, N.G., 2014. Food security: the challenge of increasing wheat health and the importance of not compromising food safety. *Ann. Appl. Biol.* 164, 354–372.
- Du, Z., Wang, Z., Wu, S., Zhang, F., Liu, R., 2020. Geographically neural network weighted regression for the accurate estimation of spatial non-stationarity. *Int. J. Geogr. Inf. Sci.* 1–25.
- Fotheringham, A.S., Charlton, M.E., Brunsdon, C., 1998. Geographically weighted regression: a natural evolution of the expansion method for spatial data analysis. *Environ. Plan. A* 30, 1905–1927.
- Fotheringham, A.S., Crespo, R., Yao, J., 2015. Geographical and temporal weighted regression (GTWR). *Geogr. Anal.* 47, 431–452.
- Gao, Y., Cheng, J., Meng, H., Liu, Y., 2019. Measuring spatio-temporal autocorrelation in time series data of collective human mobility. *Geo-spatial Inform. Sci.* 22, 166–173.
- Ghimire, R., Ghimire, B., Mesbah, A.O., Sainju, U.M., Idowu, O.J., 2019. Soil health response of cover crops in winter wheat–fallow system. *Agron. J.* 111, 2108–2115.
- Goldberg, Y., 2016. A primer on neural network models for natural language processing. *J. Artif. Intell. Res.* 57, 345–420.
- Gu, Y., Brown, J.F., Verdin, J.P., Wardlow, B., 2007. A five-year analysis of MODIS NDVI and NDWI for grassland drought assessment over the central Great Plains of the United States. *Geophys. Res. Lett.* 34.
- Guo, Y., Tang, Q., Gong, D.-Y., Zhang, Z., 2017. Estimating ground-level PM2.5 concentrations in Beijing using a satellite-based geographically and temporally weighted regression model. *Remote Sens. Environ.* 198, 140–149.
- Hagenauer, J., Helbich, M., 2021. A geographically weighted artificial neural network. *Int. J. Geogr. Inf. Sci.* 1–21.
- Han, J., Zhang, Z., Cao, J., Luo, Y., Zhang, L., Li, Z., Zhang, J., 2020. Prediction of winter wheat yield based on multi-source data and machine learning in China. *Remote Sens.* 12, 236.
- Herbek, J., Lee, C., 2009. A Comprehensive Guide to Wheat Management in Kentucky. US Department of Agriculture, M. Scott Smith, Director, Cooperative Extension Service, University of Kentucky College of Agriculture, Lexington, and Kentucky State University, Frankfort.
- Holman, J.D., Schlegel, A.J., Thompson, C.R., Lingenfelter, J.E., 2011. Influence of precipitation, temperature, and 56 years on winter wheat yields in western Kansas. *Crop Manag.* 10, 1–10.
- Huang, B., Wu, B., Barry, M., 2010. Geographically and temporally weighted regression for modeling spatio-temporal variation in house prices. *Int. J. Geogr. Inf. Sci.* 24, 383–401.
- Huang, J., Wang, X., Li, X., Tian, H., Pan, Z., 2013. Remotely sensed rice yield prediction using multi-temporal NDVI data derived from NOAA's AVHRR. *PLoS One* 8, e70816.
- Imran, M., Zurita-Milla, R., Stein, A., 2013. Modeling crop yield in west-African Rainfed agriculture using global and local spatial regression. *Agron. J.* 105, 1177–1188.
- Imran, M., Stein, A., Zurita-Milla, R., 2015. Using geographically weighted regression kriging for crop yield mapping in West Africa. *Int. J. Geogr. Inf. Sci.* 29, 234–257.
- 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, V.R., Kazula, M.J., Coulter, J.A., Naeve, S.L., y Garcia, A.G., 2020. In-season weather data provide reliable yield estimates of maize and soybean in the US central Corn Belt. *Int. J. Biometeorol.* 1–14.
- Kang, Y., 2018. County-Level Corn Yield Forecast with Deeping Learning and Satellite Data.
- Kang, Y., Ozdogan, M., Zhu, X., Ye, Z., Hain, C., Anderson, M., 2020. Comparative assessment of environmental variables and machine learning algorithms for maize yield prediction in the US Midwest. *Environ. Res. Lett.* 15, 064005.
- Kern, A., Barcza, Z., Marjanović, H., Árendás, T., Fodor, N., Bónis, P., Bognár, P., Lichtenberger, J., 2018. Statistical modelling of crop yield in Central Europe using climate data and remote sensing vegetation indices. *Agric. For. Meteorol.* 260, 300–320.
- Kim, S.-R., Prasad, A.K., El-Askary, H., Lee, W.-K., Kwak, D.-A., Lee, S.-H., Kafatos, M., 2014. Application of the Savitzky–Golay filter to land cover classification using temporal MODIS vegetation indices. *Photogramm. Eng. Remote. Sens.* 80, 675–685.
- Knott, C.A., 2016. Identifying key wheat growth stages. *Plant Soil Sci. Present.* 5.
- Kuwata, K., Shibasaki, R., 2016. Estimating corn yield in the United States with MODIS EVI and machine learning methods. *ISPRS Ann. Photogramm. Remote Sens. Spat. Inf. Sci.* 3, 131–136.
- Leung, Y., Mei, C.-L., Zhang, W.-X., 2000. Statistical tests for spatial nonstationarity based on the geographically weighted regression model. *Environ. Plan. A* 32, 9–32.
- Li, T., Shen, H., Yuan, Q., Zhang, L., 2020. Geographically and temporally weighted neural networks for satellite-based mapping of ground-level PM2.5. *ISPRS J. Photogramm. Remote Sens.* 167, 178–188.
- Lin, C.-H., Wen, T.-H., 2011. Using geographically weighted regression (GWR) to explore spatial varying relationships of immature mosquitoes and human densities with the incidence of dengue. *Int. J. Environ. Res. Public Health* 8, 2798–2815.

- Ma, Y., Zhang, Z., Kang, Y., Özdogan, M., 2021. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. *Remote Sens. Environ.* 259, 112408.
- Manjunath, K.R., Potdar, M.B., Purohit, N.L., 2002. Large area operational wheat yield model development and validation based on spectral and meteorological data. *Int. J. Remote Sens.* 23, 3023–3038.
- Maselli, F., Rembold, F., 2001. Analysis of GAC NDVI data for cropland identification and yield forecasting in Mediterranean African countries. *Photogramm. Eng. Remote. Sens.* 67, 593–602.
- Matsumura, K., Gaitan, C.F., Sugimoto, K., Cannon, A.J., Hsieh, W.W., 2015. Maize yield forecasting by linear regression and artificial neural networks in Jilin, China. *J. Agric. Sci.* 153, 399–410.
- Mishra, A., Hansen, J.W., Dingkuhn, M., Baron, C., Traoré, S.B., Ndiaye, O., Ward, M.N., 2008. Sorghum yield prediction from seasonal rainfall forecasts in Burkina Faso. *Agric. For. Meteorol.* 148, 1798–1814.
- Moran, P.A.P., 1950. Notes on continuous stochastic phenomena. *Biometrika* 37, 17–23.
- Nadkarni, P.M., Ohno-Machado, L., Chapman, W.W., 2011. Natural language processing: an introduction. *J. Am. Med. Inform. Assoc.* 18, 544–551.
- Niazian, M., Sadat-Noori, S.A., Abdipour, M., 2018. Modeling the seed yield of Ajowan (*Trachyspermum ammi* L.) using artificial neural network and multiple linear regression models. *Ind. Crop. Prod.* 117, 224–234.
- O'brien, R.M., 2007. A caution regarding rules of thumb for variance inflation factors. *Qual. Quant.* 41, 673–690.
- Oettli, P., Sultan, B., Baron, C., Vrac, M., 2011. Are regional climate models relevant for crop yield prediction in West Africa? *Environ. Res. Lett.* 6, 014008.
- Pan, Y., Li, L., Zhang, J., Liang, S., Zhu, X., Sulla-Menashe, D., 2012. Winter wheat area estimation from MODIS-EVI time series data using the crop proportion phenology index. *Remote Sens. Environ.* 119, 232–242.
- Pede, T., Mountakis, G., Shaw, S.B., 2019. Improving corn yield prediction across the US Corn Belt by replacing air temperature with daily MODIS land surface temperature. *Agric. For. Meteorol.* 276, 107615.
- Ren, J., Chen, Z., Zhou, Q., Tang, H., 2008. Regional yield estimation for winter wheat with MODIS-NDVI data in Shandong, China. *Int. J. Appl. Earth Obs. Geoinf.* 10, 403–413.
- 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.
- 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. For. Meteorol.* 284, 107886.
- Sciencing, 2017. What Are the Six Stages of the Life Cycle of a Wheat Plant? Available online: <https://sciencing.com/the-parts-of-a-wheat-plant-12211988.html> (accessed on January, 13, 2021).
- Sentelhas, P.C., Dalla Marta, A., Orlandini, S., Santos, E.A., Gillespie, T.J., Gleason, M.L., 2008. Suitability of relative humidity as an estimator of leaf wetness duration. *Agric. For. Meteorol.* 148, 392–400.
- Sharma, S., 2017. Activation Functions in Neural Networks. Towards Data Science, p. 6.
- Shen, Z., Odene, M., Okhrin, O., 2018. Adaptive local parametric estimation of crop yields: implications for crop insurance rate making. *Eur. Rev. Agric. Econ.* 45, 173–203.
- Shewry, P.R., Hey, S.J., 2015. The contribution of wheat to human diet and health. *Food Energy Secur.* 4, 178–202.
- Shiu, Y.-S., Chuang, Y.-C., 2019. Yield estimation of Paddy Rice based on satellite imagery: comparison of global and local regression models. *Remote Sens.* 11, 111.
- Shook, J., Wu, L., Gangopadhyay, T., Ganapathysubramanian, B., Sarkar, S., Singh, A.K., 2018. Integrating genotype and weather variables for soybean yield prediction using deep learning. *Agronomy Publications* 542 [https://lib.dr.iastate.edu/agron\\_pubs/542](https://lib.dr.iastate.edu/agron_pubs/542).
- Son, N.T., Chen, C.F., Chen, C.R., Minh, V.Q., Trung, N.H., 2014. A comparative analysis of multitemporal MODIS EVI and NDVI data for large-scale rice yield estimation. *Agric. For. Meteorol.* 197, 52–64.
- Song, W., Jia, H., Huang, J., Zhang, Y., 2014. A satellite-based geographically weighted regression model for regional PM2.5 estimation over the Pearl River Delta region in China. *Remote Sens. Environ.* 154, 1–7.
- Stewart Fotheringham, A., Charlton, M., Brunsdon, C., 1996. The geography of parameter space: an investigation of spatial non-stationarity. *Int. J. Geogr. Inf. Syst.* 10, 605–627.
- Svyrydov, A., Kuchuk, H., Tsiapa, O., 2018. Improving Efficiency of Image Recognition Process: Approach and case Study. IEEE, pp. 593–597.
- Tack, J., Barkley, A., Nalley, L.L., 2015. Effect of warming temperatures on US wheat yields. *Proc. Natl. Acad. Sci.* 112, 6931–6936.
- Tobler, W.R., 1970. A computer movie simulating urban growth in the Detroit region. *Econ. Geogr.* 46, 234–240.
- USDA, 2020a. United States Department of Agriculture Foreign Agricultural Service. Available online: <https://ipad.fas.usda.gov/cropexplorer/cropview/> (accessed on November, 3, 2020).
- USDA, 2020b. United States Department of Agriculture National Agricultural Statistics Service. Available online: <https://www.nass.usda.gov/> (accessed on November, 3, 2020).
- USDA, 2020c. United States Department of Agriculture National Agricultural Statistics Service. Available online: [https://www.nass.usda.gov/Charts\\_and\\_Maps/Crop\\_Progress\\_&\\_Condition/](https://www.nass.usda.gov/Charts_and_Maps/Crop_Progress_&_Condition/) (accessed on May, 26, 2020).
- Vicente-Serrano, S.M., Camarero, J.J., Olano, J.M., Martín-Hernández, N., Peña-Gallardo, M., Tomás-Burguera, M., Gazol, A., Azorin-Molina, C., Bhuyan, U., El Kenawy, A., 2016. Diverse relationships between forest growth and the normalized difference vegetation index at a global scale. *Remote Sens. Environ.* 187, 14–29.
- Vocke, G., Ali, M., 2013. US Wheat Production Practices, Costs, and Yields: Variations across Regions.
- Wang, Y., Zhang, Z., Feng, L., Du, Q., Runge, T., 2020. Combining multi-source data and machine learning approaches to predict winter wheat yield in the conterminous United States. *Remote Sens.* 12, 1232.
- Wheeler, D., Tiefelsdorf, M., 2005. Multicollinearity and correlation among local regression coefficients in geographically weighted regression. *J. Geogr. Syst.* 7, 161–187.
- Wu, B., Li, R., Huang, B., 2014. A geographically and temporally weighted autoregressive model with application to housing prices. *Int. J. Geogr. Inf. Sci.* 28, 1186–1204.
- Xiao, G., Zhang, Q., Li, Y., Wang, R., Yao, Y., Zhao, H., Bai, H., 2010. Impact of temperature increase on the yield of winter wheat at low and high altitudes in semiarid northwestern China. *Agric. Water Manag.* 97, 1360–1364.
- Yi, Y., Yang, D., Chen, D., Huang, J., 2007. Retrieving crop physiological parameters and assessing water deficiency using MODIS data during the winter wheat growing period. *Can. J. Remote. Sens.* 33, 189–202.
- You, Q., 2020. Determining paddy field spatiotemporal distribution and temperature influence using remote sensing in Songnen plain, northeastern China. *Arab. J. Geosci.* 13, 1–12.
- Yuan, W., Zheng, Y., Piao, S., Ciais, P., Lombardozzi, D., Wang, Y., Ryu, Y., Chen, G., Dong, W., Hu, Z., 2019. Increased atmospheric vapor pressure deficit reduces global vegetation growth. *Sci. Adv.* 5, eaax1396.
- Zhai, L., Li, S., Zou, B., Sang, H., Fang, X., Xu, S., 2018. An improved geographically weighted regression model for PM2.5 concentration estimation in large areas. *Atmos. Environ.* 181, 145–154.