Article

Winter Wheat Yield Estimation Based on Optimal Weighted Vegetation Index and BHT-ARIMA Model

Qiuzhuo Deng, Mengxuan Wu, Haiyang Zhang, Yuntian Cui, Minzan Li and Yao Zhang *

Key Laboratory of Smart Agriculture System Integration, Ministry of Education, China Agricultural University, Beijing 100083, China; 2018307160221@cau.edu.cn (Q.D.); 2017320201159@cau.edu.cn (M.W.);
b202030816178@cau.edu.cn (H.Z.); 2018308130115@cau.edu.cn (Y.C.); limz@cau.edu.cn (M.L.)
* Correspondence: zhangyao@cau.edu.cn

Abstract: This study aims to use remote sensing (RS) time-series data to explore the intrinsic relationship between crop growth and yield formation at different fertility stages and construct a high-precision winter wheat yield estimation model applicable to short time-series RS data. Sentinel-2 images were acquired in this study at six key phenological stages (rejuvenation stage, rising stage, jointing stage, heading stage, filling stage, filling-maturity stage) of winter wheat growth, and various vegetation indexes (VIs) at different fertility stages were calculated. Based on the characteristics of yield data continuity, the RReliefF algorithm was introduced to filter the optimal vegetation index combinations suitable for the yield estimation of winter wheat for all fertility stages. The Absolutely Objective Improved Analytic Hierarchy Process (AOIAHP) was innovatively proposed to determine the proportional contribution of crop growth to yield formation in six different phenological stages. The selected VIs consisting of MTCI(RE2), EVI, REP, MTCI(RE1), RECi(RE1), NDVI(RE1), NDVI(RE3), NDVI(RE2), NDVI, and MSAVI were then fused with the weights of different fertility periods to obtain time-series weighted data. For the characteristics of short time length and a small number of sequences of RS time-series data in yield estimation, this study applied the multiplexed delayed embedding transformation (MDT) technique to realize the data augmentation of the original short time series. Tucker decomposition was performed on the block Hankel tensor (BHT) obtained after MDT enhancement, and the core tensor was extracted while preserving the intrinsic connection of the time-series data. Finally, the resulting multidimensional core tensor was trained with the Autoregressive Integrated Moving Average (ARIMA) model to obtain the BHT-ARIMA model for wheat yield estimation. Compared to the performance of the BHT-ARIMA model with unweighted time-series data as input, the weighted time-series input significantly improves yield estimation accuracy. The coefficients of determination ($R^2$) were improved from 0.325 to 0.583. The root mean square error (RMSE) decreased from 492.990 to 323.637 kg/ha, the mean absolute error (MAE) dropped from 350.625 to 255.954, and the mean absolute percentage error (MAPE) decreased from 4.332% to 3.186%. Besides, BHT-ARMA and BHT-CNN models were also used to compare with BHT-ARIMA. The results indicated that the BHT-ARIMA model still had the best yield prediction accuracy. The proposed method of this study will provide fast and accurate guidance for crop yield estimation and will be of great value for the processing and application of time-series RS data.

Keywords: winter wheat; yield; Sentinel-2; time series; AOIAHP; Hankel tensor; ARIMA

1. Introduction

As one of the most widely planted crops globally, wheat yield prediction in large areas is essential to ensure food security and maintain sustainable agricultural development [1,2]. Traditional yield estimation methods mainly include statistical field surveys or sampling inspections, which are time-consuming and require a lot of manpower and material resources [3,4]. As a non-destructive emerging technology, remote sensing (RS) technology...
has the characteristics of high timeliness, wide coverage, and low cost [2], and is widely used in crop growth monitoring and yield prediction [5].

In recent decades, the use of assimilation methods to couple crop growth models with RS data has become one of the development trends in yield estimation. The crop growth model integrated multidisciplinary research (including crop physiology, ecology, meteorology, soil science, agronomy, etc.) to quantitatively and dynamically predict crop yield [6–14]. However, incomplete acquired localization data in practical applications will lead to serious bias and instability in crop field forecasting. Besides, the complexity of the agricultural system determines the uncertainty of assimilation, and a large number of data calculations also limit the accuracy of large-area yield estimation [15].

To simplify the calculation and improve the application of RS in regional yield estimation, some researchers have tried methods using the characteristic multispectral information and its derivatives, including characteristic vegetation indices (VI) based on one specific phenological period to construct a yield prediction model. However, the limited amount of information for a given growth period makes it difficult to track the entire process of yield formation during crop growth, and thus lacks accuracy and robustness [16–26].

The time-series data were arranged in the order of the size of the time stamp, which contains the dynamic information during a certain period. In recent years, the application of time-series data in agriculture has been more extensive, including crop mapping, yield estimation, etc. [27,28]. In crop yield estimation, the time-series RS data contain dynamic information on the growth and development of the crop throughout the growing season [29–31]. Using time-series data to estimate yields has become one of the research hot spots. Among the existing time-series forecasting (TSF) methods, Long Short-Term Memory (LSTM) [32], as a sort of deep learning method, has been widely used for crop yield estimation and prediction by its sequence modeling using a long-time memory function [33]. However, there is a large error in the parallel processing of multiple temporal data. In addition, Convolutional Neural Networks (CNN) [34] have also been increasingly used in yield estimation due to the shared convolutional kernels, which can achieve automatic feature extraction at the convolutional layer and greatly improve the efficiency of yield prediction. However, a large amount of information will be lost in the pooling layer [35]. Besides, the performance of the TSF models related to deep learning suffers badly when the amount of data is not enough, limiting these models’ use for yield estimation based on time-series RS data. Focusing on the shortages of the existing TSF models mentioned above, the Autoregressive Integrated Moving Average (ARIMA) [36] method could be used in yield estimation due to its simplicity and good statistical properties. In addition, in order to consider the intrinsic correlation between time-series data and further capture the interactions between nutrient and structural parameters with yield formation during crop growing, multiple delayed embedding transform (MDT) [37], an emerging technique, is used to perform tensor decomposition to obtain higher-order block Hankel tensor (BHT) [38]. The hybrid method combining the ARIMA and BHT can reduce information loss during tenderization, capture the intrinsic correlation of time series, solve insufficient experimental data to a certain extent, and has great potential in predicting winter wheat yield [39].

This study aims to propose a high-precision winter wheat yield estimation method applicable to short time-series RS data. Firstly, the optimal vegetation indices group suitable for yield estimation during six key growth periods of winter wheat was obtained based on the Regressional ReliefF (RReliefF) algorithm. Then, the Absolutely Objective Improved Analytic Hierarchy Process (AOIAHP) algorithm was innovatively proposed to determine the contribution weights of the different growth stages to winter wheat yield accumulation. For short time-series RS data prediction, the BHT-ARIMA was firstly applied in yield estimation to solve the problem of insufficient sampled data. The paper is organized as follows. Section 2 introduces the experiment design/data used in this study and describes the details regarding the proposed Absolutely Objective Improved Analytic Hierarchy Process and BHT-ARIMA model, etc. Results of selected optimal VIs, contribution ratio
determination, and the winter wheat yield estimation are all described in Section 3. The comparison of the BHT-ARIMA model with other methods in yield estimation is discussed in Section 4. The paper concludes in Section 5 with a summary of the results.

2. Materials and Methods

2.1. Study Region

Hengshui City (115°10′–116°34′ E, 37°03′–38°23′ N), Hebei Province, China, was chosen as the study region. The region is part of the North China Plain, an alluvial plain of the Yellow River, Huaihe River, and Haihe River [40]. It is an important food production area in North China. In this region, winter wheat is the primary crop, and the main crop planting pattern is the rotation of winter wheat and summer maize [41]. Winter wheat was selected as a research crop in yield estimation.

2.2. Experiment Design

Field data measurement was conducted on the winter wheat only during the harvest time. Twenty-two fields, each of which has an area larger than 1 km$^2$, were selected and examined. Five plots were deployed diagonally in each field with distance of 50 m to each other and 50 m from the edge of a field, as shown in Figure 1. Then, 110 plots were selected for the experiment. Besides, the coordinates of the experimental plots were recorded by a Global Positioning System (GPS, Trimble GeoExplorer 6000 Series GeoXH, Trimble Navigation, Ltd., Sunnyvale, CA, USA) to accurately obtain corresponding band information from RS images.

![Experiment locations](image)

Figure 1. Experiment locations.

2.3. Winter Wheat Yield Data

The growth of the farmland within the pixel scale of the Sentinel-2 was roughly uniform, so the yield per unit area (1 m$^2$) was chosen as a representative of the point, and the distribution of winter wheat within one plot is shown in Figure 1. From each 1 m$^2$ plot, all ears of winter wheat were collected. According to the standard method of obtaining the dry weight of the winter wheat ears, the ears were heated to 105 °C for killing to inactivate enzymes to reduce the effect on dry weight and dried at 80 °C until a constant weight was reached in the laboratory. Next, the dried grains were manually collected from the ears, and the final dry weight of the grains was recorded with the unit of g. In addition, the unit was then converted from g/m$^2$ to kg/ha.
2.4. Sentinel-2 Image Processing

The whole growth period of winter wheat occurs from early March to late May. The selected periods of this study are from rejuvenation to filling-maturity. Predominantly cloud-free Sentinel-2 images covering the whole territory during the key phenological phases (Level 1C Top-of-Atmosphere reflectance products) were downloaded from the Copernicus Open-Access Hub (https://scihub.copernicus.eu/dhus/#/home, accessed on 1 July 2020), as shown in Table 1. Only one Sentinel-2 image was used for each phenological phase. Details of the bands used in this study can be found in Table 2. All bands were atmospherically corrected using the Sen2Cor processor tool, and the bands at 20 m resolution were rescaled to 10 m before being stacked [18,42].

<table>
<thead>
<tr>
<th>Time</th>
<th>March</th>
<th>April</th>
<th>May</th>
</tr>
</thead>
<tbody>
<tr>
<td>E-Mar</td>
<td>M-Mar</td>
<td>L-Mar</td>
<td>E-Apr</td>
</tr>
<tr>
<td>Phenological Phases</td>
<td>Rejuvenation</td>
<td>Rising</td>
<td>Jointing</td>
</tr>
<tr>
<td>Dates of Sentinel-2 images (Day/Month/Year)</td>
<td>9 March 2017</td>
<td>29 March 2017</td>
<td>18 April 2017</td>
</tr>
</tbody>
</table>

Table 1. Sentinel-2 images covering dates and the phenology of winter wheat.

<table>
<thead>
<tr>
<th>Spectral Band</th>
<th>Wavelength (nm)</th>
<th>Spatial Resolution (m)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Band 2 BLUE</td>
<td>496.6</td>
<td>10</td>
</tr>
<tr>
<td>Band 3 GREEN</td>
<td>560</td>
<td>10</td>
</tr>
<tr>
<td>Band 4 RED</td>
<td>664.5</td>
<td>10</td>
</tr>
<tr>
<td>Band 5 RED EDGE1</td>
<td>703.9</td>
<td>20</td>
</tr>
<tr>
<td>Band 6 RED EDGE2</td>
<td>740.2</td>
<td>20</td>
</tr>
<tr>
<td>Band 7 RED EDGE3</td>
<td>782.5</td>
<td>20</td>
</tr>
<tr>
<td>Band 8 NIR</td>
<td>835.1</td>
<td>10</td>
</tr>
</tbody>
</table>

Table 2. Central wavelength and spatial resolution for the Sentinel-2 bands used in this study.

2.5. Data Analysis

2.5.1. RReliefF

In order to quickly and effectively select the optimal VIs suitable for winter wheat yield estimation and their importance ranking from the VI candidates, the RReliefF algorithm [43], as an extended variant of Relief [44], was adopted to screen the optimal VIs with high efficiency. Unlike Relief, which is a typical binary classification algorithm, the RReliefF could cope with the feature selection with continuous target variables. The calculation workflow is shown below.

RReliefF first sets ranking factors $R_{dy}$, $R_{dj}$, $R_{dy/dj}$, and $R_i$ equal to 0. Then, the algorithm iteratively selects a random observation $x_r$, finds the $k$-nearest observations to $x_r$, and updates all the intermediate ranking factors, as follows [45]:

\[ R_{dy}^i = R_{dy}^{i-1} + \Delta_y(x_r, x_q) \cdot d_{rq} \]
\[ R_{dj}^i = R_{dj}^{i-1} + \Delta_j(x_r, x_q) \cdot d_{rq} \]
\[ R_{dy/dj}^i = R_{dy/dj}^{i-1} + \Delta_y(x_r, x_q) \cdot \Delta_j(x_r, x_q) \cdot d_{rq} \]

where $R_{dy}$ is the ranking factor of having different values for the response $y$ (i.e., the actual yields), $R_{dj}$ is the ranking factor of having different values for the predictor $F_j$, and $R_{dy/dj}$ is the ranking factor of having different response values and different values for the predictor $F_j$. The $i$ and $i-1$ superscripts denote the iteration step number. $\Delta y(x_r, x_q)$ is the difference in the value of the continuous response $y$ between observations $x_r$ and $x_q$, shown
as Equation (4). Let \( y_r \) denote the value of the response for observation \( x_r \), and let \( y_q \) denote the value of the response for observation \( x_q \). \( \Delta y(x_r, x_q) \) is the difference in the value of the predictor \( F_j \) between observations \( x_r \) and \( x_q \), shown as Equation (5). Let \( x_{rq} \) denote the value of the \( j \)th predictor for observation \( x_r \), and let \( x_{qr} \) denote the value of the \( j \)th predictor for observation \( x_q \).

\[
\Delta y(x_r, x_q) = \frac{|y_r - y_q|}{\max(y) - \min(y)} \quad (4)
\]

\[
\Delta_j(x_r, x_q) = \frac{|x_{rj} - x_{qj}|}{\max(F_j) - \min(F_j)} \quad (5)
\]

\( d_{rq} \) is a distance function of the form, shown as Equation (6), and the distance is subject to the scaling, \( \delta_{rq} \), shown as Equation (7):

\[
d_{rq} = \frac{\delta_{rq}}{\sum_{i=1}^{k} \delta_{ri}} \quad (6)
\]

\[
\delta_{rq} = e^{-(\text{rank}(r,q))^2} \quad (7)
\]

where \( \text{rank}(r,q) \) is the position of the \( q \)th observation among the nearest neighbors of the \( r \)th observation, sorted by distance, and \( k \) is the number of nearest neighbors.

Finally, the RReliefF calculates the predictor ranking factor \( R_j \) after fully updating all the intermediate ranking factors. \( m \) is the number of iterations.

\[
R_j = \frac{R_{dy\wedge d_j}}{R_{dy}} - \frac{R_{d_j} - R_{dy\wedge d_j}}{m - R_{dy}} \quad (8)
\]

In this study, the RReliefF algorithm evaluated candidate VIs’ importance to the measured yields from the 110 sampling plots. Subsequently, the importance ranking of VIs can be obtained, which were taken as a standard of selecting optimal vegetation index combinations, as shown in Figure 2. A higher ranking of the VI obtained by RReliefF means that the VI has a closer relation with the yield formation, which laid a foundation for the contribution weights estimation and yield prediction to be discussed in Section 2.5.3.

In this study, the top 10 features were selected for the following weight estimation and yield prediction.

![Figure 2](image-url)

**Figure 2.** Method of obtaining the optimal vegetation indexes using the RReliefF algorithm.

### 2.5.2. Improved Analytic Hierarchy Process

Different phenological stages of winter wheat have different effects on the accumulation of winter wheat yield. It is crucial to determine the weights of each phenological stage scientifically and effectively using the results derived from the mathematical analysis of limited data.
The improved analytic hierarchy process (IAHP) is a weighting method that is optimized from the analytic hierarchy process (AHP) for solving fuzzy problems with incomplete information. A comparison matrix needs to be constructed at the beginning [3]. Compared with AHP, IAHP does not need a consistency test, avoiding adjusting the comparison matrix multiple times and improving efficiency and accuracy [3,46,47]. More importantly, IAHP uses the “three-scale method” to construct the comparison matrix, which means the elements in the matrix are just integers from 0 to 2, and these integers imply the degree of importance between the elements [47]. The values of the elements are determined by the subjective analysis of experts. In this study, IAHP only needed the ranking of the importance of the phenological periods to construct the comparison matrix, significantly reducing the amount of information needed to determine the weights and leveraging the results of mathematical data analysis.

2.5.3. Absolutely Objective Improved Analytic Hierarchy Process

IAHP makes it easier for experts to judge the importance of the comparison to construct the comparison matrix. However, it still requires experts to compare the importance of two stages affected by subjective factors to a certain extent [3,46,47]. The Absolutely Objective Improved Analytic Hierarchy Process (AOIAHP) is innovatively proposed in this paper to completely eliminate the influence of subjective factors. In AOIAHP, the objective data analysis determined the comparison matrix be constructed in the first step of IAHP. In this case, the AOIAHP improved the accuracy of weight calculation.

As shown in Figure 3, the selected ten VI time-series data points of six key phenological phases were used as the predictor variable vector to form the input matrices (110 × 10), and the yield data measured by different plots were used as a response vector (110 × 1). The importance ranking of six key phenological phases of winter wheat based on a single VI can be obtained after RRelieff analysis. Subsequently, the rankings were averaged among the selected ten VIs to obtain a total ranking vector of the six key phenological phases.

![Figure 3](https://example.com/figure3.png)

**Figure 3.** Method of obtaining the optimal weighted vegetation index group using AOIAHP of the study.

Based on the composite ranking of six key phenological phases, the comparison matrix, \( A(a_{ij}) \), shown as Equation (9), can be scientifically and objectively constructed:

\[
A(a_{ij}) = \begin{pmatrix}
    a_{11} & \cdots & a_{16} \\
    \vdots & \ddots & \vdots \\
    a_{61} & \cdots & a_{66}
\end{pmatrix}
\]

(9)

The value of \( a_{ij} \) is the comparison result related to the ranking of the importance of six key phenological stages of winter wheat. The values of \( i \) and \( j \) are both from 1 to 6,
corresponding to the six key phenological wheat periods, as shown in Table 1, and the integer values of $a_{ij}$ are from 0 to 2, corresponding to period $i$ being less important than period $j$, period $i$ being equally important as period $j$, and period $i$ being more important than period $j$, respectively.

Then, the importance coefficients vector, $r$, of winter wheat growth stages can be calculated using the elements in the comparison matrix, $A(a_{ij})$, shown as Equation (10):

$$r_j = \sum_{i=1}^{6} a_{ij}, j = 1, 2, \cdots, 6$$

where the subscript $j$ for $r$ indicates six key phenological periods of the winter wheat.

The judgment matrix, $B(b_{ij})$, can be constructed based on the importance coefficients of winter wheat key phenological stages, obtained by Equation (11):

$$B(b_{ij}) = \begin{pmatrix} b_{11} & \cdots & b_{16} \\ \vdots & \ddots & \vdots \\ b_{61} & \cdots & b_{66} \end{pmatrix}$$

$$b_{ij} = \begin{cases} \frac{r_i - r_j}{r_{\text{max}} - r_{\text{min}}} \times \left( \frac{r_{\text{max}} - 1}{r_{\text{max}} - 1} + 1 \right), & r_i \geq r_j, \\ \left( \frac{r_i - r_j}{r_{\text{max}} - r_{\text{min}}} \times \left( \frac{r_{\text{max}} - 1}{r_{\text{max}} - 1} + 1 \right) \right)^{-1}, & r_i < r_j \end{cases}, i = 1, 2, \cdots, 6; j = 1, 2, \cdots, 6$$

where $r$ takes on different values depending on the subscript, and $r_{\text{max}}$ and $r_{\text{min}}$ are the maximum and minimum value of importance coefficients obtained by Equation (10), respectively.

Finally, the quasi-optimal matrix, $Q(q_{ij})$, can be obtained based on elements in $B(b_{ij})$, shown as Equation (13), obtained by Equation (14):

$$Q(q_{ij}) = \begin{pmatrix} q_{11} & \cdots & q_{16} \\ \vdots & \ddots & \vdots \\ q_{61} & \cdots & q_{66} \end{pmatrix}$$

$$q_{ij} = 10^{\frac{1}{5} \sum_{k=1}^{6} q_{ij}^{k}}, i = 1, 2, \cdots, 6; j = 1, 2, \cdots, 6$$

where the value of $q_{ij}$ can be calculated by Equation (14) using the elements of the matrix $B(b_{ij})$.

The weight vector, $W$, of six key phenological phases can be obtained based on the maximum eigenvalue and the maximum eigenvector, $d$, of the quasi-optimal matrix $Q(q_{ij})$. The elements of $W$ can be obtained by Equation (15), where $\sum_{k=1}^{6} w_k = 1$, and $k$ refers to 6 key phenological phases:

$$w_k = \frac{d_k}{\sum_{i=1}^{6} d_i}, k = 1, 2, \cdots, 6$$

Finally, with the obtained weights of the six phenological phases, the optimal weighted vegetation index matrix ($110 \times 10 \times 6$) can be constructed, as shown in Figure 3.

### 2.6. BHT-ARIMA Model

The BHT-ARIMA model could fully exploit the intrinsic linkage of the time series and achieve augmentation of short time-series data, effectively solving the problem of simultaneous prediction of multiple short time-series data points [48]. The specific steps of modeling are as follows [48]:

1. **Obtain the high-order multidimensional short time-series data using MDT.** As is shown in Figure 4, the ten calculated optimal weighted vegetation index time-series data points of six key phenological phases from different experimental plots, whose
total number is \( I \) (110 herein), are decomposed in time dimension with the time decomposition number parameter, \( T \), to obtain the decomposed time-series data, \( \{ \chi \}_I \times T \). High-order time-series data, \( \{ \hat{\chi}_I \}_I \times T \times T = \{ \hat{\chi}_I, \hat{\chi}_2, \cdots, \hat{\chi}_T \} \), are obtained by stretching the decomposed time-series data, \( \{ \chi \}_I \times T \), with the replication matrix, which is determined by the parameter \( \tau \). The relationship between \( T \) and \( \hat{T} \) is shown in Equation (16):

\[
\hat{T} = T - \tau + 1
\]

(16)

Figure 4. Method of obtaining the high-order multidimensional short time-series data using MDT. (a) depicts the first step that all layers of the optimal weighted VI group were decomposed in the time dimension to get \( \{ \chi \}_I \times T \). (b) depicts the step that \( \{ \chi \}_I \times T \) was stretched by the replication matrix to get \( \{ \chi \}_I \times T \times T \) shown in (c).

(2) **Obtain core tensors from the BHT by Tucker decomposition** [49]. With the BHT, we compute its \( d \)-order differences to obtain a \( M \)-order tensor, \( \{ \Delta^d \hat{\chi}_I \} = \{ \Delta^d \hat{\chi}_1, \Delta^d \hat{\chi}_2, \cdots, \Delta^d \hat{\chi}_T \} \) [50]. After Tucker decomposition, \( \{ \Delta^d \hat{\chi}_I \} \) is decomposed into the product of a core tensor, \( \{ \Delta^d \hat{G}_I \} = \{ \Delta^d \hat{G}_1, \Delta^d \hat{G}_2, \cdots, \Delta^d \hat{G}_T \} \), and \( M \) orthogonal projection matrices, \( \{ \hat{U}^{(m)} \} = \{ \hat{U}^{(1)}, \hat{U}^{(2)}, \cdots, \hat{U}^{(M)} \} \), which are shown as below:

\[
\Delta^d \hat{G}_t = \Delta^d \hat{\chi}_t \times_1 \hat{U}^{(1)^T} \times_2 \cdots \times_m \hat{U}^{(m)^T} \\
\text{s. t. } \hat{U}^{(m)^T} \hat{U}^{(m)} = I
\]

(17)

(18)

where \( m = 1, 2, \cdots, M \), and \( \{ \hat{U}^{(m)} \} \) are the projection matrices, which could maximally preserve the temporal continuity between core tensors, and \( \{ \Delta^d \hat{G}_t \} \) are the low-rank core tensors, which could represent the important original information of Hankel and the intrinsic linkage of the time-series data.

(3) **Tensorized ARIMA model establishment.** To retain the temporal correlations among core tensors, a \( (p; d; q) \)-order ARIMA model is constructed in the tensor form, which could be used to connect the current core tensor, \( \{ \Delta^d \hat{G}_t \} \) and the previous core tensors, \( \{ \Delta^d \hat{G}_{t-1}, \Delta^d \hat{G}_{t-2}, \cdots, \Delta^d \hat{G}_{t-p} \} \), as in Equation (19). \( \{ \alpha_t \} \) and \( \{ \hat{\beta}_t \} \) are the coefficients of the autoregressive model (AR) and the moving average model (MA), respectively, which were estimated from the core tensors and the residual time-series data [39,51]. \( \{ \hat{\varepsilon}_{t-i} \} \) are the random errors of past \( q \) observations, and \( \{ \hat{\varepsilon}_t \} \) is the forecast error at the current time point, which should be minimized to optimal zero.

\[
\Delta^d \hat{G}_t = \sum_{i=1}^{p} \alpha_i \Delta^d \hat{G}_{t-i} - \sum_{i=1}^{q} \hat{\beta}_i \hat{\varepsilon}_{t-i} + \hat{\varepsilon}_t
\]

(19)
where $\{\Delta d^t G_t\}$ is input into the tensorized ARIMA model to forecast a new core tensor, $\{\Delta d^t \hat{G}_{t+1}\}$. Back projection used $\{\hat{U}^{(m)}\}$ to obtain $\{\Delta d^t \hat{\chi}^T_{t+1}\}$ to get the predicted values at the $\hat{T}+1$ time point for all the time-series data simultaneously, which is shown in Equation (20). Finally, inverse differencing and inverse-MDT were applied to obtain the prediction result $\{\hat{\chi}^T_{t+1}\}$.

$$\Delta d^t \hat{\chi}^T_{t+1} = \Delta d^t \hat{G}_{t+1} \times_{1} \hat{U}^{(1)}^{T} \cdots \times_{m} \hat{U}^{(m)}^{T} \tag{20}$$

In this study, the BHT-ARIMA model was applied to estimate the yield of the winter wheat based on short RS time-series data. Five key model parameters required special attention when predicting winter wheat yields using the BHT-ARIMA model, as shown in Table 3. The number of time-series data points of single VI ($I$) and the number of selected optimal VIs were determined by the original dataset. The number of time-series data points of single VI ($I$) is the amount of time-series data for each single VI. There were 110 sets of time-series data representing 110 experimental plots for each single VI. The meaning of the number of selected optimal VIs is obvious. It represents the number of VIs selected to be input into the BHT-ARIMA. The number of selected optimal VIs was set to 10, as 10 VIs had the best prediction performance and reduced the complexity of the model calculation, which was proven in Section 3.3.2. The time decomposition number parameter ($T$) and the replication matrix parameter ($r$) were determined manually according to the model characteristics. The time decomposition number parameter ($T$) was related to time-series segmentation, an integer that is limited to an integer multiple of 10. The replication matrix parameter ($r$) was an integer less than the time decomposition number parameter ($T$). The number of TS layers ($\hat{T}$) was calculated from the time decomposition number parameter ($T$) and the replication matrix parameter ($r$), as shown in Equation (16). The specific parameter values and notes are summarized in Table 3.

### Table 3. Setting of BHT-ARIMA parameters.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Value</th>
<th>Note</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of time-series data points of single VI ($I$)</td>
<td>110</td>
<td>Determined by the original dataset</td>
</tr>
<tr>
<td>Number of selected optimal VIs</td>
<td>10</td>
<td>Determined by the original dataset</td>
</tr>
<tr>
<td>Time decomposition number parameter ($T$)</td>
<td>10</td>
<td>Determined manually, integer multiples of 10</td>
</tr>
<tr>
<td>Replication matrix parameter ($r$)</td>
<td>5</td>
<td>Determined manually, integers less than $T$</td>
</tr>
<tr>
<td>Number of TS layers ($\hat{T}$)</td>
<td>6</td>
<td>Calculated by Equation (16)</td>
</tr>
</tbody>
</table>

Although the BHT-ARIMA model utilized complex data processing techniques such as tensor decomposition, this algorithm did not need to copy data or pack existing data, but wrote data directly to the tensor, and no additional copies were created for both in-process and out-of-process execution. Therefore, this algorithm did not require high hardware configuration. The algorithm’s arithmetic power requirement was 30 g of semi-precision, which can be run 783 times per second according to the theoretical value, and the peak bandwidth was about 540 GB/s under the common RTX 2070 graphics card environment, so it only took a few minutes to obtain the prediction results of this model.

### 2.7. Prediction Accuracy Evaluation

In this study, the coefficient of determination ($R^2$), mean absolute error (MAE), mean absolute percentage error (MAPE), and root mean square error (RMSE) were used to determine the predictive performance of the model. The $R^2$, MAE, MAPE, and RMSE were defined as follows:

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

$$\text{MAE} = \frac{1}{n} \sum_{i=1}^{n} |\hat{y}_i - y_i|$$

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

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (\hat{y}_i - y_i)^2}$$
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - y_i'|
(22)

MAPE = \frac{1}{n} \sum_{i=1}^{n} \frac{|y_i - y_i'|}{y_i} \times 100%
(23)

RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - y_i')^2}
(24)

where \( n \) represents the number of samples, \( y_i \) is the observed yield value, \( y_i' \) is the predicted yield, and \( \overline{y} \) is the average observed yield.

3. Results and Analysis

3.1. Growth Analysis of Winter Wheat Based on Optimal VIs during the Whole Growth Period

3.1.1. Initial Vegetation Index Candidates

VIs shown in Table 3 were initially chosen as candidates for the following selection. All these VIs have been proven to have a close correlation with crop nutritional parameters (such as chlorophyll, nitrogen, etc.) and structural parameters (leaf area index (LAI), etc.). VIs, including NDVI [17,18,20], GNDVI [21], MTCI [22], WDRVI [24,25], etc., have often been used directly to build linear regression models for yield estimation. In addition, some characteristic VIs, including REP [23] and SAVI [52], etc., have proven to be highly correlated with crop nutrition parameters. Besides, VIs including NDVI [53,54], WDRVI [55,56], etc., have been applied to measure structural parameters (LAI, etc.). These indexes can also be used to reflect crop yield from the aspect of nutrition. Three red-edge bands in Sentinel-2 were used in the study, so VIs including NDRE, MTCI, and CI\(_{RE}\) with red-edge bands (abbreviated as RE) were calculated using different red-edge bands, in order to take full advantage of multiple red-edge bands in Sentinel-2. A total of 18 VIs were obtained for the following selection. The specific calculation equations using the bands shown in Table 2 are shown in the middle column in Table 4.

Table 4. Initial vegetation index candidates.

<table>
<thead>
<tr>
<th>Vegetation Index</th>
<th>Description</th>
<th>References</th>
</tr>
</thead>
<tbody>
<tr>
<td>Normalized Difference Vegetation Index (NDVI)</td>
<td>((\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED}))</td>
<td>[57]</td>
</tr>
<tr>
<td>Normalized Difference Red-Edge Index (NDRE)</td>
<td>((\text{NIR} - \text{RE}) / (\text{NIR} + \text{RE}))</td>
<td>[58]</td>
</tr>
<tr>
<td>Green Normalized Difference Vegetation Index (GNDVI)</td>
<td>((\text{NIR} - \text{GREEN}) / (\text{NIR} + \text{GREEN}))</td>
<td>[59]</td>
</tr>
<tr>
<td>Wide Dynamic Range Vegetation Index (WDRVI)</td>
<td>((0.2 \times \text{NIR} - \text{RED}) / (0.2 \times \text{NIR} + \text{RED}))</td>
<td>[60]</td>
</tr>
<tr>
<td>Enhanced Vegetation Index (EV1)</td>
<td>(2.5 \times (\text{NIR} - \text{RED}) / (\text{NIR} + 6 \times \text{RED} + 7 \times \text{BLUE} - 1))</td>
<td>[61]</td>
</tr>
<tr>
<td>Two-band Enhanced Vegetation Index (EV2)</td>
<td>((\text{NIR} - \text{RED}) / (\text{NIR} + 2.4 \times \text{RED} + 1))</td>
<td>[62]</td>
</tr>
<tr>
<td>MERIS Terrestrial Chlorophyll Index (MTCI)</td>
<td>((\text{NIR} - \text{RE}) / (\text{NIR} - \text{RED}))</td>
<td>[63]</td>
</tr>
<tr>
<td>Green Chlorophyll Vegetation Index (GCVI)</td>
<td>((\text{NIR} / \text{GREEN}) - 1)</td>
<td>[64]</td>
</tr>
<tr>
<td>Chlorophyll Index Red-Edge (CI(_{RE}))</td>
<td>((\text{NIR} / \text{RE}) - 1)</td>
<td>[64]</td>
</tr>
<tr>
<td>Modified Soil-Adjusted Vegetation Index (MSAVI)</td>
<td>(\frac{1}{2} \left(2 \times \text{NIR} + 1 - \sqrt{2} (2 \times \text{NIR} + 1)^2 - 8 \times (\text{NIR} - \text{RED})\right))</td>
<td>[65]</td>
</tr>
<tr>
<td>Optimization Soil-Adjusted Vegetation Index (OSAVI)</td>
<td>(1.16 \times (\text{NIR} - \text{RED}) / (\text{NIR} + \text{RED} + 0.16))</td>
<td>[66]</td>
</tr>
<tr>
<td>Red-Edge Position Index (REP)</td>
<td>(700 + 40 \times ((\text{RED} + \text{RE3}) / 2 - \text{RE1}) / (\text{RE2} - \text{RE1}))</td>
<td>[67]</td>
</tr>
</tbody>
</table>

3.1.2. Optimal VI Selection to Monitor the Growth of Winter Wheat

Among the VI candidates, the optimal VIs for yield estimation during the whole growth period were selected through the RRelieff algorithm and shown in Table 5, which are sensitive to the chemical and biophysical parameters, which determined the photosynthetic capacity and efficiency and further affected yield formation of winter wheat [68]. As is shown in this table, the red-edge band of Sentinel-2 was fully utilized. According to the published literature, MTCI was closely related to the chlorophyll content of crops, which is...
the basic element for photosynthesis [63,69]. Besides, MTMI is highly sensitive to changes in biomass during the whole growth period of winter wheat [70]. REP strongly correlates with a wide range of foliar N concentrations during the whole growth period, influencing photosynthesis and yield formation [23,71]. These are consistent with the stable and high ranking of MTMI and REP by RRReliefF. NDVI is one of the most widely used ViSs to evaluate crop growth and monitor the nutrients needed for crop growth, development, and yield estimation. Besides, NDVI(RE), calculated from the red-edge bands, red edge1, red edge2, and red edge3, delays the saturation trend during the late growth period [70]. Additionally, EVI was proposed to handle the saturation problem of NDVI with improved sensitivity in high-biomass regions and improved vegetation monitoring through a de-coupling of the canopy background signal and a reduction in atmosphere influences [68,72]. SAVI and its optimized variants (MSAVI, OSAVI) increased vegetation reflection while suppressing soil background effects in images to improve monitoring accuracy [65,66]. The above reviews verified that the combination of selected ViSs could give full play to their respective advantages in crop growth monitoring and yield forecasting during the whole growth period of winter wheat.

Table 5. Optimal ViSs for field estimation during the whole growth period.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Vegetation Index</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>MTMI(RE2)</td>
</tr>
<tr>
<td>2</td>
<td>EVI</td>
</tr>
<tr>
<td>3</td>
<td>REP</td>
</tr>
<tr>
<td>4</td>
<td>MTMI(RE1)</td>
</tr>
<tr>
<td>5</td>
<td>RECI(RE1)</td>
</tr>
<tr>
<td>6</td>
<td>NDVI(RE1)</td>
</tr>
<tr>
<td>7</td>
<td>NDVI(RE3)</td>
</tr>
<tr>
<td>8</td>
<td>NDVI(RE2)</td>
</tr>
<tr>
<td>9</td>
<td>NDVI</td>
</tr>
<tr>
<td>10</td>
<td>MSAVI</td>
</tr>
</tbody>
</table>

3.2. Contribution Ratio Determination to Winter Wheat Yield Accumulation

Based on AOIAHP, the contribution ratios of six key growth stages can be obtained. Figure 5 presents the bar chart of the contribution ratios of six key phenological phases, showing that the jointing period had the highest contribution ratio, followed by the rising, heading, filling, rejuvenation, and lastly, filling-maturity period.

![Figure 5. Contribution ratios of phenological phases in the study.](image)

In the rejuvenation period, yellow leaves of winter wheat turned green as the canopy structure and chemical parameter content of winter wheat increased. From the rising period to jointing, chemical and biophysical parameters influencing photosynthesis continued to increase, increasing contribution ratios to yield. From the jointing period to the heading
period, the growth and development of winter wheat dramatically increased, with the root group and the leaf area reaching the maximum, and the wheat ears beginning to differentiate. These stages are the key periods for determining the number of grains per ear, and it is also the consolidation period for the effective number of ears per hectare and the weight. After the heading period, starch, protein, and accumulated organic matter produced by photosynthesis are stored in the grain through assimilation, still determining the accumulation of winter wheat.

3.3. Winter Wheat Yield Prediction and Analysis

3.3.1. Winter Wheat Yield Prediction Based on Optimal Weighted Vegetation Indexes and the BHT-ARIMA Model

This research used six phenological phases weighted VI time series to establish the BHT-ARIMA model to predict winter wheat yield. The 110 samples were divided into 2 parts for model establishment, where 88 samples were selected randomly for use as the calibration subset, and the remaining 22 were used as the independent validation subset. The $R^2$, RSME, MAE, and MAPE for the relationship between the observed and predicted yields in the validation dataset were calculated to evaluate the performance of the prediction of the BHT-ARIMA model. The input of the tensorized ARIMA model (i.e., BHT-ARIMA) has to meet the data requirements of the ARIMA model. The final BHT-ARIMA model only outputs ten reasonable data points after eliminating the unreasonable data. As shown in Figure 6, the $R^2$ reached 0.583 and the RSME, MAE, and MAPE of this model were 323.637 kg/ha, 255.954, and 3.186%, respectively. This model performed better when the observed yields were larger than 8000 kg/ha. The good performance shows that the BHT-ARIMA model has capability in short time-series forecasting in agriculture.

![Figure 6. The $R^2$, RSME, MAE, and MAPE for the relationship between the observed and predicted yields based on the weighted optimal VI group. The dashed line is the 1:1 line.](image)

The vegetation index information was constant for each sample when the model was trained. To demonstrate whether the model was overfitted on the VIs, we trained the model with randomly selected yield samples from 80% of all samples, using other samples for cross-validation. The cross-validation is to verify whether the model is overfitted. With the increase in the number of verifications, overfitting can likely occur if the difference between training and validation errors is significant. In this case, the $R^2$ was used as the only metric to measure the cross-validation results. This validation was performed ten times. As we can
see from Figure 7, $R^2$ fluctuated around 0.5. The result indicated no significant overfitting, which implies that the model is generalized and can be used to estimate winter wheat yield. However, only ten reasonable results were obtained after data tensorization, the sample pool was reduced, which also had some influence on the validation process, and there were some limitations of this method.

![Figure 7. The relationship between the $R^2$ value and the number of verifications.](image)

3.3.2. Influence of the Number of Vegetation Indexes on the Performance of the BHT-ARIMA Model

In order to analyze the effect of the number of vegetation indices we selected on the model prediction performance, 18 matrices ranging in size from $110 \times 1 \times 6$ to $110 \times 18 \times 6$ were used sequentially as inputs to the BHT-ARIMA model. The $R^2$, RSME, MAE, and MAPE were calculated to assess the performance of the BHT-ARIMA model, as shown in Figure 8. Figure 8a shows that $R^2$ basically had an increasing trend, with the increase of the number of the selected indexes less than 10, and stabilized when the number was larger than 10. Figure 8b–d show that the RSME, MAE, and MAPE had approximately the same trend. They declined with high fluctuations and had larger values before ten. When the number of the selected indexes was greater than ten, the values of all three parameters gradually leveled off.

However, if too many indexes were selected, for each additional set of indexes, the BHT-ARIMA model needed to perform the process of tensor transformation, decomposition, and inversion again, which made the calculation cycle longer, increased the burden of the system, and reduced the efficiency of the model and redundancy.

Therefore, ten VIs selected by the RReliefF algorithm were used as model inputs to obtain better prediction results while saving computational cycles and maximizing model efficiency. Besides, the results indicated that the ten selected sets of vegetation index can represent valid information related to yield formation and effectively reduce the redundancy of information from the full vegetation index.
3.3.3. Comparisons of Performance of Weighted Vegetation Indexes and Unweighted Vegetation Indexes in the BHT-ARIMA Model

In order to verify the effect of the calculated weights on different phenological periods for winter wheat yield accumulation, two datasets of winter wheat VIs were established, including the dataset used in Section 3.3.1, i.e., the weighted optimal vegetation index group, and the dataset of the unweighted optimal vegetation index group. The samples were divided into two groups: 80% were the calibration group and the remaining 20% were the validation group. Figure 8 shows the scattered plots of the predicted yield and real winter wheat yield, which were predicted based on the unweighted and weighted optimal VI groups. Since the tensorized ARIMA model specifies non-zero input values, the model has only 10 sets of output values after automatically eliminating the unreasonable values. Figure 9 shows that the predicted yield was positively correlated with the observed yield. Figure 9a,b are the plots of the relationship between the predicted yield and the observed yield using unweighted and weighed VI groups, respectively. The $R^2$ of the calibrated model with weighted VI reached 0.583, higher than that of the model with unweighted VI, whose $R^2$ reached 0.325. The weighted model’s RSME, MAE, and MAPE were 323.637 kg/ha, 255.954, and 3.186%, respectively, smaller than those of the unweighted model, with 492.990 kg/ha, 350.625, and 4.332%, respectively.

The results indicate different degrees of influence on winter wheat yield for each phenological stage of winter wheat. The selected vegetation index combined with the contribution of key phenological stages can predict the yield of winter wheat well, with good robustness and realism.

Figure 8. The relationship between the (a) $R^2$, (b) RSME, (c) MAE, and (d) MAPE and the number of vegetation indexes.

(a) (b)
(c) (d)
with 695.276, 8.503%, and 791.545 kg/ha, respectively. The experimental data, including yields based on the (a) unweighted and (b) weighted optimal VI groups. The dashed line is the 1:1 line.

3.3.4. Comparisons of Performance of BHT-ARMA, BHT-CNN, and BHT-ARIMA Models in Winter Wheat Yield Estimation

To verify the advantages of the BHT-ARIMA model in predicting winter wheat yield using short time-series RS data, BHT-ARMA and BHT-CNN models were applied to compare the performance in winter wheat yield estimation among the BHT-ARMA, BHT-CNN, and BHT-ARIMA models. BHT-ARMA and BHT-CNN models are tensorized from the original autoregressive moving average (ARMA) and CNN models. The CNN network is divided into five convolutional layers, four pooling layers, and two fully connected layers. Optimal weighted VI time-series data consisting of 10 selected VIs were used to train the BHT-ARIMA, BHT-ARMA, and BHT-CNN models and further evaluate the yield estimation accuracies of the trained models. The $R^2$, RMSE, MAE, and MAPE for the observed and predicted yields’ relationship were calculated. The yield prediction results using the three methods mentioned above are shown in Figure 10.

As we can see from Figure 9, the overall performance of the BHT-ARIMA model was better than that of the BHT-ARMA and BHT-CNN models. The $R^2$ of the BHT-ARIMA model reached 0.583, higher than that of the BHT-ARMA and BHT-CNN models, with 0.302 and 0.350, respectively. Moreover, the MAE, MAPE, and RMSE of the BHT-ARIMA model, which were 284.366, 3.498%, and 327.358 kg/ha, respectively, were all lower than the BHT-ARMA model, with 711.993, 8.562%, and 958.730 kg/ha, and the BHT-CNN model with 695.276, 8.503%, and 791.545 kg/ha, respectively. The experimental data, including $R^2$, RMSE, MAE, and MAPE, are summarized in Table 6.

The results of the model performance comparison showed that the BHT-ARIMA model outperformed traditional deep learning networks and time-series prediction models in solving prediction problems for short time series of agricultural remote sensing, especially for multiple short time-series data prediction problems. Especially, we can see intuitively from Table 6 that, compared with BHT-ARMA and BHT-CNN, the $R^2$ of BHT-ARIMA had a dramatic rise from around 0.3 to around 0.6, while RMSE, MAE, and MAPE had a sharp decline. This may be caused by insufficient time-series data used in this study. Traditional prediction models require large, extensive data, limiting their application to yield predictions with fewer data. The BHT-CNN model was trained with a large amount of information that may be lost in the pooling layer, and the training results tended to converge on local minima, ignoring the local-to-whole correlation. The BHT-ARMA may not perform autoregressive learning, and only used data from the 10 sets of VIs with larger
weights to predict. These limitations mentioned above became obvious when the input data were insufficient. In contrast, the BHT-ARIMA model made full use of the intrinsic linkage of the time series, automatically enabling data expansion for short time series.

![Graph showing R^2, RSME, MAE, and MAPE for BHT-ARMA, BHT-CNN, and BHT-ARIMA models.](image)

**Figure 10.** The $R^2$, RSME, MAE, and MAPE for the relationship between the observed and predicted yields for the (a) BHT-ARMA, (b) BHT-CNN, and (c) BHT-ARIMA models. The dashed line is the 1:1 line. Since these tensorized models specify non-zero input values, the models have only 10 sets of output values after automatically eliminating the unreasonable values.

<table>
<thead>
<tr>
<th>Model</th>
<th>$R^2$</th>
<th>RMSE (kg/ha)</th>
<th>MAE</th>
<th>MAPE (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>BHT-ARMA</td>
<td>0.302</td>
<td>958.730</td>
<td>711.993</td>
<td>8.562</td>
</tr>
<tr>
<td>BHT-CNN</td>
<td>0.350</td>
<td>791.545</td>
<td>695.276</td>
<td>8.503</td>
</tr>
<tr>
<td>BHT-ARIMA</td>
<td>0.583</td>
<td>327.358</td>
<td>284.366</td>
<td>3.498</td>
</tr>
</tbody>
</table>

**Table 6.** The $R^2$, RSME, MAE, and MAPE of predictions using the BHT-ARMA, BHT-CNN, and BHT-ARIMA models.

### 3.3.5. Regional Winter Wheat Yield Estimation in Hengshui

In Section 3.3.1, the BHT-ARIMA model effectively estimated winter wheat yield based on short time-series VIs and actual yields, with $R^2$ reaching 0.583 from 110 experimental plots. In this section, the trained BHT-ARIMA model was used to estimate the regional...
winter wheat yields in Hengshui City, Hebei Province, based on the image of the winter wheat plant area. The data of the plant area of winter wheat in Hengshui City are from the published literature [40]. The prediction results are shown in Figure 11. Most predicted winter wheat yields in Hengshui City were between 7500 and 10,000 kg/ha, which was in line with the actual yield data in Section 2.3. In the west, mid-south, and east of Hengshui City, the predicted yields were higher than those of other areas in Hengshui City. Especially, the middle and south of Shenzhou, the east of Fucheng, and the east and south of Jingxian County were the higher-yield regions. These results were in general agreement with the spatial variation trend of the field experiment. The above results prove that the BHT-ARIMA model has strong spatial generalization in winter wheat yield estimation.

Figure 11. Predicted regional winter wheat yields in Hengshui City using the BHT-ARIMA model.

4. Discussion

Previous studies have explored correlations between multitemporal VIs and crop yields based on Random Forest (RF), Support Vector Machines (SVM), etc. However, these traditional deep learning-based methods require large, extensive data, limiting their application to yield predictions with fewer data [73–76]. Therefore, the BHT-ARIMA model applied in yield estimation has greater potential than other methods. In this study, BHT was obtained and combined with the ARIMA model, which can reduce information loss during tensorization, capture the intrinsic correlation of time series, and solve the problem of insufficient experimental data. The BHT-ARIMA model outperformed the BHT-ARMA and BHT-CNN models in estimating wheat yields, as mentioned in Section 3.3.4. In Table 7, we have summarized several published RS methods for winter wheat yield estimation, which all used multi-temporal time-series data. At the satellite scale, the Agriculture Crop Photosynthesis Model (ACPM) was proposed in [40] for the winter wheat aboveground biomass (DAM) and yield estimation, with $R^2$ around 0.49. It is worth noting that the measured yield data used in [40] were the same as the yield data in this study. In [41], SVM was applied to predict winter wheat yield based on ground-measured time-series spectral data. Its $R^2$ was larger than that of BHT-ARIMA, which may be caused by the larger scale of satellite data. In [77], based on UAV-scale RS time-series data, the PLS, SVM, and RF
methods achieved yield prediction with $R^2$ of 0.45, 0.50, and 0.55, respectively, which were all less than the $R^2$ value of BHT-ARIMA.

**Table 7.** Comparisons of different RS methods for winter wheat yield estimation.

<table>
<thead>
<tr>
<th>Model</th>
<th>Scale of Time-Series Data</th>
<th>$R^2$</th>
<th>RMSE (t/ha)</th>
<th>Reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>PLS</td>
<td>UAV</td>
<td>0.45</td>
<td>1.84</td>
<td>[77]</td>
</tr>
<tr>
<td>SVM</td>
<td>UAV</td>
<td>0.50</td>
<td>1.75</td>
<td>[77]</td>
</tr>
<tr>
<td>RF</td>
<td>UAV</td>
<td>0.55</td>
<td>1.66</td>
<td>[77]</td>
</tr>
<tr>
<td>SVM</td>
<td>Ground</td>
<td>0.62</td>
<td>0.35</td>
<td>[41]</td>
</tr>
<tr>
<td>ACPM</td>
<td>Satellite</td>
<td>0.49</td>
<td>0.59</td>
<td>[40]</td>
</tr>
<tr>
<td>BHT-ARIMA</td>
<td>Satellite</td>
<td>0.58</td>
<td>0.32</td>
<td>This study</td>
</tr>
</tbody>
</table>

The results showed the potential and advantages of BHT-ARIMA in the estimation of yield based on multi-temporal RS time-series data. However, it should be noted that in the tensorization process, the inversion of the tensor generated some outliers that did not meet the data requirements of the ARIMA model. In this case, some unreasonable data were lost automatically before being input into the ARIMA model, which impacts prediction accuracy.

5. Conclusions

This study aimed to use RS time-series data, explore the intrinsic relationship between crop growth and yield formation at different fertility stages, and construct a high-precision winter wheat yield estimation model applicable to short time-series data.

The ten optimal VIs, consisting of MTCI(RE2), EVI, REP, MTCI(RE1), RECI(RE1), NDVI(RE1), NDVI(RE3), NDVI(RE2), NDVI, and MSAVI, which were suitable for yield estimation at six key phenological phases, were obtained by the RReliefF algorithm. These selected VIs were not only relevant to yield formation mechanisms, efficiently predicting winter wheat yield, but also effectively reduced redundancy and computational effort.

Furthermore, AOIAHP was proposed to quantify the different effects of winter wheat growth on yield accumulation at different fertility stages. The jointing period had the highest contribution ratio to winter wheat yield accumulation, followed by rising, heading, filling, rejuvenation, and lastly, filling-maturity. The contribution ratios of the six key phenological stages could be used as weights to be combined with the selected VIs to predict winter wheat yields.

Moreover, the BHT-ARIMA model was first applied to predict winter wheat yield with short time-series data. Compared with the BHT-CNN and BHT-ARMA models, the BHT-ARIMA model achieved the best prediction results, with $R^2$ reaching 0.583. The MAE, MAPE, and RMSE of the BHT-ARIMA model were lower than those of the BHT-ARMA and BHT-CNN models.

In this study, the BHT-ARIMA model improved the accuracy of winter wheat yield prediction based on weighted short RS time-series data, and its spatial generalization was proven. To further validate the generalization of the BHT-ARIMA model, in future work, we will validate the model for yield estimation in different regions, at different times, and for different crops to explore the universality and robustness of the model. As for the BHT-ARIMA model itself, we will improve the prediction accuracy of BHT-ARIMA, reducing the impact of data reduction due to the automatic elimination of the unreasonable values. This study provides a fast and accurate guide for crop yield prediction and will be of great value for the processing and applying of short RS time-series data.

**Author Contributions:** Conceptualization, Y.Z. and M.L.; methodology, Q.D. and M.W.; software, M.W.; validation, H.Z. and Y.C.; formal analysis, Q.D. and M.W.; investigation, Q.D., Y.Z. and M.W.; resources, Y.Z.; data curation, Q.D. and H.Z.; writing—original draft preparation, Q.D. and M.W.; writing—review and editing, Y.Z., Q.D. and M.W.; visualization, Q.D.; supervision, Y.Z. and M.L.; project administration, Y.Z. and Q.D.; funding acquisition, Y.Z. All authors have read and agreed to the published version of the manuscript.
**Funding:** This paper was supported by the National Key Research and Development Program of China (2019YFE0125500) and the National Natural Science Foundation of China (41801245).

**Data Availability Statement:** Not applicable.

**Conflicts of Interest:** The authors declare no conflict of interest.

**References**


