Agricultural and Forest Meteorology 356 (2024) 110183

Contents lists available at ScienceDirect

Agricultural and Forest Meteorology

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

Attention mechanism-based deep learning approach for wheat yield estimation and uncertainty analysis from remotely sensed variables

Huiren Tian [a], Pengxin Wang [a,*], Kevin Tansey [b], Jie Wang [a], Wenting Quan [c], Junming Liu [d]

[a] College of Information and Electrical Engineering, China Agricultural University, Beijing 100083, PR China
[b] School of Geography, Geology and the Environment, University of Leicester, Leicester LE1 7RH, UK
[c] Shaanxi Provincial Meteorological Bureau, Xi'an 710014, PR China
[d] College of Land Science and Technology, China Agricultural University, Beijing 100193, PR China

ARTICLE INFO

Keywords:
Feature extraction
Attention mechanism-based multi-level crop network (AMCN)
Uncertainty
Yield estimation

ABSTRACT

Rapid and accurate crop yield estimation is an imperative aspect of agricultural planning that is important for crop management, food security and commodity trading. There are many related factors affecting wheat yield and the relationship between them and the yield is complicated, with nonlinear spatial-temporal characteristics that are difficult to describe accurately with mathematical functions. Deep learning models can fit complex nonlinear functions efficiently and transform input data into high-dimensional features automatically. However, the feature learning process does not produce transparent information. There has been considerable evidence that the ability of attention mechanism for modeling interpretation has been demonstrated in many fields. Therefore, an attention mechanism-based multi-level crop network (AMCN) was proposed to estimate the county level wheat yield based on remote sensing data and meteorological data. To explore the difference in spatio-temporal feature extraction ability under parallel and series structures when combining CNN with LSTM, we designed the AMCN models with two forms of structure, one is a parallel module of LSTM and CNN (AMCN1) and the other is a serial connection module between LSTM and CNN (AMCN2). Our results showed that the AMCN1 model provided an improved estimation accuracy as compared to that of the AMCN2 model. We also found remote sensing data contributed significantly to crop yield estimation mainly at the late growth stages, meteorological data provided additional information mainly at the early growth stage. We assessed the estimated uncertainty using Monte Carlo dropout, and the results indicated that the uncertainty level decreased gradually as the growth stages proceeded. In addition, extreme events such as drought and uneven distribution characteristics of the samples were associated with much higher estimated uncertainties. The study highlighted that the proposed model provided more accurate yield estimations by taking advantage of multi-level crop networks while considering the uncertainty involved in model estimations.

1. Introduction

Wheat is one of China's most important cereal crops, accounting for approximately 18% of the global wheat production (Cao et al., 2020; Li et al., 2021). As the human population size expands, so does the demand of agricultural production. Increasing wheat production in order to supply the projected population increase with wheat. In this challenging agricultural environment, it is imperative that wheat growth status is monitored in various geographical locations and environmental circumstances (Weiss et al., 2020). Therefore, the accuracy and timely communication of information about wheat development and production is essential to ensure national food security and help make better decisions.

Remote sensing has the advantage of providing repeated information without destroying the crop, which can be utilized to provide useful information for monitoring crop growth (Franch et al., 2019; Zhang et al., 2021). A wide range of vegetation indices (VIs) from remote sensing data have been used to monitor crop growth and estimate yields (Becker-Reshef et al., 2010; Holzman et al., 2014; Guan et al., 2017; Kamir et al., 2020). In addition, there is no doubt that environmental factors play a significant role in crop growth. Extreme weather events (e.g. drought and frost) result in a range of physiological impacts on wheat and significant yield reductions. Therefore, climatic variables can provide necessary information about wheat yield, which can be used as

* Corresponding author at: China Agricultural University, P. O. Box 116, East Campus, Qinghua East Road No. 17, Haidian, Beijing 100083, PR China.
E-mail address: wangpx@cau.edu.cn (P. Wang).

https://doi.org/10.1016/j.agrformet.2024.110183
Received 15 May 2022; Received in revised form 21 February 2023; Accepted 29 July 2024
Available online 2 August 2024
0168-1923/© 2024 Elsevier B.V. All rights are reserved, including those for text and data mining, AI training, and similar technologies.

features to estimate wheat yield. However, only using remote sensing data or meteorological data for estimating yield not be sufficient to estimate yield. Prior studies have shown yield estimation models that incorporate both meteorological and satellite data perform better than models that draw from just one type of information (Balaghi et al., 2008; Newlands et al., 2014; Franch et al., 2015). Recognizing that multi-source data (remote sensing data and meteorological data) improve yield estimates by providing complementary information, this study will concentrate on the use of vegetation temperature condition index (VTCI), leaf area index (LAI), temperature, and precipitation at main growth stages to estimate wheat yield.

A number of studies have been conducted on yield estimation based on remote sensing data and meteorological data in recent decades. In traditional yield estimation models, empirical statistical models and process-based crop simulation models have been used to estimate crop yield. The main feature of process-based crop models is to simulate crop growth dynamics on a daily basis at a field scale and provide a deep understanding of the physiological process (Tao et al., 2009). This type of model, however, requires high-quality and detailed inputs, including soil properties, crop varieties, meteorological data, and crop management strategies. At the regional scale, it is difficult to fully obtain the parameters needed by these models, so process-based crop models are limited in how widely they can be applied. (Lobell, 2013; Aghighi et al., 2018; Filippi et al., 2019). Also, considering the complexity and nonlinearity of crop growth and grain production processes, which are influenced by environmental interactions (Wolanin et al., 2020). Thus, linear regression models cannot effectively handle non-linear and complex relationships between the predictors and the response variables (Khanal et al., 2018).

Machine learning (ML) regression models, as an immediate successor to statistical methods, are shown to outperform the traditional statistical regression methods. Recently, deep learning (DL), which is a subfield of machine learning characterized by multiple stacked layers that transforms the raw input data into higher and more abstract representations, has gained considerable attention in earth observation and geosciences (LeCun et al., 2015; Reichstein et al., 2019; Sagan et al., 2021). DL methods have superior performance due to the ability to automatically learn and extract hierarchical features from the data and have been applied for estimating crop yield (You et al., 2017; Jiang et al., 2019; Schwalbert et al., 2020; Zhang et al., 2022). Convolutional Neural Networks (CNNs) as the most representative supervised DL model, have outperformed most algorithms (Zhang et al., 2016). You et al. (2017) proposed a 2D histogram CNN model for soybean yield prediction for 11 states in the U.S. by using MODIS products of surface reflectance, land surface temperature and land cover from 2003 to 2015 that account for 75% soybean yield prediction. Based on the work by You et al. (2017), Terliksiz and Altýlar (2019) proposed a 3D CNN for soybean yield prediction using the spatial, temporal and spectral dimensions of remote sensing images. The results showed that the yield prediction accuracy of 3D CNN is significantly better than that of 2D Histogram CNN. The CNN model can better deal with patterns across space, but it cannot fully extract the crop phenological characteristics reflected by the time-series remote sensing data. The Long Short-Term Memory (LSTM) neural network was developed by replacing the unit of the ordinary Recurrent Neural Network (RNN) with a special unit known as a memory cell, which enhances the long-term memory capacity of the neural networks. Therefore, LSTM has become an important model for analyzing changes in time series data. Based on the LSTM models, Jiang et al. (2019) used meteorological data and remote sensing data with the corn growth stages as the time series to estimate the county level corn yield. In the analysis, it was found that the LSTM neural network model could identify implicit relationships in the sequence data and achieve accurate county level crop yield estimation as well as provide robust yield estimation under extreme weather conditions.

There is no doubt that CNNs are typically used for capturing the spatial dependencies of data, while LSTMs are used to facilitate the process of capturing time dependencies using sequence data. Of note is that, most of the previous studies only employed CNN or LSTM and cannot fully consider the advantages of CNN and LSTM simultaneously (Sun et al., 2019). In addition, deep learning models have the "black box" property and tend to sacrifice interpretability and transparency for estimation performance, understanding why a model makes the estimations can help build trust (Reichstein et al., 2019; Burke et al., 2021). It is essential to develop a data-driven model to describe how information is learned through deep learning by measuring how input data contributes to the final output. As a result of the limitations outlined above, we focus on establishing a framework that not only takes advantage of both CNN and LSTM, which benefit from DL in terms of improvements in the model performance, but we specifically aim to visualize and interpret the underlying drivers of yield estimation by introducing into attention mechanism, hence placing confidence in the model. Moreover, most DL models mainly focus on providing more accurate estimation yield results without providing any information about uncertainty, which is essential for calculating confidence intervals in the estimation and helping farmers to make corresponding measures (Ma et al., 2021). Therefore, we aim to quantitatively evaluate the uncertainty of the yield estimation model through Monte Carlo dropout.

An attention mechanism-based multi-level crop network (AMCN) model has been developed as a tool for estimating wheat yield at the county level by combining remote sensing data and meteorological data in this study. The proposed deep learning framework was designed to parallel and series architectures, called AMCN1 and AMCN2 respectively, for refining the extraction of temporal and spatial features as well as improving understanding of the effects of input variables on crop yield. This study aims to address the following questions: (1) How does AMCN1 perform in terms of estimating wheat yields at the county level as compared to AMCN2? (2) How does the proposed model perform in terms of robustness and adaptability at the sampling sites level? (3) How does feature learning work in deep learning methods? (4) What is the most effective way to obtain the estimation uncertainty of the proposed model?

2. Materials and methods

2.1. Study area

Research is being conducted in the Guanzhong Plain, about 400 km long from east to west and 80 km wide from north to south, lies in the central part of Shaanxi Province, PR China (106°22′-110°24′E, 33°57′-35°39′N) (Fig. 1). The climate is a continental monsoon climate with four distinct seasons, with cold and dry winters, scarce precipitation and frequent spring droughts in spring, warm and rainy summers, and large temperature variations in autumn, along with the second-highest precipitation after summer among the four seasons. Precipitation occurs on average between 500 and 700 mm a year and temperatures usually range between 6 and 13 °C. The Guanzhong Plain is characterized by cultivated land, forest land, water bodies, construction land, and bare land. It is estimated that 70% of the land in the region is cultivated, primarily in the form of grain fields, orchards and vegetable fields. Winter wheat is the main crop grown in grain fields before summer. Sowing winter wheat in early October and harvesting it in early June are typical processes. Winter wheat's green-up stage occurs between early March and mid-March, the jointing stage occurs between late March and mid-April, the heading-filling stage occurs between late April and early May, and the milk maturity stage occurs between mid-May and late May. The plain is characterized by a wide variation in precipitation and irrigation. The west has abundant water resources, while the north and northeast suffer from low precipitation. Additionally, topography, geomorphology, as well as other factors contribute to the frequency of droughts in this region. Currently, drought has become the main constraint to an increase in agricultural production and income in the plain (Xie et al., 2017). As a result, it is very critical to monitor the

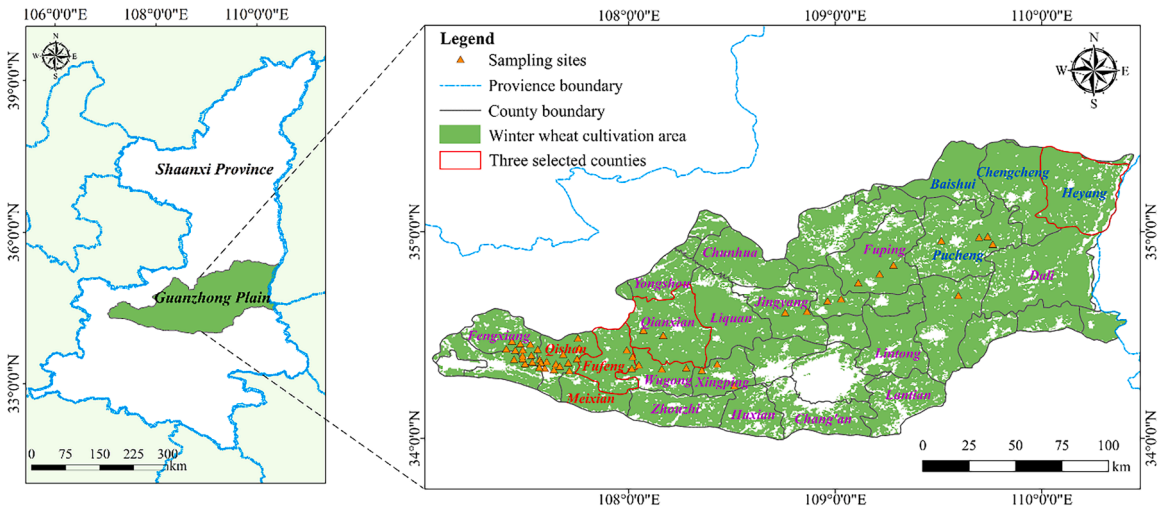

Fig. 1. Spatial distribution of wheat cultivation areas and sampling sites in the Guanzhong Plain. County names in red denote high-yield counties, purple denote medium-yield counties, and blue denote low-yield counties.

growth of winter wheat during the main growth period in the study area.

2.2. Datasets and preprocessing

2.2.1. Remote sensing data

During the period from 2007 to 2020, we collected Moderate Resolution Imaging Spectroradiometer (MODIS) Land Surface Temperature (LST) data (MYD11A1) with a spatial resolution of 1 km, surface reflectance data (MYD09GA) with a spatial resolution of 500 m, and LAI data (MCD15A3H) with a spatial resolution of 500 m, which were obtained from NASA's Earth Observing System Data and Information System website (http://reverb.echo.nasa.gov/reverb/). MODIS data products were converted into Lambert azimuthal projection using the MODIS Reprojection Tools (MRT) developed by NASA. In particular, the VTCI was calculated using the MYD09GA and MYD11A1 products. The 10-day interval LST and Normalized difference vegetation index (NDVI) products were generated for the study years using the maximum value composite (MVC) approach at the different growth stages. The minimum value compositing approach is used to produce the multiyear minimum value composite LST products using each year's maximum value composite LST products at a 10-day interval, which is referred to as the multiyear maximum-minimum value composite LST products (Sun et al., 2008). As a result, VTCI time series data at intervals of 10 days were generated based on formulas (1)-(3) (Wang et al., 2001; Sun et al., 2008; Wan et al., 2004):

$$VTCI = \frac{LST_{\max}(NDVI_i) - LST(NDVI_i)}{LST_{\max}(NDVI_i) - LST_{\min}(NDVI_i)} \quad (1)$$

$$LST_{\max}(NDVI_i) = a + bNDVI_i \quad (2)$$

$$LST_{\min}(NDVI_i) = a' + b'NDVI_i \quad (3)$$

where $LST(NDVI_i)$ is the land surface temperature of one pixel whose NDVI value is $NDVI_i$, $LST_{\max}(NDVI_i)$ and $LST_{\min}(NDVI_i)$ are the warm and cold edges, i.e., the maximum and minimum LST values of the pixels that have the same NDVI value in a study region, respectively, and coefficients a, b, a' and b' are determined by the scatter plots of the LST and NDVI. Following the division of winter wheat into its main growth stages, the average value of the multi-day VTCI contained within each growth stage is taken pixel by pixel as the 10-day VTCI values corresponding to each growth stage.

The time series LAI data generated with the MCD15A3H product have a 4-day temporal resolution. It exhibits irregular fluctuations due to the influence of clouds. This influence obscures the seasonal characteristics of crop growth that are reflected in the time series LAI data (Yuan et al., 2011). In order to address this issue, the upper envelope Savitzky-Golay filtering was employed to smooth the LAI product on a per-pixel basis (Savitzky and Golay, 1964; Xun et al., 2019). The variation in LAI curves was stable following filtering, which is commensurate with the actual growth laws of winter wheat. As demonstrated by Tian et al. (2020), during Julian day 61-161 in 2017, the effect of filtering on a survey site was illustrated. We defined maximum values of LAI over a 10-day interval as LAI values at 10-day intervals, and maximum values of LAI included in a growing period as LAI values in that growing period. Due to the different spatial resolutions of VTCI and LAI, resampling VTCI to 500 m spatial resolution was performed according to the nearest neighbor method.

2.2.2. Meteorological data

Daily precipitation and temperature data were obtained by the meteorological stations in 22 counties in the study period, which extends from March to May of 2007-2020. On the basis of the daily precipitation and temperature data, average precipitation and average temperature for each of the four growth stages in each county were determined to keep the same format as the remotely sensed variables. With the help of ArcGIS software, the station-based meteorological data were interpolated to sampling site-based data using the Kriging method. Interpolated meteorological data have a spatial resolution of 500 m.

2.2.3. Yield data

The Shaanxi Rural Yearbooks were used to collect the county level winter wheat yield data for the study area between 2007 and 2020. Our model and analysis samples were selected from counties that had winter wheat yields exceeding 2000 kg/ha and records spanning over 10 years to ensure the reliability of the data. Moreover, winter wheat planting areas are sparse and scattered in the north, southwest, and urban area of the Guanzhong Plain (Fig. 1), and some of them are located in an area of transition between the plain and its surroundings. Accordingly, yield data from the counties in these areas are not included in this study. Therefore, after filtering, 22 counties were available.

Field experiments were conducted at 50 sampling sites for the purpose of measuring the yield at the sampling site level (Fig. 1) during the main growth stages in 2021. At each sampling site, 4 individuals selected 3 sampling squares, each having an area of 60 cm × 60 cm. In total, twelve 60 cm × 60 cm sampling squares from each sampling site were collected as indicated in Fig. 2. Following is a summary of the detailed descriptions of measured yield data for sampling sites. First, the

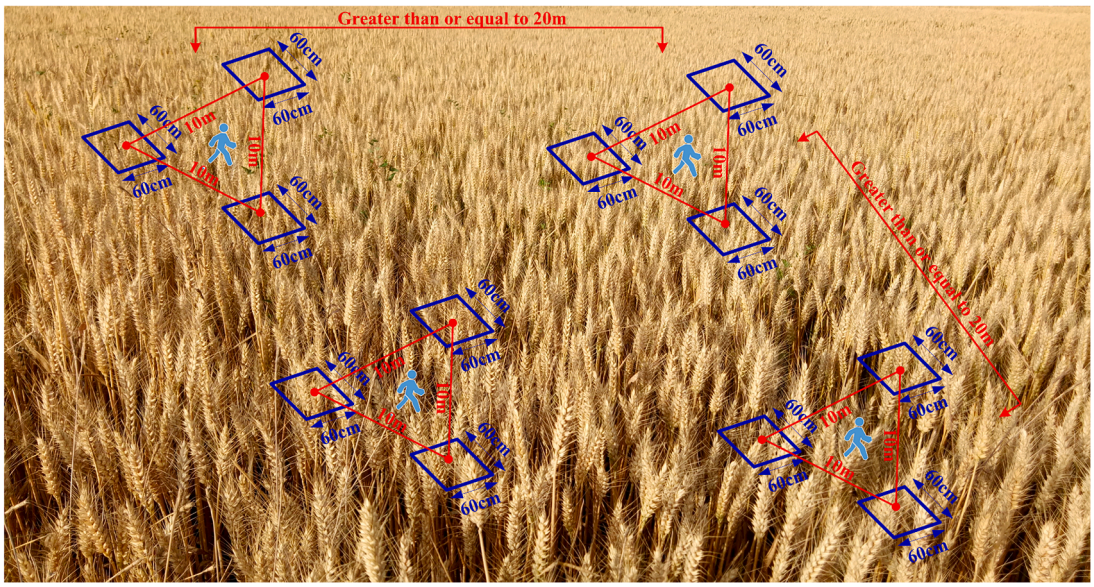

Fig. 2. Diagram of distribution of personnel and sampling square.

number of winter wheat plants with spikes within each sampling square was counted. Then, 10 spikes were randomly selected from each sampling square as the samples. Last, winter wheat spikes of the samples were threshed and dried for 24 hours, and the wheat yield at the sampling site level was calculated based on the weight of grain, the area of the sampling squares and the number of samples within the sampling squares.

2.2.4. Cropland data

Given the length of time that yield data has been collected, it is difficult to determine the classification map for winter wheat annually. It is estimated that the area planted with winter wheat in the Guanzhong Plain has accounted for more than 90% of the total planting area. In accordance with the planting statistics for winter wheat from the National Bureau of Statistics (http://www.stats.gov.cn/), it has been observed that the area sown with winter wheat in China has remained relatively stable over the last 10 years with relatively few fluctuations. In addition, some studies have demonstrated that cultivated land data can serve as a mask for regional yield estimation (Shao et al., 2015; Wang et al., 2020). Accordingly, the MODIS land cover type product (MCD12Q1, MODIS/Terra + Aqua Land Cover Type Yearly L3 Global 500 m SIN Grid) was chosen as the crop mask for identifying winter wheat planting areas. The classification scheme adopted in this study is that of International Geosphere-Biosphere Programme (IGBP) classification. Given that winter wheat in the study area is the dominant crop type, land cover type 14 (farmland/natural vegetation) corresponds roughly to the areas under cultivation for winter wheat. To obtain a distribution map of the winter wheat cultivation areas at the county level of the plain, the extracted cultivation areas for winter wheat are overlaid with the county level vector boundaries of the plain.

2.2.5. Data preprocessing

In accordance with the map of the winter wheat cultivation areas in the plain, the average VTCI and LAI values for pixels in the planting area within each county are taken as the VTCI and LAI values for that county. The datasets for this study are made up of remote sensing data, meteorological data and yield data (Table 1). Among them, remote sensing data (VTCI and LAI) and meteorological data (average precipitation and average temperature) during the wheat main growth stages are fed into the model as sequential vector inputs. Winter wheat yields are used as target vectors. The county level yield data were used for deep learning model training and testing, as well as the development of winter wheat yield estimation models, and the sampling site level yield data were used to validate the generalization performance of the constructed winter wheat yield estimation model. As the processed data are different in dimensions and dimensional units, this will have an adverse effect on the results of data analysis. Min-max normalization was used to eliminate dimensional interaction between variables, with the exception of VTCI (VTCI already falls within the range of 0-1), which also accelerated model convergence.

Table 1
Detailed information of all the datasets in this study.

Data Type	Data Source	Full Name	Spatial Resolution (m)	Temporal Resolution (d)
Remote sensing data	Aqua-MODIS MYD09GA	Surface reflectance	500	1
	Aqua-MODIS MYD11A1	Land surface temperature	1000	1
	Terra+Aqua-MODIS MCD15A3H	Leaf area index	500	4
Cropland data	Terra+Aqua-MODIS MCD12Q1	Land cover type	500	365
Meteorological data	Shaanxi Provincial Meteorological Bureau	Temperature Precipitation	500	1
County level yield data	Shaanxi Rural Yearbooks	/	/	365
Sampling site level yield data	Field experiments	/	/	365

2.3. Methods

2.3.1. Fundamentals of CNN, LSTM and the attention mechanism

A conventional CNN model consists of an input layer, convolutional layers, pooling layers, fully connected layers, and an output layer (LeCun et al., 1998). Specifically, the convolutional layer is responsible for extracting features, and the convolution kernel is responsible for integrating the features across channels to reduce the parameters of the convolution kernels. The downsampling of the convolution output occurs on the pooling layer to preserve the strong features, eliminate the

weak features, and reduce the number of parameters to avoid over-fitting. The four characteristics behind the CNN structure are local connections, shared weights, pooling, and the use of many layers (LeCun et al., 2015). These characteristics reduce the number of network parameters, simplify network parameter selection, and assist with regularization.

LSTM is a time-series neural network derived from a recurrent neural network that effectively addresses the problem of gradient disappearance. By introducing a gate function, it is possible to control the transmission of historical information and excavate the time-series variation laws of relatively long intervals and delays (Hochreiter and Schmidhuber, 1997). There are four interactive units in an LSTM cell, including a memory cell, an input gate, a forget gate, and an output gate. The memory unit is the heart of the network and is responsible for the transmission of information. The input gate controls how much of the network's input is stored in the unit state at any given time. Based on the input information and the memory unit, the output gate determines the content to be output. The forget gate determines the amount of the previous cell state that is saved to the current state of a cell.

Experts and scholars put forward the attention mechanism based on the research into human attention, which is key to the efficient distribution of resources during the information processing process (Chaudhari et al., 2020). The attention mechanism gives enough attention to key information by means of probability allocation and achieves the extraction of important features, thereby optimizing the model and making more accurate judgments.

2.3.2. Designed deep neural network

According to the above basic principles, we designed an attention mechanism-based multi-level crop network (AMCN) model in this study that used multiple layers to hierarchically learn features from remote sensing data and meteorological data. In the AMCN model, two types of structures are included.

(1) The parallel module of LSTM and CNN (AMCN1)

This model is based on the standard CNN module that serves as a dimension reduction layer for coarse-grained features extraction, adding parallel the LSTM branches to extract the time dependence of effective features, and can avoid memory loss and gradient dispersion caused by too long step length to a certain extent. Followed by the attention mechanism layer, which processes the difference in importance of short sequence features that are ignored by the existing structure to extract important fine-grained features. Through the AMCN1 model, the fusion of coarse and fine-grained features is realized, and the time series data is fully characterized. The developed AMCN1 model is depicted in the upper branch of Fig.3. Detailed descriptions are provided below for each layer. The CNN module is composed of multiple layers of one-dimensional networks stacked, including a convolutional layer, a max-pooling layer and a dense layer. Specifically, the CNN module begins with a one-dimensional convolutional (Conv1D) layer with a kernel size of 3 and a filter of 32. Next, a maximum pooling layer is applied with 2 kernels. It should be noted that the strides are set to 1, and the padding forms are set to "SAME" in the convolutional layer and pooling layer. The number of units is set to 32 in the dense layer. In the LSTM module, the number of hidden units in the LSTM cell and the dense layer is configured to 64 and 32, respectively. The features extracted from the output of the CNN module and LSTM module are then connected to a concatenate layer, followed by an attention mechanism layer. Following the output of the attention mechanism layer, the multi-dimensional array is flattened into a one-dimensional array, which is then sent into a dropout layer with a 0.3 dropout probability. To a certain extent, dropout is effective in alleviating the occurrence of over-fitting and achieving regularization. Finally, a dense layer is used to output results.

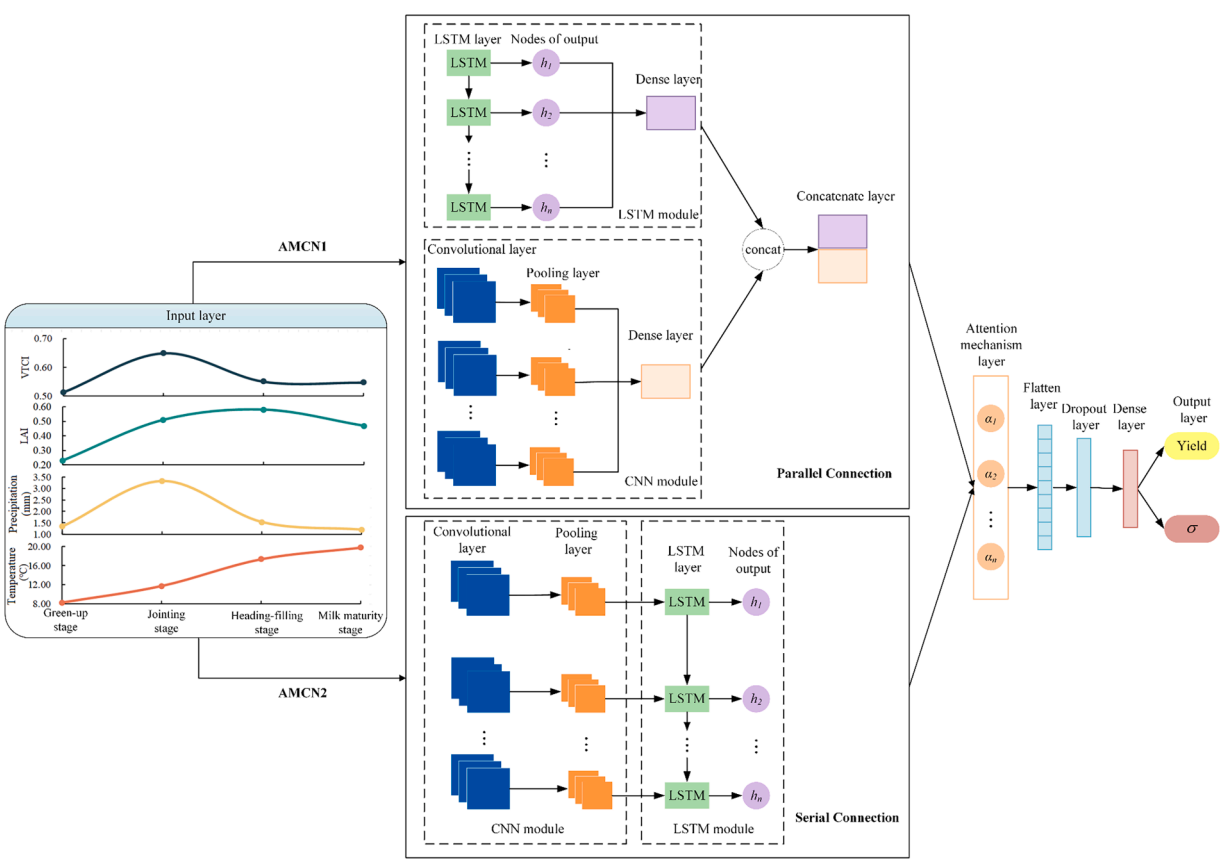

Fig. 3. The architecture of the AMCN1 model and AMCN2 model.

(2) The serial connection module between LSTM and CNN (AMCN2)

The developed AMCN2 model is shown in the lower branch of Fig.3, starts with the standard CNN module, which uses convolution operations to produce the feature map through the activation function and obtain the local correlation between the data. The CNN module follows a similar architecture to CNN in the AMCN1 model. After the standard CNN module, the extracted features are fed into an LSTM layer in which 64 hidden units to learn the time dependence relationship between the data, and a dropout layer with a 0.3 probability between the CNN module and LSTM module. Then the attention mechanism layer is subsequently added after the LSTM layer to increase the influence of significant time series features and minimize the influence of non-important features in the model, which effectively corrects for the model's failure to distinguish between the importance of features in time series. Afterwards, the output is flattened to a long vector, which is then sent into a dropout layer with a 0.3 dropout probability. Finally, the output result is obtained through the dense layer.

In this study, the AMCN models were implemented on Tensorflow 2.2 with Python version 3.8. Several experiments have been conducted to determine the optimum parameters for AMCN models. Initially, the learning rate was set to 0.0001, and the number of training iterations was set to 200 epochs with a batch size of 16. The Rectified Linear Unit (ReLU) was used as the activation function and Adam was selected as the optimizer for updating the trainable parameters after each epoch.

2.3.3. Uncertainty analysis

The AMCN model used in this study can not only make yield estimation but also provide a quantitative assessment of the estimation uncertainty. Due to the confidence interval of estimating yield in advance can help farmers to take corresponding measures and policy makers to make corresponding policies, which can maximize profits and reduce losses (Lobell et al., 2009). A Monte Carlo dropout method was employed for quantifying the estimation of uncertainty, which could be viewed as a Bayesian approximation (Gal and Ghahramani, 2016). The specific implementation method is: the model not only uses dropout during training, but also turns on dropout during testing, and performs multiple forward operations to obtain a set of estimated values, which is the result of multiple sampling from the model approximate posterior, then calculate the standard deviation and the mean of these outputs to get the uncertainty of the model and the unbiased estimation of the yield, respectively. The confidence interval measures the degree of uncertainty surrounding the effect estimate, which is composed of a lower limit and an upper limit. The width of the confidence interval indicates the level of the precision of the effect estimate. A narrower confidence interval indicates a more precise estimate, and vice versa, a wider interval indicates a less precise estimate. The confidence interval can be calculated using equation (4):

$$CI = \overline{X} \pm Z \times \frac{\sigma}{\sqrt{N}} \tag{4}$$

where Z is the Z-value for the chosen confidence level, $\overline{X}$ is the sample mean, σ is the standard deviation, and N is the sample size. We chose a 95% confidence level for this study.

2.3.4. Experiment settings and model evaluation

A total of 80% of the dataset was used for training, and 20% for testing. Through five-fold cross-validation with the training dataset, the critical hyper-parameters for each model were determined based on the highest coefficient of determination (R^2), the lowest root mean square error (RMSE) and the lowest mean absolute percentage error (MAPE). Our research questions were addressed by designing different sets of experiments, as follows: the first set of experiments compared the performance of the AMCN1 model with the AMCN2 model, to determine the most appropriate yield estimation model at the county level, thereby answering the question "How does AMCN1 perform in comparison with AMCN2?". Based on the above the most appropriate yield estimation model at the county level, the second set of experiments was conducted at sampling sites under different disaster situations, which answered the "How does the proposed model perform in terms of robustness and adaptability at the sampling sites level?". The third set of experiments was to evaluate what input variables were important for the wheat yield estimation by using Leave-One-Covariate-Out (LOCO), which is a popular approach for determining the importance of features (Hooker and Mentch, 2019). Finally, we further analyzed the estimated uncertainty based on two scenarios: (1) timeline with the different growth stages and (2) datasets with different yield levels. Specifically, we divided counties into three categories based on the measured yields (Fig. 1), among which the counties with yields greater than 5500 kg/ha are high-yield counties, the counties with yields less than 3500 kg/ha are low-yield counties and the counties with yields between 3500 and 5500 kg/ha are medium-yield counties.

In order to assess the performance of the models in estimating yield, R^2, RMSE and MAPE between estimated yield and measured yield in the test set were considered. In addition, the optimal model was identified by the highest R^2 and the lowest RMSE of 100 runs in the test set. An evaluation of the model's generalization and scalability was also conducted using a five-fold cross-validation.

3. Results and analysis

3.1. Performance of model at the county level

In the following section, we first compared the estimation performance of AMCN models with two different connection methods, and the scatter plots were presented in Fig. 4. Compared to the AMCN2 model, the AMCN1 model corresponded to an increase in R^2 of 0.03 and decreases in RMSE and MAPE of 22.97 kg/ha and 0.23%, respectively. The AMCN1 model can explain 68% of the variance in the official yield records, which performed better than that of the AMCN2 model. Focusing on the scatter density for official yield records less than 3500 kg/ha (low-yield counties), we find that the AMCN1 model showed improvement in low-yield counties compared with the AMCN2 model, indicating that the AMCN1 model was able to capture the important features of low-yield samples. Besides, it was notable that the AMCN models underestimated the yield in counties with high yields. There were relatively fewer high-yield samples in the unbalanced training set, which resulted in modeling challenges due to the unbalanced distribution of sample samples (Ma et al., 2021; Leroux et al., 2019).

We then presented the estimation performance of two models in Fig. 5 by using the five-fold cross-validated method to further determine the best yield estimation model at the county level, which answered the first research question "How does AMCN1 perform compared with AMCN2?". Over five testing datasets, the developed AMCN1 model again demonstrated the greatest agreement. The AMCN1 model outperformed the AMCN2 model in terms of estimation accuracy, with an average R^2 greater than 0.60 and an average RMSE below 550 kg/ha and stable performance in the testing dataset for all folds. Such stable performance can be attributed to the parallel structure of the model. This enables it to simultaneously focus on the spatial and temporal features of the input sequence, preventing the loss of critical features. In the following analysis, we utilized the AMCN1 model to conduct experiments and present results in light of the results of the AMCN1 model performance provided above.

3.2. Sampling sites scale yield estimation and validation

We further conducted the second group experiment at sampling sites under different disaster situations in 2021 by applying the AMCN1 model, with the performance shown in Fig. 6. The results clearly showed that in different disaster scenarios (i.e., slight lodging, severe lodging,

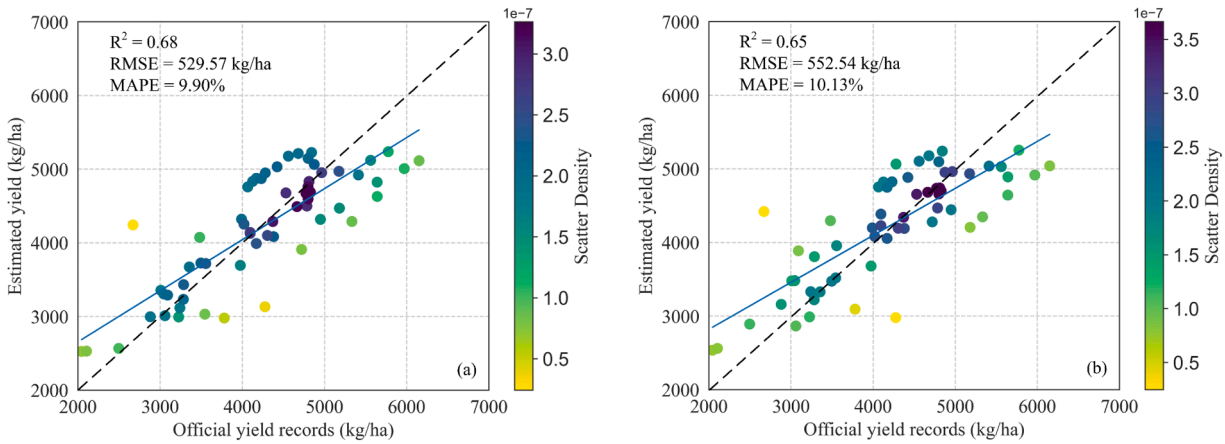

Fig. 4. Scatter plots of yield estimation in the testing dataset at the county level by different models (a) AMCN1 model and (b) AMCN2 model.

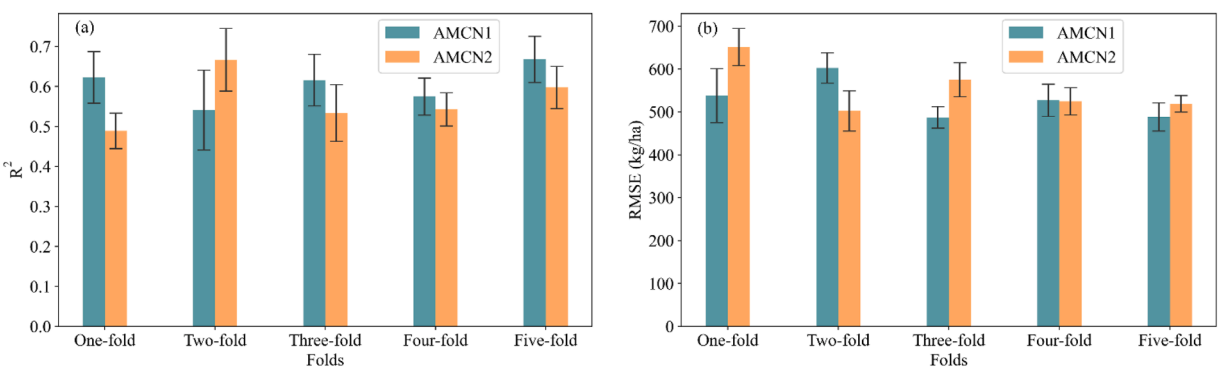

Fig. 5. The performance of the AMCN1 model and AMCN2 model in the testing dataset by using five-fold cross-validated method. The error bars are one standard deviation of estimated R^2 and RMSE from 10 ensembles by randomly dividing training and testing datasets.

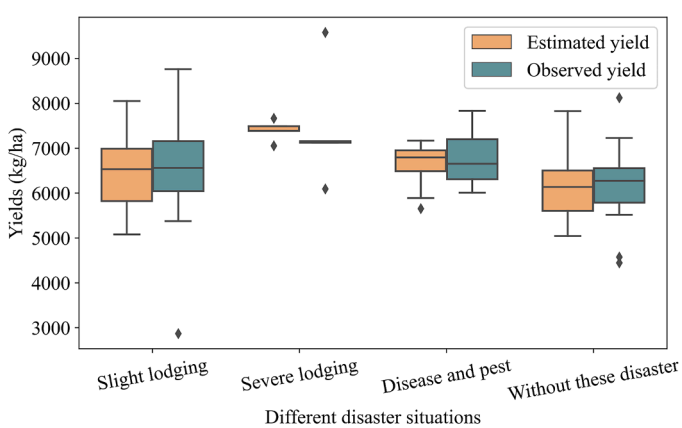

Fig. 6. Comparisons between estimated yield and observed yield of the AMCN1 model at sampling sites with different disaster situations.

disease and pest, without these disasters) the AMCN1 model yield estimation performance varied widely. When wheat is not affected by these disasters, the estimated results of the model outperformed the others suffering from different degrees of disaster. In particular, the model performed the worst at severely lodged sampling sites. It is likely due to the reason that disaster situations are not incorporated into the training record, the AMCN1 model is constrained by the information about disasters. On the other hand, minimum and maximum yields (6094 kg/ha and 9581 kg/ha) of observed yields differed greatly in Fig. 6, which leads to challenges in modeling their big variations. Therefore, the out-of-sample estimations for severe lodging did not have a good performance. In ground experiments, we found the damage by stripe rust and pest occurred at the sampling sites were more serious, which resulted in similar performance with severe lodging sampling sites. For sampling sites with a small amount of lodging, the AMCN1 model can achieve estimation accuracy with an R^2 of 0.5. That is, the accuracy of the model is closely related to the degree of damage suffered by the sampling sites. The values of RMSE and R^2 were also calculated (Fig. 7) for different disaster situations for assessment more intuitively of model accuracy. The estimation accuracy with R^2 of 0.62 and RMSE of 566.64 kg/ha in the sampling sites where these disasters did not occur, showing that the

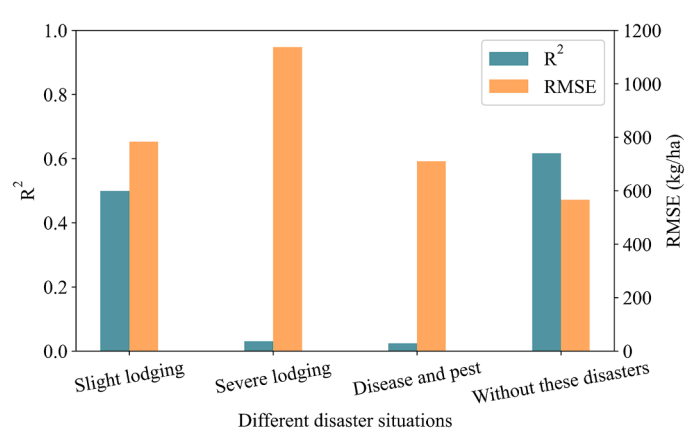

Fig. 7. Performance of the AMCN1 model at sampling sites with different disaster situations.

AMCN1 model exhibited reasonable generalization and scalability among the sampling sites, thus answering the second research question.

3.3. Feature importance analysis

As shown in Fig. 8, we explored the relative contributions of each variable in estimating yield for the AMCN1 model. We determined the relative feature importance by using the change in the RMSE resulting from dropping a particular type of feature value from the feature set. All relative importance values were normalized so that the sum equaled 100%. The results indicated that LAI was the most influential variable, and in particular, that it had a significant impact on yield at the heading-filling stage and milk maturity stage, since LAI acted as a comprehensive measurement of crop growth and was recognized as a valuable tool for estimating crop yield. Winter wheat grows vigorously at the heading-filling stage, when the leaves and stems gradually enlarge and the photosynthesis also gradually improves. The accumulation of photosynthetic products mainly determines the grain weight and has a very important impact on the final yield (Xie et al., 2017). A high ranking was also given to VTCI at the jointing stage, indicating that they are capable of explaining crop yield variations well. The reason was that water stress has a more significant effect on winter wheat at the jointing stage, and water deficiency in soil will significantly affect the accumulation rate of dry matter in roots, stems, leaves and plants, thereby affecting winter wheat growth and yield. It was followed by the temperature at the green-up stage, since it can provide unique information about weather properties that influence crop growth.

Furthermore, as the temporal progression, the model performance was shown in Fig. 9 by using the LOCO method, i.e. delete VTCI, delete LAI, delete precipitation and delete temperature. Focusing on the deleted meteorological data (i.e. precipitation and temperature) progressed over time, there is a general pattern that occurs with the progression of the growth stages, the increase in RMSE at the green-up stage (early stage) was higher than at other stages (late stages) when deleted meteorological data, while the deleted remote sensing data has exactly the reversed pattern. It is clear from this result that the model can learn spatial patterns of the meteorological variables, thereby capturing some spatial yield patterns at an early growth stage. However, the LAI at an early stage is homogeneously low in space and provides insufficient information about the spatial pattern, and less profound signals of crop growth could be captured by the model at an early growth stage. As the growth stages progress, it becomes increasingly apparent that strong correlations exist between the remote sensing data and yield, and the model gradually learned temporal patterns of input variables. Since the past state of these variables usually presents future cause-effect relationships, the proposed model that is able to learn patterns based on the sequence in which data are collected can potentially outperform models that provide static views of data (Schwalbert et al., 2020). Information accumulating has a greater impact on wheat yields, which further supports this statement.

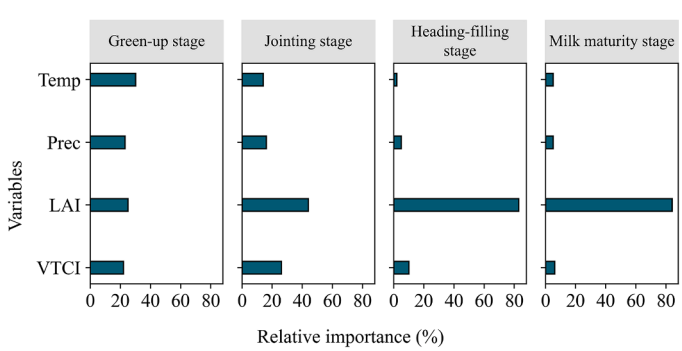

Fig. 8. The relative importance of variables at different growth stages. "Prec" means precipitation and "Temp" means temperature.

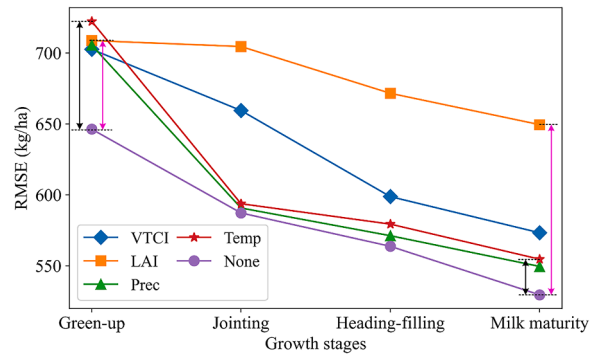

Fig. 9. The performance of the model using the LOCO method as the growth stages progress. "VTCI" means delete VTCI, "LAI" means delete LAI, "Prec" means delete precipitation, "Temp" means delete temperature and "None" means no variables are deleted. The black line indicates the benefits of temperature. The pink line indicates the benefits of LAI.

3.4. Estimated uncertainty analysis

The last group of experiments was conducted, with results shown in Fig. 10. We first analyzed the estimated uncertainty against the growth stages. The results (Fig. 10(a)) indicated that the estimated uncertainty level gradually decreased against the timeline with the growth stages. This was mainly because the model incorporated the cumulative effects of dynamics on crop growth and development that capture more temporal features, resulting in more confidence in its estimation. Therefore, the uncertainty associated with the estimation gradually decreased through the growth stages.

We further analyzed the estimated uncertainty in datasets with different yield levels. The results (Fig. 10(b)) showed that the estimated yield in high yield dataset remained at a relatively higher uncertainty level with about 120 kg/ha compared with medium yield dataset and low yield dataset. It is likely that this result is due to the inherent noise of observation data and uneven characteristic distribution of the sample data. Also, this corresponded to the results obtained in Section 3.1 regarding the lower yield estimation accuracy of the model in high-yield counties. That is, the performance of the model is associated with estimated uncertainty level. The following figure (Fig. 11) provided a more detailed analysis of this regard. We randomly select three counties, which are high-yield county, medium-yield county, and low-yield county, respectively according to the level of yield during 2007-2020 to assess the estimated uncertainty. It was notable that the estimations of high-yield county in most years had the highest uncertainty, and the estimated yield enveloped in a larger confidence interval than medium-yield county and low-yield county. As the confidence interval narrowed, the model had more confidence in its estimation, which implied a higher accuracy of the estimated crop yield. A higher uncertainty pattern occurred in 2013 when had extremely low accumulated precipitation during growth stages compared with other years. During extreme events such as droughts, crop yield estimation models may be adversely affected. Therefore, the uncertainty of the estimated crop yield along with the associated confidence interval offers the farmers and agriculture sectors the opportunity to make informed decisions and plan for the growing season while taking into consideration the risk of crop production failure (Abbaszadeh et al., 2022).

4. Discussion

4.1. Advantages of our framework

In this study, we developed AMCN models with two different structures AMCN1 and AMCN2 for county level wheat yield estimation. Our analysis found that the AMCN1 model performed better than the

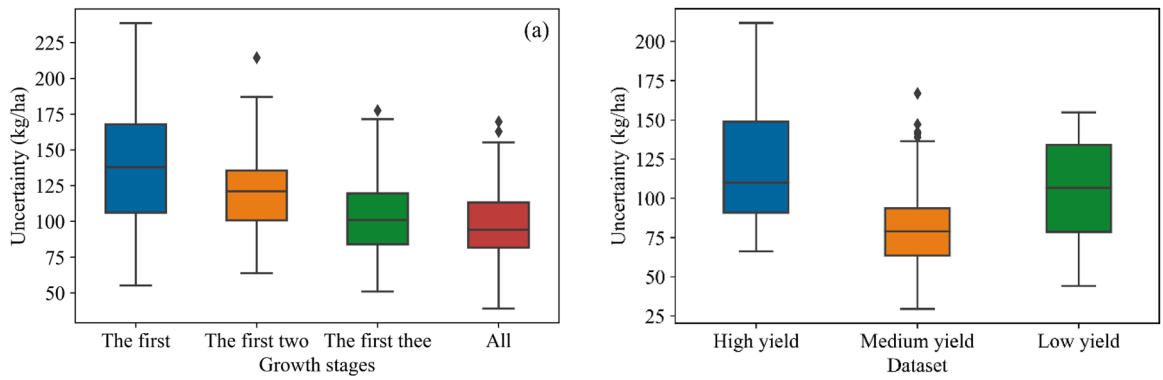

Fig. 10. Box plot of estimated uncertainty by Monte Carlo dropout based on two scenarios (a) timeline with the different growth stages and (b) datasets with different yield levels.

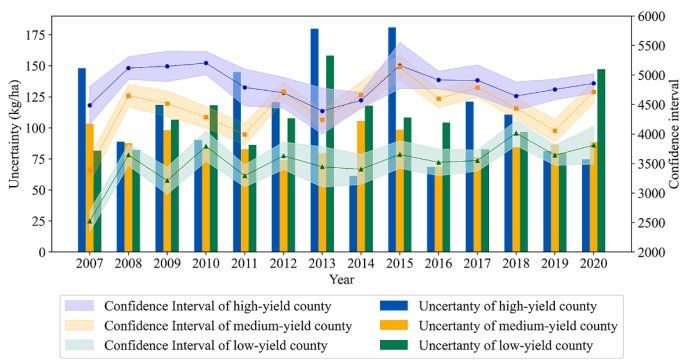

Fig. 11. The estimated uncertainty of different yield level counties during 2007-2020 by Monte Carlo dropout and the estimated yield within the 95% confidence interval. The points indicate the median of the estimated distribution.

AMCN2 model. We could explain that the advantage of the AMCN1 model lies in its parallel structure of the CNN module and LSTM module, both of which extract spatial-temporal features. The back end is the attention mechanism, which extracts the fine-grained features from spatial-temporal features. AMCN2 model extracts features via serial connection between CNN module and LSTM module limited the potential power of the neural network and missed some information. The performance at the sampling site level indicated that the proposed model provided a promising tool for regional- and site- scale crop yield estimation. However, the model has a bad estimation performance at sampling sites that suffered from severe disasters. In view of the challenge of generalizing data-driven models to such disaster scenarios, it is advisable to use a combination of a deep learning model and a process-based crop model, which is able to simulate yield responses under varied environmental conditions and help enhance the generalizability of the deep learning model (Ma et al., 2021). Our yield estimation model was data-driven, therefore, the data quality played a significant role in determining the accuracy of our model. Currently, in the field of action recognition in videos, new network architectures have been developed that are capable of dealing with spatial-temporal variability (He et al., 2019). Compared with our model, the potential of these models for applications in crop yield estimation can be further explored. Moreover, the model structure has been optimized to maximize the interpretation of the feature learning process for novel data-driven models on the basis of improving the model's accuracy.

4.2. Comparing the contributions of input variables to yield

Input feature importance evaluation showed that among the input variables, the LAI at the heading-filling stage and milk maturity stage have a significant contribution to improving the accuracy of crop yield estimates, which is in line with our previous research on this topic (Tian et al., 2020). On the one hand, LAI is closely related to photosynthetic capacity, respiration and transpiration of the crop, and is an important indicator of the growth, development and health of the crop. On the other hand, winter wheat mainly undergoes reproductive growth during the heading-filling stage and milk maturity stages, which mainly determine the grain quality and have a greater correlation with yield than other stages. Then followed by the VTCI at the jointing stage which is a water-sensitive period in wheat growth development, as well as the main stage of root, stem and leaf growth in winter wheat. Therefore, irrigation at the jointing stage results in high grain yields and efficient water use (Xie et al., 2017). At the milk maturity stage, the spike grain structure has already developed and shows a strong tolerance to water deficit, so the impact of water deficit at this stage on the grain yield is relatively small. During the general consistency analysis, it was found that importance evaluation of input variables was an invaluable tool for assessing the reliability of the model and for representing the redistribution of refined and integrated information in both the temporal and spatial dimensions. Also, with the progression of the growth stages, meteorological data contributed more to crop yield estimation at an early stage than at peak and late stages, while remote sensing data has exactly the reversed pattern (Fig. 9), which largely explains the effects of meteorological data are already reflected in crop growth during peak and late stages observed in remote sensing data (Wolanin et al., 2020). In the early growth period, the wheat seedlings had just begun to grow and the growth characteristics were not yet apparent. However, the meteorological data could provide some external characteristics affecting the growth of winter wheat, so it showed a greater contribution. After the jointing stage, the stem nodes and leaves of winter wheat grow quickly, and the growth characteristics gradually become apparent. The model can effectively capture the growth characteristics and their changes over time from the remotely sensed variables, and the impact of these characteristics on winter wheat yield is greater than that of meteorological data on winter wheat yield. At the same time, the important information derived from remote sensing data is accumulated over growth stages through the LSTM cell in the hidden layers. Therefore, the hidden features extracted by the model at peak and late stages drive the model to assign higher attention weights to these features. As Cai et al. (2019) reported that meteorological data and satellite data provide unique and overlapping information for yield prediction, and weather information contributes significantly to wheat yield prediction across the entire growing period. As well as biotic factors beyond what can be captured by satellite VIs and soil properties, there are some factors that can also contribute to explaining more variability of crop yield, and while it is beyond the scope of the current study to investigate these factors, addressing them might be of interest as a valuable future

endeavor if the appropriate data becomes available.

4.3. Uncertainties involved in model estimations

The developed framework provided accurate estimates of the crop yield along with their associated uncertainty. Generally, a significant portion of crop yield uncertainty is caused by the deep neural networks, network parameterization, and observation noise. In this study, a five-fold cross-validated was adopted to optimize the model parameters, thereby reducing uncertainty from parameterization to a certain extent. For the network model structure, the estimated uncertainty was investigated by using MC-dropout. The findings demonstrated that the level of uncertainty gradually decreased as the growth stages progressed. Meanwhile, the unbalanced distribution of the sample data would further raise the estimation's level of uncertainty. Nevertheless, some inherent noise from observation data, such as remote sensing data, meteorological data and measured yield data could still exist throughout the entire growth process. Land cover type product was chosen as the crop mask for identifying winter wheat planting areas in this study. As a result, the uncertainty in generating data for winter wheat growing areas is increased due to mixed pixels (from other crops). The data uncertainty is propagated forward during training, while the model error is propagated backward; the parameters of the deep learning model are randomly selected by MC-dropout to obtain multiple sets of estimates. The uncertainty of the estimation is obtained based on error propagation and ensemble estimation. Therefore, the uncertainty in the data is reflected in the weights of the model parameters during training. In addition, more accurate crop classification maps can significantly improve the quality of dataset. Therefore, future study will focus on improving crop classification accuracy to reduce data uncertainty, resulting in more accurate crop yield estimates. Despite the fact that we do not directly take into account the uncertainty associated with the meteorological data, a higher uncertainty pattern occurred in 2013 when had extremely low accumulated precipitation during growth stages in selected three counties. The median of the estimated yield in the high-yield county is lower than the measured yield, which is consistent with the performance of the underestimated high yield. That is, the median of the estimations along with its associated confidence interval is closely related to the accuracy of yield estimation. Deep learning for quantifying the estimated uncertainty is not a common practice in the literature, with only a small number of studies investigating this issue. For further study, using remote sensing data of high-quality, such as Sentinel-2 and Bayesian models for the integration of outputs of multiple deep neural networks may help reduce the uncertainty, and also provide probabilistic estimates (Zhang et al., 2021; Abbaszadeh et al., 2022).

5. Conclusions

Based on the multi-level crop network, we integrated VTCI, LAI and meteorological data to estimate the county level wheat yield. The results indicated that AMCN1 outperformed AMCN2, and further demonstrated that AMCN1 at sampling sites with better growth conditions displayed reasonable generalizability and stability. By using the LOCO method, we identified VTCI at the jointing stage as well as LAI at the heading-filling stage and milk maturity stage as important variables. Specifically, meteorological data mainly contribute to yield estimation during the early growth stages, while remote sensing data contribute more to yield estimation during the peak-late growth stages. In addition, we also explored the estimated uncertainty and found that gradual temporal progression and more time-series features can reduce the level of uncertainty. Also, extreme drought and uneven distribution characteristics of the training data would increase the estimated uncertainty. This study developed a robust deep learning model to achieve more accurate and reliable wheat yield estimations while accounting for the model's uncertainties, and this model is intended to be applicable to different crops and provide uncertainties involved in yield estimations.

CRediT authorship contribution statement

Huiren Tian: Conceptualization, Methodology, Data curation, Writing – original draft. **Pengxin Wang:** Supervision, Conceptualization, Methodology, Data curation, Resources, Writing – review & editing, Funding acquisition. **Kevin Tansey:** Resources, Writing – review & editing, Funding acquisition. **Jie Wang:** Methodology, Data curation. **Wenting Quan:** Resources, Data curation. **Junming Liu:** Resources, Data curation.

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.

Data availability

The authors do not have permission to share data.

Acknowledgements

This work was supported by the National Natural Science Foundation of China under Grant 42171332. This work was supported by a Royal Society-Newton Mobility Grant (UK).

References

Abbaszadeh, P., Gavahi, K., Alipour, A., Dep, P., Moradkhani, H., 2022. Bayesian multi-modeling of deep neural nets for probabilistic crop yield prediction. Agric. For. Meteorol. 314, 108773 https://doi.org/10.1016/j.agrformet.2021.108773.

Aghighi, H., Azadbakht, M., Ashourloo, D., Shahrabi, S., Radiom, S., 2018. Machine learning regression techniques for the silage maize yield prediction using time-series images of Landsat 8 OLI. IEEE J. Sel. Top. Appl. Earth Obs. Remote Sens. 11 (12), 4563–4577. https://doi.org/10.1109/JSTARS.2018.2823361.

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. https://doi.org/10.1016/j.jag.2006.12.001.

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. https://doi.org/10.1016/j.rse.2010.01.010.

Burke, M., Driscoll, A., Lobel, D.B., Ermon, S., 2021. Using satellite imagery to understand and promote sustainable development. Science 371 (6535), eabe8628. https://doi.org/10.1126/science.abe8628.

Cai, Y., Guan, K., Lobell, D., Potgieter, A., Wang, S., Peng, J., Xu, T., Asseng, S., Zhang, Y., You, L., Peng, B., 2019. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. Agric. For. Meteorol. 274, 144–159. https://doi.org/10.1016/j.agrformet.2019.03.010.

Cao, J., Zhang, Z., Tao, F., Zhang, L., Luo, Y., Han, J., Li, Z., 2020. Identifying the contributions of multi-source data for winter wheat yield prediction in China. Remote Sens 12 (5), 750. https://doi.org/10.3390/rs12050750.

Chaudhari, S., Polatkan, G., Ramanath, R., Mithal, V., 2020. An Attentive Survey of Attention Models. https://doi.org/10.48550/arXiv.1904.02874 arXiv preprint. arXiv: 1904.02874.

Filippi, P., Jones, E.J., Wimalathunge, N.S., Somarathna, P.D.S.N., Pozza, L.E., Ugbaje, S. U., Jephcott, T.G., Paterson, S.E., Whelan, B.M., Bishop, T.F.A., 2019. An approach to forecast grain crop yield using multi-layered, multi-farm data sets and machine learning. Precis. Agric. 20 (5), 1015–1029. https://doi.org/10.1007/s11119-018-09628-4.

Franch, B., Vermote, E.F., Becker-Reshef, I., Claverie, M., Huang, J., Zhang, J., Justice, C., Sobrino, J.A., 2015. Improving the timeliness of winter wheat production forecast in the United States of America, Ukraine and China using MODIS data and NCAR growing degree day information. Remote Sens. Environ. 161, 131–148. https://doi.org/10.1016/J.RSE.2015.02.014.

Franch, B., Vermote, E.F., Skakun, S., Roger, J.C., Becker-Reshef, I., Murphy, E., Justice, C., 2019. Remote sensing based yield monitoring: application to winter wheat in United States and Ukraine. Int. J. Appl. Earth Obs. Geoinf. 76, 112–127. https://doi.org/10.1016/j.jag.2018.11.012.

Gal, Y., Ghahramani, Z., 2016. Dropout as a Bayesian approximation: representing model uncertainty in deep learning. In: International Conference on Machine Learning, pp. 1050–1059. https://doi.org/10.48550/arXiv.1506.02142.

Guan, K., Wu, J., Kimball, J.S., Anderson, M.C., Frolking, S., Li, B., Hain, C.R., Lobell, D. B., 2017. The shared and unique values of optical, fluorescence, thermal and microwave satellite data for estimating large-scale crop yields. Remote Sens. Environ. 199, 333–349. https://doi.org/10.1016/j.rse.2017.06.043.

He, D., Zhou, Z., Gan, C., Li, F., Liu, X., Li, Y., Wen, S., 2019. Stnet: local and global spatial-temporal modeling for action recognition. In: AAAI Conf. Artif. Intell, 33, pp. 8401–8408. https://doi.org/10.1609/aaai.v33i01.33018401.

Hochreiter, S., Schmidhuber, J., 1997. Long short-term memory. Neural Comput. 9 (8), 1735–1780. https://doi.org/10.1162/neco.1997.9.8.1735.

Holzman, M.E., Rivas, R., Piccolo, M.C., 2014. Estimating soil moisture and the relationship with crop yield using surface temperature and vegetation index. Int. J. Appl. Earth Obs. Geoinf. 28, 181–192. https://doi.org/10.1016/j.jag.2013.12.006.

Hooker, G., Mentch, L., 2019. Please Stop Permuting Features: an Explanation and Alternatives. https://doi.org/10.48550/arXiv.1905.03151 arXiv preprint. arXiv: 1905.03151.

Jiang, H., Hu, H., Zhang, R., Xu, J., Xu, J., Huang, J., Wang, S., Ying, Y., Lin, T., 2019. A deep learning approach to conflating heterogeneous geospatial data for corn yield estimation: a case study of the US Corn Belt at the county level. Glob. Change Biol. 1–13. https://doi.org/10.1111/gcb.14885.

Kamir, E., Waldner, F., Hochman, Z., 2020. Estimating wheat yields in Australia using climate records, satellite image time series and machine learning methods. ISPRS J. Photogramm. Remote Sens. 160, 124–135. https://doi.org/10.1016/j.isprsjprs.2019.11.008.

Khanal, S., Fulton, J., Klopfenstein, A., Douridas, N., Shearer, S., 2018. Integration of high resolution remotely sensed data and machine learning techniques for spatial prediction of soil properties and corn yield. Comput. Electron. Agric. 153, 213–225. https://doi.org/10.1016/j.compag.2018.07.016.

LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. Nature 521 (7553), 436–444. https://doi.org/10.1038/nature14539.

LeCun, Y., Bottou, L., 1998. Gradient-based learning applied to document recognition. Proc. IEEE 86 (11), 2278–2324. https://doi.org/10.1109/5.726791.

Leroux, L., Castets, M., Baron, C., Escorihuela, M.J., Bégué, A., Lo, S., 2019. Maize yield estimation in West Africa from crop process-induced combinations of multi-domain remote sensing indices. Eur. J. Agron. 108, 11–26. https://doi.org/10.1016/j.eja.2019.04.007.

Li, L., Wang, B., Feng, P., Wang, H., He, Q., Wang, Y., Liu, L., Li, Y., He, J., Feng, H., Yang, G., Yu, Q., 2021. Crop yield forecasting and associated optimum lead time analysis based on multi-source environmental data across China. Agric. For. Meteorol. 308-309, 108558 https://doi.org/10.1016/j.agrformet.2021.108558.

Lobell, D.B., 2013. The use of satellite data for crop yield gap analysis. Field Crop. Res. 143, 56–64. https://doi.org/10.1016/j.fcr.2012.08.008.

Lobell, D.B., Cassman, K.G., Field, C.B., 2009. Crop yield gaps: their importance, magnitudes, and causes. Annu. Rev. Environ. Resour. 34 (1), 179–204. https://doi.org/10.1146/annurev.environ.041008.093740.

Ma, Y., Zhang, Z., Kang, Y.H., Özdoğan, M., 2021. Corn yield prediction and uncertainty analysis based on remotely sensed variables using a Bayesian neural network approach. Remote Sens. Environ. 259, 112408 https://doi.org/10.1016/j.rse.2021.112408.

Newlands, N.K., Zamar, D.S., Kouadio, L.A., Zhang, Y., Chipanshi, A., Potgieter, A., Toure, S., Hill, H.S.J., 2014. An integrated, probabilistic model for improved seasonal forecasting of agricultural crop yield under environmental uncertainty. Front. Environ. Sci. 2, 17. https://doi.org/10.3389/fenvs.2014.00017.

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., Prabhat, 2019. Deep learning and process understanding for data-driven Earth system science. Nature 566, 195–204. https://doi.org/10.1038/s41586-019-0912-1.

Sagan, V., Maimaitijiang, M., Bhadra, S., Maimaitiyiming, M., Brown, D.R., Sidike, P., Fritschi, F.B., 2021. Field-scale crop yield prediction using multi-temporal WorldView-3 and PlanetScope satellite data and deep learning. ISPRS J. Photogramm. Remote Sens. 174, 265–281. https://doi.org/10.1016/j.isprsjprs.2021.02.008.

Savitzky, A., Golay, M.J.E., 1964. Smoothing and differentiation of data by simplified least squares procedures. Anal. Chem. 36 (8), 1627–1639. https://doi.org/10.1021/ac60214a047.

Schwalbert, R.A., Amado, T., Corassa, G., Pott, L.P., Prasad, P.V.V., Ciampitti, I.A., 2020. Satellite-based soybean yield forecast: integrating machine learning and weather data for improving crop yield prediction in southern Brazil. Agric. For. Meteorol. 284, 107886 https://doi.org/10.1016/j.agrformet.2019.107886.

Shao, Y., Campbell, J.B., Taff, G.N., Zheng, B., 2015. An analysis of cropland mask choice and ancillary data for annual corn yield forecasting using MODIS data. Int. J. Appl. Earth Obs. Geoinf. 38, 78–87. https://doi.org/10.1016/j.jag.2014.12.017.

Sun, J., Di, L., Sun, Z., Shen, Y., Lai, Z., 2019. County-level soybean yield prediction using deep CNN-LSTM model. Sensors 19 (20), 4363. https://doi.org/10.3390/s19204363.

Sun, W., Wang, P., Zhang, S., Zhu, D., Liu, J., Chen, J., Yang, H., 2008. Using the vegetation temperature condition index for time series drought occurrence monitoring in the Guanzhong Plain, PR China. Int. J. Remote Sens. 29, 5133–5144. https://doi.org/10.1080/01431160802036557.

Tao, F., Zhang, Z., Liu, J., Yokozawa, M., 2009. Modelling the impacts of weather and climate variability on crop productivity over a large area: a new process-based model development, optimization, and uncertainties analysis. Agric. For. Meteorol. 149, 1266–1278. https://doi.org/10.1016/j.agrformet.2009.02.015.

Terliksiz, A.S., Altýlar, D.T., 2019. Use of deep neural networks for crop yield prediction: a case study of soybean yield in Lauderdale County, Alabama, USA. In: 2019 8th International Conference on Agro-Geoinformatics, pp. 1–4. https://doi.org/10.1109/Agro-Geoinformatics.2019.8820257.

Tian, H., Wang, P., Tansey, K., Zhang, S., Zhang, J., Li, H., 2020. An IPSO-BP neural network for estimating wheat yield using two remotely sensed variables in the Guanzhong Plain. PR China. Comput. Electron. Agric. 169, 105180 https://doi.org/10.1016/j.compag.2019.105180.

Wan, Z., Wang, P., Li, X., 2004. Using MODIS Land surface temperature and normalized difference vegetation index products for monitoring drought in the southern Great Plains, USA. Int. J. Remote Sens. 25, 61–72. https://doi.org/10.1080/0143116031000115328.

Wang, P., Gong, J., Li, X., 2001. Vegetation temperature condition index and its application for drought monitoring. Geomat. Inform. Sci. Wuhan Univ. 26 (5), 412–418. https://doi.org/10.13203/j.whugis2001.05.007 in Chinese with English abstract.

Wang, X., Huang, J., Feng, Q., Yin, D., 2020. Winter wheat yield prediction at county level and uncertainty analysis in main wheat-producing regions of China with deep learning approaches. Remote Sens. 12 (11), 1744. https://doi.org/10.3390/rs12111744.

Weiss, M., Jacob, F., Duveiller, G., 2020. Remote sensing for agricultural applications: a meta-review. Remote Sens. Environ. 236, 111402 https://doi.org/10.1016/j.rse.2019.111402.

Wolanin, A., Mateo-García, G., Camps-Valls, G., Gómez-Chova, L., Meroni, M., Duveiller, G., Liang, Y., Guanter, L., 2020. Estimating and understanding crop yields with explainable deep learning in the Indian Wheat Belt. Environ. Res. Lett. 15, 024019 https://doi.org/10.1088/1748-9326/ab68ac.

Xie, Y., Wang, P., Bai, X., Khan, J., Zhang, S., Li, L., Wang, L., 2017. Assimilation of the leaf area index and vegetation temperature condition index for winter wheat yield estimation using Landsat imagery and the CERES-Wheat model. Agric. For. Meteorol. 246, 194–206. https://doi.org/10.1016/j.agrformet.2017.06.015.

Xun, L., Wang, P., Li, L., Wang, L., Kong, Q., 2019. Identifying crop planting areas using Fourier-transformed feature of time series MODIS leaf area index and sparse-representation-based classification in the North China Plain. Int. J. Remote Sens. 40, 2034–2052. https://doi.org/10.1080/01431161.2018.1492181.

You, J., Li, X., Low, M., Lobell, D., Ermon, S., 2017. Deep Gaussian process for crop yield prediction based on remote sensing data. In: 31st AAAI Conf. Artif. Intell, 31, pp. 4559–4565. https://aaai.org/ocs/index.php/AAAI/AAAI17/paper/view/14435/14067.

Yuan, H., Dai, Y., Xiao, Z., Ji, D., Wei, S., 2011. Reprocessing the MODIS leaf area index products for land surface and climate modelling. Remote Sens. Environ. 115 (5), 1171–1187. https://doi.org/10.1016/j.rse.2011.01.001.

Zhang, J., Tian, H., Wang, P., Tansey, K., Zhang, S., Li, H., 2022. Improving wheat yield estimates using data augmentation models and remotely sensed biophysical indices within deep neural networks in the Guanzhong Plain, PR China. Comput. Electron. Agric. 192, 106616 https://doi.org/10.1016/j.compag.2021.106616.

Zhang, L., Zhang, Z., Luo, Y., Cao, J., Xie, R., Li, S., 2021. Integrating satellite-derived climatic and vegetation indices to predict smallholder maize yield using deep learning. Agric. For. Meteorol. 311, 108666 https://doi.org/10.1016/j.agrformet.2021.108666.

Zhang, L., Zhang, L., Du, B., 2016. Deep learning for remote sensing data: a technical tutorial on the state of the art. IEEE Geosc. Rem. Sen. M. 4 (2), 22–40. https://doi.org/10.1109/MGRS.2016.2540798.