# Improving Estimation of Winter Wheat Nitrogen Status Using Random Forest by Integrating Multi-Source Data Across Different Agro-Ecological Zones

Yue Li[1], Yuxin Miao[2], Jing Zhang[3], Davide Cammarano[4], Songyang Li[5], Xiaojun Liu[1], Yongchao Tian[1], Yan Zhu[1], Weixing Cao[1*] and Qiang Cao[1*]

[1] MARA Key Laboratory for Crop System Analysis and Decision Making, Jiangsu Key Laboratory for Information Agriculture, National Engineering and Technology Center for Information Agriculture, MOE Engineering and Research Center for Smart Agriculture, Collaborative Innovation Center for Modern Crop Production Co-sponsored by Province and Ministry, Nanjing Agricultural University, Nanjing, China, [2] Precision Agriculture Center, Department of Soil, Water and Climate, University of Minnesota, St. Paul, MN, United States, [3] Department of Crop and Soil Sciences, University of Georgia, Tifton, GA, United States, [4] Department of Agroecology, Aarhus University, Tjele, Denmark, [5] Department of Land Surveying and Geo-Informatics, The Hong Kong Polytechnic University, Kowloon, China

Timely and accurate estimation of plant nitrogen (N) status is crucial to the successful implementation of precision N management. It has been a great challenge to non-destructively estimate plant N status across different agro-ecological zones (AZs). The objective of this study was to use random forest regression (RFR) models together with multi-source data to improve the estimation of winter wheat (*Triticum aestivum* L.) N status across two AZs. Fifteen site-year plot and farmers' field experiments involving different N rates and 19 cultivars were conducted in two AZs from 2015 to 2020. The results indicated that RFR models integrating climatic and management factors with vegetation index ($R^2$ = 0.72–0.86) outperformed the models by only using the vegetation index ($R^2$ = 0.36–0.68) and performed well across AZs. The Pearson correlation coefficient-based variables selection strategy worked well to select 6–7 key variables for developing RFR models that could achieve similar performance as models using full variables. The contributions of climatic and management factors to N status estimation varied with AZs and N status indicators. In higher-latitude areas, climatic factors were more important to N status estimation, especially water-related factors. The addition of climatic factors significantly improved the performance of the RFR models for N nutrition index estimation. Climatic factors were important for the estimation of the aboveground biomass, while management variables were more important to N status estimation in lower-latitude areas. It is concluded that integrating multi-source data using RFR models can significantly improve the estimation of winter wheat N status indicators across AZs compared to models only using one vegetation index. However, more studies are needed to develop unmanned aerial vehicles and satellite remote sensing-based machine learning models incorporating multi-source data for more efficient monitoring of crop N status under more diverse soil, climatic, and management conditions across large regions.

**Keywords:** precision nitrogen management, machine learning, environmental variables, management practices, variable selection, nitrogen nutrition index

# INTRODUCTION

Winter wheat (*Triticum aestivum* L.), a major staple food crop, is vital to global food security and sustainable agriculture. A great challenge in winter wheat production is to optimize nitrogen (N) management to achieve high crop yield and high N use efficiency (NUE) under different soil landscapes and weather conditions across large regions. Precision N management (PNM) has the potential to overcome this challenge by matching N application and crop demand (Cao et al., 2015; Cammarano et al., 2021). Timely and accurate estimation of plant N status is crucial to the successful implementation of PNM (Li et al., 2022). The N nutrition index (NNI) is considered as a reliable N status indicator, which can be calculated as the ratio of the actual N concentration to the critical N concentration ($N_c$) (Lemaire et al., 2008).

Instead of destructive sampling and laborious analysis, active canopy sensors have been used for the estimation of crop N status based on spectral reflection and absorption properties of the crop canopy. These active canopy sensors have their own light sources and are not affected by environmental light conditions (Cao et al., 2017a). Vegetation indices (VIs) derived from these sensors and simple linear regression (SLR) have been commonly used for crop estimation (Cao et al., 2015; Bonfil, 2017). However, SLR models generally do not work well when applied in other areas or years (Munoz-Huerta et al., 2013). It is hypothesized that adding climatic and management factors can provide complementary information and improve the estimation of crop N status compared to approaches that only use VIs. Though prior studies have investigated the contributions of climatic and remote sensing data on wheat yield prediction at regional scales (Cai et al., 2019), there were limited studies that explored their contributions to the estimation of N status. Besides, diverse external factors, such as climatic and management factors, show different patterns owing to latitude and longitude.

Stepwise multiple linear regression has been commonly used for plant N status estimation using crop sensing data alone or together with other ancillary data (Miao et al., 2009; Dong et al., 2021). Recently, machine learning (ML) algorithms have been increasingly employed to combine multiple VIs or VIs with genetics, environmental, and management information to predict crop N status due to their capabilities to deal with both linear and nonlinear relationships (Zha et al., 2020; Li et al., 2022). Random forest, developed by Breiman (2001), is a representative ML algorithm with a good performance by averaging an ensemble of trees without overfitting problems (Rhee and Im, 2017; Zhang et al., 2019). It has been widely used for regression and classification applications (Wang et al., 2016), including crop N status prediction (Han et al., 2019; Lu et al., 2019; Zha et al., 2020; Li et al., 2022).

Studies developing models using crop sensing and multi-source ancillary data with ML algorithms for winter wheat N status prediction across different agro-ecological zones (AZs) are still limited. Therefore, the objective of this study was to evaluate the performance of random forest regression (RFR) models for winter wheat N status prediction across two AZs using active crop sensor data together with climatic and management information.

# MATERIALS AND METHODS

## Study Area and Experiment Design

Fifteen site-year winter wheat field experiments were conducted in agro-ecological zone 1 (AZ1, 37°43'N and 117°13'E, in Laoling County of Shandong Province) and agro-ecological zone 2 (AZ2, 33°05'N and 119°53'E, in Xinghua City of Jiangsu Province) (**Figure 1A**), with different patterns of precipitation distribution and monthly average temperature (**Figure 1B**). The soil types of AZ1 and AZ2 were sandy loam soil and loam soil, respectively.

The design includes seven plot experiments (Exp. 1–7) and eight farmers' field experiments (Exp. 8–15) involving different N rates, sowing dates (SD), and seeding rates (SR). The N fertilizers of Exp. 1–3 were applied with 40% as basal N before sowing, except for farmer's management (FM), which applied 50% as basal N. The remaining N was applied at the stem elongation (SE) stage. The N fertilizers of Exp. 4–5 were applied at the rate of 50% of total N rates before sowing and 50% at the SE stage, respectively. The differences between Exp. 6–7 and Exp. 4–5 are listed in **Table 1**. With a randomized complete block design, the treatments of plot experiments were carried out with three repetitions. In addition to plot experiments, eight farmers' field experiments (Exp. 8-15) were conducted across the Nanxia village in Shandong Province to compare different N management strategies. In each farmers' field experiment, there were three treatments: FM, regional optimum N management (RONM), and PNM. RONM, integrating high-yielding and high-NUE, was managed with 81 kg N ha$^{-1}$ as basal N and 138 kg N ha$^{-1}$ as topdressing N (Zhou et al., 2017). For PNM, the basal N rate was the same as RONM, and topdressing N was applied according to an active canopy sensors-based algorithm (Cao et al., 2017b). Other managements were the same as RONM. Irrigation was done two times at the sowing and SE stages in AZ1 and once after sowing in AZ2. All N source was urea. For all plots, wheat was not limited by phosphate and potash fertilizers and was kept free of pests and diseases during the growing seasons.

## Data Acquisition and Calculation

### VIs Datasets

The portable active sensor RapidSCAN CS-45 (Holland Scientific Inc., Lincoln, Nebraska, USA) was used to collect canopy reflectance and VIs data, including the default normalized difference vegetation index (NDVI) and the normalized difference red edge (NDRE). It was carried by hand at a constant speed to collect sensor readings at a sensor-to-canopy distance of approximately 0.7–0.9 m. The average values from three rows in each plot were then calculated to represent the plot. The NDRE was selected to monitor the N status to avoid the potential saturation effect of NDVI under the high aboveground biomass (AGB) and plant N uptake (PNU) conditions (Li et al., 2010). The NDRE was calculated as follows:

$$NDRE = (NIR_{780} - RE_{730}) / (NIR_{780} + RE_{730}) \quad (1)$$

where $NIR_{780}$ and $RE_{730}$ refer to near-infrared (780nm) and red-edge (730nm) waveband reflectance, respectively.

**FIGURE 1 |** Locations of the study region **(A)**, mean precipitation (mm) and temperature (°C) of agro-ecological zone1 (AZ1, from 2015 to 2018) and agro-ecological zone 2 (AZ2, from 2017 to 2020) from sowing to sensing date **(B)**.
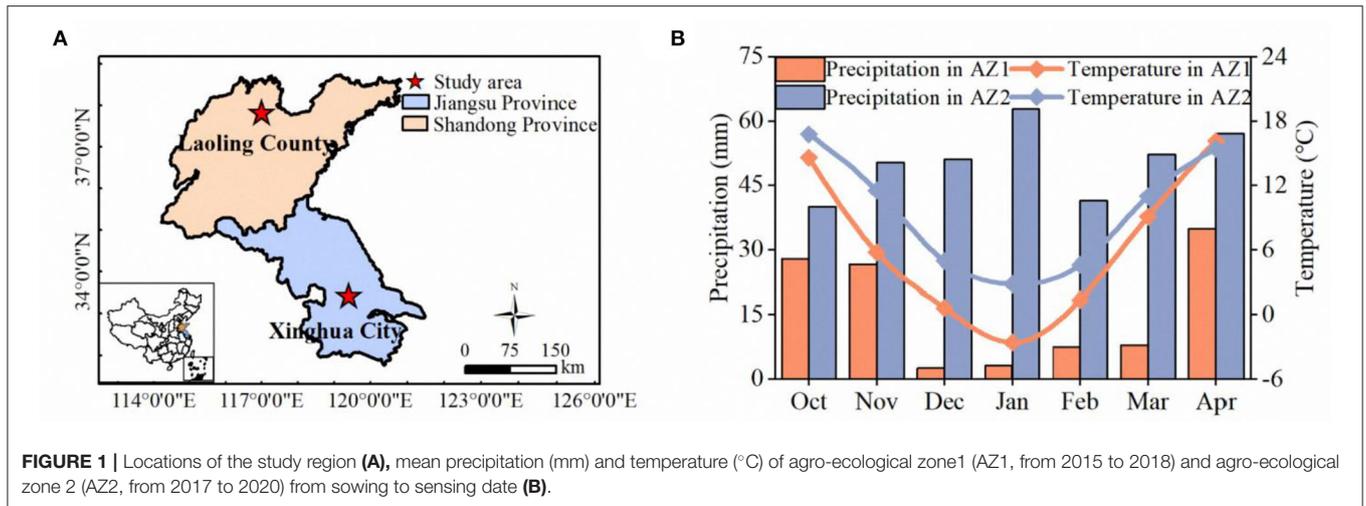
**TABLE 1 |** Basic information of the field experimental design.

| Sites | Experiments and years | Cultivars | N rates | P rates | K rates | Seeding rates | Sowing dates | Sensing dates |
|---|---|---|---|---|---|---|---|---|
| | | | (kg ha⁻¹) | | | | | |
| AZ1 | Exp.1 (2015) | JM22, LY502 | 0, 120, 180, 240, 300 | 120 | 75 | 187.4 | Oct.30th | Apr.17th |
| | Exp.2 (2016) | JM22, LX77 | 0, 120, 180, 240, 300, FM | | | | Oct.19th | Apr.2nd |
| | Exp.3 (2017) | JM22, SN29 | | | | | Oct.20th | Apr.12th |
| AZ2 | Exp.4 (2017); Exp.5 (2018) | ZM12, YM23, NM13 | 0, 90, 180, 270, 360 | 105 | 135 | 133.8 | Nov.8th; Nov.1st | Apr.1st; Mar.16th |
| AZ2 | Exp.6 (2018); Exp.7 (2019) | YM23 | 0, 180, 240, 300 | 105 | 135 | 103.5, 155.3, 207.0; 112.3, 168.4, 224.6 | Nov.4th, Nov.24th, Dec.1st; Nov.1st, Nov.15th, Dec.24th | Mar.18th, 23th, Apr.4th; Mar.14th, 14th, Apr.1st |
| AZ1 | | JM22 | FM: 280.5; | 120; | 36; | 300; | Oct.dd#; | Mar.26th |
| | | | RONM: 219.0; | 112 | 27 | 165 | Oct.dd#; | |
| | Exp.8 (2016); | | PNM1: 199.9; | | | | Oct.5th; | |
| | Exp.9 (2016); | | PNM2: 188.4; | | | | Oct.2nd; | |
| | Exp.10 (2016); | | PNM3: 208.2; | | | | Oct.13th; | |
| | Exp.11 (2016); | | PNM4: 205.0; | | | | Oct.11th; | |
| | Exp.12 (2016); | | PNM5: 213.2; | | | | Oct.12th; | |
| | Exp.13 (2016); | | PNM6: 209.4; | | | | Oct.13th; | |
| | Exp.14 (2016); | | PNM7: 200.2; | | | | Oct.11th; | |
| | Exp.15 (2016) | | PNM8: 182.5 | | | | Oct.12th | |

*FM, RONM, and PNM1-8 refer to farmers' management, regional optimum N management, and precision N management, according to Cao et al. (2017b), respectively. JM22, LY502, LX77, SN29, ZM12, YM23, and NM13 refer to Jimai22, Luyuan502, Liangxing77, Shannong29, Zhenmai12, Yangmai23 and Ningmai13, respectively.*
*AZ1 and AZ2 mean agro-ecological zone 1 and agro-ecological zone 2.*
*dd# means sowing date of FM and RONM varied according to the eight farmers while being the same with PNM, respectively.*

## Agronomic Datasets

The plants were sampled right after sensing data was collected at the SE stage (**Table 1**), which is the start of rapid crop growth and N uptake and the stage important for making in-season N recommendations (Meng et al., 2013).

In AZ1, plant samples were collected from randomly selected 1 m by 0.3 m areas in each plot and rinsed with water, and the roots were removed. The plant samples were placed in an oven at 105°C for 30 min and dried at 80°C to constant weight for the determination of AGB, which was converted to the unit of t ha⁻¹ based on the row spacing. The sub-samples were ground to pass through a 1 mm sieve in a Wiley mill, and the plant N concentration was determined using the Kjeldahl method (Nelson and Sommers, 1973). The PNU was calculated

by multiplying AGB by plant N concentration (Lu et al., 2017). According to the definition of NNI, $N_c$ was calculated from the critical N dilution curves developed by Yue et al. (2012) in AZ1.

$$N_c \;=\; 4.15 \;\times\; AGB^{-0.38} \tag{2}$$

where AGB refers to the aboveground biomass (t ha$^{-1}$).

Compared to AZ1, twenty plants in each plot were sampled randomly and averaged as the mean value of the plot in AZ2 experiments. The critical N dilution curve developed by Zhao et al. (2012) was used to calculate the $N_c$.

$$N_c \;=\; 4.33 \;\times\; AGB^{-0.45} \tag{3}$$

where AGB refers to the aboveground biomass (t ha$^{-1}$).

### Climatic Datasets

The primary climatic variables of AZ1 and AZ2 were obtained from the China Meteorological Data Service Center (http://data.cma.cn) and the Xinghua weather station, respectively. The climatic variables from the date of sowing to sensing included seasonal total precipitation (PPT), seasonal maximum, minimum, and mean temperature ($T_{max}$, $T_{min}$, $T_{mean}$), growing degree days (GDD) (Wang et al., 2014), abundant and well-distributed rainfall (AWDR) (Bean et al., 2018), shannon diversity index (SDI), relative humidity (HU), and solar radiation (RAD) calculated using the Hybrid-Maize model (Yang et al., 2004). $T_{max}$ and $T_{min}$ were determined as the maximum value of the daily maximum temperature ($D_{max}$) and the minimum value of the daily minimum temperature ($D_{min}$), respectively. GDD, SDI, and AWDR were calculated as follows:

$$GDD \;=\; \sum \left( (D_{max} + D_{min}) \,/\, 2 \,-\, T_{base} \right) \tag{4}$$

$$SDI \;=\; \sum \left( (-p_i \times \ln(p_i)) \,/\, \ln(d) \right) \tag{5}$$

$$AWDR \;=\; PPT \;\times\; SDI \tag{6}$$

where $D_{max}$ = daily maximum temperature (up to 30 °C), $D_{min}$ = daily minimum temperature, and $T_{base}$ = base temperature (0 °C). $p_i$ is the ratio of daily precipitation to $PPT$, and $d$ is the days from sowing to sensing.

### Management Datasets

Management practices included SD, SR, and basal N (BN) (shown in **Table 1**).

## Data Analysis

The data were analyzed using the following steps: (1) establishing the SLR models only using NDRE; (2) identifying optimal variables based on variable selection strategies; (3) training and validating the RFR models using selected variables by updating hyperparameters and cross-validation; (4) comparing and evaluating the robustness of different models; (5) determining the variable contributions in different AZs, and (6) exploring the effective combinations of multi-source data for N status estimation (**Figure 2**).

### Strategies of Variable Selection

Two proposed variable selection strategies were evaluated in this study. The first strategy was the Pearson correlation coefficient (PCC), a common method to identify the variables with the highest statistical significance. In the previous studies, it was commonly adopted to quantify the degrees of correlation between variables (Li et al., 2019; Hamrani et al., 2020). Taking AGB, PNU, and NNI as dependent variables, climatic and management data as candidate independent variables, and NDRE as a fixed independent variable, the PCC was used to select the optimal variables from the candidate independent variables within the same group to remove the compound effects. It should be emphasized that the climatic data were divided into two sub-groups: temperature-related variables (GDD, RAD, $T_{mean}$, $T_{min}$, and $T_{max}$), and water-related variables (PPT, SDI, AWDR, and HU). Then, two candidate independent variables would be selected from the temperature-related, water-related group, and management groups, respectively. The selection process included the following steps: (1) selecting the variables with the maximum absolute correlation with dependent variables in each group as the first group of candidate independent variables; (2) selecting the second candidate independent variables that had smaller correlation coefficients (<0.5 threshold in this study) with the first group of candidate independent variables and the maximum correlation with the dependent variables (Wang et al., 2020).

For the second strategy, the variable importance ranked by the RFR (VIRRFR) was also adopted to select the variables (Zhu et al., 2021). It was performed specifically by (1) ranking the variable importance by the RFR; (2) adding the variables iteratively into the RFR from highest to lowest according to the importance score order until all independent variables were included; (3) determining optimal variables corresponding to the models' coefficients of determination ($R^2$), root mean square error (RMSE), and relative error (RE) based on the cross-validation. The above processes were implemented to explore optimal variables for estimating AGB, PNU, and NNI across AZ1, AZ2, and the two AZs (AZ1+2), respectively.

$$RMSE \;=\; \left( \left( \sum_{n=1}^{n} (E_i - O_i)^2 \right) / n \right)^{1/2} \tag{7}$$

$$RE\ (\%) \;=\; \left( \left( \left( \sum_{n=1}^{n} (E_i - O_i)^2 \right) / n \right)^{1/2} \right) / O'_i, \tag{8}$$

Where $E_i$ and $O_i$ are the estimated and the observed values of AGB, PNU, or NNI, respectively. $O'_i$ is the mean of the observed values of the AGB, PNU, or NNI. $n$ is the number of samples.

To compare these two strategies, relative $R^2$ ($RR^2$), relative RMSE (RRMSE), relative RE (RRE), and the relative number of variables (RN) were calculated by normalizing them using the maximum values of $R^2$, RMSE, RE and the number of variables, respectively.

### Identifying Appropriate Combinations of Multi-Source Data for N Status Estimation

To determine the effects of different sources of data on N status estimation, distinct combinations of multi-source data were explored, which consisted of (1) only NDRE; (2) both NDRE and climatic data (NDRE+C); (3) both NDRE and management

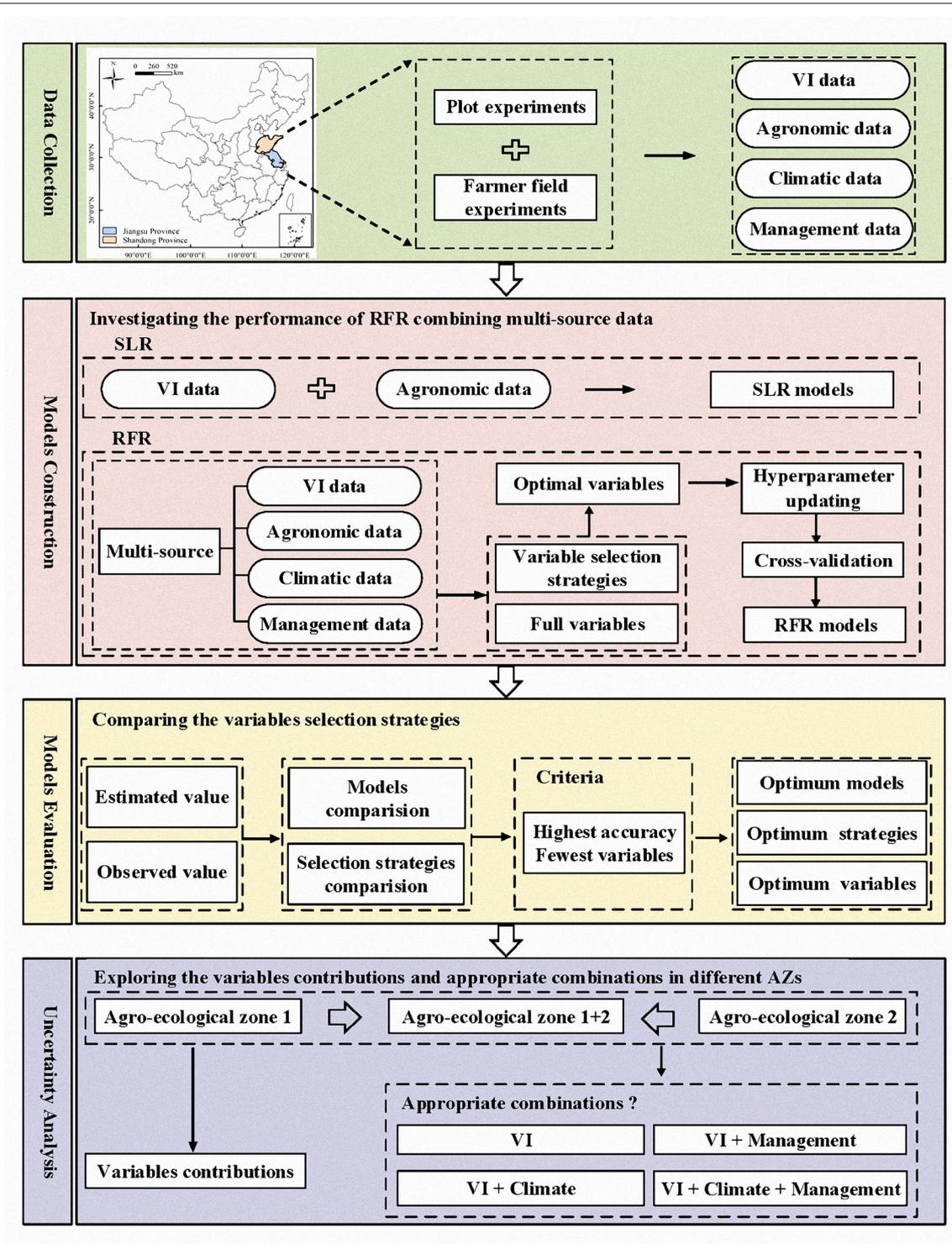**FIGURE 2 |** Workflow of this study.

data (NDRE+M), and (4) full sources of data (NDRE+C+M). The models with NDRE only serve as a reference for further assessment of the benefit of adding climatic and/or management data. The decrease of RMSE (dRMSE) was defined as the ratio of the RMSE difference obtained from the multi-source data ($RMSE_{multi-source}$) and NDRE only ($RMSE_{NDRE}$) to the

$RMSE_{NDRE}$. This was calculated as follows:

$$dRMSE\ (\%) = (RMSE_{multi\text{-}source} - RMSE_{NDRE})$$
$$/\ RMSE_{NDRE} \qquad (9)$$

where $RMSE_{NDRE}$ and $RMSE_{multi-source}$ are RMSE of NDRE only and combinations of multi-source data, respectively.

### Modeling Process

Microsoft Excel (Microsoft Corporation, Redmond, Washington, USA) was used to establish and select SLR models with the highest $R^2$. The Scikit-learn library (in Python 3.9.5) was utilized to establish the RFR models (Abraham et al., 2014). Three hyper-parameters, namely, the number of decision trees (n_estimators, from 10 to 300 at intervals of 50), the maximum depth (max_depth, from 2 to 20 at intervals of 2), and the minimum number of samples to spilt (min_samples_split, from 2 to 12 at intervals of 2) were tuned using 5-fold cross-validated grid search (Abraham et al., 2014).

### Evaluating Strategies and Performance Metrics

The total data of AZ1, AZ2, and AZ1+2 were split randomly into 70% for training and 30% for the test. Meanwhile, 5-fold cross-validation was implemented due to its simplicity, universality, and efficiency in reducing the over-fitting problem (Arlot and Celisse, 2010), and its mean values were used to represent the predictive performance. The accuracy of trained models was evaluated using the test dataset with $R^2$, RMSE, and RE.

## RESULTS

### Variability of Winter Wheat N Status Indicators

A total of 396 observations were obtained, with 277 for training and 119 for the test (**Table 2**). PNU showed the highest variation on training dataset (CV = 60.53, 51.04 and 52.29% in AZ1, AZ2 and AZ1+2, respectively), followed by AGB (CV = 50.2, 38.53, and 41.56% in AZ1, AZ2 and AZ1+2, respectively) and NNI (33.67, 34.75, and 32.35% in AZ1, AZ2, and AZ1+2, respectively). Similar results were presented in the test dataset. In general, experimental data is suitable to evaluate the predictive performance for N status estimation.

### Estimating N Status Indicators Using Simple Linear Regression Models

The performance of SLR models for estimating N status indicators varied slightly across the different AZs (**Figure 3**). Comparatively, NDRE explained 59% of AGB variation, 62% of PNU, and 57% of NNI in AZ1, which performed better than AZ2 with 36, 52, and 47%, respectively. The performance of models in AZ1+2 was similar to that in AZ2 (44% for AGB, 54% for PNU, and 46% for NNI). The validation results confirmed that models for AZ1 (**Figures 4A,D,G**) had lower or similar RMSE and RE than models for AZ2 (**Figures 4B,E,H**) or AZ1+2 (**Figures 4C,F,I**), respectively.

### Selection of Important Variables

For the PCC, the correlation and corresponding p-values were analyzed between candidate independent variables and dependent variables. Notably, there were both positive and negative correlations in each group (**Figure 5**). For the temperature-related variables, GDD was selected first due to its highest correlation with AGB in AZ1, and RAD was selected for its high correlation with AGB but lower correlation with GDD (<0.5). Similarly, PPT and SDI were selected from the water-related variables, while BN and SR from the management variables (**Figure 5A**). The above steps were repeated for AGB, PNU, and NNI in AZ2 (**Figure 5B**) and AZ1+2 (**Figure 5C**), respectively. Overall, six variables were selected in both AZ1 and AZ2 and five variables in AZ1+2.

For the VIRRFR, cross-validation accuracy improved significantly within the top two or three variables and then fluctuated slightly (**Figure 6**). Variables were determined by taking the highest $R^2$, lowest RMSE, and RE into account. For example, when the AGB model achieved the largest $R^2$ (0.71), the lowest RMSE (0.71 t ha$^{-1}$) and RE (25.3%) in AZ1, the 10 variables in the models were selected as the optimal variables (**Figure 6A**). In the same way, the number of selected variables for AGB (13 in AZ2 and 11 in AZ1+2, **Figures 6B,C**), PNU (5 in AZ1, AZ2, and 12 in AZ1+2, **Figures 6D–F**), and NNI (9 in AZ1, 11 in AZ2, and 12 in AZ1+2, **Figures 6G–I**) were determined.

### Estimating N Status Indicators by Random Forest Regression Models
#### Models Construction

The RFR models were trained using full variables and selected variables based on the above-mentioned strategies and then were evaluated to determine the optimal models based on cross-validation (**Figure 7**). Slight differences in $R^2$, RMSE, and RE were observed between the models with variable selection and full variables, while PCC showed the superiority due to the least variables with the same accuracy except for the PNU estimation models in AZ1 and AZ2. Thus, the variables selected by the PCC were considered optimal and analyzed in the next part. It should also be pointed out that similar performance of models was shown in AZ1+2 compared to AZ1 and AZ2, indicating the consistency of constructed models.
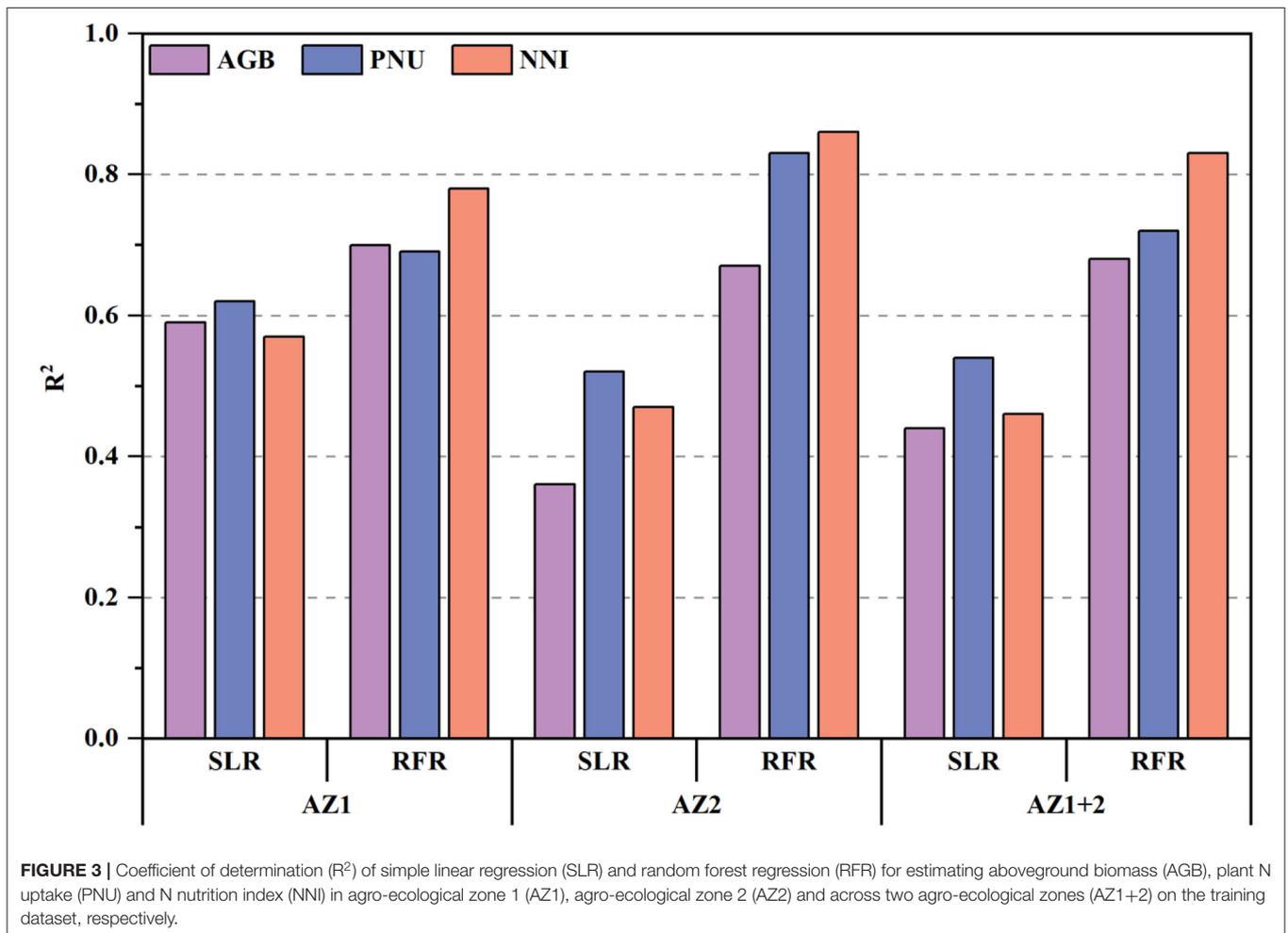
The performance of RFR models for N status estimation was consistently better than SLR models. All the $R^2$ values of RFR models were significantly higher than the corresponding $R^2$ values of SLR models, and the most obvious improvement appeared in AZ2, followed by AZ1 (**Figure 3**). Consistently, the AZ1+2 matched the above result. In terms of N status indicators, the NNI estimation models performed the best ($R^2 = 0.78$–0.83), followed by PNU ($R^2 = 0.69$–0.83) and AGB ($R^2 = 0.67$–0.70).

#### Models Evaluation

In general, the observed N status indicators were more related to the estimated ones obtained from RFR (**Figure 8**) than SLR models (**Figure 4**). Moreover, the improvement of RFR models over SLR models was larger in AZ2 than in AZ1 (**Figures 8A,B,D,E,G,H**). On the other hand, the validation

**TABLE 2 |** Descriptive statistics of winter wheat aboveground biomass (AGB), plant N uptake (PNU), and N nutrition index (NNI) for training and test dataset in different agro-ecological zones.

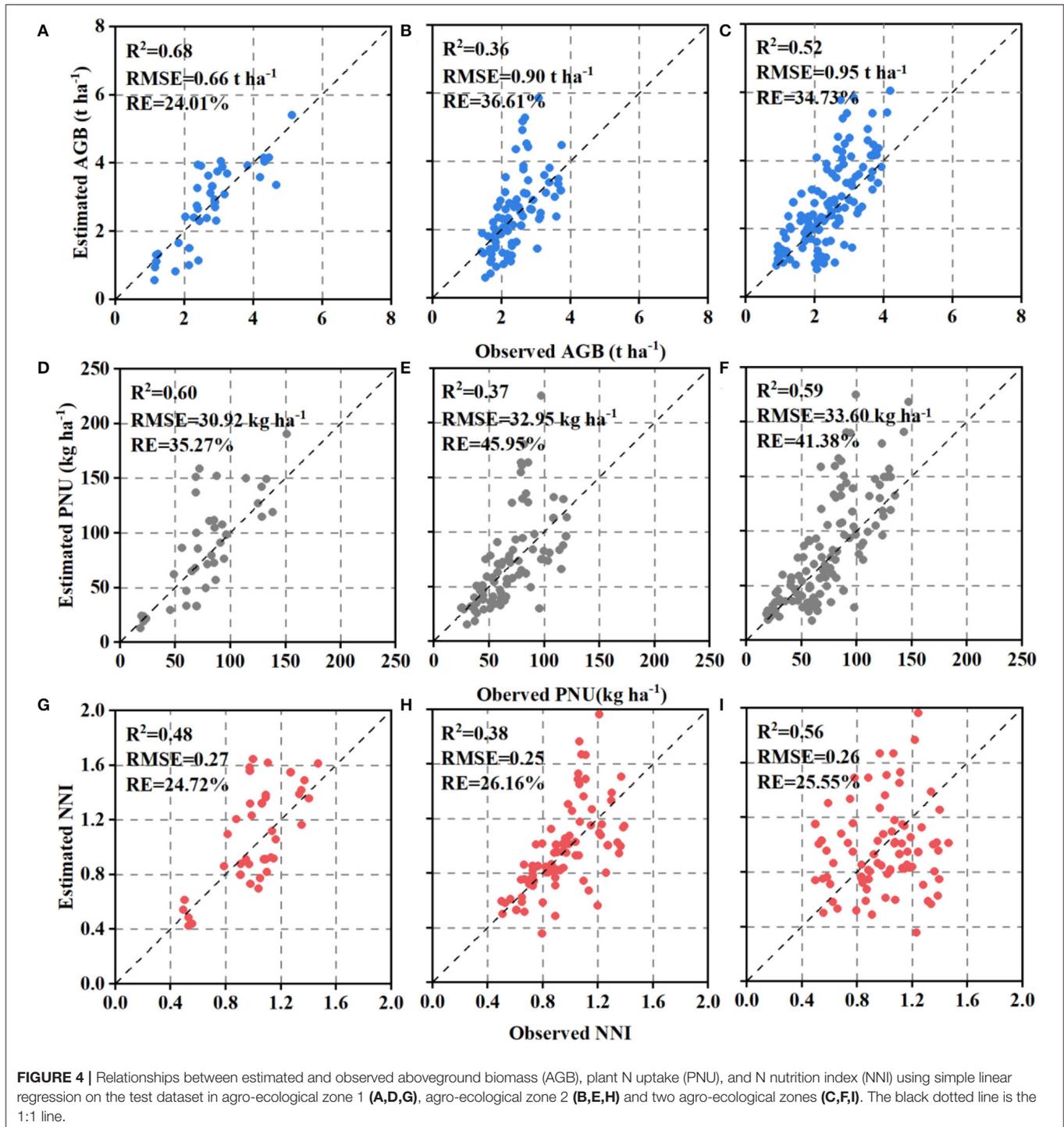| | Agro-ecological zone 1 | | | | | Agro-ecological zone 2 | | | | | Agro-ecological zone 1+2 | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Min | Max | Mean | SD | CV(%) | Min | Max | Mean | SD | CV(%) | Min | Max | Mean | SD | CV(%) |
| **Training dataset** | | | (n = 88) | | | | | (n = 189) | | | | | (n = 277) | | |
| AGB (t ha⁻¹) | 0.83 | 6.04 | 2.79 | 1.40 | 50.20 | 0.81 | 5.77 | 2.66 | 1.03 | 38.53 | 0.56 | 5.68 | 2.63 | 1.09 | 41.56 |
| PNU (kg ha⁻¹) | 15.82 | 233.63 | 86.33 | 52.25 | 60.53 | 17.87 | 190.35 | 79.34 | 40.45 | 51.04 | 12.50 | 233.63 | 79.68 | 41.66 | 52.29 |
| NNI | 0.42 | 1.92 | 1.05 | 0.35 | 33.67 | 0.28 | 1.87 | 1.03 | 0.36 | 34.75 | 0.42 | 1.92 | 1.04 | 0.34 | 32.35 |
| **Test dataset** | | | (n = 38) | | | | | (n = 81) | | | | | (n = 119) | | |
| AGB (t ha⁻¹) | 0.56 | 5.40 | 2.77 | 1.19 | 42.88 | 0.58 | 5.85 | 2.47 | 1.13 | 45.74 | 0.81 | 6.04 | 2.74 | 1.29 | 47.46 |
| PNU (kg ha⁻¹) | 12.50 | 190.39 | 87.67 | 46.75 | 53.33 | 15.40 | 224.69 | 71.70 | 41.52 | 57.91 | 17.87 | 224.69 | 81.18 | 50.16 | 61.78 |
| NNI | 0.43 | 1.64 | 1.08 | 0.36 | 33.60 | 0.36 | 1.96 | 0.97 | 0.32 | 33.04 | 0.28 | 1.96 | 1.01 | 0.38 | 38.24 |



**FIGURE 3 |** Coefficient of determination ($R^2$) of simple linear regression (SLR) and random forest regression (RFR) for estimating aboveground biomass (AGB), plant N uptake (PNU) and N nutrition index (NNI) in agro-ecological zone 1 (AZ1), agro-ecological zone 2 (AZ2) and across two agro-ecological zones (AZ1+2) on the training dataset, respectively.

results indicated the estimation accuracy varied with N status indicators, with NNI having the best result (**Figures 8G–I**).

## Importance of Different Variables

The importance of variables selected by the PCC for estimating N status indicators was explored by the RFR models (**Figure 9**). The results showed that NDRE was consistently viewed as the most important variable except for the estimation of NNI in AZ1, and BN was considered important for all N status indicators. It should be noted that the important variables differed with the AZs and N status indicators. In AZ1, SDI and RAD were ranked as the top two variables for AGB, PPT and BN for PNU, and PPT and BN for NNI (except for NDRE). The water-related variables were consistently identified to be the vital variables (**Figures 9A–C**).
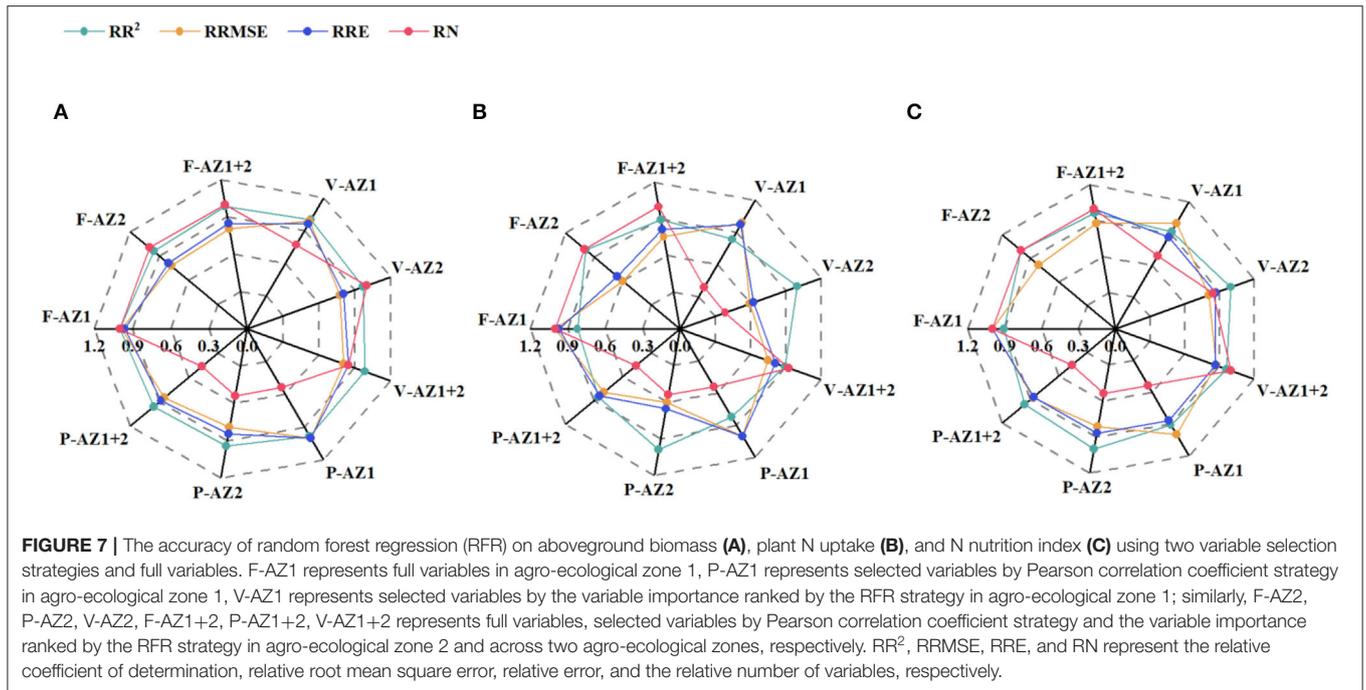
**FIGURE 4 |** Relationships between estimated and observed aboveground biomass (AGB), plant N uptake (PNU), and N nutrition index (NNI) using simple linear regression on the test dataset in agro-ecological zone 1 **(A,D,G)**, agro-ecological zone 2 **(B,E,H)** and two agro-ecological zones **(C,F,I)**. The black dotted line is the 1:1 line.

Compared to AZ1, temperature-related variables were also of great importance to AZ2 (**Figure 9D**). Generally, water and temperature-related variables were listed as the most critical variables across AZs (**Figures 9G–I**). With respect to N status indicators, temperature-related variables were more important for AGB (**Figures 9D,G**), while water-related variables were more important for PNU and NNI (**Figures 9B,C,E,F**).

## Multi-Source Data Performance

Different combinations of variables were further investigated and evaluated for the estimation of N status indicators based on dRMSE. The results indicated that the more the sources, the better the model's performance (**Figure 10**). Specifically, the dRMSE of two-source data (NDRE+C or NDRE+M) was more significant than that of single-source data (NDRE), and full

**FIGURE 5 |** Pearson correlation coefficient (PCC) analysis in agro-ecological zone 1 **(A)**, agro-ecological zone 2 **(B)** and across two agro-ecological zones **(C)**. "****", "***", and "*" note the significant correlation at 0.001, 0.01, and 0.05, respectively. GDD, RAD, T$_{max}$, T$_{min}$, T$_{mean}$, PPT, SDI, AWDR, HU, SD, SR, BN, AGB, PNU, and NNI refer to growing degree days, solar radiation, seasonal maximum, minimum, and mean temperature, seasonal total precipitation, shannon diversity index, abundant and well-distributed rainfall, relative humidity, sowing dates, seeding rates, basal N, aboveground biomass, plant N uptake, and N nutrition index, respectively.



**FIGURE 6 |** The accuracy of aboveground biomass (AGB), plant N uptake (PNU), and N nutrition index (NNI) with variables selected by the random forest regression in agro-ecological zone 1 **(A,D,G)**, agro-ecological zone 2 **(B,E,H)** and across two agro-ecological zones **(C,F,I)**.

**FIGURE 7 |** The accuracy of random forest regression (RFR) on aboveground biomass **(A)**, plant N uptake **(B)**, and N nutrition index **(C)** using two variable selection strategies and full variables. F-AZ1 represents full variables in agro-ecological zone 1, P-AZ1 represents selected variables by Pearson correlation coefficient strategy in agro-ecological zone 1; V-AZ1 represents selected variables by the variable importance ranked by the RFR strategy in agro-ecological zone 1; similarly, F-AZ2, P-AZ2, V-AZ2, F-AZ1+2, P-AZ1+2, V-AZ1+2 represents full variables, selected variables by Pearson correlation coefficient strategy and the variable importance ranked by the RFR strategy in agro-ecological zone 2 and across two agro-ecological zones, respectively. $RR^2$, RRMSE, RRE, and RN represent the relative coefficient of determination, relative root mean square error, relative error, and the relative number of variables, respectively.

sources of data achieved the best results, with 33.72–40.95% for AGB, 38.56–46.99% for PNU, and 48.54–52.50% for NNI. The combination of NDRE+C produced better results than those of NDRE+M (23.55, 27.21, and 42.83% vs. 13.62, 3.42, and −2.1% for AGB, PNU, and NNI, respectively), indicating that climatic data make more important contributions to the estimation of N status indicators in AZ1. On the contrary, the combination of NDRE+M performed better than NDRE+C in AZ2. Adding the climatic variables produced the greater dRMSE for AGB in AZ2 than that in AZ1, while the opposite was true for NNI. For PNU, the performances of adding the climatic variables were similar in AZ1 and AZ2. In AZ1+2, the contributions of the climatic data to NNI were greater than PNU.

## DISCUSSION

### Comparison of Different Regression Models

The SLR models are commonly used to estimate plant N status and guide variable N applications (Li et al., 2019; Wang et al., 2019). However, models using NDRE could only explain 48–68% of N status indicators variabilities for AZ1, 36–38% for AZ2, and 52–59% for AZ1+2 based on the test dataset (**Figure 4**). Those results agreed with the findings of wheat N status estimation only using VIs across farmer's fields in a village (Chen et al., 2019). The performance difference of these models between AZ1 and AZ2 may be attributed to the different experimental treatments (**Table 1**) and climatic and management factors (**Figure 1B**). Consequently, it is necessary to combine climatic and management factors with NDRE in RFR models to improve the estimation accuracy. In agreement with Cummings et al. (2021) and Wang et al. (2021), the multi-variable RFR
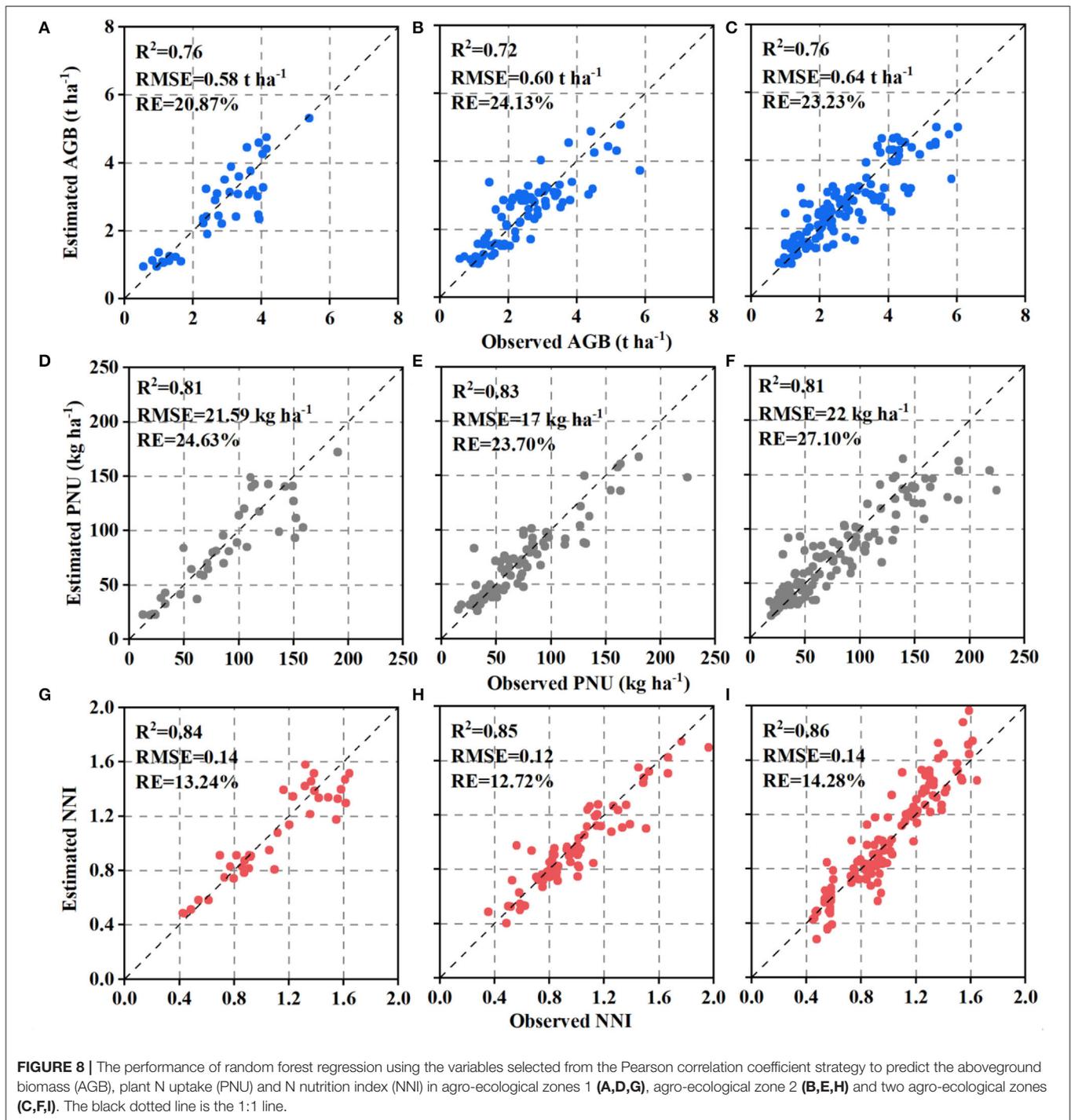
models performed better than SLR models (**Figure 3**), because the RFR models included multi-source data and could analyze nonlinear and complex relationships (Wang et al., 2021).

### Comparison of Variable Selection Strategies

Generally, variable selection prior to model construction could reduce the data dimension and measurement while differentiating meaningful variables from the noise and improving calibration efficiency (Heremans et al., 2015; Feng et al., 2020). Two variable selection strategies, the PCC and VIRRFR, were adopted in this study. In **Figure 7**, the performance of full variables models was not better than those of using fewer variables with variable selection, demonstrating the effectiveness of variable selection. Similar results were observed in the previous studies (Cai et al., 2019; Zhang et al., 2019; Wang et al., 2020). The PCC showed more potential than VIRRFR to estimate N status, except for the estimation of PNU in AZ1 and AZ2. The variables were selected based on the PCC strategy from a statistical analysis point of view, resulting in more informative and diverse variables. Many variables were correlated with each other at the $p < 0.001$ significance level but still selected in the VIRRFR strategy (data not shown), such as $T_{max}$ and $T_{mean}$ of AZ1 (**Figure 5**).

### Comparison of Uncertainties for Two Agro-Ecological Zones

Consistent with prior studies, NDRE, a useful proxy reflecting plant growth and development, has been deemed the most important variable for crop N status estimation (Osco et al., 2020; Colaço et al., 2021). Nonetheless, it is climatic and management factors that affect winter wheat growth and

**FIGURE 8 |** The performance of random forest regression using the variables selected from the Pearson correlation coefficient strategy to predict the aboveground biomass (AGB), plant N uptake (PNU) and N nutrition index (NNI) in agro-ecological zones 1 **(A,D,G)**, agro-ecological zone 2 **(B,E,H)** and two agro-ecological zones **(C,F,I)**. The black dotted line is the 1:1 line.

development. Accordingly, this study focused on the values of climatic and management factors for wheat N status estimation in different AZs. It is not surprising that these factors had different contributions in different AZs and for different N status indicators. Water-related variables played a more important role in AZ1, while both water-related and temperature-related variables were important in AZ2 (**Figures 9D–F**). Indeed, unlike AZ1, with a dry climate and sufficient sunshine, AZ2 belongs to a subtropical monsoon climate and is characterized by abundant precipitation and relatively less sunshine. In addition, the contributions of temperature-related variables to improve the estimation accuracy of AGB were greater than those of water-related variables, which were more correlated to PNU and NNI. A possible reason is that temperature has a greater effect on crop yield that is determined by the accumulation and redistribution of dry matter (Lee and Tollenaar, 2007) than rainfall (Schlenker
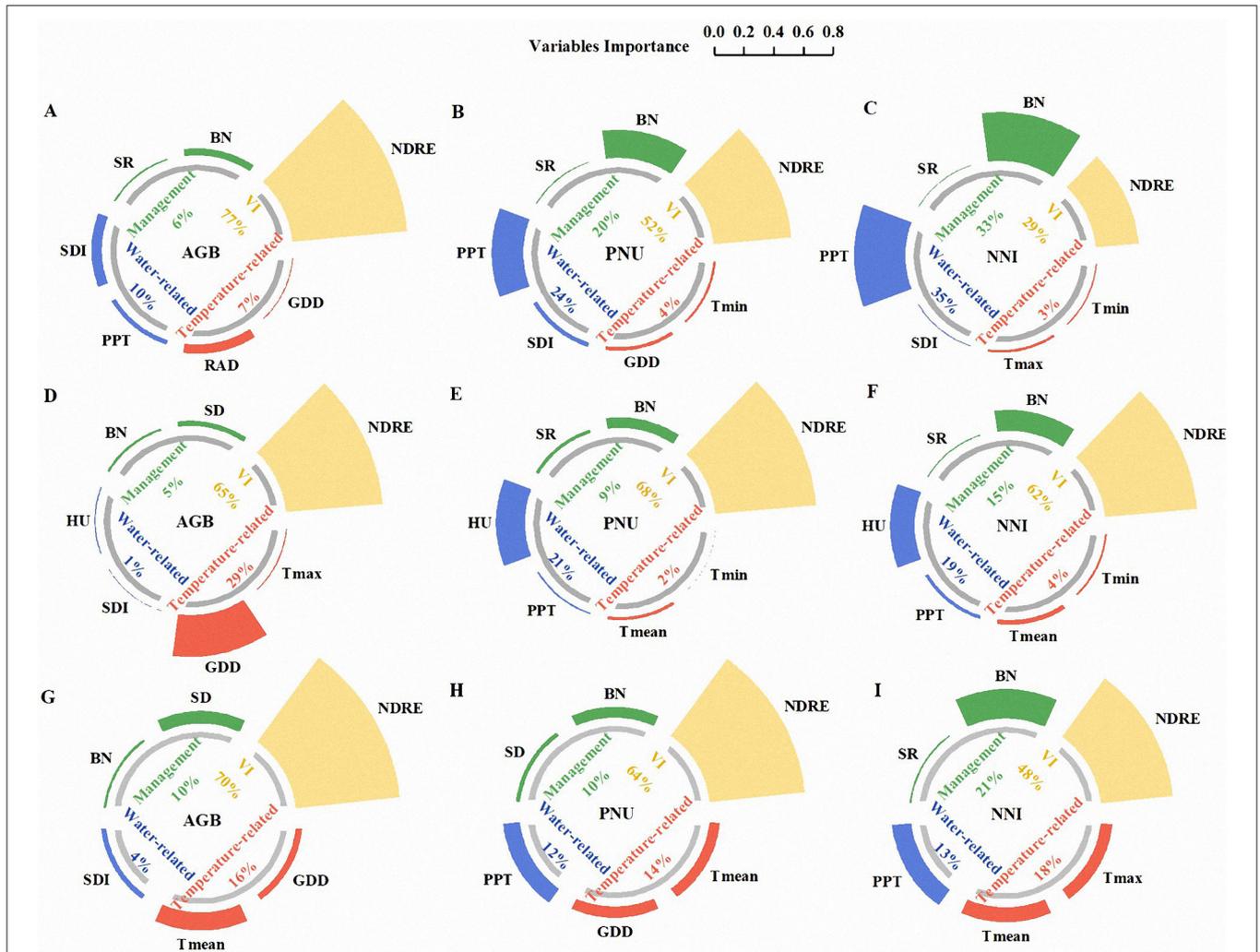
**FIGURE 9 |** The importance of variables of NDRE, climatic, and management factors for estimating the aboveground biomass, plant N uptake, and N nutrition index in agro-ecological zone 1 **(A–C)**, agro-ecological zone 2 **(D–F)**, and across two agro-ecological zones **(G–I)**. GDD, RAD, $T_{max}$, $T_{min}$, $T_{mean}$, PPT, SDI, HU, SD, SR, BN, AGB, PNU, and NNI refer to growing degree days, solar radiation, seasonal maximum, minimum, and mean temperature, seasonal total precipitation, shannon diversity index, relative humidity, sowing dates, seeding rates, basal N, aboveground biomass, plant N uptake, and N nutrition index, respectively.
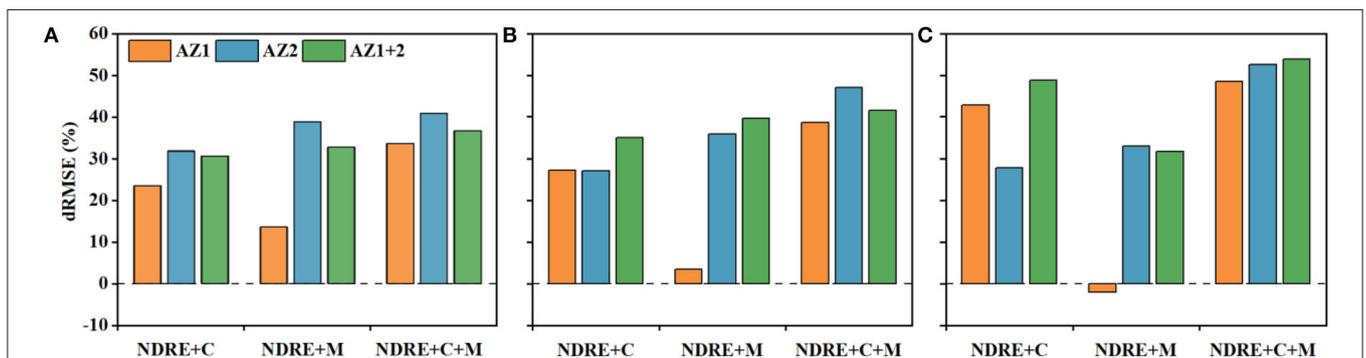


**FIGURE 10 |** The performance of the decrease of RMSE (dRMSE, %) of multi-source data for the aboveground biomass **(A)**, plant N uptake **(B)**, and N nutrition index **(C)** by combining climatic (C) factors and management (M) factors compared to normalized difference red edge (NDRE) only in agro-ecological zone 1 (AZ1), agro-ecological zone 2 (AZ2) and across two agro-ecological zones (AZ1+2).

and Lobell, 2010), because the dry matter of plants is formed by intercepting solar radiation and physiological and biochemical processes (Xue et al., 2002). Another possible reason is that the crop root system influenced by the water directly or indirectly is critical to absorbing nutrients from the soil, like N (Walsh et al., 2012; Sharma et al., 2018).

Overall, the results substantiated the viewpoint that integrating more sources of data will lead to better estimation performance (Chen et al., 2021; Gua et al., 2021; Liu et al., 2021; Li et al., 2022). Regarding their combinations, the results varied with sites and N status indicators. As for AGB, adding climatic factors improved the model performance more in AZ2 than in AZ1. Metabolic imbalances induced by lower average temperature during the vegetative stages could retard crop germination and plant growth (Ye et al., 2020). The difference of phenological information (SE stage, **Table 1**) would change AGB accumulation because it could, directly and indirectly, influence photosynthesis and respiration (Gua et al., 2021). On the contrary, adding the climatic factors improved the estimation of NNI in AZ1 more than AZ2, which could be caused by the influence of the length of the growth period or the responses of winter wheat cultivars to climate variability (Tao et al., 2014). In AZ1+2, the contributions of the management factors to PNU were greater than climatic factors because PNU showed a positive correlation to N rates (Egan et al., 2019; Sandaña et al., 2021). The sowing date in AZ2 was later than that in AZ1, leading to lower AGB accumulation, the number of tillers and plant height due to low temperature in the vegetative stages and higher temperature in the reproductive stages (Fazily, 2021). In contrast, a suitable sowing date can make full use of natural resources, such as light, heat, and water, enhancing plant population and facilitating the accumulation and uptake of N before winter. Different variables could interact with each other to influence crop growth and N status. For example, higher seeding rates could make up for stagnant tiller development, which would benefit cultivars with fewer tillers (Staggenborg et al., 2003).

## Applications and Future Work

In this study, the estimation models across two AZs showed similar results to that of a single AZ. Remarkably, it demonstrated the robustness and application of the established RFR approach.

While the results of this study were encouraging, there is still room for improvement. A total of 13 variables were considered in this study, which may be insufficient for constructing a reliable model for wider applications due to limitations in data amount and quality, model representativeness, and the number of key variables (Chlingaryan et al., 2018). Although the active canopy sensors are promising for evaluating N status, as indicated in this study, unmanned aerial vehicles and satellite remote sensing are more efficient for larger regions or farms (Huang et al., 2017; Chen et al., 2019). In this regard, follow-up research is needed to expand the scale of the study regions, supplement more climatic and management factors that directly or indirectly influence winter wheat growth, and develop more applicable N estimation models using unmanned aerial vehicles or satellite remote sensing images.

## CONCLUSION

The present work explored the adaptation and feasibility of RFR integrating NDRE, climatic, and management factors for N status estimation compared with SLR using only NDRE. The results indicated that RFR yielded better stability and higher accuracy. Variable selection was essential to construct effective models using fewer variables to achieve similar results compared to models using full variables, and PCC was confirmed to be an effective approach. Besides, dominant variables varied with AZs and N status indicators. At higher latitudes, climatic factors were more important to N status estimation, especially water-related factors. In addition, climatic factors significantly improved the performance of NNI estimation. Although climatic factors were crucial to AGB estimation, management factors were more important for N status estimation at lower latitudes. More studies are needed to develop unmanned aerial vehicles and satellite remote sensing-based ML models incorporating multi-source data for more efficient monitoring of crop N status under more diverse soil, climatic, and management conditions across large regions.

## DATA AVAILABILITY STATEMENT

The original contributions presented in the study are included in the article/supplementary material, further inquiries can be directed to the corresponding author.

## AUTHOR CONTRIBUTIONS

QC, WC, and YM conceived and designed the experiments. YL and QC performed the experiments, analyzed the data, and wrote the original manuscript. QC, YM, DC, JZ, and SL reviewed and revised the manuscript. All authors read and approved the final manuscript.

## FUNDING

## ACKNOWLEDGMENTS

# REFERENCES

Abraham, A., Pedregosa, F., Eickenberg, M., Gervais, P., Mueller, A., Kossaifi, J., et al. (2014). Machine learning for neuroimaging with scikit-learn. *Front. Neuroinform.* 8, 14. doi: 10.3389/fninf.2014.00014

Arlot, S., and Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Stat. Surv.* 4, 40–79. doi: 10.1214/09-ss054

Bean, G. M., Kitchen, N. R., Camberato, J. J., Ferguson, R. B., Fernandez, F. G., Franzen, D. W., et al. (2018). Improving an active-optical reflectance sensor algorithm using soil and weather information. *Agron. J.* 110, 1–11. doi: 10.2134/agronj2017.12.0733

Bonfil, D. J. (2017). Monitoring wheat fields by RapidScan: accuracy and limitations. *Adv. Anim. Biosci.* 8, 333–337. doi: 10.1017/s2040470017000589

Breiman, L. (2001). Random forests. *Mach. Learn.* 45, 5–32. doi: 10.1023/A:1010933404324

Cai, Y., Guan, K., Lobell, D., Potgieter, A. B., Wang, S., Peng, J., et al. (2019). Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches. *Agric. For. Meteorol.* 274, 144–159. doi: 10.1016/j.agrformet.2019.03.010

Cammarano, D., Basso, B., Holland, J., Gianinetti, A., Baronchelli, M., and Ronga, D. (2021). Modeling spatial and temporal optimal N fertilizer rates to reduce nitrate leaching while improving grain yield and quality in malting barley. *Comput. Electron. Agricult.* 182, 105997. doi: 10.1016/j.compag.2021.105997

Cao, Q., Miao, Y., Feng, G., Gao, X., Li, F., Liu, B., et al. (2015). Active canopy sensing of winter wheat nitrogen status: An evaluation of two sensor systems. *Comput. Electron. Agricult.* 112, 4–67. doi: 10.1016/j.compag.2014.08.012

Cao, Q., Miao, Y., Feng, G., Gao, X., Liu, B., Liu, Y., et al. (2017a). Improving nitrogen use efficiency with minimal environmental risks using an active canopy sensor in a wheat-maize cropping system. *Field Crops Res.* 214, 365–372. doi: 10.1016/j.fcr.2017.09.033

Cao, Q., Miao, Y., Li, F., Gao, X., Liu, B., Lu, D., et al. (2017b). Developing a new Crop Circle active canopy sensor-based precision nitrogen management strategy for winter wheat in North China Plain. *Precis. Agric.* 18, 2–18. doi: 10.1007/s11119-016-9456-7

Chen, X., Feng, L., Yao, R., Wu, X., Sun, J., and Gong, W. (2021). Prediction of maize yield at the city level in China using multi-source data. *Remote Sens.* 13, 146. doi: 10.3390/rs13010146

Chen, Z., Miao, Y., Lu, J., Zhou, L., Li, Y., Zhang, H., et al. (2019). In-season diagnosis of winter wheat nitrogen status in smallholder farmer fields across a village using unmanned aerial vehicle-based remote sensing. *Agronomy.* 9, 619. doi: 10.3390/agronomy9100619

Chlingaryan, A., Sukkarieh, S., and Whelan, B. (2018). Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review. *Comput. Electron. Agricult.* 151, 61–69. doi: 10.1016/j.compag.2018.05.012

Colaço, A. F., Richetti, J., Bramley, R. G. V., and Lawes, R. A. (2021). How will the next-generation of sensor-based decision systems look in the context of intelligent agriculture? A case-study. *Field Crops Res.* 270, 108205. doi: 10.1016/j.fcr.2021.108205

Cummings, C., Miao, Y., Paiao, G. D., Kang, S., and Fernández, F. G. (2021). Corn nitrogen status diagnosis with an innovative multi-parameter crop circle phenom sensing system. *Remote Sens.* 13, 401. doi: 10.3390/rs13030401

Dong, R., Miao, Y., Wang, X., Chen, Z., and Yuan, F. (2021). Improving maize nitrogen nutrition index prediction using leaf fluoresence sensor combined with environmental and management variables. *Field Crops Res.* 269, 108180. doi: 10.1016/j.fcr.2021.108180

Egan, G., Mckenzie, P., Crawley, M., and Fornara, D. A. (2019). Effects of grassland management on plant nitrogen use efficiency (NUE): evidence from a long-term experiment. *Basic Appl Ecol.* 41, 33–43. doi: 10.1016/j.baae.2019.10.001

Fazily, T. (2021). Effect of Sowing dates and seed rates on growth and yield of different wheat varieties: a review. *Int. J. Adv. Agri. Sci. Technol.* 8, 10–26. doi: 10.47856/ijaast.2021.v08i3.002

Feng, L., Zhang, Z., Ma, Y., Du, Q., Williams, P., Drewry, J., et al. (2020). Alfalfa yield prediction using UAV-based hyperspectral imagery and ensemble learning. *Remote Sens.* 12, 2018. doi: 10.3390/rs12122028

Gua, Y., Fua, Y., Hao, F., Zhang, X., Wu, W., Jin, X., et al. (2021). Integrated phenology and climate in rice yields prediction using machine learning methods. *Ecol. Indicators* 120, 106935. doi: 10.1016/j.ecolind.2020.106935

Hamrani, A., Akbarzadeh, A., and Madramootoo, C. A. (2020). Machine learning for predicting greenhouse gas emissions from agricultural soils. *Sci. Total Environ.* 741:140338. doi: 10.1016/j.scitotenv.2020.140338

Han, L., Yang, G., Dai, H., Xu, B., Yang, H., Feng, H., et al. (2019). Modeling maize above-ground biomass based on machine learning approaches using UAV remote-sensing data. *Plant Methods.* 15, 10. doi: 10.1186/s13007-019-0394-z

Heremans, S., Dong, Q., Zhang, B., Bydekerke, L., and Van Orshoven, J. (2015). Potential of ensemble tree methods for early-season prediction of winter wheat yield from short time series of remotely sensed normalized difference vegetation index andin situmeteorological data. *J. Appl. Remote Sens.* 9, 097095. doi: 10.1117/1.Jrs.9.097095

Huang, S., Miao, Y., Yuan, F., Gnyp, M., Yao, Y., Cao, Q., et al. (2017). Potential of rapidEye and worldView-2 satellite data for improving rice nitrogen status monitoring at different growth stages. *Remote Sens.* 9, 227. doi: 10.3390/rs9030227

Lee, E. A., and Tollenaar, M. (2007). Physiological basis of successful breeding strategies for maize grain yield. *Crop Sci.* 47, S-202–S-215. doi: 10.2135/cropsci2007.04.0010IPBS

Lemaire, G., Jeuffroy, M.-H., and Gastal, F. (2008). Diagnosis tool for plant and crop N status in vegetative stage. *Eur. J. Agron.* 28, 614–624. doi: 10.1016/j.eja.2008.01.005

Li, D., Miao, Y., Ranson, C. J., Bean, G. M., Kitchen, N. R., Fern?ndez, F. G., et al. (2022). Corn nitrogen nutrition index prediction improved by integrating genetic, environmental, and management factors with active canopy sensing using machine learning. *Remote Sens.* 14, 394. doi: 10.3390/rs14020394

Li, F., Zhang, H., Jia, L., Bareth, G., Miao, Y., and Chen, X. (2010). Estimating winter wheat biomass and nitrogen status using an active crop sensor. *Intell. Autom. Soft Comput.* 16, 1221–1230.

Li, S., Yuan, F., Ata-Ui-Karim, S. T., Zheng, H., Cheng, T., Liu, X., et al. (2019). Combining color indices and textures of UAV-based digital imagery for rice LAI estimation. *Remote Sens.* 11, 1763. doi: 10.3390/rs11151763

Liu, S., Jin, X., Nie, C., Wang, S., Yu, X., Cheng, M., et al. (2021). Estimating leaf area index using unmanned aerial vehicle data: shallow vs. deep machine learning algorithms. *Plant Physiol.* 187, 1551–1576. doi: 10.1093/plphys/kiab322

Lu, J., Miao, Y., Shi, W., Li, J., and Yuan, F. (2017). Evaluating different approaches to non-destructive nitrogen status diagnosis of rice using portable RapidSCAN active canopy sensor. *Sci. Rep.* 7, 14073. doi: 10.1038/s41598-017-14597-1

Lu, N., Zhou, J., Han, Z., Li, D., Cao, Q., Yao, X., et al. (2019). Improved estimation of aboveground biomass in wheat from RGB imagery and point cloud data acquired with a low-cost unmanned aerial vehicle system. *Plant Methods.* 15, 17. doi: 10.1186/s13007-019-0402-3

Meng, Q., Yue, S., Chen, X., Cui, Z., Ye, Y., Ma, W., et al. (2013). Understanding dry matter and nitrogen accumulation with time-course for high-yielding wheat production in China. *PLoS ONE.* 8, 1–9. doi: 10.1371/journal.pone.0068783

Miao, Y., Mulla, D. J., Randall, G. W., Vetsch, J. A., and Vintila, R. (2009). Combining chlorophyll meter readings and high spatial resolution remote sensing images for in-season site-specific nitrogen management of corn. *Precis. Agric.* 10, 45–62. doi: 10.1007/s11119-008-9091-z

Munoz-Huerta, R. F., Guevara-Gonzalez, R. G., Contreras-Medina, L. M., Torres-Pacheco, I., Prado-Olivarez, J., and Ocampo-Velazquez, R. V. (2013). A review of methods for sensing the nitrogen status in plants: advantages, disadvantages and recent advances. *Sensors.* 13, 10823–10843. doi: 10.3390/s130810823

Nelson, D. W., and Sommers, L. E. (1973). Determination of total nitrogen in plant material. *Agron. J.* 65, 109–112. doi: 10.2134/AGRONJ1973.00021962006500010033X

Osco, L. P., Junior, J. M., Ramos, A. P. M., Furuya, D. E. G., Santana, D. C., Teodoro, L. P. R., et al. (2020). Leaf nitrogen concentration and plant height prediction for maize using UAV-based multispectral imagery and machine learning techniques. *Remote Sens.* 12, 1–17. doi: 10.3390/rs12193237

Rhee, J., and Im, J. (2017). Meteorological drought forecasting for ungauged areas based on machine learning: using long-range climate forecast and remote sensing data. *Agric. For. Meteorol.* 237, 105–122. doi: 10.1016/j.agrformet.2017.02.011

Sandaña, P., Lobos, I. A., Pavez, P. B., and Moscoso, C. J. (2021). Nitrogen nutrition index and forage yield explain nitrogen utilization efficiency in hybrid ryegrasses under different nitrogen availabilities. *Field Crops Res.* 265, 108101. doi: 10.1016/j.fcr.2021.108101

Schlenker, W., and Lobell, D. B. (2010). Robust negative impacts of climate change on African agriculture. *Environ. Res. Lett.* 5, 014010. doi: 10.1088/1748-9326/5/1/014010

Sharma, L. K., Bali, S. K., Zaeen, A. A., Baldwin, P., and Franzen, D. W. (2018). Use of rainfall data to improve ground-based active optical sensors yield estimates. *Agron. J.* 110, 1561–1571. doi: 10.2134/agronj2017.12.0696

Staggenborg, S. A., Whitney, D. A., Fjell, D. L., and Shroyer, J. P. (2003). Seeding and nitrogen rates required to optimize winter wheat yields following grain sorghum and soybean. *Agron. J.* 95, 253–259. doi: 10.2134/agronj2003.2530

Tao, F., Zhang, Z., Xiao, D., Zhang, S., Rötter, R. P., Shi, W., et al. (2014). Responses of wheat growth and yield to climate change in different climate zones of China, 1981–2009. *Agric. For. Meteorol.* 189, 91–104. doi: 10.1016/j.agrformet.2014.01.013

Walsh, O. S., Klatt, A. R., Solie, J. B., Godsey, C. B., and Raun, W. R. (2012). Use of soil moisture data for refined GreenSeeker sensor based nitrogen recommendations in winter wheat (Triticum aestivum L.). *Precis. Agric.* 14, 343–356. doi: 10.1007/s11119-012-9299-9

Wang, L., Zhou, X., Zhu, X., Dong, Z., and Guo, W. (2016). Estimation of biomass in wheat using random forest regression algorithm and remote sensing data. *The Crop J.* 4, 212–219. doi: 10.1016/j.cj.2016.01.008

Wang, X., Miao, Y., Dong, R., Chen, Z., Guan, Y., Yue, X., et al. (2019). Developing active canopy sensor-based precision nitrogen management strategies for maize in Northeast China. *Sustainability.* 11, 706. doi: 10.3390/su11030706

Wang, X., Miao, Y., Dong, R., Zha, H., Xia, T., Chen, Z., et al. (2021). Machine learning-based in-season nitrogen status diagnosis and side-dress nitrogen recommendation for corn. *Eur. J. Agron.* 123, 126193. doi: 10.1016/j.eja.2020.126193

Wang, X., Wan, Q., Fan, J., Su, L., and Shen, X. (2014). Logistic model analysis of winter wheat growth on China's Loess Plateau. *Can. J. Plant Sci.* 94, 1471–1479. doi: 10.4141/cjps2013-293

Wang, Y., Zhang, Z., Feng, L., Du, Q., and Runge, T. (2020). Combining multi-source data and machine learning approaches to predict winter wheat yield in the Conterminous United States. *Remote Sens.* 12, 1232. doi: 10.3390/rs12081232

Xue, Q., Soundararajan, M., Weiss, A., Arkebauer, T. J., and Stephen Baenziger, P. (2002). Genotypic variation of gas exchange parameters and carbon isotope discrimination in winter wheat. *J. Plant Physiol.* 159, 891–898. doi: 10.1078/0176-1617-00780

Yang, H., Dobermann, A., Cassman, K. G., and Walters, D. T. (2004). *A Simulation Model for Corngrowth and Yiled*. Lincoln: Nebraska Cop. Ext.CD 9, Univ. of Nebraska.

Ye, Z., Qiu, X., Chen, J., Cammarano, D., Ge, Z., Ruane, A. C., et al. (2020). Impacts of 1.5 °C and 2.0 °C global warming above pre-industrial on potential winter wheat production of China. *Eur. J. Agron.* 120, 126149. doi: 10.1016/j.eja.2020.126149

Yue, S., Meng, Q., Zhao, R., Li, F., Chen, X., Zhang, F., et al. (2012). Critical nitrogen dilution curve for optimizing nitrogen management of winter wheat production in the North China Plain. *Agron. J.* 104, 523–529. doi: 10.2134/agronj2011.0258

Zha, H., Miao, Y., Wang, T., Li, Y., Zhang, J., Sun, W., et al. (2020). Improving unmanned aerial vehicle remote sensing-based rice nitrogen nutrition index prediction with machine learning. *Remote Sens.* 12, 215. doi: 10.3390/rs12020215

Zhang, L., Zhang, Z., Luo, Y., Cao, J., and Tao, F. (2019). Combining optical, fluorescence, thermal satellite, and environmental data to predict county-level maize yield in China using machine learning approaches. *Remote Sens.* 12, 21. doi: 10.3390/rs12010021

Zhao, B., Yao, X., Tian,Yongchao, L.iu, X., Cao, W., and Zhu, Y. (2012). Accumulative nitrogen deficit models of wheat aboveground part based on critical nitrogen concentration. *Chin. J. Appl. Ecol.* 23, 3141–3148. doi: 10.13287/j.1001-9332.2012.0470

Zhou, L., Chen, G., Miao, Y., Zhang, H., Chen, Z., Xu, L., et al. (2017). Evaluating a Crop Circle active sensor-based in-season nitrogen management algorithm in different winter wheat cropping systems. *Adv. Anim. Biosci.* 8, 364–367. doi: 10.1017/s2040470017000292

Zhu, X., Guo, R., Liu, T., and Xu, K. (2021). Crop yield prediction based on agrometeorological indexes and remote sensing Data. *Remote Sens.* 13, 2016. doi: 10.3390/rs13102016

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

**Publisher's Note:** All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

# APPENDIX

**TABLE A1 |** Symbols and acronyms used in the article.

| | |
|---|---|
| AGB | aboveground biomass |
| AWDR | abundant and well-distributed rainfall |
| AZ1 | agro-ecological zone 1 |
| AZ1+2 | two agro-ecological zones |
| AZ2 | agro-ecological zone 2 |
| BN | basal N |
| $D_{max}$ | daily maximum temperature |
| $D_{min}$ | daily minimum temperature |
| dRMSE | the decrease of RMSE |
| FM | farmer's management |
| GDD | growing degree days |
| HU | relative humidity |
| N | Nitrogen |
| $N_c$ | critical N concentration |
| NDRE | normalized difference red edge |
| NDRE+C | both NDRE and climatic data |
| NDRE+C+M | full sources of data |
| NDRE+M | both NDRE and management data |
| NNI | N nutrition index |
| PCC | Pearson correlation coefficient |
| PNM | precision N management |
| PNU | plant N uptake |
| PPT | seasonal total precipitation |
| RAD | solar radiation |
| RFR | random forest regression |
| RN | the relative number of variables |
| RONM | regional optimum N management |
| $RR^2$ | relative $R^2$ |
| RRE | relative RE |
| RRMSE | relative RMSE |
| SDI | shannon diversity index |
| SE | stem elongation |
| SLR | simple linear regression |
| $T_{base}$ | base temperatures |
| $T_{max}$ | seasonal maximum temperature |
| $T_{mean}$ | seasonal mean temperature |
| $T_{min}$ | seasonal minimum temperature |
| VIRRFR | the variable importance ranked by the RFR |
| VIs | vegetation indices |