agronomy

MDPI

Article

NDVI-Driven Dynamic Granger–Ramanathan Model by UAV Hyperspectral and Thermal Infrared Imagery for Leaf Water Content Estimation in Winter Wheat

Menglei Dai [1,2,†], Limin Gu [1,†], Baoyuan Zhang [2,3], Wen Jin [1], Pengzhao Liu [1], Baozhong Yin [4], Chong Zhang [5], Xiaohe Gu [2] and Wenchao Zhen [1,*]

[1] State Key Laboratory of North China Crop Improvement and Regulation/College of Agronomy, Hebei Agricultural University/Key Laboratory of North China Water-Saving Agriculture, Ministry of Agriculture and Rural Affairs, Baoding 071001, China; menglei_dai@pgs.hebau.edu.cn (M.D.); gulimin@hebau.edu.cn (L.G.); jinwen@hebau.edu.cn (W.J.); liupz0819@foxmail.com (P.L.)
[2] Research Center of Information Technology, Beijing Academy of Agriculture and Forestry Sciences, Beijing 100097, China; baoyuan_zhang@yeah.net (B.Z.); guxh@nercita.org.cn (X.G.)
[3] College of Agriculture, Nanjing Agricultural University, Nanjing 210095, China
[4] College of Plant Protection, Hebei Agricultural University, Baoding, 071001, China; yinbaozhong@hebau.edu.cn
[5] Hebei Dadi Seed Industry Co., Ltd., Shijiazhuang 052300, China; zhangchong0309@163.com
* Correspondence: wenchao@hebau.edu.cn
† These authors contributed equally to this work.

Abstract

Leaf water content (LWC) is a key indicator of winter wheat water status, but the reliability of UAV hyperspectral and thermal infrared predictions varies across samples, limiting the effectiveness of fixed-weight fusion. To improve LWC estimation by integrating hyperspectral and thermal infrared data, this study developed a dynamic Granger–Ramanathan averaging (GRA) method. Sensitive bands were selected using the successive projections algorithm, and random forest, XGBoost, and LightGBM models were constructed for each data source. Generalized additive models were used to characterize relationships between canopy–air temperature difference (ΔT), normalized difference vegetation index (NDVI), and single-source prediction errors, enabling dynamic adjustment of fusion weights. Random forest performed best among the single-source models, with test-set R^2 values of 0.7871 and 0.6275 for hyperspectral and thermal infrared data, respectively. Conventional GRA achieved an R^2 of 0.7645, whereas ΔT- and NDVI-driven dynamic GRA increased R^2 to 0.8207 and 0.8224, respectively. The NDVI-driven model achieved the lowest RMSE (0.0646) and MAE (0.0497). Applied to 720 samples, it achieved an R^2 of 0.8507, an RMSE of 0.0591, and an MAE of 0.0451. These findings demonstrate the potential of crop-status-driven dynamic weighting to improve multisource LWC estimation in winter wheat.

Keywords: dynamic model averaging; generalized additive model; successive projections algorithm; canopy–air temperature difference; vegetation index

Check for updates

Academic Editors: Junying Chen and Gniewko Niedbała

Received: 19 September 2026
Revised: 10 October 2026
Accepted: 23 October 2026
Published: 24 October 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

Leaf water content (LWC) is an important indicator of crop water status, and its variation is closely related to plant water deficit, transpiration regulation, and growth status. Conventional LWC measurements generally rely on sampling and laboratory analysis, which are destructive, labor-intensive, and spatially limited, making them unsuitable for large-scale and continuous crop water monitoring [1]. With the development of UAV remote

sensing, rapid estimation of crop water status using high-resolution remote sensing data has become an important topic in precision agriculture [2].

Hyperspectral and thermal infrared remote sensing provide information on crop water status from different perspectives. Hyperspectral data contain continuous narrow-band spectral information and can capture spectral variations related to vegetation water content, structure, and biochemical status [3]. Thermal infrared data provide canopy temperature and its spatial variation, offering another type of information for characterizing crop water status [4]. A single sensor can be used to estimate crop water-related parameters, but its predictions can be affected by crop growth stage, environmental conditions, and sensor-specific information. Therefore, fusing information from different sensors provides a potential approach for improving the stability of water parameter estimation [5].

Multi-source remote sensing information can generally be fused at the data, feature, and model-output levels. At the model-output level, prediction models are first developed separately for different sensors, and their predictions are then combined. This approach can avoid the high dimensionality and redundancy associated with directly integrating large numbers of heterogeneous variables. Granger and Ramanathan proposed an improved forecast combination method in 1984, in which multiple predictions are linearly combined with a constant term to obtain the combined prediction [6]. Granger–Ramanathan (GRA) model averaging was subsequently applied to hydrologic forecasting [7], digital soil mapping [8], and multi-sensor spectral analysis [9,10]. GRA can make use of complementary information among model predictions and has achieved better predictions than individual models for some soil properties and environmental variables. In recent years, GRA has also been applied to agricultural remote sensing and crop water parameter estimation. For example, GRA has been compared with other multi-sensor fusion methods for soil property estimation [11], and hyperspectral and thermal infrared information have been used to build single-sensor models whose predictions were fused with GRA to estimate canopy water content in winter wheat, achieving higher accuracy than the individual hyperspectral and thermal infrared models [12]. These studies show that model-output-level multi-sensor fusion can make better use of differences among sensors and provide an effective approach for estimating crop water parameters. However, most existing GRA applications use fixed weights, meaning that the same model combination coefficients are applied to the entire sample set [13]. Fixed weights assume that the relative predictive abilities of models from different sensors remain generally stable across sample conditions. However, prediction errors in crop remote sensing often depend strongly on crop status. As growth stage, canopy structure, and water status change, the responses of different sensors to the target variable may also change. Therefore, global fixed weights cannot adequately describe changes in the reliability of single-source models across samples. Approaches that extract dynamic environmental features from time-series remote sensing data and combine them with model averaging have also been used for cropland soil organic matter mapping [14]. These studies provide a basis for dynamic model averaging, but studies on dynamic GRA for coordinated LWC estimation using UAV hyperspectral and thermal infrared data remain limited.

Therefore, this study aimed to develop a dynamic method for estimating winter wheat leaf water content (LWC) by fusing UAV hyperspectral and thermal infrared information and dynamically adjusting the contributions of multi-source remote sensing models using crop-status-related driving factors. The specific objectives were to (1) develop single-feature-source LWC inversion models using hyperspectral and thermal infrared features, compare the predictive performance of different base learners, and identify the best model for each feature source; (2) establish a fixed-weight GRA model based on the predictions of the single-feature-source models, and further introduce ΔT and NDVI as dynamic driving factors to characterize the relationships between the driving factors and single-source model

prediction errors using GAM, thereby enabling continuous adjustment of model reliability and fusion weights; and (3) compare the estimation performance of conventional GRA, ΔT-driven dynamic GRA, and NDVI-driven dynamic GRA, identify the best fusion model, and apply it to map the spatiotemporal distribution of LWC.

2. Materials and Methods

2.1. Experimental Design

Field experiments were conducted in Xinji, Hebei Province, China, during the 2022–2023 and 2023–2024 winter wheat growing seasons. The site has a warm-temperate, semi-humid continental monsoon climate characterized by distinct seasons, dry and windy springs, hot and rainy summers, and cold, relatively dry winters. Precipitation is unevenly distributed throughout the year and is concentrated mainly in summer. According to the annual climate reports issued by the Shijiazhuang Meteorological Bureau, the regional mean annual precipitation was 590.5 mm in 2022, 741.4 mm in 2023, and 555.2 mm in 2024. These values indicate considerable interannual variability in regional precipitation during the experimental period. However, local precipitation conditions at the experimental site may differ from the regional averages. GY 2018 was used in 2022–2023, while GY 2018 and JM 22 were used in 2023–2024. Multiple water and nitrogen supply levels were established to increase sample diversity and capture water and nitrogen supply regimes representative of different locations. Irrigation was defined according to crop water requirement (CWR), with four irrigation levels: 0 CWR (no irrigation), 50% CWR, 75% CWR, and 100% CWR, denoted W0, W50, W75, and W100, respectively. Adjacent plots were separated by buffer walkways, and drip irrigation was used to minimize water movement between plots and reduce interference among irrigation treatments. The experimental soil was classified as loam. Five N application levels were used: 0, 180, 240, 300, and 360 kg N ha^{-1}, denoted N0, N1, N2, N3, and N4, respectively. The N3 treatment represented the conventional N application rate in the local area, whereas N0 served as the zero-N control. N1 and N2 represented moderate and mild N deficiency, respectively, while N4 represented excessive N application. The field experiment employed a two-factor orthogonal experimental design involving irrigation amount and nitrogen application rate, with three replicates for each treatment. The plot dimensions were 4.5 m × 6 m in the 2022–2023 growing season and 3.5 m × 4 m in the 2023–2024 growing season. The spatial arrangement of the experimental treatments is shown in Figure 1. Nitrogen fertilizer was applied at a 1:1 ratio before sowing and at the jointing stage, with topdressing applied by broadcasting. Phosphorus and potassium fertilizers were both applied at 120 kg ha^{-1} before sowing. The planting density was 375 plants per sqm.

Other field management practices, including pest and disease control and other routine agronomic measures, followed standard practices for high-standard winter wheat production. Before sowing, the field was deeply plowed and prepared to provide suitable soil conditions for crop establishment. Before winter dormancy, the wheat plants were rolled to promote root development, strengthen seedlings, and conserve soil moisture. During the regreening stage, a small amount of fertilizer and irrigation water were applied to support crop recovery and growth, while ensuring that these supplementary inputs did not substantially alter the experimental treatment gradients. Pesticides were applied as needed to control pests and diseases. At the jointing stage, fertilizer and irrigation water were applied quantitatively to each plot according to the designated experimental treatments. During flowering and grain filling, irrigation was carefully scheduled according to crop water requirements. Foliar applications, primarily consisting of monopotassium phosphate, insecticides, and fungicides, were used to supplement potassium, support grain filling,

and control major pests and diseases, including aphids, wheat blossom midges, powdery mildew, and rust.

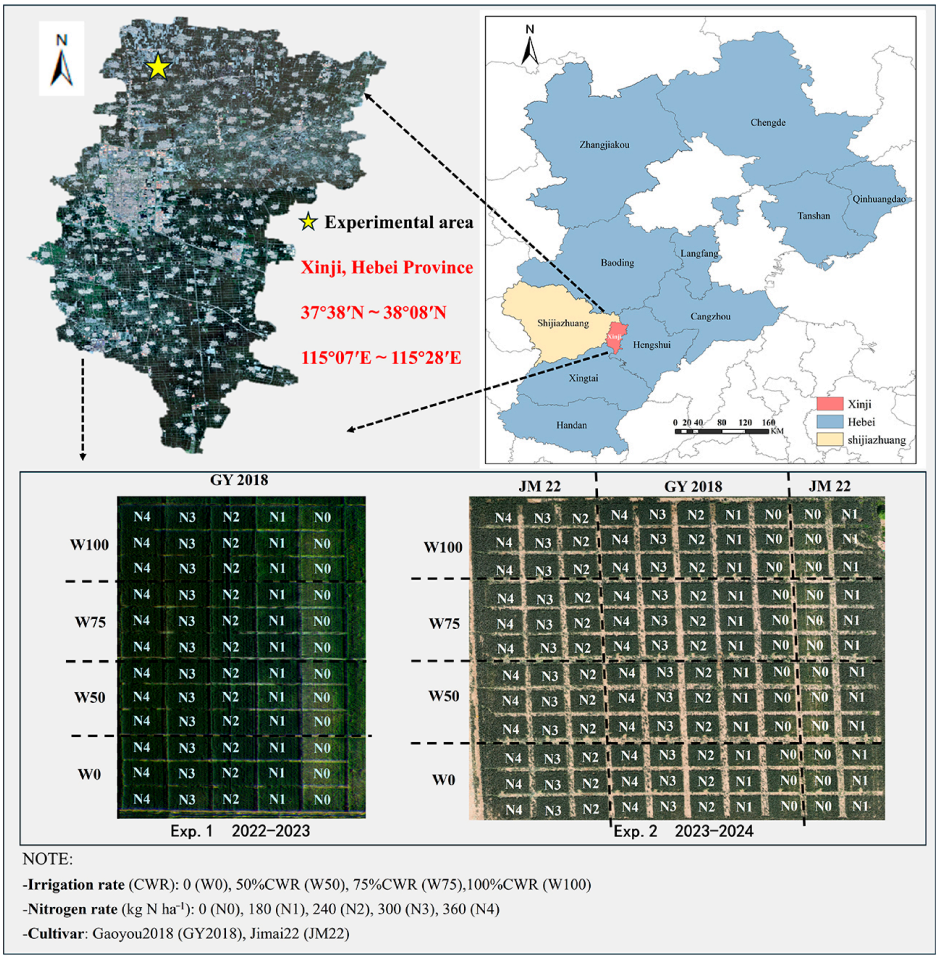


Figure 1. Geographic locations of the regional experimental sites.

2.2. Data Collection Test

UAV imagery was acquired on seven dates during the two consecutive winter wheat growing seasons (Table 1), with a total of 720 samples. Leaf water content was measured synchronously on the ground during each observation period, and UAV imaging and ground measurements were conducted simultaneously for all plots.

UAV imagery: Images were acquired under clear, cloud-free conditions between 10:00 and 14:00 using a DJI M300 UAV equipped separately with a Cubert X20 Pro hyperspectral imaging camera (Cubert GmbH, Ulm, Germany) and a WIRIS Pro Sc thermal imaging camera (Workswell, Prague, Czech Republic) to obtain hyperspectral and thermal infrared images of winter wheat canopies (Table 2). Before hyperspectral image acquisition, a calibration panel and lens cap were used for white-reference and dark-current correction to reduce the effects of differences in illumination among weather conditions on the reflectance spectra. Hyperspectral images were mosaicked using Cubert Utils Touch software (https://cubert-hyperspectral.com/en/ultris-x20-plus/, accessed on 22 October 2026), while thermal infrared images were mosaicked using Pix4Dmapper 4.5.6. The images were then converted to the required format and processed for feature extraction in ENVI 5.3.

Table 1. Dates of data collection and corresponding growth stages.

	Growth Stages	Measuring Date	Days of the Year	Number of Plots
Exp. 1 2022–2023	Early stage of grain filling	8 May	128 in 2023	60
	Middle stage of grain filling	13 May	133 in 2023	60
Exp. 2 2023–2024	Boosting stage	16 Apr	107 in 2024	120
	Anthesis stage	26 Apr	117 in 2024	120
	Anthesis to grain filling	2 May	123 in 2024	120
	Early stage of grain filling	8 May	129 in 2024	120
	Middle stage of grain filling	14 May	135 in 2024	120

Table 2. Parameters of the hyperspectral and thermal cameras.

	Hyperspectral Camera		Thermal Camera		
	Angle of field	35°	Viewing angle	68°	
	Digital resolution	12 Bit	Frame rate	30 Hz	
	Bandwidth (nominal)	4 nm	Sensitivity	0.03 °C	
	Image resolution	1886 × 1886	Image resolution	1266 × 1010	
	Wavelength range	350–1002 nm	Wavelength range	7.5–13.5 µm	

LWC: A representative and relatively uniform area was selected from each plot for destructive sampling. The sampling area covered two 0.5 m rows (row spacing, 15 cm; 0.5 m × 0.15 m × 2 = 0.15 m^2). The samples were immediately brought indoors for organ separation, and the collected leaves were grouped by treatment. Leaf fresh weight (LFW) was first measured. The leaves were then placed in an oven at 105 °C for 30 min for deactivation and subsequently dried at 80 °C for 24 h to constant weight. The final dry weight was recorded as leaf dry weight (LDW).

$$LWC = (LFW - LDW)/LFW \quad (1)$$

2.3. Data Processing

All scripts used in this study were developed in a Python 3.11 environment and executed using the PyCharm 2022.3.2 interpreter. The versions of the software packages are listed in Table 3.

Table 3. Versions of Python software packages used in this study.

Name	Version	Name	Version	Name	Version
click	8.4.2	pip	22.3.1	rasterio	1.4.4
joblib	1.5.3	pandas	3.0.0	openpyxl	3.1.5
numpy	2.4.2	shapely	2.1.2	geopandas	1.1.2
scikit-learn	1.8.0	matplotlib	3.10.8	skl-migrator	0.21.1

2.3.1. Dimensionality Reduction in Hyperspectral Reflectance

The original hyperspectral reflectance data contained 164 consecutive bands, which may include multiple features with similar explanatory ability for LWC. These redundant features increase computational burden and may reduce model accuracy because of multicollinearity. The SPA algorithm was therefore used to extract a sensitive band subset from the original reflectance data. SPA is a forward iterative search method that starts with one feature and adds one new variable at each iteration until the specified number of variables, N, is reached. The variable number corresponding to the minimum root mean square error (RMSE) is selected. The SPA procedure used n_splits = 10, shuffle = True, and random_state = 42 during the iterations.

2.3.2. Extraction of Canopy Temperature Features

The data used in this study included thermal infrared images (.tif) and vector boundary files (.shp). Thermal infrared images were used to extract land surface temperature information, while vector boundary files defined the spatial extent of the study area. The coordinate reference system (CRS) of the vector files was kept consistent with that of the thermal infrared images. Open-source Python libraries, including rasterio, geopandas, numpy, and pandas, were used to automate image processing and temperature feature extraction, including image reading and loading, image clipping and temperature statistics, temperature feature calculation, and output generation. The extracted temperature features included:

(1) temp_average: mean canopy temperature within a plot, which provides the basis for calculating other temperature features.
(2) temp_range: difference between the maximum and minimum canopy temperatures within a plot.
(3) temp_amended: temperature regulation capacity of the plants within a plot.

$$temp_amended = (T_0 - temp_average)/T_0 \quad (2)$$

Bare soil temperature at the study site during UAV image acquisition.

(4) CTSD (temperature standard deviation): standard deviation of valid temperature values within the study area, representing spatial temperature variability.

$$TSD = \sqrt{\frac{1}{(n-1)}\sum_{i=1}^{n}(T_i - temp_average)^2} \quad (3)$$

n is the total number of valid pixels within the plot canopy, and each pixel has an actual temperature value.

(5) CTCV (temperature coefficient of variation): ratio of temperature standard deviation to mean temperature, used to evaluate relative variability.

$$CTCV = CTSD/temp_average \quad (4)$$

2.3.3. Base Learner Models and Accuracy Assessment

Random forest (RF), extreme gradient boosting (XGBoost), and Light Gradient Boosting Machine (LightGBM) were used to construct single-source LWC inversion models. Hyperspectral and thermal infrared features were used separately as input variables, with measured LWC as the dependent variable, to develop LWC prediction models for the two feature sources. RF builds multiple regression trees and integrates their predictions to capture nonlinear relationships between single-source features and LWC (n_estimators

= 10, max_depth = 4, and random_state = 41). XGBoost sequentially fits the prediction residuals of the previous model and adds regression trees to improve prediction accuracy (n_estimators = 20, learning_rate = 0.03, max_depth = 5, subsample = 0.8, colsample_bytree = 0.8, and random_state = 41). LightGBM uses a gradient boosting framework to construct regression trees iteratively and reduce prediction errors (n_estimators = 20, learning_rate = 0.03, max_depth = 5, random_state = 42, verbose = −1, verbosity = −1, and force_col_wise = True). The predictive performance of the three base learners was compared for the hyperspectral and thermal infrared feature sources to identify the best base learner for each source. Their predictions were then used as inputs to the subsequent GRA fusion model.

The coefficient of determination (R^2), root mean square error (RMSE), and mean absolute error (MAE) were used to evaluate the predictive accuracy and stability of all estimation models, as follows:

$$R^2 = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \quad (5)$$

$$RMSE = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(y_i - \overline{y}_i)^2} \quad (6)$$

$$MAE = \frac{1}{n}\sum_{i=1}^{n}\left|y_i - \overline{y}_i\right| \quad (7)$$

y_i is the measured value, $\hat{y}_i$ is the predicted value, and $\overline{y}_i$ is the average value.

R^2 is the square of the correlation coefficient between predicted and measured values, with higher values indicating stronger predictive ability. RMSE and MAE measure the deviations between predicted and measured values, with lower values indicating smaller deviations.

2.3.4. GRA Strategy

Granger–Ramanathan Averaging (GRA) is a model averaging method that linearly combines predictions from multiple models. Its basic principle is to use complementary information among model predictions and determine the combination coefficients by regression to obtain the final prediction. In this study, LWC inversion models were first developed separately using hyperspectral and thermal infrared features, and the predictions of the two best single-feature-source models were denoted as y_HS and y_TI, respectively. Based on these predictions, measured LWC was used as the dependent variable and the predictions of the two single-feature-source models were used as independent variables to establish a conventional fixed-weight GRA fusion model

$$\mathrm{LWC} = b + \omega_{HS} y_{HS} + \omega_{TI} y_{TI} \quad (8)$$

where b is the intercept, and the two terms represent the fusion coefficients for the hyperspectral and thermal infrared models, respectively.

Because the prediction errors of the hyperspectral and thermal infrared models can vary across sample conditions, a dynamic weighting strategy was further developed based on conventional GRA. First, the absolute prediction errors of the two single-feature-source models were calculated separately

$$\mathrm{e} = |y_0 - y| \quad (9)$$

where LWC is the measured value and the prediction term represents the predicted value. The canopy–air temperature difference (ΔT) and NDVI were then used separately as driving factors. GAM was used to characterize the nonlinear relationships between the driving factors and the prediction errors of the single-feature-source models, and the corresponding model reliability was calculated from the expected prediction error

$$R_{HS} = 1/f_{HS}(\gamma) \quad (10)$$

$$R_{TI} = 1/f_{TI}(\gamma) \tag{11}$$

where the first term represents the driving factor, and the other terms represent the corresponding model reliabilities. The reliabilities were further normalized to obtain dynamic weights

$$\omega_{HS} = R_{HS}/(R_{HS} + R_{TI}), \tag{12}$$

$$\omega_{TI} = R_{TI}/(R_{HS} + R_{TI}). \tag{13}$$

The ΔT- and NDVI-driven dynamic GRA models were constructed separately and compared with conventional fixed-weight GRA to evaluate the effects of different fusion strategies on LWC estimation in winter wheat.

3. Results and Analysis

3.1. Responses of UAV Features to Different Water and Nitrogen Supply Conditions

During SPA dimensionality reduction, RMSECV decreased rapidly as the number of selected variables increased (Figure 2a). When the number of variables reached four, RMSECV decreased to 0.089. The decrease then became much smaller, remaining near 0.0885 with 5–8 variables. When nine variables were selected, RMSECV further decreased to approximately 0.085, the lowest value during the selection process. After increasing the number of variables to 10, RMSECV increased slightly. Therefore, nine sensitive bands were finally selected for modeling.

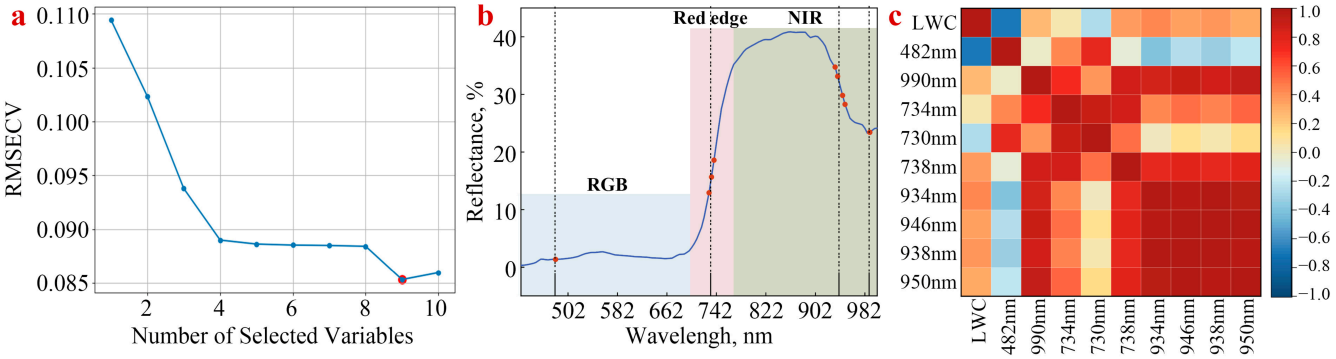


Figure 2. Dimensionality reduction in hyperspectral reflectance and selection of sensitive bands ((**a**) variation in RMSECV during SPA dimensionality reduction; (**b**) distribution of sensitive bands; (**c**) correlations among sensitive bands).

The nine selected bands were 482, 730, 734, 738, 934, 938, 946, 950, and 990 nm (Figure 2b). The 482 nm band is located in the visible blue region, 730–738 nm is concentrated in the red-edge region, and 934–990 nm is mainly in the near-infrared region. The selected features did not evenly cover the entire 350–1002 nm spectral range but were concentrated in several regions with clear spectral variation, including the blue, red-edge, and near-infrared regions.

The nine selected bands showed different levels of correlation (Figure 2c). Strong positive correlations were observed among 730, 734, and 738 nm, and among 934, 938, 946, and 950 nm. The 990 nm band was also strongly correlated with several near-infrared and red-edge bands. In contrast, 482 nm was relatively weakly correlated with most selected bands and was negatively correlated with LWC. LWC was generally positively correlated with the 730–738 and 934–990 nm bands, whereas 482 nm showed the opposite correlation direction. The resulting nine-band combination retained information from different spectral regions while reducing substantial redundancy in the original hyperspectral data.

Under different irrigation levels, canopy temperature variability and temperature range showed clear responses to water supply (Figure 3). CTSD increased from approximately 0.19 °C under W100 to approximately 0.24 °C under W75 and W50 and further increased to approximately 0.28 °C under W0. CTCV also increased gradually from approximately 0.009 to 0.013. Meanwhile, TMP_Range was approximately 1.7, 1.3, and 1.5 °C under W100, W75, and W50, respectively, and increased to approximately 2.1 °C under W0. TMP_Amended remained generally within 4.2–4.8 °C, and the differences among irrigation treatments were smaller than those in TMP_Range. Nitrogen supply had a more pronounced effect on canopy temperature features. As the N level decreased from N4 to N0, CTSD generally increased from approximately 0.29 °C to 0.84 °C, while CTCV increased from approximately 0.013 to 0.041. TMP_Range remained approximately 1.7–2.1 °C under N4–N1 but increased sharply to approximately 7.3 °C under N0. TMP_Amended increased gradually from approximately 3.8 °C under N4 to approximately 4.8 and 5.5 °C under N2 and N1, respectively, and was approximately 5.4 °C under N0. All canopy temperature features could distinguish water and nitrogen supply conditions to some extent, but their responses to the treatment gradients differed in magnitude.

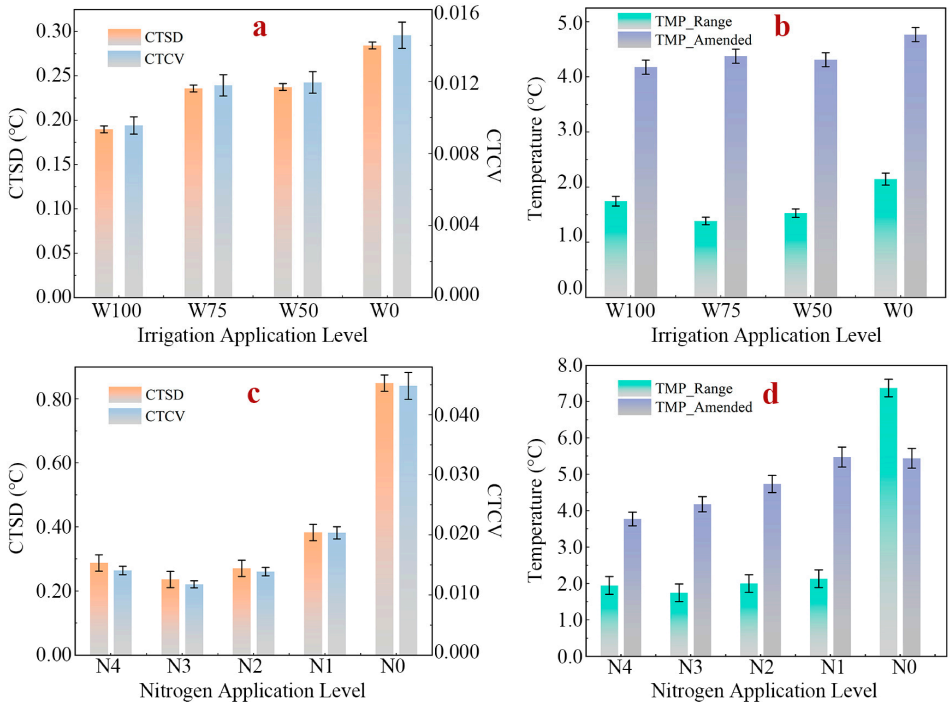


Figure 3. Responses of canopy temperature features to different water and nitrogen supply conditions ((**a**) CTSD and CTCV under different irrigation levels; (**b**) TMP_Range and TMP_Amended under different irrigation levels; (**c**) CTSD and CTCV under different nitrogen application rates; (**d**) TMP_Range and TMP_Amended under different nitrogen application rates).

3.2. Single-Feature-Source Modeling Results

For the hyperspectral feature source, the RF model had a clearly higher R^2 than XGBoost and LightGBM (Figure 4). Their RMSE values were approximately 0.071, 0.097, and 0.101, respectively, and their MAE values were approximately 0.052, 0.068, and 0.069. For the thermal infrared feature source, RF also achieved the highest R^2, approximately 0.63, whereas XGBoost and LightGBM achieved approximately 0.36 and 0.37, respectively. The RF RMSE was approximately 0.094, lower than the approximately 0.12 obtained by XGBoost and LightGBM, and its MAE decreased from approximately 0.085–0.086 to approximately 0.072. Therefore, RF was used as the base predictor for both feature sources in the sub-

sequent fusion model. The hyperspectral RF model had higher overall accuracy than the thermal infrared RF model, particularly in R^2, with values of approximately 0.78 and 0.63, respectively. Its RMSE and MAE were also lower than those of the thermal infrared model. The two independent feature sources therefore provided complementary base models with different predictive abilities for subsequent GRA fusion.

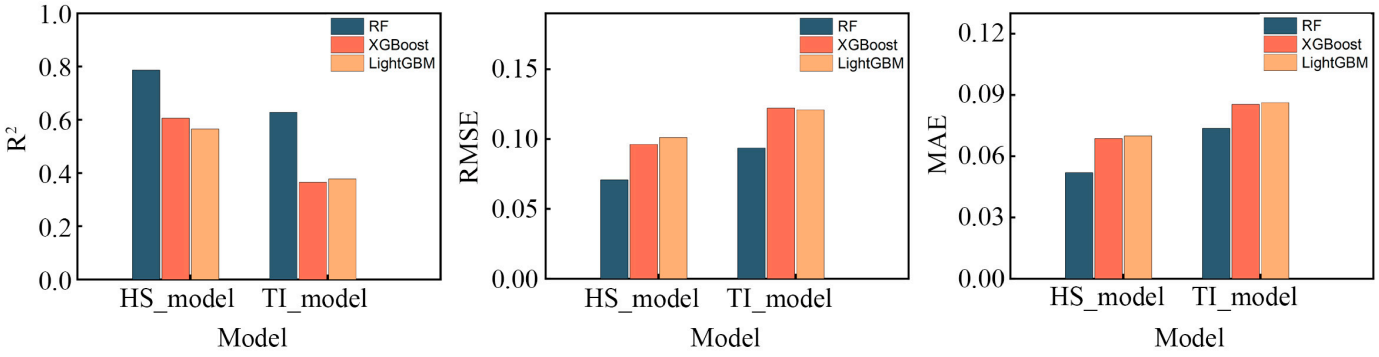


Figure 4. Prediction accuracy of single-feature-source models using different base learners.

Predictions from the hyperspectral RF model in both the training and test sets followed the variation in measured LWC (Figure 5). Most samples were close to the 1:1 line, particularly in the medium-to-high LWC range, where the points showed a clear linear pattern. However, some samples deviated from the 1:1 line at approximately 0.4–0.65, and some test samples were clearly underestimated. Hyperspectral model predictions mainly ranged from approximately 0.1 to 0.8, with relatively concentrated predictions for high-LWC samples. The thermal infrared RF model also captured the overall variation in LWC, but its scatter was more dispersed than that of the hyperspectral model. In particular, for measured LWC of approximately 0.4–0.8, predictions were more concentrated, with some samples clustered around 0.6–0.7 while measured values covered a wider range. The prediction-error distributions further highlighted differences between the two feature sources. Most absolute prediction errors were within 0–0.10, although both models had a small number of large errors. High-error points occurred in both the hyperspectral training and test sets, whereas the thermal infrared model had more errors above 0.10, with some reaching approximately 0.20–0.27. Overall, the hyperspectral model had a more concentrated error distribution at lower error levels, whereas the thermal infrared model showed a wider error range.

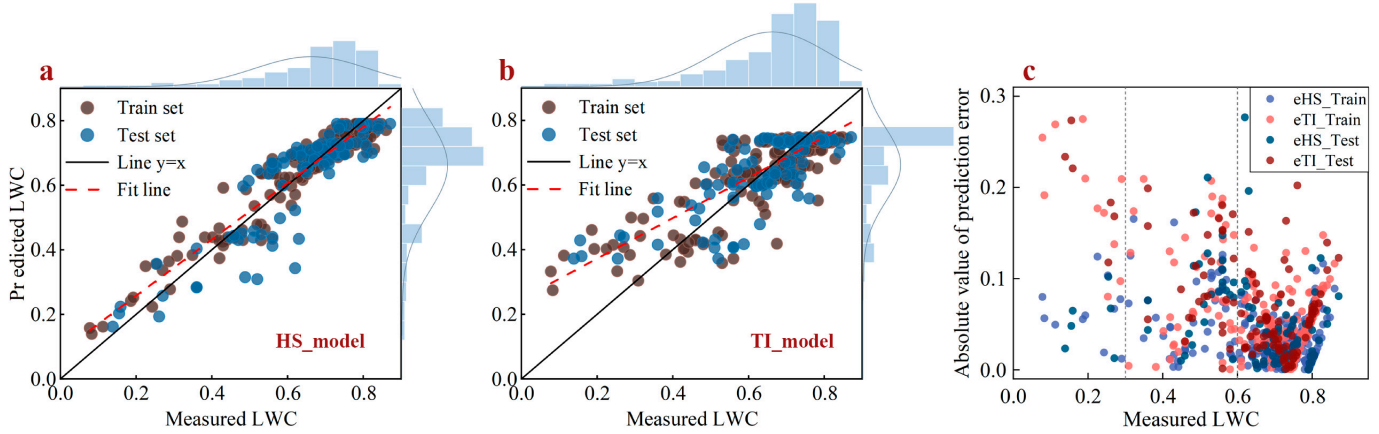


Figure 5. Predictions and error distributions of the best single-feature-source models ((**a**,**b**) scatter plots of predictions from the best models; (**c**) distribution of prediction errors).

3.3. Conventional GRA (Fixed-Weight) Fusion Results

The conventional GRA model fitted with fixed weights was LWC = −0.0679 + 1.0300y_HS + 0.0693y_TI. Based on the coefficient magnitudes, conventional GRA assigned a much larger fixed coefficient to the HS prediction than to the TI prediction (Table 4). The HS coefficient was close to 1, whereas the TI coefficient was only 0.0693. Meanwhile, the confidence interval of the HS coefficient was entirely above zero, whereas that of the TI coefficient crossed zero. The fixed-weight fusion was therefore dominated by the hyperspectral prediction, while TI provided only a small fixed correction.

Table 4. Model parameter statistics.

Coefficient	Estimate	Std_Error	t_Value	*p*_Value	CI_Lower	CI_Upper
const	−0.0679	0.0169	−4.0162	7.83×10^{-5}	−0.1012	−0.0346
y HS	1.0300	0.0439	23.4727	4.76×10^{-65}	0.9436	1.1164
y TI	0.0693	0.0511	1.3567	0.1761	−0.0313	0.1700

As shown in Figure 6, the training and test samples generally followed the fitted line, but some samples deviated clearly from the 1:1 line, particularly in the medium LWC range. The test-set R^2 was 0.7645. Compared with the hyperspectral single-feature-source model, conventional GRA did not further improve test-set R^2; instead, R^2 decreased from 0.7871 to 0.7645. RMSE increased from 0.0707 to 0.0744, and MAE increased from 0.0520 to 0.0536. Although fixed-weight fusion increased training-set R^2 from 0.9056 to 0.9116, this improvement was not reproduced in the independent test set.

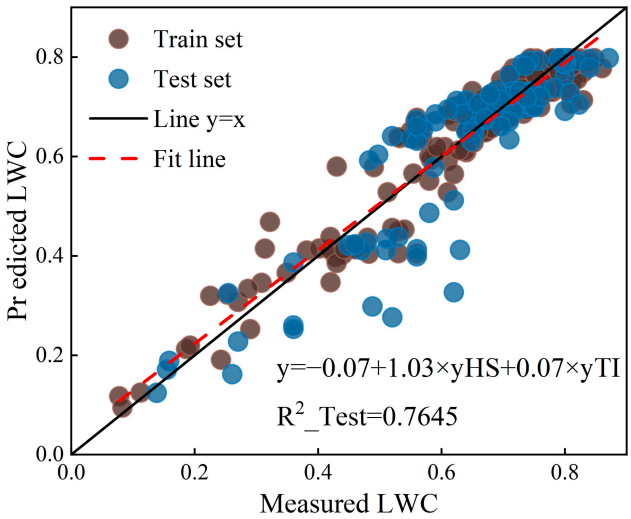


Figure 6. Scatter plot of predictions from the conventional GRA fusion model.

3.4. Dynamic GRA (Dynamic-Weight) Fusion Results

3.4.1. Decomposition Based on the Two Driving Factors

After the 720 samples were ordered by data sequence, both ΔT and NDVI showed clear stage-related changes, but their patterns differed substantially (Figure 7). ΔT remained low for most samples, mainly within approximately 1–5 °C, but several clear high-value peaks occurred, with some samples exceeding 10 °C and reaching approximately 25 °C. These high values were not persistent but appeared as isolated peaks across different sample intervals. NDVI showed a more continuous range of variation. During the first approximately 250 samples, NDVI remained around 0.85–0.95. It began to decrease around samples 250–350 and reached clearly lower levels after approximately sample 350. It then generally

declined from approximately 0.6 to 0.2–0.3, followed by a marked increase around sample 600. After that, NDVI remained around 0.6–0.9, with larger fluctuations. Compared with ΔT, NDVI showed stronger continuity and interval structure, whereas ΔT was characterized by a low-value background with local high-value peaks. The two driving factors therefore covered different patterns of sample variation.

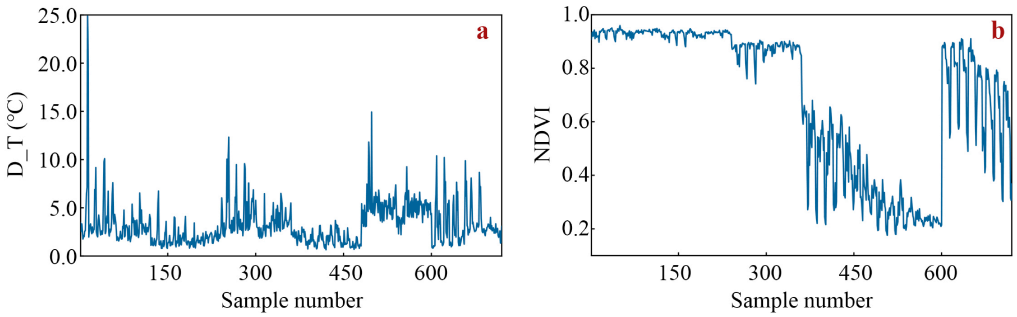


Figure 7. Variation in ΔT (**a**) and NDVI (**b**) across samples.

Figure 8a,b show the relationships between HS and TI prediction errors and ΔT and NDVI, together with the GAM fits. The actual absolute errors of the HS model varied substantially across samples. Most samples were around 0–0.06, but several high-error points exceeded 0.10, with the maximum close to 0.16. In contrast, the expected-error curves from the ΔT-GAM and NDVI-GAM varied within a narrower range, generally around 0.03–0.04. The two GAM curves did not completely overlap, with differences in their local magnitude and position. The actual absolute errors of the TI model varied over an even wider range, with some exceeding 0.20 and the maximum close to 0.28. Similar to the HS model, the GAM-fitted expected-error curves were substantially smoother than the actual error sequence, with the main variation range around 0.05–0.07. The ΔT-GAM and NDVI-GAM curves deviated clearly in some sample intervals. For example, around samples 160–170, NDVI-GAM showed a relatively high local value, whereas the actual errors remained scattered.

Figure 8c,d show the dynamic weights derived from the two driving factors. The HS weights calculated from NDVI were mainly concentrated at approximately 0.60–0.70, with a central level of approximately 0.64. The corresponding TI weights were mainly within 0.30–0.40, with a central level of approximately 0.36. Both weights changed gradually across samples, with clear deviations at only a few positions. The dynamic weights derived from ΔT showed a much wider range. Most HS weights were approximately 0.55–0.70, but in some samples they decreased to approximately 0.34 or increased above approximately 0.80. TI weights were mainly within 0.30–0.45, with local extremes of approximately 0.20–0.65. In comparison, the NDVI-driven weight curves were more continuous, and a clear difference between HS and TI weights was maintained throughout the samples.

From both the expected-error and weight perspectives, NDVI-driven dynamic weights remained mainly around an HS:TI ratio of approximately 0.64:0.36 while retaining continuous sample-level adjustment. ΔT-driven weights showed stronger local fluctuations. Both dynamic strategies moved beyond the fixed-coefficient form of conventional GRA, allowing the contributions of HS and TI to vary with sample status.

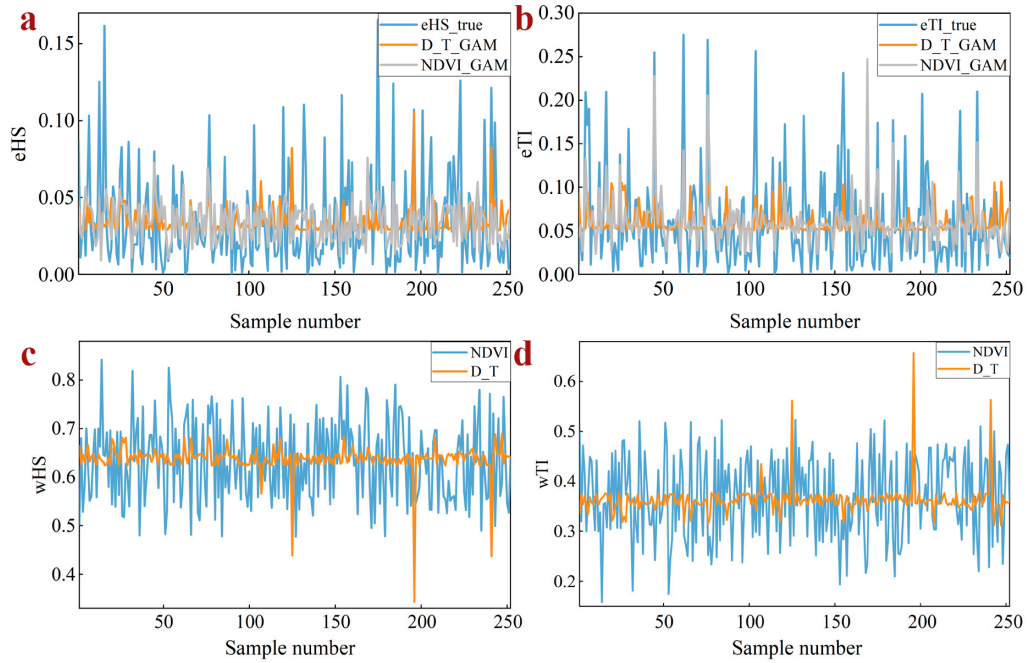


Figure 8. Error decomposition and dynamic weights ((**a**,**b**) decomposition of prediction errors of the two single-feature-source models based on the two driving factors; (**c**,**d**) dynamic weights of the two single-feature-source models derived from the two driving factors).

3.4.2. Comparison of Model Accuracy Among Fusion Strategies

The training- and test-set accuracies differed substantially among models (Table 5). Among the single-feature-source models, the HS model performed better than the TI model for all three evaluation metrics and provided the main predictive information in the fusion framework. Conventional GRA achieved training-set R^2 = 0.9116, RMSE = 0.0443, and MAE = 0.0329, representing an increase of 0.0060 in R^2 and decreases of 0.0015 and 0.0012 in RMSE and MAE, respectively, compared with the HS single model. However, its test-set R^2 was only 0.7645, lower than the 0.7871 of the HS model, while test-set RMSE and MAE were 0.0744 and 0.0536, respectively, both higher than those of the HS model. Thus, fixed-weight fusion produced a small improvement in the training set but not in the test set.

Table 5. Accuracy comparison of all models.

	Train Set			Test Set		
	R^2	RMSE	MAE	R^2	RMSE	MAE
HS	0.9056	0.0458	0.0341	0.7871	0.0707	0.0520
TI	0.7080	0.0806	0.0601	0.6275	0.0936	0.0736
GRA	0.9116	0.0443	0.0329	0.7645	0.0744	0.0536
GRA_ΔT	0.8793	0.0518	0.0393	0.8207	0.0649	0.0501
GRA_NDVI	0.8843	0.0507	0.0382	0.8224	0.0646	0.0497

After introducing ΔT-driven dynamic weights, test-set R^2 increased to 0.8207, which was 0.0336 higher than the HS single model and 0.0562 higher than conventional GRA. RMSE decreased from 0.0707 for the HS model to 0.0649, and MAE decreased from 0.0520 to 0.0501. With NDVI-driven dynamic weights, test-set R^2 reached 0.8224, with RMSE and MAE of 0.0646 and 0.0497, respectively. These were the highest R^2 and lowest error values among the compared models. Relative to conventional GRA, test set R^2 increased by 0.0579, RMSE decreased by 0.0098, and MAE decreased by 0.0039. Relative to the HS single model, R^2 increased by 0.0353, RMSE decreased by 0.0061, and MAE decreased by 0.0023.

The training-set R^2 was 0.8843, with RMSE and MAE of 0.0507 and 0.0382, respectively. The HS single model already had relatively high predictive ability; conventional GRA did not improve test-set performance; both ΔT- and NDVI-driven strategies increased test-set R^2 above 0.82 and reduced RMSE and MAE, with the NDVI-driven strategy achieving R^2 = 0.8224, RMSE = 0.0646, and MAE = 0.0497.

3.4.3. Overall Prediction Results of the Best Model

Predictions from the NDVI-driven dynamic GRA model were generally distributed along the 1:1 line and showed a clear linear relationship with measured LWC (Figure 9). Samples with low LWC were mainly distributed around 0.15–0.45, whereas medium-to-high LWC samples were concentrated around 0.55–0.80. Predictions generally increased with measured LWC, although some scatter remained in certain ranges. The overall fit achieved R^2 = 0.8507, RMSE = 0.0591, and MAE = 0.0451. The marginal distribution of predictions showed that most predictions were concentrated between 0.55 and 0.80, consistent with the overall LWC distribution. The fitted line was slightly above the 1:1 line at low LWC and gradually fell below the 1:1 line at higher LWC, indicating some compression of the prediction range across the full LWC range.

Compared with the test-set results in Table 4, evaluation using all 720 samples in Figure 9 increased R^2 from 0.8224 to 0.8507, reduced RMSE from 0.0646 to 0.0591, and reduced MAE from 0.0497 to 0.0451. The all-sample results showed the same metric ordering as the independent test results, and the NDVI-driven dynamic GRA maintained a high level of consistency between predictions and measured LWC across the full sample set.

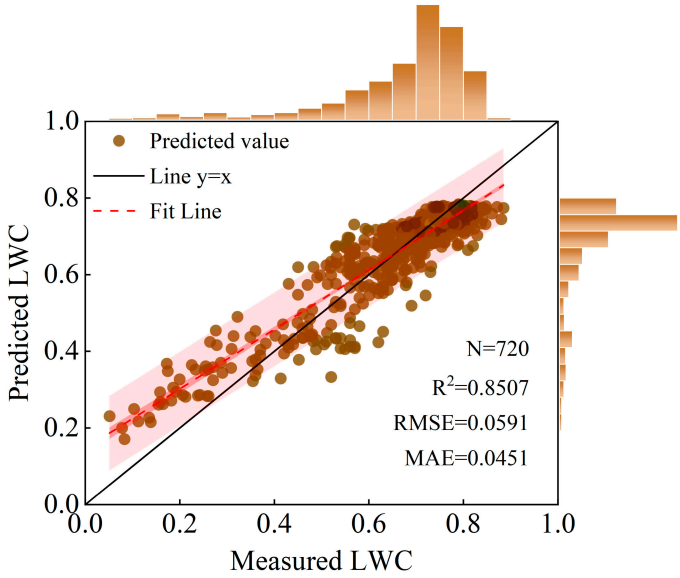


Figure 9. Final predictions for all samples by the best model, the NDVI-driven dynamic GRA model.

Figure 10 illustrates the spatiotemporal distribution of LWC predicted by the NDVI-driven dynamic GRA model across observation dates, nitrogen treatments, and cultivar combinations. The prediction maps showed clear variations in LWC levels among observation periods and treatment combinations. For the 2023 observations on DOY 128 and 133, predicted LWC was generally high, with visible differences among treatment combinations, although some groups exhibited relatively uniform values. In 2024, the predicted LWC distribution varied across observation dates. On DOY 107 and 117, differences among treatment combinations were apparent, while the later observations on DOY 123 and 129 showed further variation across nitrogen levels and cultivar combinations. In particular, some N0 and low-N treatment combinations exhibited relatively low predicted LWC. Over-

all, the prediction maps reveal temporal changes in LWC and spatial heterogeneity among treatment combinations throughout the observation periods. These patterns provide a visual representation of the model-estimated LWC distribution, while quantitative model performance is evaluated separately using measured and predicted values.

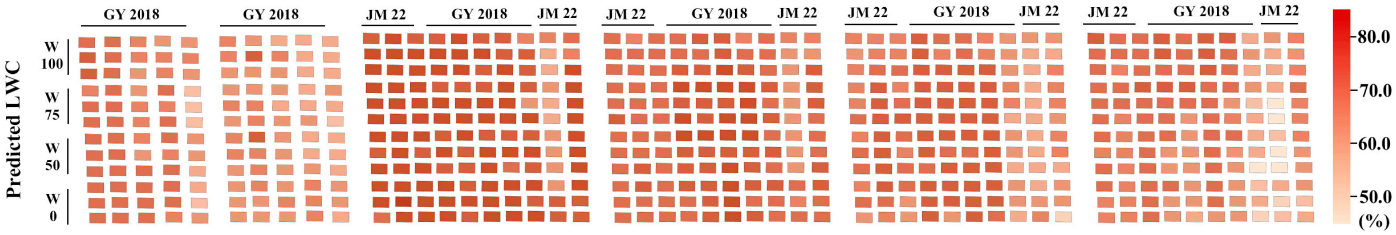


Figure 10. Comparison of the spatiotemporal distributions of predicted LWC.

4. Discussion

4.1. Differences in Predictive Ability Between Hyperspectral and Thermal Infrared Models for LWC

The hyperspectral model performed better overall than the thermal infrared model, indicating a stronger predictive relationship between hyperspectral features and LWC under the conditions of this study. Although the thermal infrared model had lower overall accuracy, its predictions were not identical to those of the hyperspectral model and still contained independent predictive information. Similar differences in performance have been reported among other sensors. Previous studies have shown that sensors differ in their predictive ability for different target properties and that whether sensor fusion improves prediction accuracy depends partly on the predictive performance of the individual sensors [9,15]. However, differences among single-source models do not mean that a weaker model has no value for fusion. In this study, the test-set R^2 of the hyperspectral RF model was 0.7871, clearly higher than the 0.6275 of the thermal infrared RF model. The error distributions also differed between the two models. Most hyperspectral errors were relatively small, although some samples showed substantial local deviations. The thermal infrared model had a wider error range and more high-error samples. Similar differences in error patterns have been observed in winter wheat LWC studies, where models based on different feature sources showed different predictive abilities and feature fusion further improved LWC estimation accuracy [16]. Therefore, this study retained the independent predictions of the hyperspectral and thermal infrared single-source models rather than directly concatenating the two types of raw variables, allowing subsequent fusion to combine the two established predictions.

The absolute prediction errors of both the hyperspectral and thermal infrared models varied substantially across samples. Most hyperspectral errors were relatively small, but some samples showed clear deviations, whereas the thermal infrared model had larger error fluctuations and higher local errors. At the same time, ΔT and NDVI varied substantially across the 720 samples. Therefore, the prediction error of a single-source model can be further represented as a function of crop status rather than assuming the same model reliability for all samples. This idea is consistent with the recent development of dynamic model averaging [17]. For crop remote sensing, using remote sensing variables that characterize changes in crop status to further explain single-source prediction errors provides a natural step from fixed model averaging toward state-dependent model averaging. In this study, GAM relationships were established separately between the absolute errors of the single-source models and ΔT or NDVI. The expected errors were then converted into model reliability and used to calculate sample-level dynamic weights. Compared with calculating weights directly from the true error of the current sample, this approach avoids

using measured LWC to determine weights for unknown samples and allows the dynamic weights to be obtained from remote sensing driving factors during practical prediction.

4.2. Advantages of the Dynamic GRA Model

The conventional GRA model fitted to the training set was LWC = −0.0679 + 1.0300yHS + 0.0693yTI. The combination coefficient of the hyperspectral model was much larger than that of the thermal infrared model, whereas the statistical significance of the thermal infrared coefficient was weaker. On the test set, conventional GRA did not outperform the hyperspectral single-source model. This shows that model combination does not necessarily improve prediction accuracy. Its performance depends on the prediction errors of the input models, error dependence, and the stability of the estimated combination weights. Previous studies have shown that estimating combination weights can introduce additional errors, so complex weighting methods do not necessarily outperform individual models or simple averages under limited sample sizes [18–21]. The benefits of combination methods are also context-dependent, and their performance is closely related to the candidate models and the method used to estimate the weights [22]. Therefore, conventional GRA can serve as a benchmark for multi-source model fusion, but fixed weights cannot describe changes in the relative reliability of single-source models across sample conditions.

Both dynamic GRA models improved on conventional fixed-weight GRA. Compared with conventional GRA, the NDVI-driven model increased test set R^2 by 0.0579 and reduced RMSE and MAE by 0.0098 and 0.0039, respectively. The key feature of dynamic modeling is that the relative contributions of different models can change rather than remaining at the same combination coefficients for all samples [23,24]. Time-varying weights have been used in hydrologic forecasting, soil moisture estimation, and environmental simulation to adjust model contributions according to predictive ability or prediction error [25,26]. When model predictive ability depends strongly on state or time, dynamic combination can make better use of complementary information among models. Unlike these approaches, the present study does not calculate weights directly from measured LWC but uses remote sensing driving factors such as ΔT and NDVI. Thus, dynamic weights can be determined directly from remote sensing information for unknown samples. The two driving factors also produced clearly different dynamic weights. The HS weights derived from NDVI were mainly concentrated around 0.60–0.70, while TI weights were mainly around 0.30–0.40. The NDVI-driven dynamic GRA showed a more stable fusion result. This difference indicates that the effectiveness of dynamic weighting depends on whether the driving variable can stably describe changes in the predictive ability of the single-source models. In this study, NDVI provided a relatively continuous representation of canopy status and showed relatively stable nonlinear relationships with the errors of the hyperspectral and thermal infrared models, resulting in smoother dynamic weights. In contrast, ΔT showed stronger local variation among samples, and its corresponding weights were more dispersed. Therefore, NDVI was more suitable as the weight-driving factor for dynamic GRA under the conditions of this study. This result indicates the relative advantage of NDVI in the present winter wheat dataset rather than suggesting that NDVI is superior to ΔT for all crops or environmental conditions. The introduction of environmental covariates does not necessarily improve prediction performance in all cases, and their effects depend on the study object and spatial scale.

The NDVI-driven dynamic GRA achieved an overall R^2 of 0.8507, RMSE of 0.0591, and MAE of 0.0451 for the 720 samples. The predictions generally followed the variation in measured LWC, but some overestimation remained at low LWC and some underestimation occurred at high LWC, indicating prediction range compression. Similar behavior is common in machine learning and model-combination prediction, particularly when

models tend to reduce errors at the extremes, causing predictions to shrink toward the main distribution range. Therefore, dynamic GRA mainly improves the combination of the two single-source models and cannot fully eliminate the systematic biases of the base learners. Nevertheless, the predicted spatial distribution of LWC from the best model showed good overall agreement with the measured results across observation dates, cultivars, and N treatments. Dynamic model averaging therefore not only improved sample-level prediction accuracy but also extended the fused LWC estimates to the spatial scale of UAV imagery.

4.3. Limitations and Future Research

Several issues in this study require further validation. First, the dynamic weighting model was developed using a specific winter wheat dataset, and the relationship between NDVI and model prediction error may be affected by cultivar, growth stage, canopy structure, and water and nitrogen status. Therefore, the current results should be interpreted as effective under the conditions of this study and cannot be directly generalized to other crops or ecological regions.

Second, this study used a single variable, NDVI or ΔT, as the weight-driving factor. Crop water status is simultaneously affected by canopy structure, radiation conditions, temperature, and nutrient status. Future studies could compare multiple driving variables or multivariable GAMs for explaining dynamic weights and test whether adding driving variables provides stable independent gains.

Third, dynamic GRA remains essentially a linear fusion of model predictions. Although the weights vary continuously with crop status, the final HS and TI predictions are still combined through linear weighting. Therefore, under more complex sensor error relationships, dynamic GRA could be further compared with other nonlinear ensemble methods. The stability of the dynamic weight relationships should also be evaluated over more years, locations, and independent experiments to assess their cross-environment applicability.

Finally, further research should explore the practical application of dynamic multi-source fusion for large-scale crop water status monitoring. Integrating UAV hyperspectral and thermal infrared observations with satellite remote sensing could help extend the proposed approach from plot-level experiments to regional monitoring. Such applications would require evaluating the effects of differences in spatial resolution, observation frequency, and sensor characteristics on model performance. Overall, these efforts could improve the generalizability and operational applicability of dynamic model averaging for winter wheat LWC estimation and support more reliable crop water status assessment.

5. Conclusions

This study addressed UAV remote sensing estimation of winter wheat leaf water content (LWC) by integrating hyperspectral and thermal infrared feature sources and developing a multi-source model fusion framework consisting of single-feature-source models, conventional fixed-weight GRA, and driving-factor-based dynamic GRA. The main conclusions are as follows:

(1) UAV hyperspectral and thermal infrared data can both be used to estimate winter wheat LWC, but their predictive abilities differ substantially. RF was the best model for both feature sources. For the hyperspectral test set, R^2, RMSE, and MAE were 0.7871, 0.0707, and 0.0520, respectively, while the corresponding values for the thermal infrared test set were 0.6275, 0.0936, and 0.0736.

(2) Conventional GRA can fuse predictions from the hyperspectral and thermal infrared at the model level, but fixed weights cannot adapt to changes in single-source predictive

ability across sample conditions. Its test-set R^2, RMSE, and MAE were 0.7645, 0.0744, and 0.0536, respectively, and it did not improve the test-set accuracy of the hyperspectral model. Global fixed-weight model averaging therefore has limitations.

(3) Using ΔT and NDVI to drive changes in fusion weights clearly improved GRA prediction performance. The NDVI-driven dynamic GRA was the best model, achieved test-set R^2, RMSE, and MAE values of 0.8224, 0.0646, and 0.0497. And its predictions showed good overall spatial agreement with measured LWC across observation dates, cultivars, and treatments.

In summary, dynamic GRA driven by remote sensing variables can further exploit changes in the reliability of single-source models across samples beyond conventional fixed-weight model averaging, enabling continuous dynamic fusion of hyperspectral and thermal infrared predictions. The NDVI-driven dynamic GRA achieved the highest overall prediction performance in the present dataset and can be used for sample-scale LWC estimation and spatial mapping at the UAV image scale.

Author Contributions: M.D.: investigation, writing—original draft preparation, visualization. L.G.: Conceptualization, writing—original draft preparation, supervision, project administration. B.Z.: methodology, investigation. W.J.: investigation, writing—review and editing. P.L.: investigation, writing—review and editing. B.Y.: methodology, writing—review and editing. C.Z.: methodology, Conceptualization. X.G.: methodology, writing—review and editing, supervision. W.Z.: methodology, writing—review and editing, supervision, funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by the National Key R&D Program of China, grant number 2023YFD2301503.

Data Availability Statement: The data presented in this study are available on request from the corresponding author. As the project to which this paper belongs has not yet been accepted, the data should not be disclosed for the time being in accordance with the law, but the data processing code used can be disclosed.

Acknowledgments: We would like to acknowledge the Xinji Experimental Station of Hebei Agricultural University and the Beijing Academy of Agriculture and Forestry for their work support.

Conflicts of Interest: Author Chong Zhang was employed by the company Hebei Dadi Seed Industry Co., Ltd. The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as potential conflicts of interest. This research was funded by the National Key R&D Program of China, grant number2023YFD2301503. Not funded by commercial funding.

References

1. Sun, D.; Zhang, L.; Li, H.; Lan, W.; Tu, K.; Liu, J.; Pan, L. Non-destructive prediction of the moisture content of individual wheat kernels combining hyperspectral imaging and WGAN data augmentation algorithm. *Food Res. Int.* **2025**, *212*, 116498. [CrossRef] [PubMed]
2. Wang, J.; Zhang, S.; Lizaga, I.; Zhang, Y.; Ge, X.; Zhang, Z.; Zhang, W.; Huang, Q.; Hu, Z. UAS-based remote sensing for agricultural Monitoring: Current status and perspectives. *Comput. Electron. Agric.* **2024**, *227*, 109501. [CrossRef]
3. Wu, Y.; Yuan, S.; Zhu, J.; Tang, Y.; Tang, L. Estimation of Wheat Leaf Water Content Based on UAV Hyper-Spectral Remote Sensing and Machine Learning. *Agriculture* **2025**, *15*, 1898. [CrossRef]
4. Neinavaz, E.; Schlerf, M.; Darvishzadeh, R.; Gerhards, M.; Skidmore, A.K. Thermal infrared remote sensing of vegetation: Current status and perspectives. *Int. J. Appl. Earth Obs. Geoinf.* **2021**, *102*, 102415. [CrossRef]
5. Gerhards, M.; Schlerf, M.; Mallick, K.; Udelhoven, T. Challenges and Future Perspectives of Multi-/Hyperspectral Thermal Infrared Remote Sensing for Crop Water-Stress Detection: A Review. *Remote Sens.* **2019**, *11*, 1240. [CrossRef]
6. Granger, C.W.J.; Ramanathan, R. Improved methods of combining forecasts. *J. Forecast.* **1984**, *3*, 197–204. [CrossRef]
7. Diks, C.G.H.; Vrugt, J.A. Comparison of point forecast accuracy of model averaging methods in hydrologic applications. *Stoch. Environ. Res. Risk Assess.* **2010**, *24*, 809–820. [CrossRef]

8. Malone, B.P.; Minasny, B.; Odgers, N.P.; McBratney, A.B. Using model averaging to combine soil property rasters from legacy soil maps and from point data. *Geoderma* **2014**, *232*, 34–44. [CrossRef]
9. O'Rourke, S.M.; Stockmann, U.; Holden, N.M.; McBratney, A.B.; Minasny, B. An assessment of model averaging to improve predictive power of portable vis-NIR and XRF for the determination of agronomic soil properties. *Geoderma* **2016**, *279*, 31–44. [CrossRef]
10. Tavares, T.R.; Molin, J.P.; Javadi, S.H.; Carvalho, H.W.; Mouazen, A.M. Combined Use of Vis-NIR and XRF Sensors for Tropical Soil Fertility Analysis: Assessing Different Data Fusion Approaches. *Sensors* **2021**, *21*, 148. [CrossRef] [PubMed]
11. Javadi, S.H.; Mouazen, A.M. Data Fusion of XRF and Vis-NIR Using Outer Product Analysis, Granger–Ramanathan, and Least Squares for Prediction of Key Soil Attributes. *Remote Sens.* **2021**, *13*, 2023. [CrossRef]
12. Gao, C.; Liu, S.; Wu, P.; Wang, Y.; Wu, K.; Li, L.; Wang, J.; Liu, S.; Gao, P.; Zhao, Z.; et al. Estimation of Canopy Water Content by Integrating Hyperspectral and Thermal Imagery in Winter Wheat Fields. *Agronomy* **2024**, *14*, 2569. [CrossRef]
13. Malone, B.P.; Searle, R.D.; Tian, S.; Bishop, T.F.; Yu, Y. Improving plant-available water estimation using model averaging of national soil water models. *Front. Soil Sci.* **2025**, *5*, 1629686. [CrossRef]
14. Zhang, X.; Xue, J.; Chen, S.; Zhuo, Z.; Wang, Z.; Chen, X.; Xiao, Y.; Shi, Z. Improving model performance in mapping cropland soil organic matter using time-series remote sensing data. *J. Integr. Agric.* **2024**, *23*, 2820–2841. [CrossRef]
15. Xue, J.; Zhang, X.; Chen, S.; Lu, R.; Wang, Z.; Wang, N.; Hong, Y.; Chen, X.; Xiao, Y.; Ma, Y.; et al. The validity domain of sensor fusion in sensing soil quality indicators. *Geoderma* **2023**, *438*, 116657. [CrossRef]
16. Yang, N.; Zhang, Z.; Zhang, J.; Yang, X.; Liu, H.; Chen, J.; Ning, J.; Sun, S.; Shi, L. Accurate estimation of winter-wheat leaf water content using continuous wavelet transform-based hyperspectral combined with thermal infrared on a UAV platform. *Eur. J. Agron.* **2025**, *168*, 127624. [CrossRef]
17. Broderick, C.; Matthews, T.; Wilby, R.L.; Bastola, S.; Murphy, C. Transferability of hydrological models and ensemble averaging methods between contrasting climatic periods. *Water Resour. Res.* **2016**, *52*, 8343–8373. [CrossRef]
18. Smith, J.; Wallis, K.F. A Simple Explanation of the Forecast Combination Puzzle. *Oxf. Bull. Econ. Stat.* **2009**, *71*, 331–355. [CrossRef]
19. Claeskens, G.; Magnus, J.R.; Vasnev, A.L.; Wang, W. The forecast combination puzzle: A simple theoretical explanation. *Int. J. Forecast.* **2016**, *32*, 754–762. [CrossRef]
20. Qian, W.; Rolling, C.A.; Cheng, G.; Yang, Y. On the Forecast Combination Puzzle. *Econometrics* **2019**, *7*, 39. [CrossRef]
21. Chan, F.; Pauwels, L.L. Some theoretical results on forecast combinations. *Int. J. Forecast.* **2018**, *34*, 64–74. [CrossRef]
22. Wang, X.; Hyndman, R.J.; Li, F.; Kang, Y. Forecast combinations: An over 50-year review. *Int. J. Forecast.* **2023**, *39*, 1518–1547. [CrossRef]
23. Sun, Y.; Hong, Y.; Lee, T.-H.; Wang, S.; Zhang, X. Time-varying model averaging. *J. Econom.* **2021**, *222*, 974–992. [CrossRef]
24. Sun, Y.; Hong, Y.; Wang, S.; Zhang, X. Penalized time-varying model averaging. *J. Econom.* **2023**, *235*, 1355–1377. [CrossRef]
25. Chowdhury, S.; Sharma, A. Multisite seasonal forecast of arid river flows using a dynamic model combination approach. *Water Resour. Res.* **2009**, *45*, W10428. [CrossRef]
26. Mehrnegar, N.; Jones, O.; Singer, M.B.; Schumacher, M.; Bates, P.; Forootan, E. Comparing global hydrological models and combining them with GRACE by dynamic model data averaging (DMDA). *Adv. Water Resour.* **2020**, *138*, 103528. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.