A paper from the Proceedings of the
17th International Conference on Precision Agriculture and 11th Brazilian Congress on Precision and Digital Agriculture
13-15 July 2026
Porto Alegre, Brazil

Study on the Phenological Zoning Method for Winter Wheat in the Huang-Huai-Hai Region of China

Song Xiaoyu*[a,b],Wu Qiang[a,b],Ma Yuanyuan[a,b], Zhang Jie[a,b],Dong Peixuan[a,b], Xu Xingang[a,b]

([a] Information Technology Research Center, Beijing Academy of Agriculture and Forestry Sciences, Beijing,China,100097; [b]Key Laboratory of Quantitative Remote Sensing in Agriculture of Ministry of Agriculture,Beijing, China, 100097;)

Abstract.

This study focuses on the primary winter wheat production area in the Huang-Huai-Hai (HHH) region and introduces a phenological zoning method based on multi-year heading dates as a single factor. Utilizing MODIS surface reflectance time series from 2016 to 2021, the dual logistic function was applied to extract the spatiotemporal patterns of heading dates. To construct alternative zoning schemes, K-means clustering and KNN-Ward hierarchical clustering were compared, and regional wheat masks combined with administrative boundaries were used to delineate the geographic distribution of each zone.The results reveal: (1) Temporal characteristics: Regional mean heading date (DOY) varied between 106.7 and 111.7, with a multi-year average of 110 DOY and modest interannual variation (standard deviation ≈ 1.7 days). Pixel-scale heading dates concentrated between DOY 96 and 124, with homogeneous zone-level standard deviations of 7.5–15.4 days. Heading was relatively early in 2020 (DOY 106.7) and later in 2019 (DOY 111.7). (2) Spatial gradient: A notable latitudinal gradient was observed, with zone centroids progressing from 33.4°N to 37.5°N and corresponding mean heading DOY increasing from 104 to 121, following an "earlier in the south, later in the north" pattern consistent with thermal differentiation. (3) Four phenological zones: Both methods detected a comparable 16–20 day gradient. However, KNN-Ward clustering, incorporating spatial constraints, generated more coherent and homogeneous zones (within-zone standard deviation 11.0–15.4 days): an extremely early zone (Henan–Anhui border, DOY 104.4, 33.4°N), an early zone (northern Jiangsu–eastern Henan–southwestern Shandong–northern Anhui, DOY 110.1, 34.2°N), a mid zone (central Shandong–southern Hebei, DOY 115.4, 36.6°N), and a late zone (northwestern Shandong–central–southern Hebei, DOY 120.6, 37.5°N). In contrast, K-means produced fragmented zones with poor contiguity; its Zone 4 showed severe internal heterogeneity (standard deviation 26.9 days) and geographic misplacement, rendering it unsuitable for operational management. K-means may serve as an exploratory tool but should not be applied directly for operational zoning. This study provides a lightweight methodological paradigm for crop phenological zoning under data-limited conditions, supporting regional adaptation of winter wheat management in the HHH region.

Keywords.

Winter wheat; Heading date; Phenological zoning; K-means clustering; KNN-Ward hierarchical clustering; Spatial constraint; Huang-Huai-Hai region

1. Introduction

The impact of global climate change on agricultural phenology is becoming increasingly significant. As a major producer of winter wheat, China's cultivation areas span multiple climate zones. Against the backdrop of climate change, the spatiotemporal differentiation of crop phenology has raised new scientific demands for agricultural zoning. Phenological zoning has guiding significance for variety selection, irrigation management, and pest prediction.

As one of the major wheat-producing countries in the world, the Huang-Huai-Hai region in China is the area with the most concentrated winter wheat cultivation, accounting for over 70% of the total winter wheat output in the country [1]. The wheat cultivation in this region is large in scale and extensive in scope, and it is one of the core areas for food security in China . However, there are significant climate variations and frequent extreme weather events in this region, resulting in significant temporal and spatial heterogeneity in crop growth, which poses challenges for precise estimation of large-scale yields [2]. Studies have shown that the growth process of crops is influenced by the interaction of environmental factors such as climate and soil, as well as phenological progress. The response of crops to environmental factors not only has regional differences but also changes with the phenological period. There are 5-15 day gradient differences in the growth process of winter wheat from flowering to maturity in different geographical units [3]. At the same time, the contribution of environmental factors has spatial dependence. For example, the yield of winter wheat in Denmark is highly sensitive to summer high temperatures ($P < 0.01$) [4], while the main production areas in China are constrained by water stress during the flowering period. This relationship shows significant spatial heterogeneity and non-linearity at the regional scale. Therefore, studying regional heterogeneity is particularly important for crop yield estimation and growth management.

Addressing the insufficient consideration of phenological heterogeneity in regional winter wheat yield estimation, this study conducts phenological zoning research based on multi-year phenological monitoring results in the Huang-Huai-Hai (HHH) region, employing both KNN-Ward and K-means algorithms.

2. Materials and Methods

2.1. Research Area

This study takes the Huang-Huai-Hai (HHH) region of China as the research area. Located in the warm temperate semi-humid and semi-arid continental monsoon climate zone, the HHH region features complex and diverse climate types with distinct zonal characteristics.

2.2. Data collection

2.2.1Remote sensing data

The satellite data used in this study were obtained from the Google Earth Engine (GEE) platform. MODIS imagery with an 8-day composite interval during the winter wheat filling stage from 2016 to 2021 was selected for the Huang-Huai-Hai region. The EVI vegetation index was extracted from MOD09A1 data for subsequent analysis.

2.2.2Winter Wheat Distribution Maps

In this study, winter wheat planting area data from 2016 to 2021 were obtained from the winter wheat identification dataset provided by the National Ecological Science Data Center (https://www.nesdc.org.cn/, accessed on 4 January 2025).The wheat distribution map was first binarized at a 30 m resolution, where pixel values of 1 and 0 represent wheat-growing and non-wheat-growing areas, respectively. To match the spatial resolution of other datasets, the binary map was resampled from 30 m to 500 m using the minimum pooling method [5,6]. A 500 m pixel containing any non-wheat (0) value was identified as a mixed pixel. To ensure reliable yield estimation, pixels with a fractional wheat coverage above 0.6 were classified as wheat-dominant pixels. We further analyzed the residual differences in yield estimation between pixels with coverage fractions above and below this 0.6 threshold.

2.3. Data Analysis Methods

2.3.1. Winter Wheat Heading Stage Identification

The heading stage is a crucial phase in which wheat transitions from vegetative to reproductive growth. This stage directly influences subsequent processes such as flowering, grain filling, and maturation, and plays a key role in predicting the timing of crop maturity [7]. During this stage, notable changes occur in the structure of wheat leaves and spikes, leading to increased chlorophyll absorption and light scattering, particularly manifesting as a significant rise in near-infrared reflectance. Field vegetation coverage reaches its maximum during the heading stage, and vegetation indices such as NDVI and EVI2 typically attain their peak values at this time. Consequently, the maximum vegetation index method is widely employed for monitoring the wheat heading stage [8].

In this study, MOD09A1 imagery from the GEE platform was used for phenology extraction. The time range was restricted to October 1 of the previous year to June 30 of the current year. Cloud and shadow masking were conducted using the StateQA band from MOD09A1 via bit extraction, and atmospheric correction was applied. Missing or cloud-contaminated pixels were filled using temporal interpolation. The EVI2 index was then calculated, missing values were interpolated, and the resulting time series were smoothed using the Savitzky–Golay (SG) filter to reduce atmospheric noise and residual cloud effects. Based on previous studies on winter wheat phenology extraction [9,10], the SG filter was implemented with a three-observation window (approximately 24 days) and a third-order polynomial. To account for regional variations in phenological responses, a curve-stretching method was applied to align local EVI2 maxima and minima with the regional mean curve. This normalization adjusted amplitude and temporal scale without altering the relative peak position of local curves, thereby improving the robustness and spatial consistency of heading-stage detection [11].

2.3.2. Phenological Zoning Method

(1) K-means Clustering Algorithm

K-means is one of the most widely used unsupervised machine learning algorithms for partitioning data into distinct groups. Proposed by Stuart Lloyd in 1957 and later refined by James MacQueen[12], it aims to minimize within-cluster variance by iteratively assigning data points to the nearest cluster centroid and updating centroids accordingly.In phenological zoning, K-means groups geographic pixels or stations based on similarity in phenological metrics—such as heading dates, growing season length, or vegetation index trajectories. Each cluster represents a phenological zone with homogeneous temporal characteristics, supporting regional crop management, yield prediction, and climate adaptation strategies. Its simplicity and interpretability make it a baseline method for comparing advanced clustering approaches like hierarchical or density-based techniques.

(2) Dynamic K-Nearest Neighbors

To characterize spatial autocorrelation of heading-stage DOY values, this study employed the Dynamic K-Nearest Neighbors (KNN) algorithm to construct a spatial weight matrix [13]. The KNN algorithm adaptively balances local detail capture and noise suppression, ensuring robust and efficient representation of spatial dependency [14]. Specifically, Euclidean distances were calculated between geographic units based on DOY values and spatial coordinates (Equation (1)). The number of neighbors K was then determined dynamically (Equation (2)), where the constant 6 is an empirical value commonly used to balance local spatial dependency with statistical stability [15]. This generated a binary adjacency matrix (Equation (3)), which was subsequently normalized to eliminate density biases, yielding the standardized weight matrix (Equation (4)). The resulting matrix quantifies spatial co-variation of DOY values among adjacent units, providing geographic constraints for subsequent clustering.The relevant equations are shown as follows.

$$d_{ij}=\sqrt{(s_{i1}-s_{j1})^2+(s_{i2}\text{-}s_{j2})^2} \quad (1)$$

d_{ij} is the Euclidean distance between units i and j used to measure geographical proximity s_{i1} and s_{i2} are the latitude and longitude coordinates of unit i;

$$k=\min(6, N - 1) \quad (2)$$

k is the dynamic domain size, adaptively adjusted to accommodate data of different scales; N is the number of valid observation points remaining after dynamic grid sampling;

$$w_{ij}=\begin{cases}1 & j\in N_i(k)\\ 0 & \text{other}\end{cases} \quad (3)$$

$$\tilde{w}_{ij}=\frac{w_{ij}}{\sum_{j=1}^{N} w_{ij}} \quad (4)$$

where w_{ij} is the binary adjacency relationship, indicating whether unit j is one of the k nearest neighbors of unit i; $\tilde{w}_{ij}$ is the standardized spatial weight.

(3) Improving the Ward hierarchical clustering algorithm

Based on the spatial weight matrix W_{ij}, this study proposes an improved Ward hierarchical clustering algorithm that simultaneously minimizes within-cluster DOY heterogeneity while enforcing spatial contiguity, thereby preventing the formation of “disconnected” clusters and ensuring that the resulting regions are both feature-homogeneous and geographically continuous. The algorithm employs a bottom-up agglomerative strategy, in which, at each iteration, only geographically adjacent units or clusters with the smallest DOY differences are merged. This approach minimizes within-cluster variance while maintaining the natural continuity of cluster boundaries [16]. The algorithm was defined as follows.

$$\Delta(C_a,C_b)=\frac{|C_a||C_b|}{|C_a|+|C_b|}(\bar{x}_a - \bar{x}_b)^2 \quad (5)$$

$$\bar{x}_c=\frac{|C_a|\bar{x}_a+|C_b|\bar{x}_b}{|C_a|+|C_b|} \quad (6)$$

where $\Delta(C_a,C_b)$ is the variance increment of the two clusters DOY; $|C_a|$ and $|C_b|$ represent the sample sizes of clusters a and b, respectively; $\bar{x}_a$ and $\bar{x}_b$ represent the average DOY values of clusters a and b, respectively; $\bar{x}_c$ is the average DOY value after the two clusters are merged.

2.3.3 Accuracy Evaluation Methods

To determine the optimal number of clusters, the range of cluster numbers (M) was set from 2 to 7. The Silhouette Coefficient, Calinski–Harabasz Index, and the corresponding sum of squared errors (SSEs) based on the elbow method were calculated to comprehensively evaluate the compactness and separability of the clustering results, thereby enabling the selection of the optimal cluster count.

3. Results

3.1. Wheat Heading Stage Monitoring Results

The maximum value method was used to monitor winter wheat heading stage in the Huang-Huai-Hai region from 2016 to 2021. Validation against ground phenology observations yielded R^2 = 0.48 and RMSE = 6.7 days. Regional mean DOY ranged 106.0 – 115.0, with medians of 106.0 – 116.0 and standard deviations of 10.8 – 13.6 days, showing moderate interannual variation. Spatially, DOY values displayed a distinct "north-high, south-low" gradient, indicating earlier heading in southern areas and later heading in northern areas. This latitudinal pattern provides the foundation for subsequent phenological zoning and yield estimation.

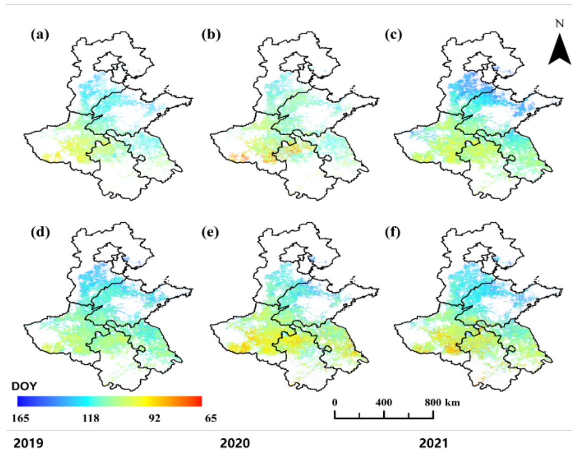

Figure 4. Winter wheat heading stage results from 2016 to 2021 : (a–f) represent the monitoring results from 2016 to 2021, respectively.

3.2. Phenological Zoning of Winter Wheat Based on the Multi-Year Heading Stage Maps

3.2.1. Zoning Parameters and Cluster Evaluation

Based on the monitoring results of the heading stage of winter wheat from 2016 to 2021, the KNN-Ward spatial constraint clustering method was used to perform phenological zoning. Cluster numbers ranging from two to seven were tested, and the resulting cluster configurations were comprehensively evaluated to determine the optimal number of clusters. The clustering outcomes for groupings of two to seven clusters are presented in Figure 5.

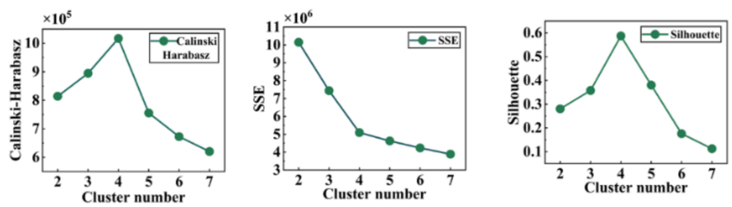

Figure 5. Clustering zoning results.

As illustrated in Figure 5, when the number of clusters is set to 4, all clustering evaluation metrics demonstrate optimal clustering performance. Specifically, the Silhouette coefficient reaches its maximum value of 0.58; the Calinski-Harabasz index peaks at 1,016,910.85 under the same clustering configuration; and the inflection point identified by the elbow method also occurs at four clusters. Taken together, these three indicators suggest that the clustering results are most effective when the number of clusters equals 4, reflecting the best balance between cluster cohesion and in-cluster separation.

3.2.2. Phenological Zoning results

Figure 6 is the phenological zoning results which indicate that K-means and KNN-Ward clustering both partition the Huang-Huai-Hai winter wheat area into four phenological tiers spanning approximately 16–17 DOY(2–3 weeks). This gradient aligns with the known latitudinal thermal pattern of the region (earlier heading in the south, later in the north). The results demonstrate that multi-year heading date alone is a viable basis for crop phenological zoning under data-limited conditions.

KNN-Ward zones exhibit uniform within-zone dispersion (std. dev. 11–15 DOY) and consistent mean–median agreement. The late-heading zone (Zone 3, mean DOY 120.6) is concentrated in northern Shandong and central Hebei (37.5°N), matching the expected spatial pattern of a northern late-heading tier. In contrast, K-means Zone 4 shows severe internal heterogeneity (std. dev. 26.9 DOY; interannual range 21.7 DOY) and is geographically misplaced in the south

(33.5°N), making it unsuitable as an independent late-heading management unit.

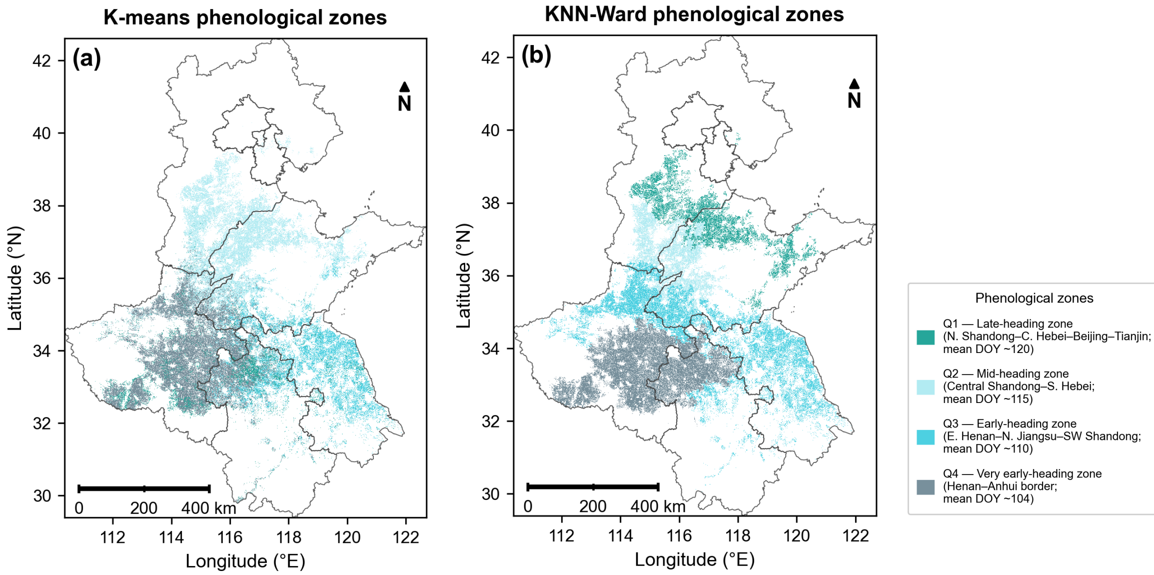

Fig 6. Phenological Zoning results of K-means(a) and KNN-Ward(b) .

K-means clustering produces highly fragmented, pixel-scattered zones with poor spatial contiguity, limiting direct use in regional agricultural management. KNN-Ward hierarchical clustering, which incorporates spatial constraints, generates contiguous, administratively meaningful zones that can be linked to provincial and municipal boundaries.

Regional mean heading date varied by approximately 5 DOY across 2016–2021, with 2020 as a notably early year (regional mean 106.7 DOY). Zone-level management strategies should retain sufficient flexibility to accommodate such interannual shifts. K-means Zone 4 showed anomalous late-heading behavior in 2017 and 2019, further confirming its unreliability as a stable phenological unit.

3.3 Conclusion

KNN-Ward hierarchical clustering is superior to K-means clustering for winter wheat phenological zoning in the Huang-Huai-Hai region. While both methods detect a comparable macro-scale phenological gradient (16 DOY span), KNN-Ward achieves better within-zone homogeneity, spatial contiguity, and geographic plausibility. K-means may serve as a rapid exploratory tool but should not be applied directly for operational zoning; if used, Zone 4 must be excluded or reclassified. This study provides a lightweight methodological framework for crop phenological zoning under data-limited conditions and supports the regional adaptation of winter wheat management — including variety layout, irrigation scheduling, and pest prediction — across the Huang-Huai-Hai production area.

Acknowledgments

This research was funded by the National Key Research and Development Program of China (2023YFB3906204).We appreciate the help from Pingping Li, Hong Chang and Weiguo Li during field data collection.

References

1. Zhao, J.; Xu, X.; Liu, S.; Jia, J.; Li, M.; Huang, H.; Zhang, G.; Zhao, C. Optimizing Wheat Prosperity: Innovative Drip Irrigation and Nitrogen Management Strategies for Enhanced Yield and Quality of Winter Wheat in the Huang-Huai-Hai Region. *Front. Plant Sci.* **2024**, *15*, 1454205. https://doi.org/10.3389/fpls.2024.1454205.

2. Sun, Y.; Zhang, S.; Tao, F.; Aboelenein, R.; Amer, A. Improving Winter Wheat Yield Forecasting Based on Multi-Source Data and Machine Learning. *Agriculture* **2022**, *12*, 571. https://doi.org/10.3390/agriculture12050571.

3. Schierhorn, F.; Hofmann, M.; Gagalyuk, T.; Ostapchuk, I.; Müller, D. Machine Learning Reveals Complex Effects of Climatic Means and Weather Extremes on Wheat Yields during Different Plant Developmental Stages. *Clim. Change* **2021**, *169*, 39. https://doi.org/10.1007/s10584-021-03272-0.

4. Kristensen, K.; Schelde, K.; Olesen, J.E. Winter Wheat Yield Response to Climate Variability in Denmark. *J. Agric. Sci.* **2011**, *149*, 33–47. https://doi.org/10.1017/S0021859610000675.

5. Gurjar, S.B.; Padmanabhan, N. Study of Various Resampling Techniques for High-Resolution Remote Sensing Imagery. *J. Indian Soc. Remote Sens.* **2005**, *33*, 113–120. https://doi.org/10.1007/BF02989999.

6. Lyons, M.B.; Keith, D.A.; Phinn, S.R.; Mason, T.J.; Elith, J. A Comparison of Resampling Methods for Remote Sensing Classification and Accuracy Assessment. *Remote Sens. Environ.* **2018**, *208*, 145–153. https://doi.org/10.1016/j.rse.2018.02.026.

7. Pan, Y.; Li, L.; Zhang, J.; Liang, S.; Hou, D. Crop Area Estimation Based on MODIS-EVI Time Series According to Distinct Characteristics of Key Phenology Phases: A Case Study of Winter Wheat Area Estimation in Small-Scale Area. *J. Remote Sens.* **2011**, *15*, 17.

8. Zhang, F.; Wu, B.; Liu, C.; Luo, Z. Methods of Monitoring Crop Phonological Stages Using Time Series of Vegetation Indicator. *Trans. CSAE* **2004**, 155–159.

9. Huang, J.; Zhao, J.; Wang, X.; Xie, Z.; Zhuo, W.; Huang, R. Extraction Method of Growth Stages of Winter Wheat Based on Accumulated Temperature and Remote Sensing Data. *Trans. Chin. Soc. Agric. Mach.* **2019**, *50*, 169–176. https://doi.org/10.6041/j.issn.1000-1298.2019.02.019.

10. Li, Y.; Zhang, C.; Luo, W.; Gao, W. Summer Maize Phenology Monitoring Based on Normalized Difference Vegetation Index Reconstructed with Improved Maximum Value Composite. *Trans. Chin. Soc. Agric. Eng.* **2019**, *35*, 159–165. https://doi.org/10.11975/j.issn.1002-6819.2019.14.020.

11. Kong, L.; Yan, H.; Bao, Y.; Chen, H. Remote Sensor Monitoring Method for Winter Wheat Growth Based on Key Development Periods. *Chin. J. Agrometeorol.* **2012**, *33*, 424–430.

12. MacQueen, J. (1967). Some methods for classification and analysis of multivariate observations. In Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability (Vol. 1, pp. 281–297). University of California Press.

13. Dudani, S.A. The Distance-Weighted k-Nearest-Neighbor Rule. *IEEE Trans. Syst. Man Cybern.* **1976**, *SMC-6*, 325–327. https://doi.org/10.1109/TSMC.1976.5408784.

14. Cover, T.; Hart, P. Nearest Neighbor Pattern Classification. *IEEE Trans. Inf. Theory* **1967**, *13*, 21–27. https://doi.org/10.1109/TIT.1967.1053964.

15. Getis, A.; Aldstadt, J. Constructing the Spatial Weights Matrix Using a Local Statistic. *Geogr. Anal.* **2004**, *36*, 90–104. https://doi.org/10.1111/j.1538-4632.2004.tb01127.x.

16. Chavent, M.; Kuentz-Simonet, V.; Labenne, A.; Saracco, J. ClustGeo: An R Package for Hierarchical Clustering with Spatial Constraints. *Comput. Stat.* **2018**, *33*, 1799–1822. https://doi.org/10.1007/s00180-018-0791-1.