Supplemental Materials:

Table S1. Statistical values of climate elements for the four sub-regions in Henan.

No.	Climate elements	Region I	Region II	Region III	Region IV
1	Mean temperature (℃)	9.4±0.1	10.4±0.4	9.4±0.4	10.2±0.8
2	Precipitation (mm)	207.2±6.2	314.0±93.8	197.7±39.5	324.1±137.8
3	Atmospheric pressure (hPa)	977.8±6.7	1003.3±6.0	1013.2±0.3	1014.3±4.0
4	Relative humidity (%)	58.1±0.2	63.9±6.2	58.9±9.9	65.5±5.9
5	Hours of sunshine (h)	5.5±0.2	4.9±0.3	5.1±0.2	5.0±0.3
6	Wind speed (m/s)	2.4±0.6	2.1±0.3	2.5±0.2	2.2±0.3

Section S1. Detailed explanation of statistical methods used in our study

In this study, we used several statistical methods. Detailed explanation of these statistical methods can be shown as follows.

1. Mann-Kendall test

A trend analysis is one of the most important measurements in studying time series data. The Mann–Kendall trend test is a widely used non-parametric tests to detect significant trends in time series, which is based on the correlation between the ranks of a time series and their time order [34]. It can test trends in a time series without requiring normality or linearity [35]. It is therefore highly recommended for general use by the World Meteorological Organization. The test statistic S is given by

$$S = \sum_{i=1}^{n-1} \sum_{j=i+1}^{n} sgn(x_j - x_i) \tag{S1}$$

where n is the number of data points, x_i and x_j are the data values in time series i and j ($j>i$), respectively and $sgn(x_j-x_i)$ is the sign function as:

$$sgn(x_j - x_i) = \begin{cases} +1, & \text{if } x_j - x_i > 0 \\ 0, & \text{if } x_j - x_i = 0 \\ -1, & \text{if } x_j - x_i < 0 \end{cases} \tag{S2}$$

The variance is computed as

$$Var(S) = \frac{n(n-1)(2n+5) - \sum_{i=1}^{m} t_i(t_i-1)(2t_i+5)}{18} \tag{S3}$$

where n is the number of data points, m is the number of tied groups and t_i denotes the number of ties of extent i. A tied group is a set of sample data having the same value. In cases where the sample size n>10, the standard normal test statistic Z_S is computed using Eq. (S4):

$$Z_S = \begin{cases} \dfrac{S-1}{\sqrt{Var(S)}}, & \text{if } S > 0 \\ 0, & \text{if } S = 0 \\ \dfrac{S+1}{\sqrt{Var(S)}}, & \text{if } S < 0 \end{cases} \quad \text{(S4)}$$

Positive values of Z_S indicate increasing trends while negative Z_S values show decreasing trends. Testing trends is done at the specific α significance level. When $|Z_S| > Z_{1-\alpha/2}$, the null hypothesis is rejected and a significant trend exists in the time series. $Z_{1-\alpha/2}$ is obtained from the standard normal distribution table. In this study, significance levels α= 0.05 were used. At the 5% significance level, the null hypothesis of no trend is rejected if $|Z_S|$>1.96.

In this study, we used the Mann–Kendall test to detect the trend of the annual winter wheat yield (AWWY) time series.

2. Sen's slope estimator

Sen (1968) [36] developed the non-parametric procedure for estimating the slope of trend in the sample of N pairs of data:

$$Q_i = \frac{x_j - x_k}{j-k} \quad \text{if } i = 1,...,N \quad \text{(S5)}$$

where x_j and x_k are the data values at times j and k (j>k), respectively. If there is only one datum in each time period, then N = n(n-1)/2, where n is the number of time periods. If there are multiple observations in one or more time periods, then N < n(n-1)/2, where n is the total number of observations.

The N values of Q_i are ranked from smallest to largest and the median of slope or Sen's slope estimator is computed as

$$Q_{med} = \begin{cases} Q_{[(N+1)/2]}, & \text{if } N \text{ is odd} \\ \dfrac{Q_{[N/2]} + Q_{[(N+2)/2]}}{2}, & \text{if } N \text{ is even} \end{cases} \quad \text{(S6)}$$

The Q_{med} sign reflects data trend reflection, while its value indicates the steepness of the trend. To determine whether the median slope is statistically different than zero, one should obtain the confidence interval of Q_{med} at specific probability.

The confidence interval about the time slope [34] can be computed as follows:

$$C_a = Z_{1-\alpha/2}\sqrt{Var(S)} \tag{S7}$$

where $Var(S)$ is defined in Eq. (3) and $Z_{1-\alpha/2}$ is obtained from the standard normal distribution table. In this study, the confidence interval was computed at one significance level (α=0.05). Then, $M^1 = (N-C_\alpha)/2$ and $M = (N+C_\alpha)/2$ are computed. The lower and upper limits of the confidence interval, Q_{min} and Q_{max}, are the M_1th largest and the (M_2+1)th largest of the N ordered slope estimates [33]. The slope Q_{med} is statistically different than zero if the two limits (Qmin and Qmax) have similar sign.

In this study, we used the Sen's slope estimator to calculate the trend magnitude of the AWWY time series.

3. Hurst method

In this study, we used Hurst's rescaled range (R/S) analysis and the corresponding Hurst Exponent [37] to detect the future trends of the AWWY time series. The basic idea of the R/S analytical method could be described as follows:

For the time series of a certain physical quantity {x(τ)} (τ =1, 2,…, n), the average value of x(τ) is

$$x_\tau = \frac{1}{\tau}\sum_{t=1}^{\tau} x(t) \tag{S8}$$

The cumulative deviation is

$$X(t,\tau) = \sum_{t=1}^{\tau}(x(t)-x_\tau) \quad 1 \le t \le \tau \tag{S9}$$

The range sequence is

$$R(\tau) = \max_{1\le t\le\tau} X(t,\tau) - \min_{1\le t\le\tau} X(t,\tau) \tag{S10}$$

The standard deviation sequence is

$$S(\tau) = \sqrt{(\frac{1}{\tau}\sum_{t=1}^{\tau}(x(t)-x\tau)^2)} \tag{S11}$$

The non-dimensional ratio R/S is defined as

$$\frac{R(\tau)}{S(\tau)} = (\alpha\tau)^H \quad \text{(S12)}$$

H is the Hurst Exponent. When H = 0.5, it means that the time series is an independent random process, which indicates that the current trend will not affect the future trend. When 0.5 < H < 1, it describes a dynamically persistent, or trend reinforcing series; with the greater the H value, the stronger the persistent. When 0 < H < 0.5, it describes an anti-persistent, or a mean reverting system; the smaller the H value, the stronger the anti-persistent [38].

4. Ensemble empirical mode decomposition

EMD decomposes nonlinear and nonstationary data series for extracting a finite number of decomposed components termed as IMFs. Each IMF shows an oscillatory pattern, which may represent physically meaningful information hidden in the original data series [39]. The IMFs should satisfy two conditions: 1) the number of extrema and zero crossings must either be equal to each other or differ at most by one in the whole data series; and 2) the mean value of the upper envelope defined by connecting all the local maxima, and the lower envelope defined by connecting all the local minima, should be zero at any point. EMD is performed by an iterative process called a sifting algorithm as follows:

(1) Find the upper and lower envelopes by connecting the local maxima ($X_u(t)$) and local minima ($X_l(t)$) using the Cubic Spline method for a given time series X(t), t=1,2,3,…,T (T is data length).

(2) Calculate the mean value between the local maxima and local minima, that is, $X_{mean}(t)=(X_u(t))+X_l(t))/2$.

(3) Obtain h(t) by extracting the $X_{mean}(t)$ from the original time series X(t), that is, $h(t)=X(t)-X_{mean}(t)$.

(4) Check whether h(t) satisfies the two conditions of IMFs or not. If h(t) is an IMF, h(t) is the first IMF of the given time series; else treat X(t) as h(t) and iterate steps (1) to (3) until h(t) satisfies the two conditions of IMFs.

(5) Define a new time series n(t) by extracting IMF(t) from the original time series X(t), that is, n(t)=X(t)-IMF(t); and the original time series X(t) is replaced by n(t).

(6) Repeat steps (1) to (5) until no more IMFs can be extracted; and the last IMF becomes the residue, r(t).

(7) Finally, the original time series X(t) can be written as Eq (S13).

$$X(t) = \sum_{i=1}^{m} IMF_i(t) + r(t) \quad \text{(S13)}$$

where m is the number of IMFs.

EMD suffers the drawbacks caused by the appearance of mode mixing, which is a specific signal retained in different IMF components. To ameliorate the drawbacks caused by the mode mixing problem, Wu and Huang (2009) [40] proposed EEMD.

EEMD is a noise-assisted analysis method that sifts data with an added ensemble of white noise signals and treats the mean as the final result [40]. A white noise signal is added to a given time series, and it is decomposed by the sifting algorithm as described above. Different white noise signals are added to each repetition. The ensemble means of the decomposed components can be obtained as the final result. This additional step improves the results over the EMD process by eliminating the chance of mode mixing and extracting improved physical meaning from the decomposed components [41].

In this study, AWWY time series can be decomposed into three periodic oscillation intrinsic mode functions (IMFs) and a trend component, and this was achieved by using ensemble empirical mode decomposition (EEMD). The variance contribution of each IMF (i.e., CIMF1, CIMF2 and CIMF3) and the trend component (CTrend) to AWWY was selected to reflect the frequency-domain characteristics of the AWWY time series.

5. Principal Components Analysis (PCA)

The PCA method is a technique applied to multivariate analysis for dimensionality reduction, emphasizing patterns on data and relations between variables and between variables and observations [43]. The original intercorrelated variables could be reduced to a small number of new linearly uncorrelated ones that explain most of the total variance [44].

Considering k variables in a given time period i, $X_{i,1}$, $X_{i,2}$, …, $X_{i,k}$, k principle components (PCs) are produced for the same time period, $Y_{i,1}$, $Y_{i,2}$, …, $Y_{i,k}$, using linear combinations of the first ones.

$$\begin{cases} Y_{i,1} = a_{11}X_{i,1} + a_{12}X_{i,2} + \ldots + a_{1k}X_{i,k} \\ Y_{i,2} = a_{21}X_{i,1} + a_{22}X_{i,2} + \ldots + a_{2k}X_{i,k} \\ \quad\quad\quad \ldots \\ Y_{i,k} = a_{k1}X_{i,1} + a_{k2}X_{i,2} + \ldots + a_{kk}X_{i,k} \end{cases} \quad \text{(S14)}$$

In the previous combinations the Y values are orthogonal and uncorrelated variables, such that $Y_{i,1}$ explains most of the variance, $Y_{i,2}$ explains the reminiscent amount of variance, and so on. The coefficients of the linear combinations are called "loadings" and represent the weights of the original variables in the PCs.

PCs extraction could be based on variance/covariance or correlation matrix of data with $\{a_{11}, a_{12}, \ldots, a_{1k}\}$ being the first eigenvector and $\{a_{k1}, a_{k2}, \ldots, a_{kk}\}$ being the eigenvector of k order. Each eigenvector includes the coefficients of the k principal component.

Finally, the amount of variance explained by the first PC is called the first eigenvalue, λ_1, the second is λ_2, so that $\lambda_1 \geq \lambda_2 \geq \lambda_3 \geq \lambda_4 \ldots \lambda_k$, since each eigenvalue represents the fraction of the total variance in the original data and explained by each component [45] so that this proportion can be calculated as $\lambda_j / \sum \lambda_j$. The analysis of the results of PCs can be focused on the eigenvalues, on the correlations between PCs and the original variables (factor loadings), or on the observation coordinates in the PC (factor scores).

In this study, the values of 13 indices in 17 cities in Henan were first calculated. Then, the principal components of 13 columns (13 indices) × 17 rows (17 cities) were analyzed using the PCA method. To reduce the dimensionality of the data, we selected the first three principal components (PCs) because their respective eigenvalues were greater than 1, and their cumulative variances exceeded 89%.

6. K-means clustering analysis

K-means is applied to divide the study area into multiple geographical clusters with homogeneous temporal patterns. K-means with Euclidean Distance (ED) has been popular during the past 60 years [46]. It is proved to be an effective, robust, and also computationally efficient approach in clustering time series data while compared to other raw-based (directly work with the raw data) [47] or model-based (indirectly work with models built on the raw data) [48] time series clustering algorithms. It explores the structure of the data at a higher level of abstraction without artificial interference. For time series clustering, it identifies clusters with

homogeneous temporal patterns through the comparison of similarity among the time series [47]. Theoretically, the cluster centroids get updated iteratively until the distances between pixels and centroids are minimized.

In this study, we obtained the corresponding score sequences according to the first three PCs. A matrix of 3 columns (3 score sequences) × 17 rows (17 cities) was clustered using the K-means method in our study in order to partition the 17 cities into *k* clusters, and the clustering result was evaluated using silhouette coefficients (SCs). The highest SC value determined the optimum number of clusters.

7. Time-lag correlation analysis

To explore the relationships between the key meteorological drought/wetness index (KMDWI) and seven atmospheric circulation indices, correlation coefficients between the KMDWI and seven atmospheric circulation indices were calculated as follows [49]:

$$r_k(x,y) = \frac{\sum_{i=1}^{n-k}[(x_i - \overline{x}_i)(y_{i+k} - \overline{y}_{i+k})]}{\sqrt{\sum_{i=1}^{n-k}(x_i - \overline{x}_i)\sum_{i=1}^{n-k}(y_{i+k} - \overline{y}_{i+k})}} \quad \text{(S15)}$$

where $r_k(x,y)$ is correlation coefficient series between the KMDWI and seven atmospheric circulation indices; n is the length of series; k is time lag; x_i is KMDWI time series; and y_{i+k} is the time series for atmospheric circulation indices, with a time lag of k. In this study, a time lag of 0-11 months was selected. According to the results of the correlation analysis, we then selected the atmospheric circulation indices with the highest correlation with the corresponding KMDWI for each month as the premonitory influencing signals.

8. Multiple linear regression method

Multiple linear regression is used to explain the relationship between a response variable and a number of explanatory variables [50]. The general form of a multiple linear regression model is the following equation with a response variable (=predicted data) $Y(t)$ and explanatory variables $X_p(t)$ [39].

$$Y(t) = \beta_0 + \sum_{p=1}^{N} \beta_p x_p(t) + \varepsilon \quad \text{(S16)}$$

where N is the number of explanatory variables, β_0 is a constant term (intercept), β_p are the regression coefficients for explanatory variables, and ε is a noise term. The method of least squares estimation is used to estimate parameters.

In this study, we used the multiple linear regression method to construct an empirical KMDWI simulation model based on the selected atmospheric circulation indices. Two indicators were used to evaluate the performance of the multiple linear regression model: the determination coefficients (R^2) and root mean square error (RMSE).