Temperature trend analysis and extreme high temperature prediction based on weighted Markov Model in Lanzhou

Zhiqiang Pang1 · Zhaoxu Wang1,2

Received: 7 January 2021 / Accepted: 20 March 2021 / Published online: 7 April 2021
© The Author(s), under exclusive licence to Springer Nature B.V. 2021

Abstract
In this study, temporal trend analysis was conducted on the annual and quarterly meteorological variables of Lanzhou from 1951 to 2016, and a weighted Markov model for extremely high temperature prediction was constructed. Several non-parametric methods were used to analyse the trend of meteorological variables. Considering that sequence autocorrelation may affect the accuracy of the trend test, we performed an autocorrelation test and carried out trend analysis for sequences with autocorrelation after removing correlation. The results show that the maximum temperature, minimum temperature and average temperature in Lanzhou all have a significant upward trend and show different performances in each season. In detail, the trend of maximum temperature in the summer is not significant, while the upward trend of minimum temperature in the winter is the most significant, which leads to more and more “warm winter” phenomenon. Finally, we construct a weighted Markov prediction model for extremely high temperature and obtain the conclusion that the prediction results by the model are consistent with the actual situation.

Keywords Meteorological variables · Trend analysis · Mann–Kendall test · Weighted Markov model

1 Introduction
The trend analysis of temperature change is one of the main aspects of climate change research and an important means to understand global climate change. The Intergovernmental Panel on Climate Change (IPCC) assessment points out that there has been a marked trend in global temperature change since the 20th century and that the frequency of extreme weather and climate events has increased. Therefore, the trend analysis of temperature change plays an important role in the detection of climate change.
In recent years, many scholars have done extensive research on the trend analysis of temperature variation in different regions. It is the main content of research to use trend analysis to explore the trend, direction, amplitude and abrupt of hydro-climatic data change. Partal (2017) and Sharma et al. (2016) explored the annual and multi-annual variation characteristics of data such as temperature and rainfall by using trend analysis method and summarized the law of climate change. China’s extreme temperature changes are generally consistent with global’s, He and Zhang (2005) analyzed climate change trends and characteristics along the Lancang River (China) according to the archival data of monthly air temperature and precipitation series. Zhai and Pan (2003) studied the extreme temperature in northern China in the past 50 years, and concluded that in northern China, both extremely minimum temperature and extremely maximum temperature tend to be warmer. In the northwest of China, extremely minimum and maximum temperatures have undergone an asymmetric growth (Shi and Zhao 2014). Liu et al. (2016) analyzed the hot weather and high temperature days in the summer of Gansu Province from 1981 to 2010, they found the summer high temperature days for nearly 30 years had an increasing trend.

In general, we are mainly concerned about two issues for the trend analysis of time series data. The first is whether there is a significant trend, the direction of trend change and the magnitude of change range. The second is whether there is a break in the time series and when the break occurs. Faced with these two problems, current research methods on trend analysis can be divided into two categories. One is trend testing with non-parametric methods. Koutsoyiannis (2003) pointed out that changes in climatic data over the past period can be reflected by the non-stationarity of the data. Therefore, the commonly used methods to solve the first kind of problems are Mann–Kendall (MK) test, Spearman’s rho test and Sen’s slope estimation. The MK test is a non-parametric trend test proposed by Mann (1945) and Kendall (1975), however, it and Spearman’s rho test can only be used to test the trend but not to estimate the amplitude of the trend. Although the amplitude estimation of the change trend can be obtained by the least square method, considering the normality assumption of the least square method and its sensitivity to extreme values, the non-parameter method of Sen’s slope (SS) estimation is proposed by Sen (1968). This method has strong robustness to the extreme value and no assumption of normality to the original data, so it has been widely used especially in the trend test and estimation of meteorological data. The annual and quarterly trends of temperature change in Serbia are studied by using the above methods in Gocic and Trajkovic (2013), and more applications can be found in Chen et al. (2009), Gocic and Trajkovic (2013), Meena (2020), Partal (2017), Sayemuzzaman and Jha (2014), Sonali and Kumar (2013) and Yacoub and Tayfur (2018). The other is the change point test method for time series, in which Pettitt’s method (Pettitt 1979) is mainly used for change point test in meteorological variable data. Yacoub and Tayfur (2018) and Partal (2017) used Pettitt’s method to test the annual change point of rainfall and temperature sequences. In this study, the above two kinds of methods are used to analyse the trend of meteorological data in Lanzhou.

In the above trend analysis articles, most studies only used the non-parametric methods to study whether there is a change trend, and few articles paid attention to the effect of sequence autocorrelation on the trend. Both the MK test and Sen’s slope estimation require time series to be serially independent which can be accomplished by using the pre-whitening technique (von Storch 1995). The trend test in this paper takes into account the influence of sequence autocorrelation and carries out trend test after removing the correlation of sequences with autocorrelation.

Markov process is used to describe random processes without aftereffect. It is used in many natural sciences such as queuing theory, biology and geography. As an improvement of
Markov model, weighted Markov model has been applied in stochastic phenomenon prediction, disaster prediction and other fields. Tian et al. (2010) applied the grey weighted Markov chain model to forecast the utility of water supply. Gong et al. (2013) used weighted Markov model to predict the low temperature risk in Nanjing. Masseran (2015) analyzed the stochastic behaviors of wind-direction data by Markov chain model. In our study, we used the method in literature Gong et al. (2013) to classify the state of the extremely high temperature series in Lanzhou and used weighted Markov to predict the occurrence of extremely high temperature.

The main purposes of this paper are as follows: (1) Analyse the variation trend of the meteorological variables in Lanzhou, conduct in-depth research on the different changes of the annual and seasonal trends and give the specific variation trend in the annual and seasonal. (2) Consider the influence of autocorrelation of time series on the trend and discuss the change trend after removing the correlation of the influential series. (3) Explore whether there are changes in meteorological variables and when they occur. (4) Prediction of extremely high temperature weather using weighted Markov model.

The remainder of this paper is organized as follows. In Sect. 2, the data source used in this paper and the form of the data are described; In Sect. 3, the common non-parametric analysis methods using trend analysis and the Markov model used for prediction are introduced. In Sect. 4, we use the methods and models in Sect. 3 to analyse and apply the data.

2 Study area and data

Lanzhou is located in the northwest of China, the geometric center of China’s continental map, and the central part of Gansu Province. The city center is located at 36°03′ north latitude and 103°40′ east longitude, with a total area of 13,085.6 square kilometers. Lanzhou has a temperate continental climate with large temperature difference and little precipitation. It has a mild climate without severe cold in winter or scorching heat in summer. In recent years, more and more people begin to pay attention to the change in temperature. Temperature change affects not only agricultural production but also the daily life of residents. Therefore, it is of great significance to analyse the temperature variation trend in Lanzhou.

In this paper, the meteorological datasets of Lanzhou from 1951 to 2016 collected by Lanzhou Central Meteorological Observatory of Gansu Province are used, which include daily temperature, rainfall, pressure and variables such as whether extreme weather occurs. First of all, basic data cleaning is carried out for the original data. We focus on analyzing the variation trend of daily temperature and rainfall. Through the analysis of the collected data, the annual average rainfall of Lanzhou is 311.2 mm, and the annual average temperature is 9.87 °C. August is the wettest month with an average rainfall of 72.2 mm, while December is the least wet month with an average rainfall of only 1 mm. The average annual temperature in July was as high as 22.7 °C, while the average annual temperature in January was as low as −5.6 °C. Detailed descriptions of temperature and rainfall are presented in the following sections.
3 Methods and model

3.1 Trend analysis

In this part, several non-parametric methods of trend analysis are introduced. It is mainly used to test whether there is a trend in meteorological data variables (such as temperature), and to estimate the size of the trend if it exists.

3.1.1 Mann–Kendall test

The Mann–Kendall (MK) test is a non-parametric rank-based test (Kendall 1975; Mann 1945), which does not require the data to be distributed normally. The test statistic \( S \), which has mean zero and a variance computed by Eq. (2), is calculated using Eq. (1):

\[
S = \sum_{i=2}^{n} \sum_{j=1}^{i-1} \text{sign}(x_i - x_j),
\]

\[
\text{Var}(S) = \frac{n(n-1)(2n+5)}{18} - \sum_{p=1}^{q} t_p(t_p-1)(2t_p+5)/18.
\]

where \( x_i \) and \( x_j \) are the data value in time point \( i \) and \( j \), \( \text{sign()} \) is the sign function. \( n \) is the number of data points, \( p \) is the number of tied groups, and \( t_p \) denotes the number of ties of extent \( p \). A tied group is a set of sample data having the same value. In cases where the sample size \( n \neq 10 \), the standard normal test statistic \( Z \) is computed by Eq. (3):

\[
Z = \begin{cases} 
\frac{S-1}{\sqrt{\text{Var}(S)}} & \text{if } S > 0; \\
0 & \text{if } S = 0; \\
\frac{S+1}{\sqrt{\text{Var}(S)}} & \text{if } S < 0.
\end{cases}
\]

Positive values of \( Z \) indicate increasing trends while negative \( Z \) values show decreasing trends. To test for monotonic trend at a significance level, the null hypothesis of no trend is rejected if the absolute value of standardized test statistic \( Z \) is greater than \( Z_{1-a/2} \).

3.1.2 Sen’s slope estimate (SS)

If a linear trend is present in a time series, the true slope (change per unit time) can be estimated using a simple non-parametric procedure developed by Sen (1968). The slope estimates of \( N \) pairs of data are first computed by

\[
Q_i = \frac{x_j - x_k}{j - k} \quad \text{for } i = 1, 2, \ldots, N.
\]

where \( x_j \) and \( x_k \) are data values at times \( j \) and \( k(j > k) \), respectively. The median of these \( N \) values of \( Q_i \) is Sen’s estimator of slope. Sen’s estimator is computed by

\[
Q_{\text{med}} = \begin{cases} 
Q_{[(N+1)/2]}, & N \text{ is odd} \\
\frac{1}{2}(Q_{[N/2]} + Q_{[(N+2)/2]}), & N \text{ is even}
\end{cases}
\]
Positive and negative signs of test statistics indicate increasing and decreasing trends.

### 3.1.3 Spearman’s rho test

Like the MK test, the rank-based non-parametric statistical SR test is commonly used to assess the significance of monotonic trends in hydro-meteorological data time series (Yacoub and Tayfur 2018). This method detects the existence and non-existence of trends in data time series. It can also identify if there is increase or decrease in the trend. In this test: the null hypothesis $H_0$ means that the given data are independent and identically distributed in time, while the alternative hypothesis $H_1$ indicates that a trend exists. The SR test statistic $D$ and the standardized test statistic $Z_{SR}$ are calculated by Eqs. (4) and (5) as follows.

$$D = 1 - \frac{6}{n(n^2 - 1)} \sum_{i=1}^{n} (R(X_i) - i)^2,$$  

(4)

$$Z_{SR} = D \sqrt{\frac{n-2}{1-D^2}}.$$  

(5)

where $R(X_i)$ is the rank of $i$th observation $X_i$ in the sample size $n$. In this test, $H_0$ is rejected and $H_1$ is accepted if $|Z_{SR}| > 2.08$ for the 5% significance level. Positive values of $Z_{SR}$ indicate trend increase, while the negative values indicate trend decrease.

### 3.1.4 Serial correlation effect

In trend analysis of time series, the correlation of time series will affect the result of trend analysis. Therefore, if there is correlation, the trend analysis should be carried out after the correlation is removed.

The lag-1 serial coefficient $r_1$ of sample data $X_i$, originally derived by Salas et al. (1985), but several recent researchers have been utilizing the same equation (Gocic and Trajkovic 2013) to compute $r_1$. It can be computed by

$$r_1 = \frac{1}{n-1} \sum_{i=1}^{n-1} (x_i - E(x_i)) \cdot (x_{i+1} - E(x_{i+1})) \cdot \frac{1}{n} \sum_{i=1}^{n} (x_i - E(x_i))^2,$$

where $E(x_i)$ is the mean of sample data, and $n$ is the number of observations in the data. According to Salas et al. (1985) and most recent study, Gocic and Trajkovic (2013) have used the following equation for testing the time series data sets of serial correlation.

$$\frac{-1 - 1.645 \sqrt{(n-2)}}{n-1} \leq r_1 \leq \frac{-1 + 1.645 \sqrt{(n-2)}}{n-1}.$$

If $r_1$ falls inside the above interval, then the time series data sets are independent observations. In cases where $r_1$ is outside the above interval, the data are serially correlated. If time series data sets are independent, then the MK test and the SS can be applied to original values of time series. If time series data sets are serially correlated, then the ‘pre-whitened’ time series may be obtained as $(x_2 - r_1x_1, x_3 - r_1x_2, \ldots, x_n - r_1x_{n-1})$ (Gocic and Trajkovic 2013; Sayemuzzaman and Jha 2014).
3.1.5 Pettitt’s change point test

The Pettitt test developed by Pettitt (1979), was used to find the existence of a main change point in the time series. This test determines a significant change in the general tendency of a time series. Test statistic, $K$ is given as:

$$K = \max_{1 \leq k \leq N} |U_k|,$$

where $U_k$ is equivalent to a Mann–Whitney statistic using for testing two samples $(x_1, x_2, \ldots, x_k)$ and $(x_{k+1}, x_{k+2}, \ldots, x_N)$. $U_k$ can be acquired as below:

$$U_k = 2 \sum_{i=1}^{k} r_i - k(N + 1),$$

where $r_i$ is the rank of the $i$th observation. A change point occurs in the series where $U_k$ reaches a maximum (Tarhule & Woo 1991). If a change point exists, then $U_k$ increases up to this change point, then begins to decrease. This point may occur several times in a time series, indicating several local change points. Significance level of the $K$ value is computed approximately as:

$$K_a = \left[ -\ln \alpha \left( \frac{N^3 + N^2}{6} \right) \right]^{0.5}.$$

If $K$ exceeds $K_a$, then the change point is statistically significant.

3.2 Weighted Markov chain modeling

The Fluctuation Index is introduced in Markov chain modeling process:

$$v_k = \frac{x^{(o)}(k) - \hat{x}^{(o)}(k)}{\hat{x}^{(o)}(k)} \cdot 100\%, (k = 1, 2, \ldots, n-m).$$

Note: taking $\hat{x}^{(o)}(k)$ as the trend value of the $(k)$th item, $x^{(o)}(k) - \hat{x}^{(o)}(k)$ indicates the $(k)$th item’s absolute fluctuation. Obviously, the greater $v_k$ is, the farther the actual value deviates from the trend value and the worse the stability is; the smaller $v_k$ is, the closer the actual value deviates from the trend value and the better the stability is.

According to $v_k$, the fluctuation index sequence $V = (v_1, v_2, \ldots, v_{n-m})$ can be divided into $s$ states. If there is $v_k \in [a_{1i}, a_{2i}]$, $i = 1, 2, \ldots, s$, the fluctuation state of the $k$th item is in State $i(E_i)$ in which $a_{1i}, a_{2i}$ denote the lower bound and upper bound of $E_i$, respectively. Thereby, state set $E = (E_1, E_2, \ldots, E_s)$ is generated according to each fluctuation index.

In order to make full usage of the recent data and to reduce the effects of random errors, here adopts multi-step transitions by constructing state transition probability matrix. The transition probability of state $i$ to state $j$ after $w$ steps acts as $p^{(w)}_{ij}$:

$$p^{(w)}_{ij} = \frac{m^{(w)}_{ij}}{M_i},$$
Note: $m_{ij}^{(w)}$ is the number of state $i$ to state $j$ after $w$ steps’ transition; $M_i$ is the total number of state $i$. The transition probability matrix after $w$ steps is composed by transition probabilities $p_{ij}^{(w)}$ after $w$ steps:

$$P^{(w)} = \begin{bmatrix}
p_{11}^{(w)} & p_{12}^{(w)} & \cdots & p_{1s}^{(w)} \\
p_{21}^{(w)} & p_{22}^{(w)} & \cdots & p_{2s}^{(w)} \\
\vdots & \vdots & \ddots & \vdots \\
p_{s1}^{(w)} & p_{s2}^{(w)} & \cdots & p_{ss}^{(w)}
\end{bmatrix}, i = 1, 2, \ldots, s$$

Next, select the last $s$ items’ states as the initial states, and the transition steps $w$ are distances between the forecasting item and each selected item. Considering the initial states $(i_1, i_2, \ldots, i_s)$ of $s$ items, respectively, the corresponding transition probability after $w$ steps $p_{i}^{(w)} = \left( p_{i1}^{(w)}, p_{i2}^{(w)}, \ldots, p_{is}^{(w)} \right), i = 1, 2, \ldots, s$ should be adopted from the transition probability matrix $P^{(w)}$. Therefore, state transition probability matrix of the $t$th forecasting data can be obtained:

$$R = \begin{bmatrix}
p_{11}^{(1)} & p_{12}^{(1)} & \cdots & p_{1s}^{(1)} \\
p_{21}^{(1)} & p_{22}^{(1)} & \cdots & p_{2s}^{(1)} \\
\vdots & \vdots & \ddots & \vdots \\
p_{i1}^{(j)} & p_{i2}^{(j)} & \cdots & p_{is}^{(j)}
\end{bmatrix}, i = 1, 2, \ldots, s$$

Determine the state of the $t$th forecasting data by choosing the corresponding state $\ast$ of max $\{ p_i = \sum_{w=1}^{w} p_{iw}^{(w)}, i = 1, 2, \ldots, s \}$ and this state’s upper bound and lower bound are $a_{1\ast}, a_{2\ast}$, respectively.

Finally, the modified forecasting data can be calculated:

$$\hat{y}(t) = \hat{x}^{(0)}(t) \cdot \left[ 1 + 0.5 \left( a_{1\ast} + a_{2\ast} \right) \right].$$

For the prediction of the $(t+1)$th item, the state of the $t$th item should be added in the state transition probability matrix. Then equal dimension processing should be implemented through reconstructing the probability matrix. Consequently, the next modified forecasting value will come out. Repeating to use of this method until all the rest forecasting values are modified.

### 4 Application

In this section, The meteorological data of Lanzhou city are analyzed by the method introduced the Sect. 3. Firstly, the trend analysis of the maximum temperature, minimum temperature, average temperature and rainfall is carried out. Secondly, weighted Markov model is established to predict high temperature disaster.

#### 4.1 Trend analysis

Non-parametric trend test was carried out on the annual highest temperature($T_{\text{max}}$), lowest temperature($T_{\text{min}}$), average temperature($T_{\text{av}}$), temperature difference(diff-T) and annual average rainfall of annual data. The analysis results are shown in the Table 1.
In Table 1, the mean, standard deviation (sd), Mann–Kendall test statistic Z, Spearman rho test statistic $Z_{SR}$, sen’s slope, trend, year of change point occurred and lag-1 serial coefficient $r_1$ of meteorological variables are analyzed. From Table 1, the mean values of annual maximum temperature, minimum temperature and average temperature are, respectively, 35.6 °C, −16.0 °C, 9.9 °C, and the sd of average temperature is the smallest, while the sd of annual minimum temperature is the largest. The mean value of the annual mean temperature difference is 12.4 °C, and the variance is small, which means it’s less volatile. The annual average rainfall was 311.3 mm, and sd is 75.84 mm, indicating that it fluctuated within a large range.

In order to make the result of trend test more reasonable, first observe the lag-1 serial coefficient $r_1$. According to the conclusion of serial correlation effect test, if $r_1 \in (-0.218, 0.187)$, the series is considered to have no correlation effect. In Table 1, there are four variables whose lag-1 serial coefficients are outside the interval, that is, the data corresponding to the variable are serially correlated. We have to consider the new sequence $\left( x_2 - r_1x_1, x_3 - r_1x_2, \ldots, x_n - r_1x_{n-1} \right)$, which is removed the correlation. After the correlation is removed, the new sequence is analyzed again. The test results are shown in Table 2.

It can be seen from the test results in Table 2, firstly, the test coefficients of the temperature series are all within the range $(-0.218, 0.187)$, indicating that the series is independent at this time, and the trend test results are valid. Secondly, according to the results of Mann–Kendall and Spearman rho test, there was no significant variation trend in the annual maximum temperature and the annual average rainfall series, while the annual minimum temperature and annual average temperature series have significant increased trend, while the temperature difference series has a significant decreased trend.

To sum up, the main reasons for the climate warming in Lanzhou in recent years are the increase of the minimum temperature and the decrease of the temperature difference. The average annual minimum temperature increases by 0.104 °C, the average temperature increases by 0.044 °C, and the temperature difference decreases by 0.026 °C.

### Table 1

| Variable  | Mean  | SD    | MK-Z  | $Z_{SR}$   | Sen’s slope | Trend | Change point | Lag-1 |
|-----------|-------|-------|-------|------------|--------------|-------|--------------|-------|
| $T_{\text{max}}$ | 35.574 | 1.542 | 2.326** | 2.472** | 0.023 | Yes | 1996 | 0.303 |
| $T_{\text{min}}$ | −16.003 | 2.660 | 5.813*** | 7.010*** | 0.104 | Yes | 1981 | 0.501 |
| $T_{\text{av}}$ | 9.874 | 0.968 | 7.394*** | 11.747*** | 0.044 | Yes | 1986 | 0.799 |
| Diff-T | 12.401 | 0.802 | −5.617*** | −7.013*** | −0.026 | Yes | 1973 | 0.554 |
| Rainfall | 311.3 | 75.84 | −0.852 | −0.908 | −0.462 | No | No | 0.063 |

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

### Table 2

| Variable  | MK-Z  | $Z_{SR}$   | Trend | Lag-1 |
|-----------|-------|------------|-------|-------|
| $T_{\text{max}}$ | 1.5796 | 1.681 | No | −0.025 |
| $T_{\text{min}}$ | 3.0288*** | 3.000*** | Yes | −0.171 |
| $T_{\text{av}}$ | 2.1796** | 2.268** | Yes | −0.161 |
| Diff-T | −2.2589** | −2.279** | Yes | −0.105 |

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
To illustrate the performance of the increased trend of temperature in different seasons, the above annual meteorological data were divided into seasons, and the non-parametric trend test statistics of each variable in different seasons were obtained. The test results are shown in Table 3.

Similarly, we first check the lag-1 serial coefficient $r_1$. Some of the Lag-1 serial coefficients in the table are outside the interval $(-0.218, 0.187)$, which indicating that there is a correlation, and the sequence needs to be de-correlated. Table 4 shown the decorrelated test results for correlated sequences.

According to Table 4, in the test of correlation coefficient, the average temperature and the temperature difference in winter are still correlated. We used the method of decorrelation again to remove the correlation before the trend test. The conclusions of trend test are shown in Table 5.

In Table 5, according to the value of correlation effect test statistic $r_1$, there is no correlation effect in the sequence, and the trend test conclusion is reliable. Synthesize the test results in Tables 3, 4 and 5, first of all, from the average level of temperature and rainfall in the four seasons, Lanzhou has distinctive seasons with the largest temperature difference in spring and the smallest in autumn. Secondly, in terms of the variation trend of meteorological variables, there is no significant variation trend of rainfall and summer maximum temperature, while other seasonal temperature variables have significant variation trend. Among the variables with obvious trend, the maximum temperature, the minimum temperature and the average temperature all show an increasing trend, while the temperature difference shows a decreasing trend.

| Season | Variable | Mean   | SD    | MK-Z  | $\text{Sr}(Z_{SR})$ | Sen’s slope | Trend | Change point | Lag-1 |
|--------|----------|--------|-------|-------|----------------------|-------------|-------|--------------|-------|
| Spring | $T_{\text{max}}$ | 31.406 | 1.517 | 2.227 | 2.180 | 0.025 | Yes | 1993 | 0.226 |
|        | $T_{\text{min}}$ | -6.474 | 2.828 | 2.397 | 2.395 | 0.042 | Yes | No | 0.114 |
|        | $T_{\text{av}}$ | 11.687 | 1.047 | 5.324 | 6.444 | 0.034 | Yes | 1991 | 0.462 |
|        | Diff-T | 13.263 | 0.864 | -3.038 | -3.237 | -0.017 | Yes | 1969 | 0.321 |
|        | Rainfall | 62.57  | 29.91 | 0.321 | 0.395 | 0.06 | No | No | 0.0039 |
| Summer | $T_{\text{max}}$ | 35.537 | 1.544 | 2.309 | 2.458 | 0.023 | Yes | 1996 | 0.301 |
|        | $T_{\text{min}}$ | 9.144  | 1.897 | 4.878 | 5.619 | 0.060 | Yes | 1997 | 0.427 |
|        | $T_{\text{av}}$ | 21.653 | 0.992 | 4.605 | 5.349 | 0.030 | Yes | 1996 | 0.573 |
|        | Diff-T | 12.701 | 0.721 | -3.553 | -3.861 | -0.017 | Yes | 1974 | 0.109 |
|        | Rainfall | 171.55 | 65.16 | -0.43 | -0.48 | -0.23 | No | No | 0.102 |
| Autumn | $T_{\text{max}}$ | 29.061 | 2.182 | 2.807 | 3.136 | 0.04 | Yes | 1986 | 0.250 |
|        | $T_{\text{min}}$ | -7.902 | 2.705 | 4.098 | 4.517 | 0.066 | Yes | 1981 | 0.105 |
|        | $T_{\text{av}}$ | 9.592  | 1.014 | 6.375 | 8.854 | 0.039 | Yes | 1986 | 0.601 |
|        | Diff-T | 11.585 | 1.233 | -2.446 | -2.528 | -0.018 | Yes | No | 0.165 |
|        | Rainfall | 71.95  | 30.93 | -1.096 | -1.18 | -0.24 | No | No | 0.035 |
| Winter | $T_{\text{max}}$ | 14.748 | 2.697 | 3.122 | 3.420 | 0.053 | Yes | 1991 | 0.123 |
|        | $T_{\text{min}}$ | -16.003 | 2.660 | 5.813 | 7.010 | 0.104 | Yes | 1981 | 0.501 |
|        | $T_{\text{av}}$ | -3.695 | 1.575 | 7.482 | 11.521 | 0.074 | Yes | 1984 | 0.686 |
|        | Diff-T | 12.040 | 1.540 | -6.198 | -8.126 | -0.057 | Yes | 1980 | 0.548 |
|        | Rainfall | 5.23  | 4.54  | 0.69 | 0.61 | 0.02 | No | No | -0.08 |

*** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$
trend. Finally, from the extent of the variation in meteorological variables, the four temperature variables all show the largest slope in winter. In detail, the average maximum temperature increasing by 0.04 °C per year, the average minimum temperature increasing by 0.104 °C per year, the average temperature increasing by 0.074 °C per year, and the temperature difference decreasing by 0.057 °C per year. Similarly, the variation degree of temperature variables in other seasons can be seen in the Sen’s slope in Table 3. The results of the Pettitt’s change-point test show that the temperature changes in different seasons are different.

Plot the seasonal variation trend of temperature variables, as shown in Fig. 1. Horizontally, the trend of the lowest temperature in each season is the most obvious. Longitudinally, each variable has a greater variation trend in winter. In a comprehensive way, the main reason for the climate warming in Lanzhou lies in the rise of the minimum temperature. Meanwhile, the variation trend of all temperature variables in winter indicates that the temperature rising in Lanzhou is mainly reflected in the “warm winter” phenomenon.

### 4.2 Extreme high temperature prediction using weighted Markov models

In the last section, we have found and analyzed the significant trend of temperature rise. In this section, we build a weighted Markov model to predict extreme high temperature weather in Lanzhou.
In the context of climate warming, the increasing frequency of disasters and the increasing losses caused by meteorological disasters have seriously affected agricultural production, social economy and regional sustainable development.

The extreme high temperature events were defined by Bonsal et al. (2001). Extreme high weather is considered when a day’s temperature is higher than the 90th percentile value of the daily maximum temperature series of 66 years was selected, and the average value of 66 years was defined as the threshold value.

After calculation, the threshold $T^* = 30.12 \, ^\circ\text{C}$. Next, calculate the days whose temperature is greater than or equal to $T^*$, and divide these days into 6 categories according to the number (one, two, three, four, five, six and above) of consecutive high temperature days, and then we construct a set of annual high temperature index, denoted by $h_{d1}, h_{d2}, h_{d3}, h_{d4}, h_{d5}, h_{d6}$, where $h_{di}$ represents the high temperature index of the consecutive $d$-days in the $i$th year ($i = 1, 2, \ldots, 66, d = 1, 2, \ldots, 6$). The high temperature index of the six categories can be calculated as follows: If the frequency of the consecutive $d$-days of high temperature in the $i$th year is $N$, then the $d$-day high temperature index in this year can be counted as $h_{di} = Nd, d = 1, 2, \ldots, 6$. For example, the frequency of the consecutive four days of high temperature in 1951 (the first year) is 4, then the 4-day index in this year is $h_{d4} = 4 \times 4 = 16$. The frequency of the consecutive 7 days of high temperature is 2, and the frequency of the consecutive 8 days is 1 in 1951, then we can calculate the 6-day high temperature index by $h_{d6} = 1 \times 7 + 1 \times 8 = 15$. By the same token, we obtain the high temperature index of all categories of Lanzhou city from 1951 to 2016 (As shown in Table 6).

![Fig. 1 The trend of temperature variables in each season](image)

In the context of climate warming, the increasing frequency of disasters and the increasing losses caused by meteorological disasters have seriously affected agricultural production, social economy and regional sustainable development.

The extreme high temperature events were defined by Bonsal et al. (2001). Extreme high weather is considered when a day’s temperature is higher than the 90th percentile value of the daily maximum temperature series of 66 years was selected, and the average value of 66 years was defined as the threshold value.

After calculation, the threshold $T^* = 30.12 \, ^\circ\text{C}$. Next, calculate the days whose temperature is greater than or equal to $T^*$, and divide these days into 6 categories according to the number (one, two, three, four, five, six and above) of consecutive high temperature days, and then we construct a set of annual high temperature index, denoted by $h_{d1}, h_{d2}, h_{d3}, h_{d4}, h_{d5}, h_{d6}$, where $h_{di}$ represents the high temperature index of the consecutive $d$-days in the $i$th year ($i = 1, 2, \ldots, 66, d = 1, 2, \ldots, 6$). The high temperature index of the six categories can be calculated as follows: If the frequency of the consecutive $d$-days of high temperature in the $i$th year is $N$, then the $d$-day high temperature index in this year can be counted as $h_{di} = Nd, d = 1, 2, \ldots, 6$. For example, the frequency of the consecutive four days of high temperature in 1951 (the first year) is 4, then the 4-day index in this year is $h_{d4} = 4 \times 4 = 16$. The frequency of the consecutive 7 days of high temperature is 2, and the frequency of the consecutive 8 days is 1 in 1951, then we can calculate the 6-day high temperature index by $h_{d6} = 1 \times 7 + 1 \times 8 = 15$. By the same token, we obtain the high temperature index of all categories of Lanzhou city from 1951 to 2016 (As shown in Table 6).

Obviously, the longer duration of high temperature lasts, the higher risk it will bring. Reference to Gong et al. (2013), in this paper, the different weights of the $d$-day high temperature index $h_{di}, d = 1, 2, 3, 4, 5, 6$, are endowed to $\omega_{di}$, where $\omega_1 = 1.0, \omega_2 = 1.05, \omega_3 = 1.1, \omega_4 = 1.15, \omega_5 = 1.2, \omega_6 = 1.25$; Thus, we obtain the set of
the high temperature weighted index, denoted by \( \{ f_{1}, f_{2}, f_{3}, f_{4}, f_{5}, f_{6} \} \), where \( f_{di} = \omega_{d} h_{di} \) indicates the high-temperature weighted index of consecutive d-days in the i\textsuperscript{th} year. Finally, we can get an annual high temperature weighted index of the i\textsuperscript{th} year through arithmetic weighted average on all categories, denoted by \( F_{i} = \sum_{d=1}^{6} \omega_{d} h_{di} \). For example, the annual high temperature weighted index of second year (1952) can be obtained by \( F_{2} = 7 \times 1.0 + 2 \times 1.05 + 12 \times 1.1 + 0 \times 1.15 + 5 \times 1.2 + 17 \times 1.25 = 49.55 \), see Table 6.

Establish classification standard, divide states: Let \( \tilde{F} = \frac{1}{66} \sum_{i=1}^{66} F_{i} \), and \( S = \sqrt{\frac{1}{66-1} \sum_{i=1}^{66} (F_{i} - \tilde{F})^{2}} \) represent the mean and the standard deviation of \( \{F_{1}, F_{2}, \ldots, F_{66}\} \), respectively. The annual high-temperature weighted indexes \( \{F_{1}, F_{2}, \ldots, F_{66}\} \) are divided into five intervals as follows: \( E_{1} = (-\infty, \tilde{F} - \alpha_{1} S] \), \( E_{2} = (\tilde{F} - \alpha_{2} S, \tilde{F} - \alpha_{3} S] \), \( E_{3} = (\tilde{F} - \alpha_{2} S, \tilde{F} + \alpha_{3} S] \), \( E_{4} = (\tilde{F} + \alpha_{2} S, \tilde{F} + \alpha_{4} S] \), \( E_{5} = (\tilde{F} + \alpha_{4} S, +\infty) \), where \( \tilde{F} = 42.376, S = 14.8 \), and we set \( \alpha_{1} = 1.0, \alpha_{2} = 0.5 \). Each interval represents a state (interval). Thus, \( \{F_{1}, F_{2}, \ldots, F_{66}\} \) actually belong to a state set \( E = \{E_{1}, E_{2}, \ldots, E_{5}\} \), see Table 7. The resulting states are shown in the last column in the Table 6.

The autocorrelation coefficient of the sequence is calculated by using the following formula (6). According to the magnitude of the correlation coefficient, the correlation coefficient of lag order 4 is selected, and the weight vector is calculated according to the formula (7), where \( l = 4 \). The correlation coefficients and weight vector are shown in Table 8.

\[
r_{w} = \frac{\sum_{i=1}^{n-w} (x_{i} - \bar{x})(x_{i+w} - \bar{x})}{\sum_{i=1}^{n} (x_{i} - \bar{x})^{2}},
\]

\[
\theta_{w} = \frac{|r_{w}|}{\sum_{w=1}^{l} r_{w}}, \omega = 1, 2, \ldots, l.
\]

The state transition probability matrix can be calculated by using the index value of extreme high temperature and the partitioned state. \( \omega \) steps state transition probability matrix is

\[
P^{(\omega)} = (p_{ij}^{(\omega)})_{5 \times 5}, p_{ij}^{(\omega)} = \frac{m_{ij}^{(\omega)}}{M_{i}}, \omega = 1, 2, 3, 4.
\]

where \( m_{ij}^{(\omega)} \) is the number of state \( i \) to state \( j \) after \( \omega \) steps’ transition; \( M_{i} \) is the number of state \( i \).

We used 2012–2015 as the initial year to predict extreme high temperatures in 2016 using a weighted Markov model. Combined with the state transition probability and the weight coefficient, the probability of the occurrence of extreme high temperature in each state in 2016 is predicted. The results are shown in Table 9. The bold figure in Table 9 means the maximum of the state transition probability based on the annual high-temperature weighted index of 2012–2015. We find that when \( \max\{P_{i}\} = 0.4570 \), we have \( i = 5 \), which means the risk of the high temperature based on the annual high-temperature weighted index of 2016 belongs to state 5. Finally, the fuzzy mathematical method mentioned in was used to estimate the extreme high temperature index. After calculation, it can be concluded that the extreme high temperature index value in 2016 is 67.01, which is still a certain gap from the real value of 75.06, indicating that this method still has space for improvement.
Table 6 The high temperature index and states

| Years | 1-day | 2-day | 3-day | 4-day | 5-day | 6-day | Index  | States |
|-------|-------|-------|-------|-------|-------|-------|--------|--------|
| 1951  | 3     | 18    | 3     | 4     | 0     | 15    | 48.55  | 3      |
| 1952  | 7     | 2     | 12    | 0     | 5     | 17    | 49.55  | 3      |
| 1953  | 9     | 8     | 3     | 4     | 10    | 21    | 63.55  | 5      |
| 1954  | 6     | 12    | 9     | 8     | 0     | 7     | 46.45  | 3      |
| 1955  | 7     | 12    | 3     | 8     | 0     | 13    | 48.35  | 3      |
| 1956  | 3     | 8     | 9     | 4     | 0     | 0     | 25.90  | 1      |
| 1957  | 12    | 2     | 6     | 8     | 0     | 6     | 37.40  | 3      |
| 1958  | 8     | 10    | 0     | 0     | 0     | 8     | 28.50  | 2      |
| 1959  | 2     | 8     | 0     | 0     | 5     | 6     | 23.90  | 1      |
| 1960  | 6     | 4     | 15    | 0     | 10    | 0     | 38.70  | 3      |
| 1961  | 6     | 6     | 3     | 4     | 10    | 12    | 47.20  | 3      |
| 1962  | 5     | 6     | 3     | 4     | 0     | 16    | 39.20  | 3      |
| 1963  | 9     | 14    | 12    | 0     | 0     | 8     | 46.90  | 3      |
| 1964  | 2     | 6     | 3     | 0     | 0     | 6     | 19.10  | 1      |
| 1965  | 5     | 6     | 6     | 0     | 5     | 6     | 31.40  | 2      |
| 1966  | 6     | 8     | 9     | 4     | 0     | 9     | 40.15  | 3      |
| 1967  | 2     | 2     | 3     | 4     | 0     | 8     | 22.00  | 1      |
| 1968  | 3     | 6     | 6     | 0     | 0     | 0     | 15.90  | 1      |
| 1969  | 6     | 14    | 9     | 8     | 0     | 0     | 39.80  | 3      |
| 1970  | 6     | 4     | 0     | 8     | 0     | 6     | 26.90  | 1      |
| 1971  | 6     | 0     | 3     | 4     | 10    | 12    | 40.90  | 3      |
| 1972  | 7     | 4     | 3     | 4     | 10    | 11    | 44.85  | 3      |
| 1973  | 4     | 0     | 12    | 0     | 0     | 11    | 30.95  | 2      |
| 1974  | 8     | 2     | 9     | 4     | 5     | 7     | 39.35  | 3      |
| 1975  | 4     | 6     | 3     | 4     | 5     | 11    | 37.95  | 3      |
| 1976  | 3     | 0     | 3     | 4     | 0     | 0     | 10.90  | 1      |
| 1977  | 3     | 4     | 3     | 0     | 0     | 19    | 34.25  | 2      |
| 1978  | 5     | 4     | 3     | 8     | 5     | 6     | 35.20  | 3      |
| 1979  | 5     | 2     | 0     | 0     | 5     | 0     | 13.10  | 1      |
| 1980  | 7     | 12    | 3     | 8     | 0     | 0     | 32.10  | 2      |
| 1981  | 3     | 8     | 3     | 4     | 0     | 21    | 45.55  | 3      |
| 1982  | 6     | 6     | 9     | 0     | 5     | 14    | 45.70  | 3      |
| 1983  | 5     | 6     | 9     | 4     | 0     | 6     | 33.30  | 2      |
| 1984  | 7     | 6     | 3     | 4     | 0     | 7     | 29.95  | 2      |
| 1985  | 4     | 6     | 0     | 8     | 5     | 8     | 35.50  | 3      |
| 1986  | 4     | 6     | 0     | 0     | 0     | 16    | 30.30  | 2      |
| 1987  | 4     | 8     | 3     | 4     | 0     | 18    | 42.80  | 3      |
| 1988  | 2     | 6     | 9     | 4     | 0     | 15    | 41.55  | 3      |
| 1989  | 13    | 6     | 0     | 0     | 0     | 9     | 30.55  | 2      |
| 1990  | 6     | 10    | 3     | 4     | 5     | 13    | 46.65  | 3      |
| 1991  | 5     | 8     | 0     | 4     | 5     | 31    | 62.75  | 5      |
| 1992  | 7     | 6     | 6     | 0     | 0     | 0     | 19.90  | 1      |
| 1993  | 8     | 2     | 6     | 0     | 5     | 0     | 22.70  | 1      |
| 1994  | 5     | 10    | 6     | 0     | 5     | 12    | 43.10  | 3      |
| 1995  | 6     | 8     | 6     | 4     | 0     | 9     | 36.85  | 3      |
| 1996  | 4     | 4     | 3     | 4     | 5     | 13    | 38.35  | 3      |
5 Result

The annual and quarterly variation trends of meteorological data variables are analyzed in detail. The results show that the annual trend changes of temperature variables are obvious. In terms of seasonal trend, the maximum temperature in summer has no obvious trend, while the seasonal trend of other variables is obvious. Generally speaking, the temperature in Lanzhou has a significant trend of rise in the past 66 years, and it has
a larger range of rise in winter. The annual trend and quarterly trend of rainfall are not obvious, but the total amount of rainfall has decreased. In detail, rainfall increases in winter and spring, and decreases in summer and fall. In addition, we use the weighted Markov model to explore the possibility of extreme high temperature, which provides a certain theoretical basis for high temperature warning and prevention.

Acknowledgements Thanks to Lanzhou Central Meteorological Observatory of Gansu Province for providing the original data for the research in this paper.

Declarations

Conflict of interests As far as we know, there are no competing economic interests or personal relationships that may affect the work described in this paper. We like to declare that the work was original research that has not been published previously, and not under consideration for publication elsewhere, in whole or in part. The data in this paper are the original observation data, and all the relevant codes are written by the author. If there is any need, you can contact the author via email.

References

Bonsal BR, Zhang X, Laea V (2001) Characteristics of daily and extreme temperatures over Canada. J Climate 14(9):1959–1976

Chen S, Kuo C, Yu P (2009) Historical trends and variability of meteorological droughts in Taiwan/ tendances historiques et variabilité des sécheresses météorologiques à taiwan. Hydrol Sci J 54(3):430–441

Gocic M, Trajkovic S (2013) Analysis of changes in meteorological variables using Mann–Kendall and Sen’s slope estimator statistical tests in Serbia. Global Planet Change 100:172–182

Gong Z, Chen C, Ge X (2013) Risk prediction of low temperature in Nanjing city based on grey weighted Markov model. Nat Hazards 71(2):1159–1180

He Y, Zhang Y (2005) Climate change from 1960 to 2000 in the Lancang river valley, China. Mt Res Dev 25(4):341–348

Kendall MG (1975) Rank correlation measures. Charles Griffin, London

Koutsoyiannis D (2003) Climate change, the Hurst phenomenon, and hydrological statistics. Hydrol Sci J 48(1):3–24

Liu W, Zhang L et al (2016) Analysis of hot weather and high temperature days in the summer of Gansu province from 1981 to 2010. J Lanzhou University (Nat Sci) 52(4):475–478

Mann HB (1945) Nonparametric tests against trend. Econometrica 13(3):245–259

Masseran N (2015) Markov chain model for the stochastic behaviors of wind-direction data. Energy Convers Manage 92:266–274
Meena DM (2020) Rainfall statistical trend and variability detection using Mann–Kendall test, Sen’s slope and coefficient of variance: a case study of Udaipur district (1957–2016). Appl Ecol Environ Sci 8(1):34–37
Partal T (2017) Multi-annual analysis and trends of the temperatures and precipitations in west Anatolia. J Water Climate Change 8(3):456–473
Pettitt AN (1979) A non-parametric approach to the change-point problem. J R Stat Soc Ser C (Appl Stat) 28(2):126–135
Salas J, Delleur J, Yevjevich V (1985) Applied modeling of hydrologic time series, Water Resources Publications
Sayemuzzaman M, Jha MK (2014) Seasonal and annual precipitation time series trend analysis in North Carolina, USA. Atmos Res 137:183–194
Sen PK (1968) Estimates of the regression coefficient based on Kendall’s tau. J Am Stat Assoc 63(324):1379–1389
Sharma CS, Panda SN, Pradhan RP, Singh A, Kawamura A (2016) Precipitation and temperature changes in eastern India by multiple trend detection methods. Atmos Res 180:211–225
Shi Y, Zhao J (2014) Study on spatial-temporal characteristics of extreme temperature in arid areas of north-west China. J Lanzhou Univ (Nat Sci) 50(4):529–533
Sonali P, Kumar DN (2013) Review of trend detection methods and their application to detect temperature changes in India. J Hydrol 476:212–227
Tian Y, Shen H, Zhang L, Lv X (2010) Utility water supply forecast via a gm(1,1) weighted Markov chain. J Zhejiang Univ Sci A 11(9):677–682
von Storch H (1995) Misuses of statistical analysis in climate research. In: von Storch H, Navarra A (eds) Analysis of climate variability. Springer, Heidelberg
Yacoub E, Tayfur G (2018) Trend analysis of temperature and precipitation in Trarza region of Mauritania. J Water Climate Change 10(3):484–493
Zhai P, Pan X (2003) Change in extreme temperature and precipitation over northern china during the second half of the 20 century. Acta Geographica Sinica 58(S1):1–10

Publisher’s Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.