Temperature Trend Analysis and Extreme High Temperature Prediction Based on Weighted Markov Model in Lanzhou

Zhiqiang Pang
Lanzhou University of Finance and Economics

Zhaoxu Wang (✉ wangzhaoxu1989@126.com)
Lanzhou University of Finance and Economics   https://orcid.org/0000-0001-9329-9846

Manuscript

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

Zhiqiang Pang · Zhaoxu Wang

Abstract In this study, temporIn 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 analyze the time trend. 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 have a significant rising trend and show different performances in each season. In detail, the maximum temperature in summer does not have a significant change trend, while the minimum temperature in winter is the most significant rising trend, which leads to more and more "warm winter" phenomenon. Finally, we construct a weighted Markov prediction model for extremely high temperatures and obtain the conclusion that the prediction results by the model are consistent with the real situation. and show different performances in each season. In detail, the maximum temperature in summer does not have a significant change trend, while the minimum temperature in winter is the most significant rising trend, which leads to more and more "warm winter" phenomenon. Finally, we construct a weighted Markov prediction model for extremely high temperatures and obtain the conclusion that the prediction results by the model are consistent with the real situation.

Z. Pang · Z. Wang
School of Statistics Lanzhou University of Finance and Economics,Lanzhou,Gansu,China
E-mail: pangzhq@lzufe.edu.cn

Z. Wang
School of Mathematics and Statistics of Qinghai Normal University,Xining,Qinghai,China
E-mail: wangzx@qhnu.edu.cn

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); 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 nonparametric 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. 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); 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); 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 analyze 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 H. (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 city, and used weighted Markov to predict the occurrence of extremely high temperature. The application of the model was tested and the conclusion shows that the prediction results were robust.

The main purposes of this paper is as follows: 1. Analyze 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 Section 2, the data source used in this paper and the form of the data are described; In section 3, the common non-parametric analysis methods using trend analysis and the Markov model used for prediction are introduced. In Section 4, we use the methods and models in Section 3 to analyze 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 13085.6 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 analyze 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 311.2mm and the annual average temperature is $9.87°C$. August is the wettest month with an average rainfall of 72.2mm, while December is the least wet month with an average rainfall of only 1mm. 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 Eqs. (1) :

$$S = \sum_{i=2}^{n} \sum_{j=1}^{i-1} sign\left(x_i - x_j\right), \qquad (1)$$

$$\mathrm{Var}(S) = \frac{n(n-1)(2n+5) - \sum_{p-1}^{q} t_p\left(t_p - 1\right)\left(2t_p + 5\right)}{18}. \qquad (2)$$

where x_i and x_j are the data value in time point i and j, *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)

$$Z = \begin{cases} \frac{S-1}{\sqrt{\text{Var}(S)}} & \text{if } S > 0; \\ 0 & \text{if } S = 0; \\ \frac{S+1}{\sqrt{\text{Var}(S)}} & \text{if } S < 0. \end{cases} \tag{3}$$

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 $Z_{1-\alpha/2}$.

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

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

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,

$$Q_{med} = \begin{cases} Q_{[(N+1)/2]}, & N \text{ is odd} \\ \frac{1}{2} \left(Q_{[N/2]} + Q_{[(N+2)/2]} \right), & N \text{ is even} \end{cases}$$

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 nonparametric 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 Equations (4) and (5) as follows.

$$D = 1 - \frac{6 \sum_{i=1}^{n} \left(R(X_i) - i \right)^2}{n \left(n^2 - 1 \right)}, \tag{4}$$

$$Z_{SR} = D\sqrt{\frac{n-2}{1-D^2}}. \qquad (5)$$

where $R(X_i)$ is the rank of ith 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

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

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 $(x_2 - r_1 x_1, x_3 - r_1 x_2, \ldots, x_n - r_1 x_{n-1})$ (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 Igg). 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:

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

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

3.2 Weighted Markov chain modeling

The Fluctuation Index is introduced in Markov chain modeling process :

$$v_k = \frac{x^{(0)}(k) - \hat{x}^{(0)}(k)}{\hat{x}^{(0)}(k)} \cdot 100\%, (k = 1, 2, \cdots, n - m).$$

Note: taking $\hat{x}^{(0)}(k)$ as the trend value of the k_{th} item, $x^{(0)}(k) - \hat{x}^{(0)}(k)$ 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 $V = (v_1, v_2, \cdots, v_{n-m})$ can be divided into s states. If there is $v_k \in [a_{1i}, a_{2i}], i = 1, 2, \cdots, s$, the fluctuation state of the kth 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 $E = (E_1, E_2, \cdots, E_s)$ 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 $p_{ij}^{(w)}$:

$$p_{ij}^{(w)} = \frac{m_{ij}^{(w)}}{M_i}.$$

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:

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

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 $(i_1, i_2, \cdots, i_s)$ of s items respectively, the corresponding transition probability after w steps $p_i^{(w)} = \left(p_{i1}^{(w)}, p_{i2}^{(w)}, \cdots, p_{is}^{(w)} \right), i = 1, 2, \cdots, s$ should be adopted from the transition probability matrix $P^{(w)}$. Therefore, state transition probability matrix of the tth forecasting data can be obtained:

$$R = \begin{bmatrix} p_{i_1 1}^{(1)} & p_{i_1 2}^{(1)} & \cdots & p_{i_1 s}^{(1)} \\ p_{i_2 1}^{(2)} & p_{i_2 2}^{(2)} & \cdots & p_{i_2 s}^{(2)} \\ \vdots & \vdots & \vdots & \vdots \\ p_{i_s 1}^{(j)} & p_{i_s 2}^{(j)} & \cdots & p_{i_s s}^{(j)} \end{bmatrix}, i = 1, 2, \cdots, s$$

Determine the state of the tth forecasting data by choosing the corresponding state $*$ of $\max \left\{ p_i = \sum_{w=1}^{s} p_{\bullet i}^w, i = 1, 2, \cdots, s \right\}$ and this state's upper bound and lower bound are a_{1*}, a_{2*} respectively.

Finally, the modified forecasting data can be calculated:

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

For the prediction of the $(t+1)$th item, the state of the tth 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 section 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

Nonparametric trend test was carried out on the annual highest temperature(T_{max}), lowest temperature(T_{min}), average temperature(T_{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°C , 9.9°C,

Table 1 Trend analysis for annual data

	mean	sd	MK-Z	Sr(Z_{SR})	Sen's slope	Trend	change point	Lag-1
T_{max}	35.574	1.542	2.326**	2.472**	0.023	Yes	1996	0.303
T_{min}	-16.003	2.660	5.813***	7.010***	0.104	Yes	1981	0.501
T_{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

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 311.3mm and sd is 75.84mm, 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 $(x_2 - r_1x_1, x_3 - r_1x_2, \ldots, x_n - r_1x_{n-1})$, which is removed the correlation. After the correlation is removed, the new sequence is analyzed again. The test results are shown in Table 2

Table 2 Trend analysis for annual data(Removed correlation)

	MK-Z	Sr(Z_{SR})	Trend	Lag-1
T_{max}	1.5796	1.681	No	-0.025
T_{min}	3.0288***	3.000***	Yes	-0.171
T_{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

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.

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.

Table 3 Trend analysis for season data

season	variable	mean	sd	MK-Z	Sr(Z_{SR})	Sen's slope	Trend	change point	Lag-1
Spring	T_{max}	31.406	1.517	2.227**	2.180**	0.025	Yes	1993	0.226
	T_{min}	-6.474	2.828	2.397**	2.395**	0.042	Yes	No	0.114
	T_{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_{max}	35.537	1.544	2.309**	2.458**	0.023	Yes	1996	0.301
	T_{min}	9.144	1.897	4.878**	5.619**	0.060	Yes	1997	0.427
	T_{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_{max}	29.061	2.182	2.807***	3.136***	0.04	Yes	1986	0.250
	T_{min}	-7.902	2.705	4.098***	4.517***	0.066	Yes	1981	0.105
	T_{av}	9.592	1.014	6.375***	8.854***	0.039	Yes	1986	0.601
	diff-T	11.585	1.233	-2.4461**	-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_{max}	14.748	2.697	3.122***	3.420***	0.053	Yes	1991	0.123
	T_{min}	-16.003	2.660	5.813***	7.010***	0.104	Yes	1981	0.501
	T_{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

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 Table 3, Table 4 and Table 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.

Table 4 Trend analysis for season data(Removed correlation)

season		MK-Z	Sr(Z_{SR})	Trend	Lag-1
Spring	T_{max}	2.151**	2.161**	Yes	-0.03
	T_{av}	2.825***	2.952***	Yes	-0.039
	diff-T	-1.795*	-1.924*	Yes	-0.075
Summer	T_{max}	1.580	1.680	No	-0.025
	T_{min}	3.046***	3.317***	Yes	-0.027
	T_{av}	2.542**	2.625**	Yes	0.003
Autumn	T_{max}	2.032**	2.218**	Yes	-0.013
	T_{av}	2.848***	3.174***	Yes	-0.141
Winter	T_{min}	3.029***	3.000***	Yes	-0.171
	T_{av}	3.278***	3.204***	Yes	-0.314
	diff-T	-2.667***	-2.666***	Yes	-0.236

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

Table 5 Trend analysis for winter data(Removed correlation)

season		MK-Z	Sr(Z_{SR})	Trend	Lag-1
Winter	T_{av}	4.235***	4.407***	Yes	0.056
	diff-T	-3.146***	-3.224***	Yes	0.049

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

Finally, from the extent of the variation in meteorological variables, the four temperature variables all show the largest slop 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 Figure 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.
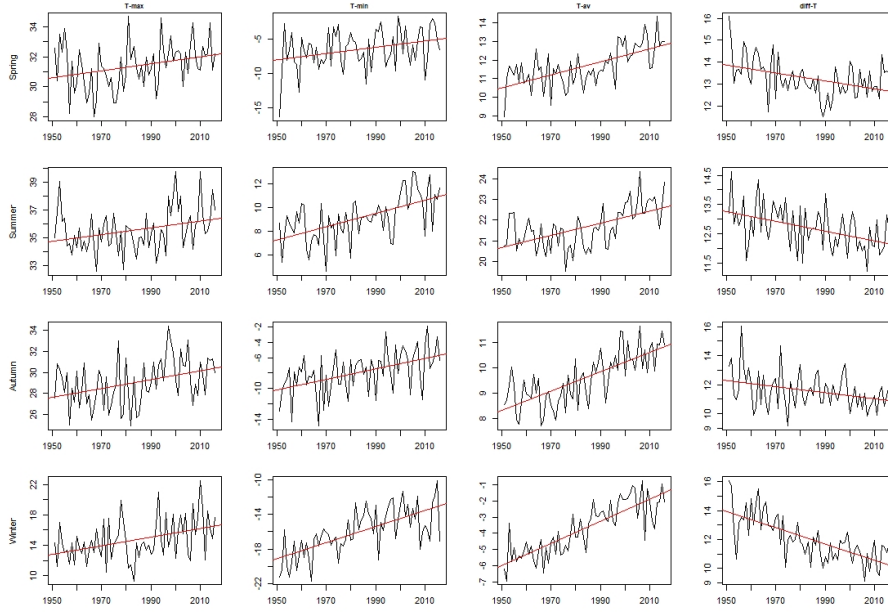
Fig. 1 Seasonal temperature trend

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 90 percent of the annual temperature. In our article, 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 $T^* = 30.12°C$. 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 $h_{1i}, h_{2i}, h_{3i}, h_{4i}, h_{5i}, h_{6i}$, where h_{di} represents the high temperature index of the consecutive d-days in the ith year($i = 1, 2, \cdots, 66, d = 1, 2, \cdots, 6$). 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 ith year is N, then the d-day high temperature index in this year can be counted as $h_{di} = Nd, d = 1, 2, \cdots, 6$. 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 $h_{41} = 4 \times 4 = 16$. 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 $h_{61} = 1 \times 7 + 1 \times 8 = 15$. 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 ω_d, 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 $\{f_{1i}, f_{2i}, f_{3i}, f_{4i}, f_{5i}, f_{6i}\}$, where $f_{di} = \omega_d h_{di}$ indicates the high-temperature weighted index of consecutive d-days in the ith year. Finally, we can get an annual high temperature weighted index of the ith 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 $\bar{F} = \frac{1}{66} \sum_{i=1}^{61} F_i$, and $S = \sqrt{\frac{1}{61-1} \sum_{i=1}^{61} (F_i - \bar{F})^2}$ represent the mean and the standard deviation of $\{F_1, F_2, \ldots, F_{66}\}$ respectively. The annual high-temperature weighted indexes $\{F_1, F_2, \ldots, F_{66}\}$ is divided into five intervals as follows: $E_1 = (-\infty, \bar{F} - \alpha_1 S], E_2 = (\bar{F} - \alpha_1 S, \bar{F} - \alpha_2 S]$, $E_3 = (\bar{F} - \alpha_2 S, \bar{F} + \alpha_2 S], E_4 = (\bar{F} + \alpha_2 S, \bar{F} + \alpha_1 S], E_5 = (\bar{F} + \alpha_1 S, +\infty)$, where $\bar{F} = 42.376, S = 14.8$, and we set $\alpha_1 = 1.0, \alpha_2 = 0.5$. Each interval represents a state (interval). Thus, $\{F_1, F_2, \cdots, F_{66}\}$ actually belong to a state set $E = \{E_1, E_2, \cdots, E_5\}$, 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,

$$[h] r_w = \frac{\sum_{i=1}^{n-w} (x_i - \bar{x})(x_{i+w} - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}, \tag{6}$$

$$\theta_w = \frac{|r_w|}{\sum_{w=1}^{l} r_w}, \omega = 1, 2, \cdots, l. \tag{7}$$

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

$$P^{(\omega)} = (p_{ij}^{\omega})_{5 \times 5}, p_{ij}^{\omega} = \frac{m_{ij}^{\omega}}{M_i}, \quad \omega = 1, 2, 3, 4.$$

where m_{ij}^{ω} is the number of state i to state j after ω 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

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
1997	2	6	6	16	0	27	67.05	5
1998	7	6	9	4	5	23	62.55	5
1999	6	8	3	8	0	11	40.65	3
2000	6	10	6	0	10	16	55.10	4
2001	6	2	6	4	0	38	66.80	5
2002	4	6	3	8	0	35	66.55	5
2003	9	8	3	4	0	7	34.05	2
2004	8	4	12	4	0	19	53.75	4
2005	5	8	18	8	0	13	58.65	5
2006	2	4	9	0	5	45	78.35	5
2007	8	6	18	4	0	8	48.70	3
2008	8	8	9	8	0	14	53.00	4
2009	5	6	15	8	5	6	50.50	4
2010	12	8	0	4	15	9	54.25	4
2011	10	8	0	0	0	36	63.40	5
2012	9	10	9	4	5	8	50.00	4
2013	8	16	6	4	5	15	60.75	5
2014	4	4	3	4	10	10	40.60	3
2015	9	8	9	0	5	15	52.05	4
2016	6	4	3	4	0	46	75.60	5

Table 7 State classification of all the annual high-temperature weighted indexes

State	Classification standard	Interval of annual
1	E_1	$(-\infty, 27.576]$
2	E_2	$(27.576, 34.976]$
3	E_3	$(34.976, 49.776]$
4	E_4	$(49.776, 57.176]$
5	E_5	$(57.176, +\infty)$

Table 8 Self-correlation coefficients and weights of 1st-4th steps

Step	w=1	w=2	w=3	w=4
r_w	0.406	0.239	0.407	0.337
θ_w	0.292	0.172	0.293	0.243

extreme high temperature in each state in 2016 is predicted. The results are shown in Table 9.

Table 9 Risk state of the high-temperature simulation in 2015 based on the high-temperature weighted index

Initial	State	Delay	Weight	State 1	State 2	State 3	State 4	State 5	Sources
2015	5	1	0.292	0.100	0.100	0.400	0.100	0.300	$P^{(1)}$
2014	4	2	0.239	0	0	0	0.333	0.667	$P^{(2)}$
2013	3	3	0.407	0.074	0.148	0.593	0.037	0.148	$P^{(3)}$
2012	5	4	0.337	0	0	0.222	0.333	0.445	$P^{(4)}$
P_i				0.0593	0.0895	0.4329	0.2363	0.4570	

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.

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.

Declare

We have no known competing financial interests or personal relationships that could have appeared to influence the work reported 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

B. R. Bonsal, X. Zhang, and L. A. et al Vincent. Characteristics of daily and extreme temperatures over canada. Journal of Climate, 14(9):1959–1976, 2001.

ShineTsung Chen, ChunChao Kuo, and PaoShan Yu. Historical trends and variability of meteorological droughts in taiwan / tendances historiques et variabilité des sécheresses météorologiques à taiwan. Hydrological Sciences Journal, 54(3):430–441, 2009.

Milan Gocic and Slavisa Trajkovic. Analysis of changes in meteorological variables using mann-kendall and sen's slope estimator statistical tests in serbia. Global and Planetary Change, 100:172 – 182, 2013. ISSN 0921-8181.

Zaiwu Gong, Caiqin Chen, and Xinming Ge. Risk prediction of low temperature in nanjing city based on grey weighted markov model. Natural Hazards, 71(2):1159–1180, 2013. ISSN 0921-030X-1573-0840.

Yunling He and Yiping Zhang. Climate change from 1960 to 2000 in the lancang river valley, china. Mountain Research and Development, 25(4): 341–348, 2005.

WenXiong Jia. Seasonal characteristics of extreme temperature changes in qilian mountains and hexi corridor during last fifty years. SCIENTIA GEOGRAPHICA SINICA, 32(11):1377–1383, 2012.

M. G. Kendall. Rank Correlation Measures. Charles Griffin, London, 1975.

Demetris Koutsoyiannis. Climate change, the hurst phenomenon, and hydrological statistics. Hydrological Sciences Journal, 48(1):3–24, 2003.

Weiping Liu, Lei Zhang, and Pengfei Tian et al. Analysis of hot weather and high temperature days in the summer of gansu province from 1981 to 2010. Journal of Lanzhou University (Natural Sciences), 52(4):475–478, 2016.

Henry B. Mann. Nonparametric tests against trend. Econometrica, 13(3): 245–259, 1945. ISSN 00129682, 14680262.

Nurulkamal Masseran. Markov chain model for the stochastic behaviors of wind-direction data. Energy Conversion and Management, 92:266–274, 2015. ISSN 01968904.

Dr. Manish Meena. 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). Applied Ecology and Environmental Sciences, 8(1):34–37, 2020. ISSN 2328-3920.

Turgay Partal. Multi-annual analysis and trends of the temperatures and precipitations in west anatolia. Journal of Water and Climate Change, 8(3): 456–473, 2017. ISSN 2040-2244-2408-9354.

A. N. Pettitt. A non-parametric approach to the change-point problem. Journal of the Royal Statistical Society. Series C (Applied Statistics), 28(2): 126–135, 1979. ISSN 00359254, 14679876.

J.D. Salas, J.W. Delleur, and V. Yevjevich. Applied Modeling of Hydrologic Time Series. Water Resources Publications, 1985.

Mohammad Sayemuzzaman and Manoj K. Jha. Seasonal and annual precipitation time series trend analysis in north carolina, united states. Atmospheric Research, 137:183–194, 2014. ISSN 01698095.

Pranab Kumar Sen. Estimates of the regression coefficient based on kendall's tau. Journal of the American Statistical Association, 63(324):1379–1389, 1968.

Chandra Shekhar Sharma, Sudhindra N. Panda, Rudra P. Pradhan, Amanpreet Singh, and Akira Kawamura. Precipitation and temperature changes in eastern india by multiple trend detection methods. Atmospheric Research, 180:211 – 225, 2016. ISSN 0169-8095.

YF Shi and Jun Zhao. Study on spatial-temporal characteristics of extreme temperature in arid areas of northwest china. Journal of Lanzhou University (Natural Sciences), 50(4):529–533, 2014.

P. Sonali and D. Nagesh Kumar. Review of trend detection methods and their application to detect temperature changes in india. Journal of Hydrology, 476:212–227, 2013.

H. J. D. Sørup, H. Madsen, and K. Arnbjerg-Nielsen. Descriptive and predictive evaluation of high resolution markov chain precipitation models. Environmetrics, 23(7):623–635, 2012. ISSN 11804009.

H. Theil. A Rank-invariant Method of Linear and Polynomial Regression Analysis. Stichting Mathematisch Centrum. Statistische Afdeling, 1950.

Yimei Tian, Hai-liang Shen, Li Zhang, and Xiang-rui Lv. Utility water supply forecast via a gm (1,1) weighted markov chain. Journal of Zhejiang University-SCIENCE A, 11(9):677–682, 2010. ISSN 1862-1775.

von Storch H. Misuses of Statistical Analysis in Climate Research. In: von
    Storch H., Navarra A. (eds) Analysis of Climate Variability. Springer, Berlin,
    Heidelberg., 1995.

Ely Yacoub and Gokmen Tayfur. Trend analysis of temperature and precipita-
    tion in trarza region of mauritania. Journal of Water and Climate Change,
    10(3):484–493, 2018.

Panmao Zhai and Xiaohua Pan. Change in extreme temperature and precipi-
    tation over northern china during the sencond half of the 20 century. Acta
    Geographica Sinica, 58(S1):1–10, 2003.

Figures
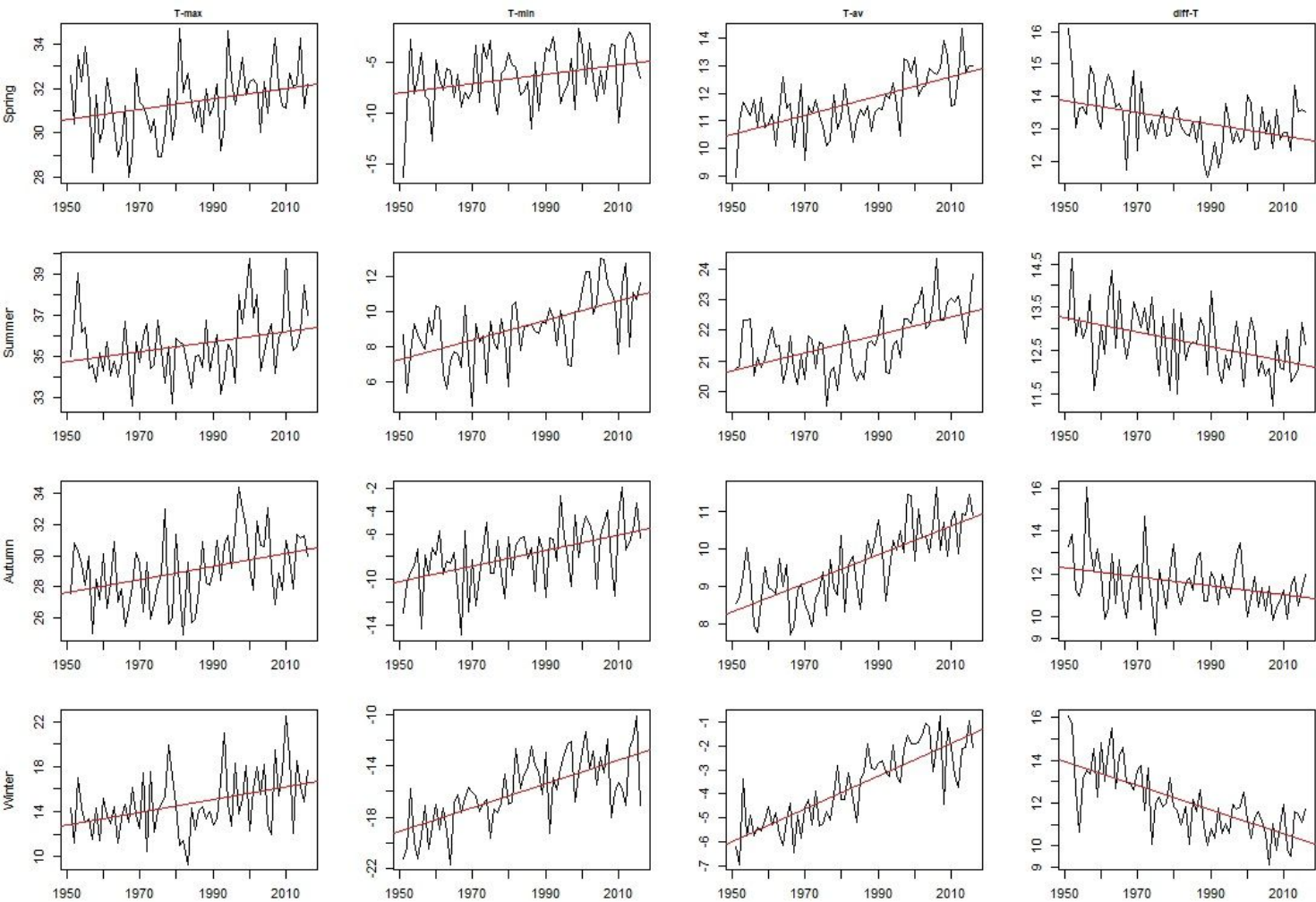
Figure 1

Seasonal temperature trend