Vol 15 No 9, September 2006
1009-1963/2006/15(09)/2180-05

Chinese Physics

© 2006 Chin. Phys. Soc.
and IOP Publishing Ltd

On the correlation of nonlinear variables containing secular trend variations: numerical experiments*

Shi Neng(施 能)[a)†], Yi Yan-Ming(易燕明)[b)],
Gu Jun-Qiang(顾骏强)[c)], and Xia Dong-Dong(夏冬冬)[d)]

[a)] *Key Laboratory of Meteorology and Environment, Nanjing University of Information Science and Technology, Nanjing 210044, China*

[b)] *Guangzhou Institute of Tropical and Oceanic Meteorology, Guangzhou 510080, China*

[c)] *Zhejiang Meteorological Science Research Institute, Hanzhou 310017, China*

[d)] *National Marine Environment Forecast Center, Beijing 100081, China*

(Received 21 March 2006; revised manuscript received 29 May 2006)

Due to global warming, the general circulation, underlying surfaces characteristics, and geophysical and meteorological elements all show evident secular trends. This paper points out that when calculating the correlation of two variables containing their own obvious secular trends, the interannual correlation characteristics between the two variables may be distorted (overestimated or underestimated). Numerical experiments in this paper show that if two variables have opposite secular trends, the correlation coefficient between the two variables is reduced (the positive correlation is underestimated, or the negative correlation is overestimated); and if the two variables have the same sign of secular trends, the correlation coefficient between the two variables is increased (the positive correlation is overestimated, or the negative correlation is underestimated). Numerical experiments also suggest that the effect of secular trends on the interannual correlation of the two variables is interchangeable, that is to say, as long as the values of the two trends are not changed, the two variables interchange their positions, and the effect of the secular trends on the interannual correlation coefficient of the two variables remains the same. If the two variables have the same-(opposite-) sign trends, the effect of secular trends on the interannual correlation coefficient is more (less) distinctive. A meteorological example is given.

Keywords: correlation analysis, secular trend, trend coefficient, numerical experiment
PACC: 9260X, 9260S, 0250, 0545

1. Introduction

Correlation analysis is frequently used in physics research[1−2] and prediction, and the correlation coefficient is always used to describe the closeness of variable correlation.[3−7] As the correlation coefficient must be calculated from finite data, it is also a stochastic variable.[8] If the correlation coefficient between two variables is large for a reduced sample, are they still correlated with each other in the infinite population? This falls into the category of the statistical test problem of correlation coefficients. For a long period of time, the Student's t test has been predominantly used as the statistical test for the correlation coefficients. However, the t test not only requires that variables should obey the joint normal distribution, but also that n observations should be mutually independent. Furthermore, the t test is not suitable for a large quantity of repeated experiments. As is well known, the time series of geophysical variables has a multiple-time scale character,[9] i.e. it may have the secular trend, interdecadal and interannual variations. The major purpose of correlation analysis is to understand the relation between the interannual changes of different variables. If a variable used to calculate the correlation coefficient contains an evident trend, i.e. it possesses a large persistency, obviously it does not satisfy the requirement that observational samples are mutually independent, and therefore the Student's t test is theoretically not applicable. If the t test is still used in this case, then the threshold of the t statistics obtained is so obviously underestimated that the two independent variables might pass the statistical test and be erroneously considered as being correlated with each other. Besides, filtering is frequently used to remove the interannual variation from time series

*Project supported by the National Natural Sciences Foundation of China (Grant Nos 40275031 and 90411008).
†E-mail: NengShi2005@163.com

of variables, yielding the interdecadally slow-varying oscillations.[10] It has been found from computational experiments that the correlation coefficient between the filtered slow-varying series of two variables is much higher than that between the original time series of the two variables.

So far as we know, paying attention to whether or not variables in correlation analysis contain trends is of practical significance. As is well known, global warming has become evident after the 1980s, and has manifested at first in the evident positive trends of the annual/seasonal mean temperatures over large-scale areas on the globe, such as northwest China, north China and northeast China, as well as in the evident trends of the characteristic quantities of the general circulation such as the intensity index of teleconnection patterns, oscillation indices, geopotential heights, the intensity of active centres, global rainfall,[11] intensity of monsoon index[12] etc. Therefore, it is possible that two variables in interannual correlation analysis both possess evident trends, thus greatly affecting the correlation analysis. The third section of the present paper performs numerical experiments to examine the effect of the trends of variables on the interannual correlation coefficient.

2. Methods-quantitative index for secular trends

In order to obtain the secular trend of a geophysical element y, we adopt the expression of linear regression in the form

$$y = a_1 + b_1 t, \quad t = 1, 2, \ldots n \text{ (year)}, \tag{1}$$

where a_1 denotes a constant and b_1 a regression coefficient. Obviously, Eq.(1) means that the element y has a trend of linear increase (decrease) when b_1 is positive (negative) in sample's size (n years). The coefficient $b_1$is referred to as the climatic tendency in Ref.[13] and expressed in the element's unit per year or 10 times the unit per decade. Since b_1 has its own unit, its magnitude cannot be compared with different factors or variables, and so even for the same element in different regions. We utilize the so-called unitless climatic trend coefficient r_{xt},[14] which is related to the regression coefficient b_1 through

$$b_1 = r_{xt}(\sigma_x/\sigma_t), \tag{2}$$

where r_{xt} is defined as the correlation coefficient between an element's series and the sequence of natural numbers of $1, 2, 3, \ldots n$, and can be tested by means of the Student's t test method or the Monte Carlo technique; σ_x and σ_t denote the standard deviations of the element y and natural number sequences, respectively. Since r_{xt} is a unitless quantity, its magnitudes can be compared and utilized to estimate the intensity of secular trends for different geophysical elements, and therefore it is applicable, in particular, to the investigation on the spatial features of a large-scale geophysical field.

3. Numerical experiments: influence of trends on the interannual correlation analysis

In the following, Monte Carlo numerical experiments are employed to calculate the influence of variable trends on the interannual correlation analysis of two variables. Our experimental scheme is as follows. First, assume that x_t and y_t are two stochastic series: i.e. their trend coefficients are 0.0, and in experiments the sample's sizes are 53 and 30 for (a) and (b), respectively. The sample correlation coefficient between x_t and y_t is certainly very close to 0.0 (theoretical value). Then, the two trend series of sample sizes n are generated, and then superposed onto stochastic series x_t and y_t, yielding series $\tilde{x}_t$ and $\tilde{y}_t$, respectively. Results show that because affected by their trends, the sample correlation coefficients of $\tilde{x}_t$ and $\tilde{y}_t$ are no longer close to 0.0. A large number of numerical experiments, more than 1000 times, wherein one or two trend coefficients of the two time series take a value between –1.0 and 1.0, have been performed, and the numerically calculated function relation between sample correlation coefficients and trend coefficients is plotted in Fig.1.

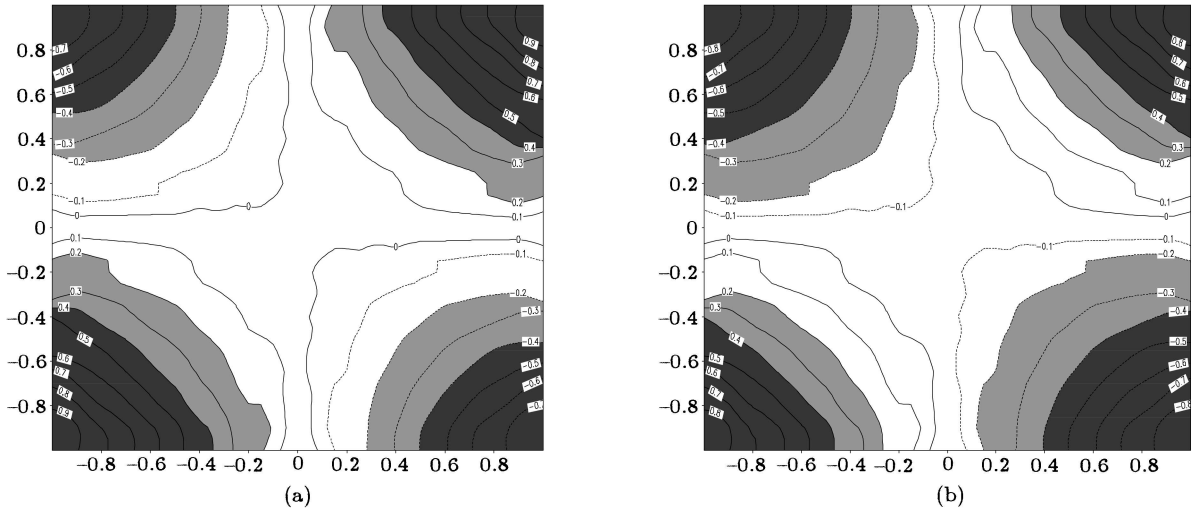


Fig.1. Influence of trends on the sample correlation coefficient of two variables. Both the trend and correlation coefficients of original variables x_t and y_t are zero, with n =53 in Fig.1(a) and n=30 in Fig.1(b). The contour value denotes the sample correlation coefficient between variables $\tilde{x}_t$ and $\tilde{y}_t$, with the negative contour dashed. The ordinate and abscissa represent the trend coefficients of $\tilde{x}_t$ and $\tilde{y}_t$, respectively.

Shown in Fig.1 are the following.

(i) The top (bottom) left and bottom (top) right corners in Figs.1(a) and 1(b) are the negative (positive) value areas, indicating that superimposing two opposite sign trends onto two variables reduces the correlation coefficient between the two variables, i.e. the positive correlation is under-estimated, while the negative correlation is over-estimated. When two same-sign trends, whether positive or negative, are superimposed onto two variables, the correlation coefficient of the two variables increases (the positive correlation is over-estimated, while the negative correlation is under-estimated).

(ii) The contours in the top (bottom) left corners and bottom (top) right corners in Figs.1(a) and 1(b) are symmetric with respect to the diagonal line, and the symmetry does not depend on the size of samples, indicating that when two same- or opposite-sign trends are superimposed onto two variables, as long as the values of the two trend coefficients are not changed, if the two variables interchange their trends, then the influence of the trends on the correlation coefficient between the two variables remains the same. However, the contours in the top and bottom parts in Figs.1(a) and 1(b) are not symmetric with respect to the median line, i.e. the absolute value of the influence of two same-sign trends on the correlation coefficient between the two variables is greater than that of two opposite-sign trends, indicating that the correlation coefficients between the two variables containing two same-sign trends is distorted more seriously than that between the two variables containing two opposite-sign trends.

(iii) It can be seen by comparing Fig.1(a) with Fig.1(b) that the influence of two opposite -sign trends (the negative correlation becomes strong or the positive correlation becomes weak) on the correlation coefficient between the two variables, when the sample size is smaller (n=30; Fig.1(b)), is larger than that when the sample size is larger (n=53; Fig.1(a)); whereas, the influence of two same-sign trends (the negative correlation becomes weak or the positive correlation becomes strong), when the sample size is smaller (n=30; Fig.1(b)), is smaller than that when the sample size is larger (n=53; Fig.1(a)).

4. A meteorological example: influence of trends on interannual correlation analysis

As is well known, the Siberian high is an important weather system in winter, which affects dominantly winter weather in China. However, in the past 50 years, the Siberian high in winter was abnormally weak, and its intensity showed an evident negative trend. According to Shi *et al*,[15] the trend coefficient of Siberian high in 1951/52–1999/2000 winter obtained is –0.68 (a distinctive weakening). Such a strong negative trend of the intensity index will certainly impact and distort its interannual correlation

with the other elements or factors. Correlations between the Siberian high intensity and winter temperature containing their own trends and with their trends removed in China are calculated. Figure 2 gives their differences of the correlation coefficients, which reflect the influence of their own trends on the interannual correlation between Siberian high intensity and winter temperature in China. There exists a small patch of positive difference values in Southwest China in Fig.2 and most parts of China are in the area of negative difference values, which agrees not only with the feature of an overall winter warming in China, and a weaker cooling in Southwest China, and also with the result inferred from numerical experiments in the third section of this paper.

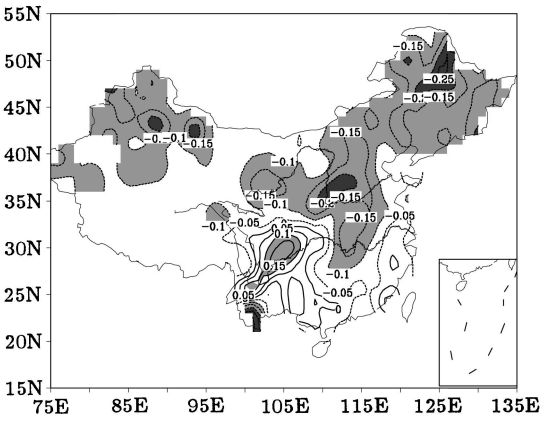


Fig.2. The influence of trends on interannual correlation analysis between Siberian high intensity and winter temperature in China. Negative contours are dashed.

5. Conclusions

This paper has demonstrated that when two variables contain obvious trends, the correlation characteristics of interannual changes of the two variables might be distorted, which has been investigated further using numerical experiments. The major results are as follows.

(i) When two variables contain their own obvious secular trends, the interannual correlation between the two variables might be distorted. Numerical experiments suggest that when two variables have opposite-sign trends, the correlation coefficients between the two variables may be erroneously underestimated (the positive correlation becomes weak, or the negative correlation becomes strong). When the two variables have same-sign trends, the correlation coefficient between the two variables may be erroneously overestimated (the positive correlation becomes strong, or the negative correlation becomes weak). This property does not depend on the size of the samples.

(ii) Results of numerical experiments show that the influence of secular trends on the interannual correlation of two variables is interchangeable. That is to say, as long as the values of the two trends are not changed, if the two variables interchange their positions, then the influence of trends on the correlation coefficient of the two variables is the same. This property also does not depend on the size of the samples.

(iii) Numerical experiments suggest that the influence of two same-sign trends of same absolute values on the interannual correlation of the two variables is slightly greater than that of two opposite-sign trends, indicating that the correlation coefficient between two variables containing two same-sign trends is distorted more seriously than that of two variables containing two opposite sign trends. This property also does not depend on the size of the samples.

(iv) Numerical experiments also indicate that the influence of two opposite-sign trends (the negative correlation becomes strong, or the positive correlation becomes weak) on the interannual correlation between the two variables, when the sample size is smaller, is slightly greater than that when the sample size is larger, whereas the influence of two same-sign trends (the negative correlation becomes weak, or the positive correlation becomes strong), when the sample size is smaller, is smaller than that when the sample size is larger.

Acknowledgments

We want to thank Professor Zhu Fucheng, Dr Zinecker, Rayco Vega and Dr Tom Robinson for valuable suggestions and comments.

References

[1] Cheng Y, Zhang X, Wu L, Mao W M and Yon L S 2006 *Acta Phys. Sin.* **55** 988 (in Chinese)

[2] Wu L W and Cheng Q S 2005 *Acta Phys. Sin.* **54** 3027 (in Chinese)

[3] Gong Z Q, Feng G L,Wan S Q and Li J P 2006 *Acta Phys. Sin.* **55** 477 (in Chinese)

[4] He W P, Feng G L, Dong W J and Li J P 2006 *Acta Phys. Sin.* **55** 969 (in Chinese)
[5] Feng G L, Hou W and Dong W J 2006 *Acta Phys. Sin.* **55** 962 (in Chinese)
[6] Dai X G, Fu C B and Wang P 2005 *Chin. Phys.* **14** 850
[7] Cheng Q H, Cao L, Xu D H and Wu D J 2005 *Chin. Phys.* **14** 1159
[8] Huang Z H, Wang X R and Zhu H 2004 *Chin. Phys.* **13** 273
[9] Zhang L J R, Li L and Shen Y P 2005 *Acta Phys. Sin.* **54** 1958 (in Chinese)
[10] Wu J R and Song S J 2004 *Chin. Phys.* **13** 1977
[11] Yang Y and Shi N 2003 *Sci. Meteorol. Sin.* **23** 253 (in Chinese)
[12] Shi N, Gu J Q, Yi Y M and Lin Z M 2005 *Chin. Phys.* **14** 844
[13] Jones P, Wigley T M L and Kelly P M 1982 *Mon. Wea. Rev.* **110** 59
[14] Shi N and Deng Z W 2000 *Meteorol. Atmos. Phys.* **73** 35
[15] Shi N, Chen H and Xiao D D 2000 *J. Trop. Meteorol.* **6** 194 (in Chinese)