

非定长相关时间序列集内平均序列的 摆动异常及其解释^{*}

刘洪滨 周江兴

(国家气候中心, 北京, 100081)

摘 要

文中对多个时间序列的平均序列中常常出现的异常摆动问题进行了探讨。结果表明, 给定时间序列集中序列个数的多少会造成其所组成的平均序列的方差的变化, 即序列个数越多, 平均序列方差越小的可能性就越大, 由此定义了方差增长率 α 。当给定方差增长率限定值 α_0 并使其满足 $\alpha \leq \alpha_0$ 时, 可以得到最小序列个数 m_0 , 且 α_0 越大或序列间平均相关系数越大, 所需要的时间序列数越少。这一结果可应用于气候变化研究中有关平均时间序列建立的研究中, 亦可运用到其他相关研究领域。

关键词: 时间序列, 方差均值, 气候变化。

1 引 言

统计学方法作为一种十分重要的分析手段已被广泛地应用于气候变化研究中, 其中取多个时间序列的平均序列以提取共同变化信息是常用的统计方法之一, 树木年轮气候学中年表的建立过程以及全球或区域平均温度序列的建立过程均采用了这一方法。在处理年代较长的气候资料及树木年轮资料时, 常常会出现这样一种情况: 早期的时间序列个数相对较少, 而随着时间的推移, 时间序列个数逐步增加, 即不同时段内平均序列取值所用的时间序列个数随时间递增。这种改变会导致平均序列的许多统计特征在不同的时间段内出现明显的差别。Wigley 等^[1]曾就少量时间序列组成的平均序列对全部时间序列组成的平均序列的变化一致性问题进行了有益的探讨, 并由此给出两个统计量 $(\bar{R}_{mM})^2$ 和 $(\bar{R}_m)^2$, 分别表示有限时间序列集 X_M (M 为序列个数) 中的有限子集 X_m ($m < M$) 当中所有序列的平均序列对有限时间序列集 X_M 中所有序列的平均序列的平均方差解释量, 以及无限时间序列集 X_∞ 中的有限子集 X_m 当中所有序列的平均序列对无限时间序列集 X_∞ 中所有序列的平均序列的平均方差解

释量。上述讨论的是部分序列平均序列对总体序列平均序列变化的逼近性问题, 即满足何种条件可以使得两条平均序列的变化足够相似。这一方法已在树木年轮气候学研究中得到了广泛的应用^[2~4]。此外, New 等^[5,6]在建立全球大陆地区气候要素格点场的工作中也考虑了与序列个数有关的代表性问题。

文中所关心的问题是, 在这种平均序列中常常出现的异常摆动问题。刘洪滨等^[7]在建立秦岭镇安鹰嘴崖样点树木年轮年表时发现, 约在 16 世纪中期年表序列的摆动幅度由大变小 (图 1), 并对应于组成年表样本序列个数的大幅度增多; 邵雪梅等^[8]在重建川西地区气候变化研究中也发现了同样的问题。

这种平均序列中出现的摆动异常显然并非所研究的客观现象本身所固有的特性, 而是由于组成平均序列的序列个数的非均一性所导致的波幅异常变化, 它会干扰人们对所研究的真实现象的理解和分析, 特别是在研究气候突变这一重要问题时, 这种干扰是不容忽视的。因此, 从理论上仔细地分析这种干扰的强度随序列个数变化的关系就显得十分重要, 这是本文试图解决的问题。

* 初稿时间: 2003 年 3 月 3 日; 修改稿时间: 2003 年 3 月 28 日。

资助课题: 国家“十五”科技攻关项目 (2001BA611B-01) 及中国科学院知识创新工程重要方向项目 (KZCX3-SW-321)。

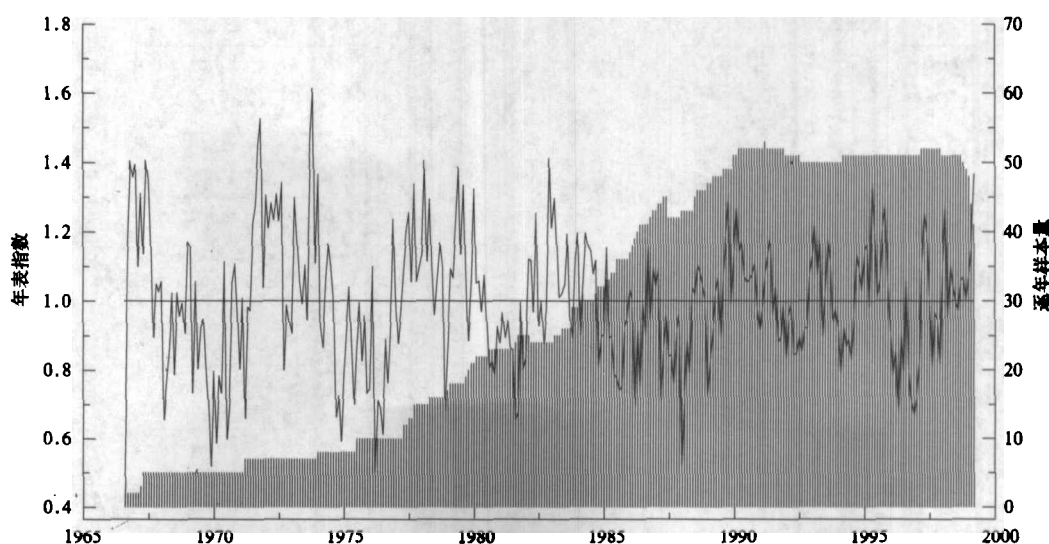


图1 陕西镇安鹰嘴崖样点树轮年表及逐年样本量

2 问题的数学描述

本文中假定用于合成平均序列的各时间序列的方差及时间序列间的相关系数等统计量在所研究的时间段内是平稳的,即在不同的时间段内时间序列的基本统计特征是不变的。因此,不同时段内平均序列的方差随时间序列个数变化的问题可以转换为在相同时段内平均序列的方差随时间序列个数变化的问题。

2.1 定义

设 X 为 m 个距平时间序列 x_i 所组成的时间序列集,时间序列长度均为 n , x_{ij} 为 x_i 中的第 j 个元素,且 x_{ij} 在 j 点上的取值服从正态分布, $\bar{x}$ 为 X 中所有 m 个时间序列的平均序列, $\bar{x}_j$ 为 $\bar{x}$ 中的第 j 个元素,即

$$\bar{x}_j = \frac{1}{m} \sum_{i=1}^m x_{ij}$$

其中, $m > 1, n > 1, i \in [1, m], j \in [1, n]$; $m'_k X$ 为 X 中 m' 个时间序列 x_i 所组成的第 k 个子集, $m'_k x_{ij}$ 为 $m'_k x_i$ 中的第 j 个元素, $m'_k \bar{x}$ 为 $m'_k x$ 中所有 m' 个时间序列的平均序列, $m'_k \bar{x}_j$ 为 $m'_k \bar{x}$ 中的第 j 个元素,即

$$m'_k \bar{x}_j = \frac{1}{m'} \sum_{i=1}^{m'} m'_k x_{ij}$$

其中, $m \geq m' \geq 1, k \in [1, C_m^{m'}], i \in [1, m']$ 。

序列 $x_i, \bar{x}, m'_k x_i$ 及 $m'_k \bar{x}$ 的方差 $S_i, S, m'_k S_i$ 及 $m'_k S$ 分别定义为:

$$S_i = \frac{1}{n} \sum_{j=1}^n x_{ij}^2 \quad (1)$$

$$S = \frac{1}{n} \sum_{j=1}^n \bar{x}_j^2 = \frac{1}{m^2 n} \sum_{j=1}^n \left(\sum_{i=1}^m x_{ij} \right)^2 \quad (2)$$

$$m'_k S_i = \frac{1}{n} \sum_{j=1}^n m'_k x_{ij}^2 \quad (3)$$

$$m'_k S = \frac{1}{n} \sum_{j=1}^n m'_k \bar{x}_j^2 = \frac{1}{m'^2 n} \sum_{j=1}^n \left(\sum_{i=1}^{m'} m'_k x_{ij} \right)^2 \quad (4)$$

X 中任意两条不同序列 x_{i_1} 和 x_{i_2} 间的相关系数 $r_{i_1 i_2}$ 为:

$$r_{i_1 i_2} = \frac{\sum_{j=1}^n x_{i_1 j} x_{i_2 j}}{\sqrt{\sum_{j=1}^n x_{i_1 j}^2} \sqrt{\sum_{j=1}^n x_{i_2 j}^2}} = \frac{\sum_{j=1}^n x_{i_1 j} x_{i_2 j}}{n \sqrt{S_{i_1}} \sqrt{S_{i_2}}} \quad (5)$$

其中, $i_1 \in [1, m], i_2 \in [1, m]$, 且 $i_1 \neq i_2$ 。

由此分别定义统计量 $\bar{S}$ 和 $\bar{r}$ 。方差均值 $m' \bar{S}$ 为固定 m' 的所有 $m'_k S (k=1, 2, \dots, C_m^{m'})$ 的平均值, 即

$$\begin{aligned} m' \bar{S} &= \frac{1}{C_m^{m'}} \sum_{k=1}^{C_m^{m'}} m'_k S \\ &= \frac{1}{m'^2 C_m^{m'} n} \left[\sum_{k=1}^{C_m^{m'}} \sum_{j=1}^n \left(\sum_{i=1}^{m'} m'_k x_{ij} \right)^2 \right] \\ &= \frac{1}{m'^2 C_m^{m'} n} \sum_{j=1}^n \left[\sum_{k=1}^{C_m^{m'}} \sum_{i=1}^{m'} m'_k x_{ij}^2 + \right. \\ &\quad \left. \sum_{k=1}^{C_m^{m'}} \sum_{i_1=1}^{m'-1} \sum_{i_2=i_1+1}^{m'} 2 m'_k x_{i_1 j} m'_k x_{i_2 j} \right] \\ &= \frac{1}{m'^2 C_m^{m'} n} \sum_{j=1}^n \left[\frac{m' C_m^{m'}}{m} \sum_{i=1}^n x_{ij}^2 + \right. \\ &\quad \left. \frac{C_m^{m'} C_m^2}{C_m^2} \sum_{i_1=1}^{m'-1} \sum_{i_2=i_1+1}^m 2 x_{i_1 j} x_{i_2 j} \right] \end{aligned}$$

$$= \frac{1}{mm'n} \sum_{j=1}^n \sum_{i=1}^m x_{ij}^2 + \frac{m'-1}{m(m-1)m'n} \cdot \sum_{j=1}^n \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m 2x_{i_1j}x_{i_2j} \quad (6)$$

可见,当 $m'=m$ 时, ${}^m\bar{S}=S$ 。方差均值 ${}^m\bar{S}$ 可被认为是固定 m' 的平均序列的方差估计。

平均相关系数 $\bar{r}$ 为 X 中所有 $r_{i_1i_2}$ 的平均值,即

$$\bar{r} = \frac{1}{C_m^2} \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m r_{i_1i_2} = \frac{2}{m(m-1)} \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m r_{i_1i_2} \quad (7)$$

可见, $\bar{r}$ 满足 $-1 \leq \bar{r} \leq 1$ 。

2.2 证明

证明 ${}^m\bar{S} \geq {}^{m'}\bar{S}$ 成立,其中 $m \geq m' \geq m'' \geq 1$, 且 $m > 1$ 。

由方差均值定义式(6)可知,

$${}^m\bar{S} = \frac{1}{mm'n} \sum_{j=1}^n \sum_{i=1}^m x_{ij}^2 + \frac{m''-1}{m(m-1)m''n} \cdot \sum_{j=1}^n \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m 2x_{i_1j}x_{i_2j} \quad (8)$$

则

$$\begin{aligned} {}^m\bar{S} - {}^{m'}\bar{S} &= \frac{m'-m''}{mm'm'n} \sum_{j=1}^n \sum_{i=1}^m x_{ij}^2 - \frac{m'-m''}{m(m-1)m'm''n} \cdot \sum_{j=1}^n \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m 2x_{i_1j}x_{i_2j} \\ &= \frac{m'-m''}{m(m-1)m'm''n} \sum_{j=1}^n \left[(m-1) \sum_{i=1}^m x_{ij}^2 - \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m 2x_{i_1j}x_{i_2j} \right] \\ &= \frac{m'-m''}{m(m-1)m'm''n} \cdot \left[\sum_{j=1}^n \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m (x_{i_1j} - x_{i_2j})^2 \right] \quad (9) \end{aligned}$$

当 X 中时间序列间各不相同,满足

$$\left[\sum_{j=1}^n \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m (x_{i_1j} - x_{i_2j})^2 \right] > 0$$

由式(9)可见, ${}^m\bar{S} \geq {}^{m'}\bar{S}$ 的充分必要条件是: $m' \geq m''$ 。式(9)表示:当给定时间序列集 X 时,方差均值 ${}^m\bar{S}$ 的大小与 m' 值有关: m' 越小,方差均值 ${}^m\bar{S}$ 越大。

由方差定义式(1)和相关系数定义式(5),式(9)可改写为

$$\begin{aligned} {}^m\bar{S} - {}^{m'}\bar{S} &= \frac{m'-m''}{m(m-1)m'm''n} \cdot \left[\sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m \sum_{j=1}^n (x_{i_1j} - x_{i_2j})^2 \right] \\ &= \frac{m'-m''}{m(m-1)m'm''n} \cdot \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m \left[\sum_{j=1}^n x_{i_1j}^2 + \sum_{j=1}^n x_{i_2j}^2 - 2 \sum_{j=1}^n x_{i_1j}x_{i_2j} \right] \\ &= \frac{m'-m''}{m(m-1)m'm''n} \cdot \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m \left[S_{i_1} + S_{i_2} - 2r_{i_1i_2} \sqrt{S_{i_1}} \sqrt{S_{i_2}} \right] \\ &= \frac{m'-m''}{m(m-1)m'm''n} \cdot \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m \left[(\sqrt{S_{i_1}} - \sqrt{S_{i_2}})^2 + 2(1-r_{i_1i_2}) \sqrt{S_{i_1}} \sqrt{S_{i_2}} \right] \quad (10) \end{aligned}$$

当 X 中所有 m 条序列 x_i 均为标准化序列时,满足 $S_{i_1} = S_{i_2} = 1$,且由平均相关系数定义式(7),式(10)

可改写为

$$\begin{aligned} {}^m\bar{S} - {}^{m'}\bar{S} &= \frac{2(m'-m'')}{m(m-1)m'm''n} \cdot \sum_{i_1=1}^{m-1} \sum_{i_2=i_1+1}^m (1-r_{i_1i_2}) \\ &= \frac{m'-m''}{m'm''n} (1-\bar{r}) \geq 0 \quad (11) \end{aligned}$$

不等式得证。

2.3 方差增长率

在标准化序列假定下,当 $m'=m$ 时,由式(11)得到

$${}^m\bar{S} - S = \frac{m-m''}{mm''n} (1-\bar{r})$$

用 m' 替换上式中的 m'' ,得到

$${}^m\bar{S} - S = \frac{m-m'}{mm'n} (1-\bar{r}) \quad (12)$$

由于当 $m'=1$ 时,满足 $S_i=1$,则 ${}^1\bar{S}=1$,得到

$$S = 1 - \frac{m-1}{m} (1-\bar{r}) \quad (13)$$

将式(13)代入式(12),得到

$${}^m\bar{S} = 1 - \frac{m'-1}{m'} (1-\bar{r}) \quad (14)$$

则

$$\frac{m'\bar{S} - S}{S} = \frac{(m - m')(1 - \bar{r})}{m'[1 + (m - 1)\bar{r}]} \equiv \alpha \quad (15)$$

其中, α 定义为方差均值 $m'\bar{S}$ 对 S 的方差增长率。当给定方差增长率限定值 α_0 并使其满足 $\alpha \leq \alpha_0$ 时, 由式(15)得到

$$\frac{(m - m')(1 - \bar{r})}{m'[1 + (m - 1)\bar{r}]} \leq \alpha_0$$

$$m' \geq \frac{1}{\frac{1 + \alpha_0}{m} + \frac{\alpha_0 \bar{r}}{1 - \bar{r}}} \equiv m'_0 \quad (16)$$

其中, m'_0 定义为满足 $\alpha \leq \alpha_0$ 的最小序列个数。可见, 方差增长率限定值 α_0 越大, m'_0 值越小; 平均相关系数 $\bar{r}$ 越大, m'_0 值越小。

当 $m \rightarrow +\infty$ 时, 则有

$$m'_0 = \frac{1 - \bar{r}}{\alpha_0 \bar{r}} \quad (17)$$

3 实例

以陕西镇安鹰嘴崖样点的树轮样本宽度序列的差值序列为例, 选取覆盖时间区间 1873~1988 年 ($n=116$) 的 40 个 ($m=40$) 差值序列组成时间序列集 X 。首先将 X 中各序列进行标准化处理, 使得各序列均满足均值为 0, 方差为 1。经计算得到平均相关系数 $\bar{r}=0.276$, 平均序列 $\bar{x}$ 的方差为 $S=0.293$ 。

讨论 $m'=5$ 和 $m'=10$ 两种情况。利用式(14)可以得到 ${}^5\bar{S}=0.420$, ${}^{10}\bar{S}=0.347$, 然后分别计算 ${}^{10}S_k$ 和 5S_k (图 2)。可见 ${}^{10}S_k$ 和 5S_k 均大致呈单峰对称分布, 且以分布图形中心的概率分布密度最大。但二者分布也存在明显差别, ${}^{10}S_k$ 的分布中心密度较 5S_k 偏大 (分布图形较“高”), 且分布较 5S_k 更为集中 (分布图形较“瘦”), 说明 ${}^{10}S_k$ 与 ${}^{10}\bar{S}$ 相近的概率大于 5S_k 与 ${}^5\bar{S}$ 相近的概率。可见, m' 取值越大, m'_S 较小的可能性越大 (图 3)。

在实际应用中, 一般需要对序列给出一个估计, 使得 $m'\bar{S}$ 与 S 的偏差小于某一限定值, 因此可以根据式(16)计算在不同给定方差增长率限定值 α_0 情

况下所需最小序列个数 (表 1)。

由表 1 可见, 随着方差增长率限定值 α_0 的增大, 所需要的序列个数也逐步减少。例如此个例中若要满足 $\alpha \leq 20\%$, 则需满足 $m' \geq 10$; 若要满足 $\alpha \leq 10\%$, 则需满足 $m' \geq 16$ 。

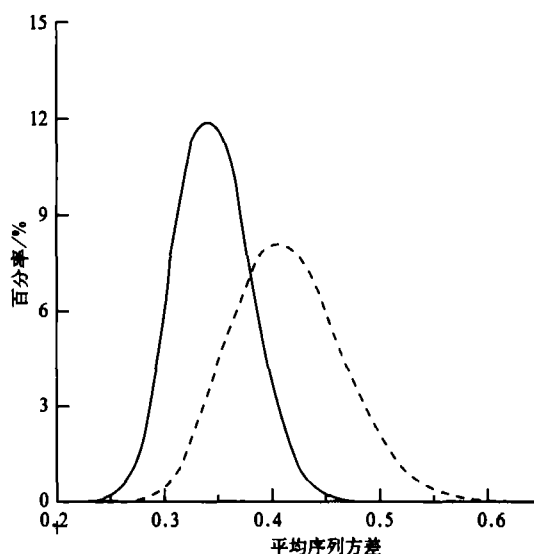


图 2 序列子集平均序列方差估计比较
(实线: $m'=10$, 虚线: $m'=5$)

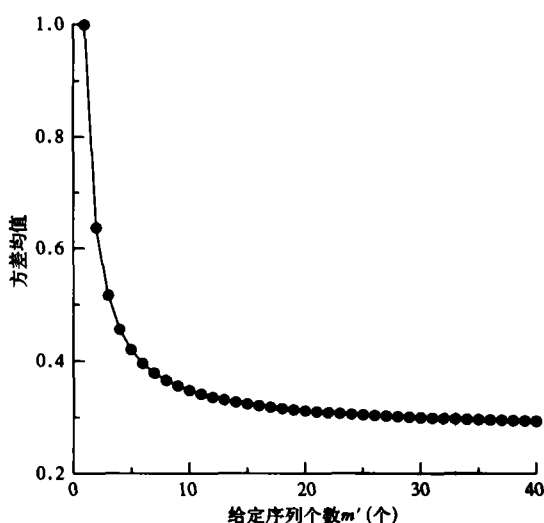


图 3 不同序列个数子集的方差均值变化

表 1 方差增长率限定值 (α_0) 及其对应的最小序列个数 (m'_0)

α_0 (%)	0	5	10	15	20	25	30	35	40	45	50
m'_0 (个)	40	22.1	15.2	11.6	9.4	7.9	6.8	6.0	5.3	4.8	4.4

4 结 论

通过以上分析,可以得出一些结论。当给定时间序列集 X 时,若 X 中各时间序列的方差及时间序列间的相关系数等统计特征在所研究的时间段内是平稳的,那么

(1) 用于得到平均序列的时间序列个数会对平

均序列的方差大小产生影响,时间序列个数越多,平均序列方差越小的可能性就越大;

(2) 给定的方差增长率限定值越小,所需时间序列个数越多;而时间序列间的相关性越小,所需时间序列个数也越多。

致谢:中国科学院应用数学研究所陈敏副研究员和中国科学院地理科学与资源研究所邵雪梅研究员给予了热情帮助和指点,在此表示衷心感谢。

参考文献

- 1 Wigley T M L, Briffa K R, Jones P D. On the average value of correlated time series, with applications in dendroclimatology and hydrometeorology. *J Clim App Meteo*, 1984,23:201~213
- 2 Hughes M K, Wu X, Shao X, et al. A preliminary reconstruction of rainfall in North-Central China since A. D. 1600 from tree-ring density and width. *Quaternary Research*, 1994,42:88~99
- 3 Shao X, Wu X. Radial growth of Huashan Pine and its response to climate. *Journal of Chinese Geography*, 1994,4(3/4):88~102
- 4 Liu Y, Ma J, Hughes M K, et al. Seasonal temperature reconstruction from central China based on tree ring data. *Palaeobotanist*, 2001, 50:89~94
- 5 New M G, Hulme M, Jones P D: Representing twentieth-century space-time climate variability. Part I: Development of a 1961 - 90 mean monthly terrestrial climatology. *J Clim*, 1999,12:829~856
- 6 New, M G, Hulme M, Jones P D. Representing twentieth-century space-time climate variability. Part II: Development of 1901 - 96 monthly grids of terrestrial surface climate. *J Clim*, 2000,13: 2217~2238
- 7 刘洪滨,邵雪梅. 采用秦岭冷杉年轮宽度重建陕西镇安 1755 年以的初春温度. *气象学报*,2000,58(2):223~233
- 8 邵雪梅,范金梅. 树轮宽资料所指示的川西过去气候变化. *第四纪研究*,1999,19(1):81~89

ON THE ABNORMAL SWING OF THE AVERAGED TIME SERIES PRODUCED FROM A NUMBER OF CORRELATED TIME SERIES

Liu Hongbin Zhou Jiangxing

(National Climate Center, Beijing 100081)

Abstract

In many climatological researches, the averaged time series can be produced from a number of time series to enhance common signals. The abnormal swing always exists in the averaged series, so that the changes of the former part of the series, with less samples, are larger than those of the latter part, with more samples. This phenomenon could be found either in tree ring chronologies produced from tree ring series with different lengths or in the averaged time series of some climatic parameters produced from records of meteorological station network.

We suppose that there is a series set defined as X including m time series ($m > 1$). So in X , there are $C_m^{m'}$ ($m \geq m' \geq 1$) series subsets ${}_k^m X$ ($k \in [1, C_m^{m'}]$), including m' time series. For each subset ${}_k^m X$, ${}_k^{m'} S$ is the variance of the averaged series of all time series in the subset. ${}^{m'} \bar{S}$ is the mean value of all variances ${}_k^{m'} S$ with m' time series. So, ${}^{m'} \bar{S}$ could be taken as the variance of the averaged time series of m' time series with maximum possibility. Term $\bar{r}$ is the mean value of each inter-series correlation coefficient in set X . The ${}^{m'} \bar{S}$ value would be getting smaller as the series number m' getting larger, namely, ${}^{m'} \bar{S} \geq {}^{m''} \bar{S}$ ($m \geq m' \geq m'' \geq 1$).

Then, variance increasing rate α and minimum series number m'_0 are define as

$$\alpha \equiv \frac{(m - m')(1 - \bar{r})}{m'[1 + (m - 1)\bar{r}]}$$

$$m'_0 \equiv \frac{1}{\frac{1 + \alpha_0}{m} + \frac{\alpha_0 \bar{r}}{1 - \bar{r}}}$$

where α_0 is the limit value of α . That is to say, when the value α_0 is given to make α satisfy $\alpha \leq \alpha_0$, the series number m' should satisfy $m' \geq m'_0$. When $m \rightarrow +\infty$, term m'_0 would be expressed as

$$m'_0 \equiv \frac{1 - \bar{r}}{\alpha_0 \bar{r}}$$

These results are applied to determine the uncertainty in any conditions when averaged time series is produced from a number of time series with different lengths. The terms defined here are useful for the researchers to make decision how long an averaged series (like tree ring chronologies) could be used with lower uncertainty.

Key words: Time series, Mean variance, Climate change.