“卫星资料应用”专题系列

气候变化趋势计算及其对观测精度的敏感性

邹晓蕾

（1 佛罗里达州立大学地球海洋大气科学系，美国； 2 南京信息工程大学资料同化与研究应用中心，南京 210044，中国）

全球变暖是众所周知的事实。然而，地球大气的变率估计还存在较大的不确定性。这里，简要描述根据资料获得气候变化趋势的一个传统估计方法——线性回归方法，并指出影响气候变化趋势估计精度的几个重要因子。

用 x 来表示想要得到其气候变化趋势的变量。假设给定了一个时间序列：$\{x_i^{obs}=x^{obs}(t_i), i=1, 2, L, N\}$，其中 t_i 表示第 i 个观测，N 是序列的总长度。再假定 N 次的观测变量的平均值为零，即$\overrightarrow{x_i^{obs}}=0$。

首先，观测资料时间序列可以表达为：$x^{obs}=x^{true}+\varepsilon$，其中$x^{true}$是真值，$\varepsilon$表示观测误差，其方差表示为$\sigma_{obs}^2$。然后把变量$x$表示为时间的线性函数：$x^{\mathrm{mod}\,el}=a(t-\bar{t})$，其中$a$是回归系数（即气候变化趋势，其值待定），$\bar{t}$是平均年。最后再把变量$x$的真值表达为：$x^{true}=x^{\mathrm{mod}\,el}+e$，其中$e$代表变量$x$的非线性变化（即自然变率），其方差表达为$\sigma_{nv}^2$。写成矩阵形式为：

$$\boldsymbol{x}^{true}=\boldsymbol{A}a+\boldsymbol{e} \tag{1}$$

$$\boldsymbol{X}^{true}=\begin{pmatrix} x_1^{true} \\ x_2^{true} \\ \boldsymbol{M} \\ x_N^{true} \end{pmatrix},\ \boldsymbol{e}=\begin{pmatrix} e_1 \\ e_2 \\ \boldsymbol{M} \\ e_N \end{pmatrix},\ \boldsymbol{A}=\begin{pmatrix} t_1-\bar{t} \\ t_2-\bar{t} \\ \boldsymbol{M} \\ t_N-\bar{t} \end{pmatrix}$$

线性回归系数a的值可以通过最小二乘估计方法得到，即要求变量x的观测值和线性回归模式值方差

$$\sigma^2(a)=\left(\boldsymbol{x}^{obs}-\boldsymbol{A}a\right)^T\left(\boldsymbol{x}^{obs}-\boldsymbol{A}a\right) \tag{2}$$

最小。假定观测误差与模型误差逐年无时间相关，观测误差与模型误差也无相关，通过求解一阶导数为零的方程：$\partial\sigma^2(a)/\partial a=0$，可以得到使误差方差最小$a^*=\min\limits_a\sigma^2(a)$的解的表达式：

$$a^*=\frac{\sum_{i=1}^{N}\left(x_i^{obs}-\overline{x^{obs}}\right)\left(t_i-\bar{t}\right)}{\sum_{i=1}^{N}\left(t_i-\bar{t}\right)^2}=\frac{\sum_{i=1}^{N}x_i^{obs}\left(t_i-\bar{t}\right)}{\frac{N^3-N}{12}} \tag{3}$$

收稿日期：2011 年 11 月 7 日

作者：邹晓蕾，Email: xzou@fsu.edu

公式（3）是根据气候数据估计气候变化趋势的表达式。在推导（3）式的过程中用到了下面的不等式（注意 $\bar{t}=(N+1)/2$）：

$$\sum_{i=1}^{N}(t_i-\bar{t})^2=\sum_{i=1}^{N}(t_i^2+\bar{t}^2-2t_i\bar{t})=\sum_{i=1}^{N}t_i^2+N\left(\frac{N+1}{2}\right)^2-2\frac{N(N+1)}{2}\frac{N+1}{2}$$

$$=\frac{N(N+1)(2N+1)}{6}-N\left(\frac{N(N+1)}{2}\right)^2=\frac{N^3-N}{12}$$

把（3）式代入（2）式可以得到根据（3）式得到的趋势估计的精度为

$$\sigma_{trend}^2\equiv\sigma^2(a^*)=\frac{12\left(\sigma_{obs}^2+\sigma_{nv}^2\right)}{N^3-N} \tag{4}$$

其中σ_{obs}^2是观测数据的误差方差，σ_{nv}^2是自然变率方差。由（4）式可见，用线性回归法方法估计得到的气候变化趋势的准确性与观测误差的大小、数据的长度和观测变量本身的自然变率大小有关。自然变率和观测误差越大，要准确推测气候变化趋势所需要的数据记录就越长。

下面我们用一个简单例子来描述气候变化趋势计算结果对观测精度的敏感性。首先，我们产生300年月平均温度的三个时间序列。这三个时间序列是这样产生的：在0.2K线性年代际变化趋势上加上不同的随机扰动，方差分别为σ_{obs}=0.1，0.3和1K。自然变率假定为0.1K。然后用不同尺度的资料根据（2）式计算线性年代际变化趋势，结果如图1所示。由图可见，如果资料精度高，正确估计线性年代际变化趋势所需要的资料尺度短。反言之， 如果资料精度低，估计线性年代际变化趋势所需要的资料尺度要长得多。线性回归方法得到的气候变化趋势的估计精度如图2所示。观测资料精度对气候变化趋势的影响显而易见。如果观测精度从1K减小到0.3K或0.1K，要达到同一趋势估计精度σ_{trend}=1.5×10^{-5}K所需要的资料长度分别为82，140和300年。

由上，得出如下结论：当资料供应者提供观测资料时必须同时提供观测精度的定量值，这对用此资料研究气候变化趋势尤为重要。当资料使用者得到一个观测资料序列时，一定要重视得到描述该观测资料的误差精度的偏差和标准差，从而大体推算出通过线性回归得到气候变化趋势的可靠估计所需要的资料序列长度。

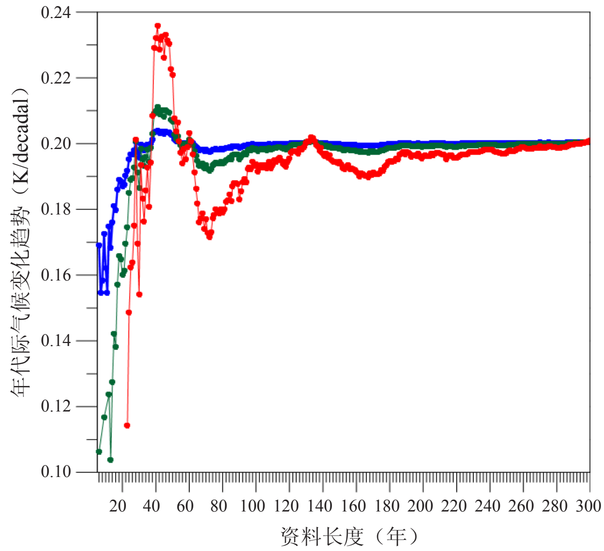

图1 根据不同资料长度得到的三个不同时间序列的年代际气候变化趋势，三个不同时间序列之间的唯一差别在于观测精度的不同：σ_{obs}=0.1K（蓝色），0.3K（绿色）和1K（红色）

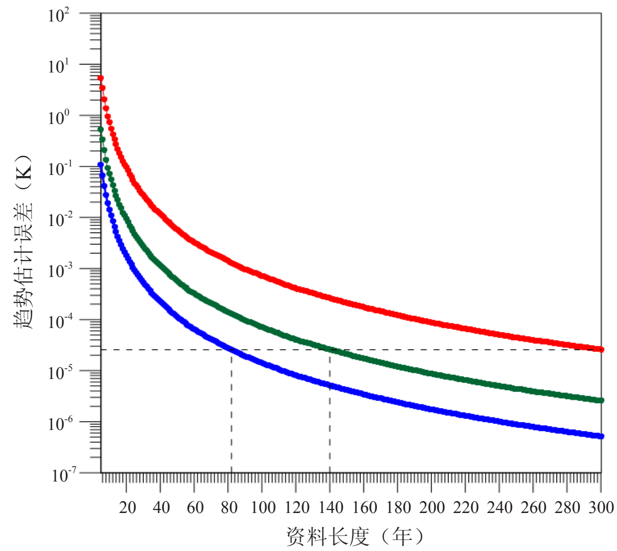

图2 根据不同资料长度得到的三个不同时间序列的年代际气候变化趋势估计误差，三个不同时间序列之间的唯一差别在于观测精度的不同：σ_{obs}=0.1K（蓝色），0.3K（绿色）和1K（红色）

Serial of Applications of Satellite Observations

Climate Trend Detection and Its Sensitivity to Measurement Precision

Zou Xiaolei[1, 2]

(1 Department of Earth, Ocean and Atmospheric Sciences, Florida State University, USA

2 Center of Data Assimilation for Research and Application, Nanjing University of Information and Science & Technology, Nanjing 210044, China)

Global warming is a well-known fact. However, large uncertainties exist in the quantitative estimate of global climate trend of the atmosphere. Here, we briefly describe a simple statistical method -- the linear regression method -- for climate trend detection using observations and point out a few factors controlling the precision of such an estimate.

Given a time series of data: $\{x_i^{obs}=x^{obs}(t_i), i=1, 2, L, N\}$, where x represents a measured variable (such as annual global mean near-surface atmospheric temperature), t_i represents the i^{th} measurement in time, and N is the total number of measurements in the time series. Assume the mean value of the variable x for the N measurements has been removed from the data, i. e., $\overrightarrow{x_i^{obs}}=0$.

Firstly, we can express the observed time series as follows: $x^{obs}=x^{true}+\varepsilon$, where x^{true}represent the truth and ε is the observation error whose variance is denoted as σ_{obs}^2. Secondly, the variable x is modeled by a linear function of time:$x^{model}=a(t-\bar{t})$, where a is the regression coefficient (e.g., the climate trend to be determined) and $\bar{t}$ represents the average year. Thirdly, the true value of the variable x is expressed as $x^{true}=x^{model}+e$, where e is the nonlinear term representing the natural variability whose variance is denoted as σ_{nv}^2. We may write the above expressions into the following matrix form

$$\boldsymbol{x}^{true}=\boldsymbol{A}a+\boldsymbol{e} \quad (1)$$

$$\boldsymbol{X}^{true}=\begin{pmatrix} x_1^{true} \\ x_2^{true} \\ \boldsymbol{M} \\ x_N^{true} \end{pmatrix}, \quad \boldsymbol{e}=\begin{pmatrix} e_1 \\ e_2 \\ \boldsymbol{M} \\ e_N \end{pmatrix}, \quad \boldsymbol{A}=\begin{pmatrix} t_1-\bar{t} \\ t_2-\bar{t} \\ \boldsymbol{M} \\ t_N-\bar{t} \end{pmatrix}$$

The linear regression coefficient a is obtained by a least-square fit, which minimizes the differences between observations and linear regression model:

$$\sigma^2(a) = \left(\boldsymbol{x}^{obs} - \boldsymbol{A}a\right)^T \left(\boldsymbol{x}^{obs} - \boldsymbol{A}a\right) \quad (2)$$

Assuming that there is no temporal error correlation for both observations and regression model, and observation and model errors are independent, one may obtain the minimum solution of $a^* = \min_a \sigma^2(a)$ by setting the first derivative to zero: $\partial\sigma^2(a)/\partial a = 0$, which gives the following expression for trend detection:

$$a^* = \frac{\sum_{i=1}^{N}\left(x_i^{obs} - \overline{x^{obs}}\right)\left(t_i - \bar{t}\right)}{\sum_{i=1}^{N}\left(t_i - \bar{t}\right)^2} = \frac{\sum_{i=1}^{N} x_i^{obs}\left(t_i - \bar{t}\right)}{\dfrac{N^3 - N}{12}} \quad (3)$$

Equation(3) is used for estimating the trend from data. It is pointed out that in the derivation of (3) we used the following equality for equating the denominators in (3) (notice $\bar{t} = (N+1)/2$):

$$\sum_{i=1}^{N}(t_i - \bar{t})^2 = \sum_{i=1}^{N}(t_i^2 + \bar{t}^2 - 2t_i\bar{t}) = \sum_{i=1}^{N} t_i^2 + N\left(\frac{N+1}{2}\right)^2 - 2\frac{N(N+1)}{2}\frac{N+1}{2}$$
$$= \frac{N(N+1)(2N+1)}{6} - N\left(\frac{N(N+1)}{2}\right)^2 = \frac{N^3 - N}{12}$$

By substituting (3) for (2) one may obtain the precision for the trend detection using (3) is equal to

$$\sigma_{trend}^2 \equiv \sigma^2(a^*) = \frac{12\left(\sigma_{obs}^2 + \sigma_{nv}^2\right)}{N^3 - N} \quad (4)$$

where σ_{obs}^2 is the observation error variance and σ_{nv}^2 represents natural variability. Based on (4) it is seen that the precision of the trend deduced from data depends on observation error (σ_{obs}^2), the length of data (N) and the natural variability (σ_{nv}^2) of the variable whose trend is under investigation. The larger the observation error and natural variability, the longer the required data record for an accurate estimate of climate trend.

An example is provided in Figs. 1-2 to show the sensitivity of climate trend calculated from data to measurement precision. Firstly, three monthly mean temperature time series are generated over a 300 hundred years period by adding three different random noises (with σ_{obs}=0.1, 0.3 and 1K) to the same climate trend of 0.2 K/decade (i.e., the truth). The natural variability is assumed σ_{obs}=0.1 K. The trends calculated by (2) from these three time series with varying lengths of data are presented in Fig. 1. It is seen that the true trend of 0.2 K/decade could be deduced from a shorter time series when observation error is smaller (σ_{obs}=0.1–0.3K). When observation error is increased to 1K, a much longer time series of data is required to deduce the decadal trend. The precision for the trend estimate (i.e., σ_{trend} in (4)) is shown in Fig. 2. It is indicated that the measurement precision has a significant impact on climate trend detection. As a consequence, the required data length increases from 82 years when σ_{obs}=0.1 K to 140 years when σ_{obs}=0.3 K, and 300 years when σ_{obs}=1 K, for detecting a climate trend on the order of 0.2 K/decade with the same precision of about σ_{trend}=1.5×10^{-5}K.

It is thus concluded that providing measurement precision along with data is extremely important for climate trend detection. In terms of obtaining and applying data for climate study, it is also important to obtain from data providers the measurement bias and precision so that the data length required for reliable climate trend detection can be estimated.

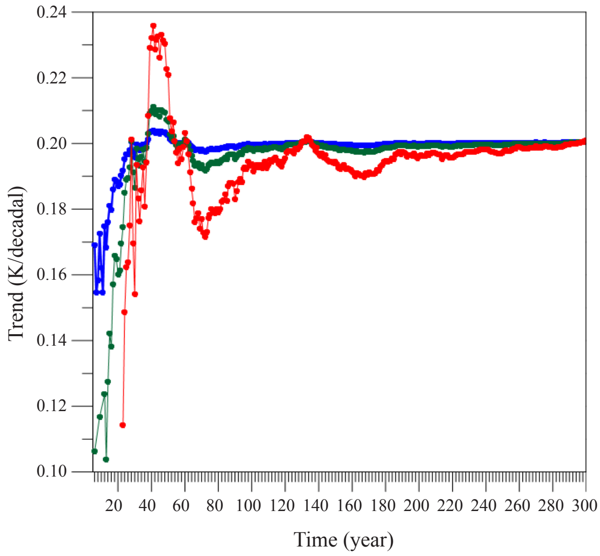

Fig. 1 Climate trend calculated from different lengths of time series with three different observation error variances: 0.1K (blue line), 0.3K (green line) and 1K (red line)

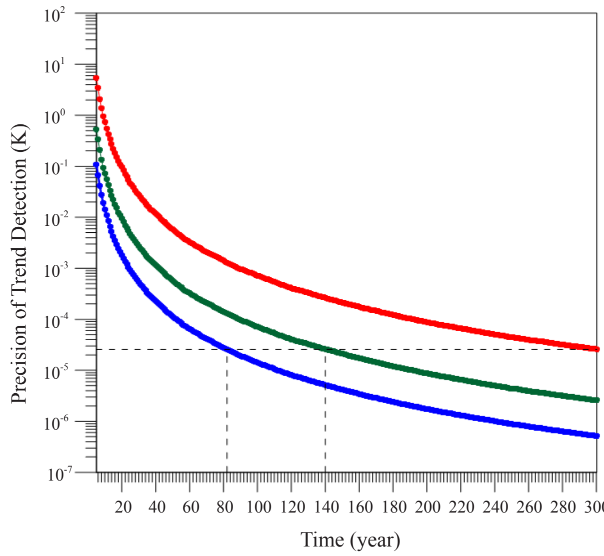

Fig. 2 Variations of σ_{trend} with respect to data length for the trends shown in Fig. 1