第 44 卷 第 3 期
2003 年 8 月

天 文 学 报
ACTA ASTRONOMICA SINICA

Vol.44 No.3
Aug., 2003

时间序列数字滤波后自由度的确定 —— 应用于日长变化与南方涛动指数的相关分析*

闫昊明 钟 敏 朱耀仲

(中国科学院测量与地球物理研究所动力大地测量学重点实验室 武汉 430077)

摘要 原始观测时间序列进行数字滤波后，其自由度将大大降低. 对各种数字滤波器都可以通过 Monte Carlo 方法模拟获得对应置信水平的相关系数临界值，从而得到时间序列滤波后的自由度. 如果滤波后的频段宽度与 Nyquist 频率之比为 Z, 那么对于理想的单边和双边滤波器，时间序列滤波后的自由度为原始值的 $Z/2$ 和 Z 倍，非理想滤波器一般不符合此规律.

关键词 **天体测量学**， Monte Carlo **方法**，**滤波**，**日长**

中图分类号： P 127

1 引言

在天文数据处理过程中，经常要用到两维和多维时间序列的相关分析，其中包括时滞互相关分析 [1−4], 从而揭示时间序列的相依关系和相关程度. 在相关分析中，经常要用到数字滤波. 采用数字滤波技术可以分离出所需频域段的信息，这对原始观测时间序列进行分析提供了较好的依据. 此外，对相关分析的统计结果进行相应的检验也是极其重要的. 数字滤波与自由度之间的关系比较复杂，滤波后原始观测时间序列的自由度将大大降低，其相应置信水平的相关系数临界值也将提高. 20 世纪 80 年代开始流行对原始观测时间序列加窗，从而减少谱泄漏，提高谱的稳定性，这种方法同样会降低原始观测数据的自由度. 因此，更好地解决时间序列数字滤波后自由度的确定，是一个有实际应用价值的问题. 这一问题，可以通过 Monte Carlo 方法模拟来精确的解决. 此方法简单、实用，对科学研究有一定的作用.

2 采用 Monte Carlo 检验法估计相关系数临界值

两个时间序列 $x_1(t)$ 和 $x_2(t)$ 的时滞互相关函数 $\rho(\tau)$ 可以由下式定义 [5]:

2002-11-20 收到原稿， 2003-01-06 收到修改稿

* 中国科学院院知识创新主要方向项目（ KZCX2-106 、 KZCX3-SW-132 ）资助

$$\rho(\tau)=\frac{\mathrm{Cov}(x_1,x_2)}{[\mathrm{Var}(x_1)\cdot\mathrm{Var}(x_2)]^{1/2}},$$

式中， $\mathrm{Cov}(x_1,x_2)$ 为两个时间序列在延迟为 τ 时刻的互协方差函数， $\mathrm{Var}(x)$ 为时间序列 x 的方差函数， $\rho(0)$ 被称为相关系数. 在相关分析时，需要进行显著性检验，以确定相关系数是否超过了给定显著水平上的临界值.

在相关分析中，还经常遇到多个时间序列之间的相关问题. 例如，太阳的耀斑可能与黑子面积、半球面黑子相对数、日面综合谱斑指数等 10 多个因素有关，此时可以通过多元相关分析来确定其相关性. 同理，此时也要进行显著性检验.

近年来，随着计算机的运算速度飞速提高， Monte Carlo 方法得到了广泛应用. 采用 Monte Carlo 方法来估计相关系数的临界值，是一个简便、易行且样本量大，可以重复计算的新方法. 应用 Monte Carlo 方法计算给定置信水平上相关系数临界值的具体步骤可以参见文 [4].

样本容量为 40 万时，采用 Monte Carlo 方法模拟置信水平分别为 90% 和 95% 的两个时间序列和多元时间序列 (三元和四元) 的相关系数临界值与文 [6] 附表 2 中的结果相差不到 0.01, 符合得很好. 这些临界值的标准偏差最大为 ±0.005, 说明运算结果是正确和稳定的.

3　数字滤波与自由度 DOF

数字滤波是提取观测资料中的有用信号，抑止测量噪声的一种平滑手段 [6]. 在天文测量数据处理中，应用比较多的数字滤波是 Vondark 滤波和 FFT 滤波.

对原始时间序列进行滤波后，得到的是部分有用信号，此信号的自由度将比原自由度大大降低. 自由度与滤波之间的内在联系是通过 Monte Carlo 方法来解决的.

在进行 Monte Carlo 方法模拟滤波与自由度之间的关系时，与计算相关系数临界值近似，不同的是需要对生成的原始正态随机序列进行相应的滤波处理. 在求出滤波后的时间序列的相关系数后，通过查相关系数给定水平上的临界值表，就可以得到对应的自由度 DOF.

3.1　Vondark 滤波与自由度 DOF

三阶 Vondark 滤波器的频率响应函数 $F(\varepsilon,T)$ 解析表达式为 [6,7]:

$$F(\varepsilon,T)=\left(1+\frac{1}{\varepsilon}\left(\frac{2\pi}{T}\right)^6\right)^{-1},$$

式中， ε 为平滑因子， T 为响应周期. 设 $F(\varepsilon,T)=A,(0<A<1)$, 则可以推出：

$$\varepsilon=\left(\frac{2\pi}{T}\right)^6\cdot\frac{A}{1-A}\quad.\tag{1}$$

$$T=2\pi\cdot\left(\frac{(1-A)\varepsilon}{A}\right)^{-1/6}.\tag{2}$$

3.1.1 高通滤波和低通滤波

高通滤波器的特点是保留观测数据中的高频信号. 设 T_c 为截断周期, 采样间隔为 ΔT, 根据公式 (1) 可以找到 $F(\varepsilon_h, T_c) = 0.01$ 时对应的平滑因子 ε_h, 然后采用此 ε_h 进行数字滤波. 此时, 滤波后数据自由度 DOF 的解析表达式为:

$$\mathrm{DOF} = \left(1 - \frac{\Delta T}{T_c}\right) \cdot (N-2)\,, \tag{3}$$

式中 N 为原始观测时间序列数据总数.

低通滤波器的特点是保留观测数据中的低频信号. 对低通滤波, 首先根据公式 (1) 找到 $F(\varepsilon_l, T_c) = 0.99$ 时对应的平滑因子 ε_l, 再采用 ε_l 进行数字滤波. 同时, 为给出数字滤波与自由度之间的解析关系, 还需要在计算出 ε_l 后, 利用 (2) 式求出 $F(\varepsilon_l, T_l) = 0.01$ 时所对应的低频零响应周期 T_l. 此时, 滤波后数据自由度 DOF 的解析表达式为:

$$\mathrm{DOF} = \frac{\Delta T}{T_l} \cdot (N-2)\,. \tag{4}$$

从表 1 中的模拟结果可以看出, Vondark 高通和低通数字滤波器确实存在公式 (3) 和 (4) 所表述的关系.

表 1 Monte Carlo 方法模拟的 Vondark 高通和低通滤波后相关系数的临界值和自由度 DOF

Table 1 Critical correlation coefficients and DOF of high and low pass filtered data for Vondark filter simulated by Monte Carlo method

Method	N	ΔT	T_c	ε	Critical correlation Coefficients at 90% Confidential levels	Critical correlation Coefficients at 95% Confidential levels	DOF
	702	30	210	7.2×10^{-12}	0.067	0.080	600
High pass	502	30	150	5.5×10^{-11}	0.083	0.098	400
	1002	30	120	2.1×10^{-10}	0.060	0.072	750
	3002	30	180	1.8×10^{-11}	0.075	0.089	500
Low pass	1602	30	120	2.0×10^{-10}	0.084	0.100	400
	402	30	120	2.0×10^{-10}	0.167	0.199	100

值得注意的是, 当高通和低通零响应周期 $T_l = 2\Delta T$ 时, 称为极限滤波. 当 $T_l < 2\Delta T$ 时, 其频率小于 Nyquist 频率, 滤波无实际意义, 不能再应用公式 (3) 和 (4).

3.1.2 带通滤波

带通滤波器的特点是使观测数据中的某一特定频带的信号被保留. 设 T_{c1} 和 T_{c2} 分为较低和较高的截断周期, 采样间隔为 ΔT. 从 3.1.1 给出的解析表达式, 似乎可以推断带通滤波后数据的自由度应为 $\mathrm{DOF}_{\mathrm{band}} = \left(\frac{\Delta T}{T_{c1}} - \frac{\Delta T}{T_{c2}}\right) \cdot (N-2)$, 但实际情况并非如此. 实际带通滤波后的自由度 DOF 在 $\mathrm{DOF}_{\mathrm{band}}$ 和 $2\times\mathrm{DOF}_{\mathrm{band}}$ 之间, 且无解析规律可以遵循. 从表 2 的结果可以看出, 对于带通数据的自由度 DOF 比 $\mathrm{DOF}_{\mathrm{band}}$ 的结果要大些, 对应于表 2, 其比例因子分别为 1.67 、 1.3 和 1.3. 为应用上的方便, 在带通滤波时, 可选

自由度 $\mathrm{DOF}_{\mathrm{band}}$ 作为滤波后的自由度，但注意此时所对应的相关系数临界值会比实际置信水平偏高．如果需要准确确定自由度，则需采用 Monte Carlo 方法进行实际模拟．

造成带通滤波后的自由度 DOF 大于 $\mathrm{DOF}_{\mathrm{band}}$ 的原因可作如下解释：设 F_1 和 F_2 分别为对应的低通滤波后的数据，则在理想的滤波情况下，带通的结果应为 $\Delta F = F_1 - F_2$，且其自由度应为 $\mathrm{DOF}_{\mathrm{band}}$．但由于 Vondark 滤波器不是理想情况，实际结果为 $\Delta F = F_1 - F_2 + \Delta$，此处的 Δ 对应于 F_2 中低通时在频率响应为 (0 – 1) 频率区间的值与 F_1 中同频率值的残差．由于此残差值与 F_2 的滤波因子有关，所以具有不确定性．但带通滤波的一个共同特点是：由于此残差值的存在，使带通滤波后的自由度略大于 $\mathrm{DOF}_{\mathrm{band}}$ 的值．在实际应用中，当截断周期远大于采样间隔 (例如，年际变化) 时，可用 $\mathrm{DOF}_{\mathrm{band}}$ 的值来近似代替滤波后数据的自由度．

表 2　Monte Carlo 方法模拟的 Vondark 带通滤波后相关系数的临界值和自由度 DOF

Table 2　Critical correlation coefficients and DOF of band pass filtered data for Vondark filter simulated by Monte Carlo method

Method	N	ΔT	T_{c1}	T_{c2}	Critical correlation Coefficients at 90% Confidential levels	Critical correlation Coefficients at 95% Confidential levels	DOF
	362	30	120	180	0.220	0.262	50
Band pass	362	30	60	180	0.132	0.158	160
	602	30	210	420	0.225	0.266	55

表 3　Monte Carlo 方法模拟的 FFT 滤波后相关系数的临界值和自由度 DOF

Table 3　Critical correlation coefficients and DOF of filtered data for FFT filter simulated by Monte Carlo method

Method	N	ΔT	T_{c1}	T_{c2}	Critical correlation Coefficients at 90% Confidential levels	Critical correlation Coefficients at 95% Confidential levels	DOF
High pass	602	5	30	—	0.082	0.098	400
	602	5	60	—	0.074	0.088	500
Low pass	602	5	30	—	0.117	0.138	200
	602	5	300	—	0.377	0.443	20
Band pass	602	5	30	120	0.134	0.158	150
	602	5	300	400	0.645	0.725	5

3.2　FFT 滤波与自由度 DOF

在天文数据处理中另一种滤波方法是频域内 FFT 滤波．这种滤波方法比较简单实用．设低通和高通的截止周期为 T_{c1}，带通滤波的截止周期分别为 T_{c1} 和 $T_{c2}(T_{c1} < T_{c2})$，数据总数为 N，采样间隔为 ΔT．通过模拟实验 (表 3)，对于 FFT 滤波后数据的自由度有：

(1) 对于低通滤波：　$\mathrm{DOF} = \dfrac{2 \cdot \Delta T}{T_{c1}}(N-2)$;

(2) 对于高通滤波：　$\mathrm{DOF} = \left(1 - \dfrac{2 \cdot \Delta T}{T_{c1}}\right)(N-2)$;

(3) 对于带通滤波： $\mathrm{DOF} = 2 \cdot \left(\dfrac{\Delta T}{T_{c1}} - \dfrac{\Delta T}{T_{c2}} \right) (N-2)$.

与 Vondark 滤波比较可知，采用 FFT 滤波后的数据自由度是相应 Vondark 滤波的 2 倍. 原因为， FFT 是双边滤波，在正频和负频区域都有贡献；而 Vondark 滤波是单边滤波. 由于 FFT 是理想的滤波器，所以其带通滤波后数据的自由度有准确的解析表达式.

4 应用举例

为了更好地说明时间序列滤波后自由度的确定是在相关分析中应用的, 在此给出日长变化与南方涛动 SOI 之间的相关分析. 关于二者之间相关的性质, 很多学者进行过研究, 可见文 [2, 3, 8]. SOI 数据来自互联网 http://www.cpc.ncep.noaa.gov/data/indices/index.html 中的相关数据，日长数据取自 COMB2000. 数据的时间索引从 1970 年初到 2000 年底，共 372 个月均值数据. 为了获得 2 至 7 年的年际信号，对原始数据进行 Vondark 带通滤波，滤波因子分别为 1.0×10^{-12} 和 1.0×10^{-17}. 根据第三节的理论，可以用 Monte Carlo 方法模拟出滤波后数据的自由度为 $\mathrm{DOF}_{\mathrm{band}} = 32$.

滤波后的数据首先经过标准化处理, 然后对两列数据进行时间域内的时滞相关分析, 结果见图 1. 从图 1 中可以看出，如果不考虑滤波后数据自由度的变化，其 95% 的置信水平为图中长虚线所示；而考虑自由度变化后的 95% 置信水平为图中短虚线所示，二者的差别是明显的. 如果不考虑自由度的变化，则会导致给出的置信水平偏低，从而可能会导致错误的结论. 从图 1 中还可以看出，二者的最大相关在时间延迟一个月处，其相关系数为 0.616, 此值超过了 95% 和 99% 置信水平的相关系数临界值 0.342 和 0.437, 说明日长和 SOI 之间存在显著相关.

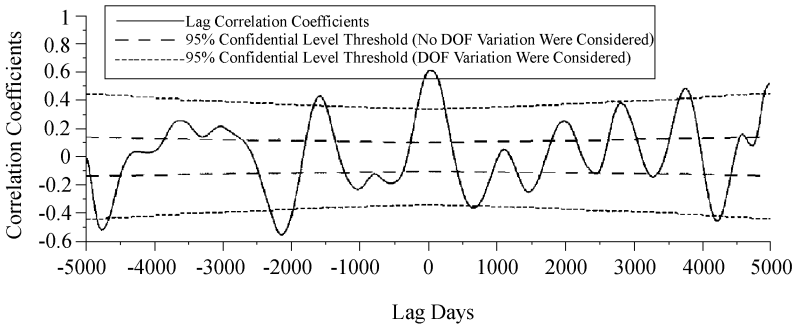

图 1 日长和 SOI 之间的时滞相关系数 (实线) 和 95% 置信水平的相关系数临界值 (考虑自由度变化 (短虚线), 未考虑自由度变化 (长虚线))

Fig. 1 Time lagged correlation coefficients between Length of Day (LOD) variation and SOI (solid line) and 95% confidential level threshold critical correlation coefficients (DOF changes were considered (short dash line), no DOF changes were considered (long dash line))

5 结论

通过研究两种不同滤波器滤波后数据的自由度与滤波参数之间关系的实验，不难得

出，对于滤波参数与滤波后的自由度之间确实存在内在联系．应用 Monte Carlo 方法可以解决滤波后数据的自由度 DOF 的确定．根据截断频率为截断周期的倒数这一关系，设滤波后的频段宽度与 Nyquist 频率之比为 Z, 则对于理想的双边滤波器和单边滤波器，可得滤波后的自由度为原始数据的 Z 和 $Z/2$ 倍，非理想滤波器一般不符合此规律．对于非理想情况下的滤波器，需要进行 Monte Carlo 模拟，以确定精确的自由度 DOF 的值．由于 Monte Carlo 方法比较简单，进行这一工作是必要的．最后，通过日长和 SOI 相关的例子，说明了只有在考虑滤波后自由度变化的情况下的置信水平才可信，如不考虑自由度的变化，可能会导致错误的结论．总之，时间序列数字滤波后的自由度可以通过 Monte Carlo 方法确定．

致谢 感谢周永宏博士在本文的完成过程中给予的帮助！

参 考 文 献

1 Chao B. J Geophys R, 1988, 93: 7709-7715
2 郑大伟，周永宏，廖新浩等．中国科学， 2000, 30: 946-952
3 韩延本，赵娟，李志安．科学通报， 2001, 46: 1858-1861
4 周永宏．博士学位论文，中国科学院上海天文台， 1997
5 杨位钦，顾岚．北京：北京理工大学出版社， 1988
6 丁月蓉．天文测量数据的处理方法．南京：南京大学出版社， 1990
7 HUANG Kuen-Yi, ZHOU Xiong. AcASN, 1981, 22: 120-130
8 ZHONG Min, ZHU Yao-Zhong, GAO Bu-Xi. AcASn, 1999, 40: 101-106

The Determination of Degrees of Freedom for Digital Filtered Time Series ——An Application in the Correlation Analysis between Length of Day Variation and SOI

YAN Hao-Ming ZHONG Min ZHU Yao-Zhong
(*Key Laboratory of Dynamic Geodesy, Institute of Geodesy and Geophysics, CAS, Wuhan* 430077)

Abstract The degree of Freedom (DOF) of digital filtered time series will be decreased rapidly. To determine the DOF of a filtered data series by different kinds of filters, Monte Carlo method can be applied and the critical correlation coefficients at different confidential levels can be simulated correctly. Assuming that the ratio between filtered frequency band and Nyquist frequency band is Z, the DOF of the filtered data series is Z and $Z/2$ times as that of its origin data series respectively, for a one-way and a two-way perfect filter. But for a non-perfect filter, this rule is not suitable.

Key words astrometry, Monte Carlo method, Filtering, length of day