

计算条件对 Hurst 指数稳定性影响分析

赵翠萍, 王海涛

(新疆维吾尔自治区地震局, 新疆 乌鲁木齐 830011)

摘要: 基于随机 poisson 分布模型和 Gutenberg-Richter 关系, 构造出 20 年尺度内 8 组样本量从 1 000 到 15 000 的模拟地震目录数据, 以讨论样本量条件、最大拟合窗长 τ_{\max} 等计算条件对地震活动频次数列 H 值计算结果的影响. 结果表明, 随着样本量的增加, H 值整体精度及稳定性增大. 当 $\tau_{\max} = 300$ 时, 表现尤为明显. 实际震例计算结果也表明, 通过合理选取计算条件, 可以较好地提取强震前 H 值的异常变化过程.

关键词: R/S 分析; Hurst 指数 H 值; 样本条件; 最大拟合窗长; 模拟地震

中图分类号: P315.63 **文献标识码:** A **文章编号:** 1000-0844(2002)01-0034-08

0 引言

Hurst 指数 H 值是描述时间结构自仿射分形特征的量, 与分维密切相关, 可以看作时间序列极差 (R)、标准差 (S) 结构的分维. 对某一时间序列进行非线性时间序列 R/S 分析即可得到 Hurst 指数. R/S 分析是英国水文学家 Hurst 于 1965 年提出的时间序列分析方法, 全名为“改变尺度范围的分析”(Rescaled range analysis)^[1]. 随着 20 世纪 80 年代以来非线性科学在地震科学中应用的发展, R/S 分析以其能够描述非线性时间序列复杂变化内在规律的优势, 在地震学中也有了重要而广泛的应用. 1997 年刘杰等^[2] 对全球、中国大陆等地区进行不同震级档次地震频度的 R/S 分析, 结果表明地震的发生在大小空间尺度之间和不同震级层次之间都具有相似的统计规律性, 地震的时间分布是自仿射分形的. 陈凌等^[3] 把 R/S 分析应用于地震能量、地震年频度、地震年发生率等分析研究中, 将小尺度范围内统计特性变化的规律外推到大的时间尺度范围. 刘长海^[4]、王碧泉^[5]、陈佩燕^[6] 和王林瑛^[1] 等人把 R/S 分析应用于地震预报研究, 以固定的计算窗长 (τ_{\min} 、 τ_{\max}) 对一个地区地震活动的能量、地震时间间隔、震级、 b 值等地震活动性指标沿时间进程滑动计算其 H 值, 从而寻找中强地震前 H 值的异常变化信息结果表明显著不同于随机预报. 作者对中国大陆 7 级强震震例以及新疆地区系统性的研究结果表明 H 值在中强震前的异常特征为“下降—低值—回升”, 地震往往发生在 H 值回升的过程. 滑动 H 值的时间进程能描述地震时间序列的演化特征^[2].

在以往地震预报实践和有关震例研究中, 人们已经发现大部分基于地震目录的地震预报方法在不同程度上显示出计算结果对样本时空窗尺度、震级区间、样本量等条件的敏感性^[7]. 应用 R/S 分析滑动计算地震活动时间序列的 Hurst 指数 H 值, 同样主要以地震目录为基础资

收稿日期: 2001-05-23

作者简介: 赵翠萍(1967—), 女(汉族), 河南许昌人, 助理研究员, 主要从事地震活动性及地震综合预报研究工作.

① 王林瑛, 等. 各种非线性地震学预报指标与地震活动性指标的相互关系及中短期预报效能的研究. “九五”地震科技公关项目, 1999.

② 赵翠萍. 非线性时间序列 R/S 分析及其在新疆地震预报中的应用研究[D]. 硕士学位论文, 2001.

料进行. 基于每个滑动时间窗内极差 (R) / 标准差 (S) 序列, H 值最终由最小二乘法拟合实现. 因此计算中样本量及最大拟合窗长 τ_{\max} 是影响 H 值计算结果整体精度及稳定性的主要因素. 在以往工作中我们可以发现由于最大拟合窗长过小, H 值时间进程涨落频繁, 极不稳定, 造成 H 值异常的可信度较低, 虚报率高. 本文在时间进程遵从 poisson 分布和震级分布遵从 G-R 关系的基础上, 模拟产生以 20 年为时间尺度的 8 组地震目录数据 (样本量从 1 000 到 15 000), 讨论不同样本量条件下不同最大窗长 τ_{\max} 值对 H 值计算结果稳定性的影响, 从而为在实际地震时间序列资料的 R/S 分析中, 合理选取计算参数提供了依据. 目的旨在提高 Hurst 指数的预报效能.

1 R/S 分析原理

R/S 分析的基本思想是改变所研究的时间尺度的大小, 研究其统计特性变化的规律, 从而可以将小时间尺度范围的规律用于大的时间尺度范围, 或者将大的时间尺度得到的规律用于小的时间尺度, 显示整体与部分之间规律的相似性, 这正是分形几何的核心思想. 其原理如下:

考虑一个时间序列 $\{\xi(t)\}$, $t = 1, 2, \dots$. 对于任何正整数 τ , 定义均值系列

$$\bar{\xi}(\tau) = \frac{1}{\tau} \sum_{t=1}^{\tau} \xi(t) \quad (1)$$

用 $X(t)$ 表示累积离差

$$X(t, \tau) = \sum_{n=1}^t \{\xi(n) - \bar{\xi}(\tau)\} \quad 1 \leq t \leq \tau \quad (2)$$

极差 R 定义为:

$$R(\tau) = \max_{1 \leq t \leq \tau} X(t, \tau) - \min_{1 \leq t \leq \tau} X(t, \tau) \quad \tau = 1, 2, \dots \quad (3)$$

均方差 S :

$$S(\tau) = \left(\frac{1}{\tau} \sum_{t=1}^{\tau} \{\xi(t) - \bar{\xi}(\tau)\}^2 \right)^{\frac{1}{2}} \quad (4)$$

Hurst 发现许多自然现象所构成时间序列的极差与方差之比 (R/S) 满足下式:

$$R/S \sim (a\tau)^H \quad (5)$$

若以 R/S 描述时间序列的统计特征, 则表明该时间序列具有自仿射统计分形特征, R/S 表现出自仿射分形结构, 其中 a 为统计常数. H 即是 Hurst 指数, 可看作是极差与标准差序列的自仿射分形结构的维数. 对上式两边取对数, 有:

$$\lg(R/S) \propto H \lg \tau + H \lg a \quad (6)$$

由此对一定时间序列用足够的样本量经最小二乘法拟合即可得到 Hurst 指数 H 值.

2 模拟地震数据的生成

地震发生的时间过程可以用多种分布模型进行模拟^[8~9]. 本文以常用的稳态 poisson 分布模型和 G-R 关系进行地震数据的生成^[10~11].

设 $N(t)$ 表示在 $[0, t]$ 区间内出现的事件的数目, $N(t_1, t_2) = N(t_1) - N(t_2)$ 表示在 $[t_1, t_2]$ 内出现的事件数. 当过程分布满足 poisson 分布模型的条件时, 发生在 $[0, t]$ 区间内的事件数目的概率分布为:

$$Pr(n) = Pr\{N(t) = n\} = \frac{(\lambda t)^n}{n!} e^{-\lambda t} \quad (7)$$

根据式(7),即可以生成不同样本量的模拟随机地震时间分布序列.

地震序列中震级的确定在 G-R 关系的基础上进行.首先对 G-R 关系,即

$$\lg N(\geq M) = a - bM \quad (8)$$

进行概率语言描述.

设地震震级为 M_0 , 则式(8)可改写为

$$\lg N(\geq M) = a - b(M - M_0) \quad (9)$$

于是震级大于 M 的地震频次为

$$N(\geq M) = 10^{a-b(M-M_0)} = A e^{-\beta(M-M_0)} \quad (10)$$

式中: $A = 10^a$, $\beta = b \log 10$. 式(10)两边同除以地震的总数目 N_T , 有

$$\frac{N(\geq M)}{N_T} = \frac{A}{N_T} e^{-\beta(M-M_0)} = A' e^{-\beta(M-M_0)} \quad (11)$$

当 $N_T \rightarrow \infty$ 时,

$$\lim_{N_T \rightarrow \infty} \frac{N(\geq M)}{N_T} = Pr(N \geq M) = A' e^{-\beta(M-M_0)} \quad (12)$$

将 $Pr(N \geq M_0) = 1$ 代入上式, 有 $A' = 1$, 故

$$Pr(N \leq M) = 1 - Pr(N \geq M) = 1 - e^{-\beta(M-M_0)} \quad (13)$$

将式(13)式对 M 求导, 则得到地震震级的概率密度函数为指数分布, 即

$$f(M) = \begin{cases} \beta e^{-\beta(M-M_0)} & M \geq M_0 \\ 0 & \text{其它} \end{cases} \quad (14)$$

考虑到震级上限 (M_1) 的限制, 则地震样本的震级分布在 $[M_0, M_1]$ 区间内, 于是对应于式(14)的震级概率密度取为截断的指数分布函数

$$f(M) = \begin{cases} \alpha e^{-\beta(M-M_0)} & M_0 \leq M \leq M_1 \\ 0 & \text{其它} \end{cases} \quad \alpha = \frac{\beta}{1 - e^{-\beta(M_1-M_0)}} \quad (15)$$

根据式(7)和式(15)式即可生成一定时间尺度内不同样本量的随机地震序列. 计算时取 b 值取为 0.75. 为了与利用实际地震资料的 H 值计算结果进行对比, 将模拟地震数据分布在 20 年(如 1980~1999 年)的时间窗内, 共生成 8 组样本供分析使用.

3 模拟地震时间序列的计算与处理分析

首先对每种样本的地震目录以固定的 30 天为窗长, 5 天为滑动步长进行频次 N 值的计算. 再对频次时间序列分别以 6 个不同的最大窗长 τ_{\max} 值逐步滑动计算 H 值. 共计算得到 $8 \times 6 = 48$ 个不同样本量、不同最大窗长 τ_{\max} 的 H 值时间进程.

为了反映 H 值计算结果的整体精度及平稳性, 表 1 中列出了每种样本量及最大窗长的 H 值时程中每个拟合 H 的相关系数分别大于 90%、95%、97% 的比例 p_1 、 p_2 、 p_3 , 以及每个 H 时间进程的离差系数 C_v 值^[2]、 H 的均值. 由于篇幅所限, 在图 1 中仅展示其中几幅时程曲线结果.

分析表 1 及图 1, 可以得到以下几点认识:

(1) 在不同样本量条件下, H 值的总体精度随着拟合窗长 τ 的增大而明显增高. 当窗长大于等于 300 时, 各条 H 值时间进程中每个拟合 H 值时间进程的相关系数大于 90% 的比例已达 97% 以上, 多数甚至达到 100%. 相关系数大于 95% 的比例除样本数为 1 000 的各组外, 其余均在 75% 以上. 结果显示窗长是影响 H 值稳定的主要原因.

表 1 8 种模拟地震样本情况下不同拟合窗长对频次计算 H 值精度影响分析

N	τ	P_1	P_2	P_3	C_v	H
1 000	50	0.763	0.531	0.371	0.320	0.770
	100	0.951	0.692	0.394	0.193	0.754
	200	0.975	0.707	0.361	0.159	0.685
	300	0.998	0.657	0.232	0.132	0.618
	400	1.00	0.732	0.139	0.105	0.587
	500	1.00	0.779	0.234	0.087	0.576
1 500	50	0.945	0.693	0.454	0.169	0.786
	100	0.967	0.764	0.513	0.148	0.711
	200	0.983	0.803	0.479	0.161	0.651
	300	0.969	0.869	0.389	0.160	0.604
	400	0.999	0.804	0.423	0.144	0.579
	500	1.00	0.695	0.347	0.127	0.558
3 000	50	0.949	0.728	0.492	0.192	0.793
	100	0.963	0.715	0.458	0.175	0.734
	200	0.945	0.739	0.437	0.195	0.686
	300	0.977	0.748	0.413	0.171	0.669
	400	0.981	0.757	0.410	0.157	0.634
	500	0.981	0.618	0.302	0.145	0.598
4 000	50	0.982	0.797	0.543	0.164	0.820
	100	0.985	0.823	0.527	0.132	0.720
	200	0.967	0.760	0.420	0.171	0.702
	300	0.973	0.780	0.537	0.142	0.687
	400	0.987	0.854	0.392	0.107	0.662
	500	1.00	0.794	0.342	0.094	0.641
6 000	50	0.977	0.844	0.618	0.160	0.832
	100	0.979	0.808	0.555	0.146	0.758
	200	1.00	0.830	0.561	0.150	0.699
	300	1.00	0.889	0.683	0.112	0.710
	400	1.00	0.941	0.739	0.078	0.717
	500	1.00	0.969	0.677	0.090	0.708
8 000	50	0.984	0.831	0.579	0.157	0.833
	100	0.976	0.728	0.382	0.170	0.733
	200	0.985	0.779	0.459	0.151	0.691
	300	1.00	0.824	0.573	0.118	0.73
	400	0.999	0.867	0.545	0.14	0.742
	500	1.00	0.938	0.660	0.13	0.747
10 000	50	0.98	0.85	0.62	0.152	0.83
	100	1.00	0.85	0.44	0.137	0.74
	200	1.00	0.85	0.60	0.127	0.69
	300	0.962	0.87	0.67	0.118	0.70
	400	0.953	0.85	0.66	0.149	0.71
	500	0.986	0.86	0.71	0.142	0.71
15 000	50	0.99	0.88	0.66	0.159	0.86
	100	0.97	0.81	0.53	0.144	0.76
	200	0.98	0.73	0.47	0.125	0.67
	300	0.98	0.72	0.37	0.099	0.645
	400	1.00	0.66	0.33	0.092	0.62
	500	1.00	0.60	0.26	0.115	0.60

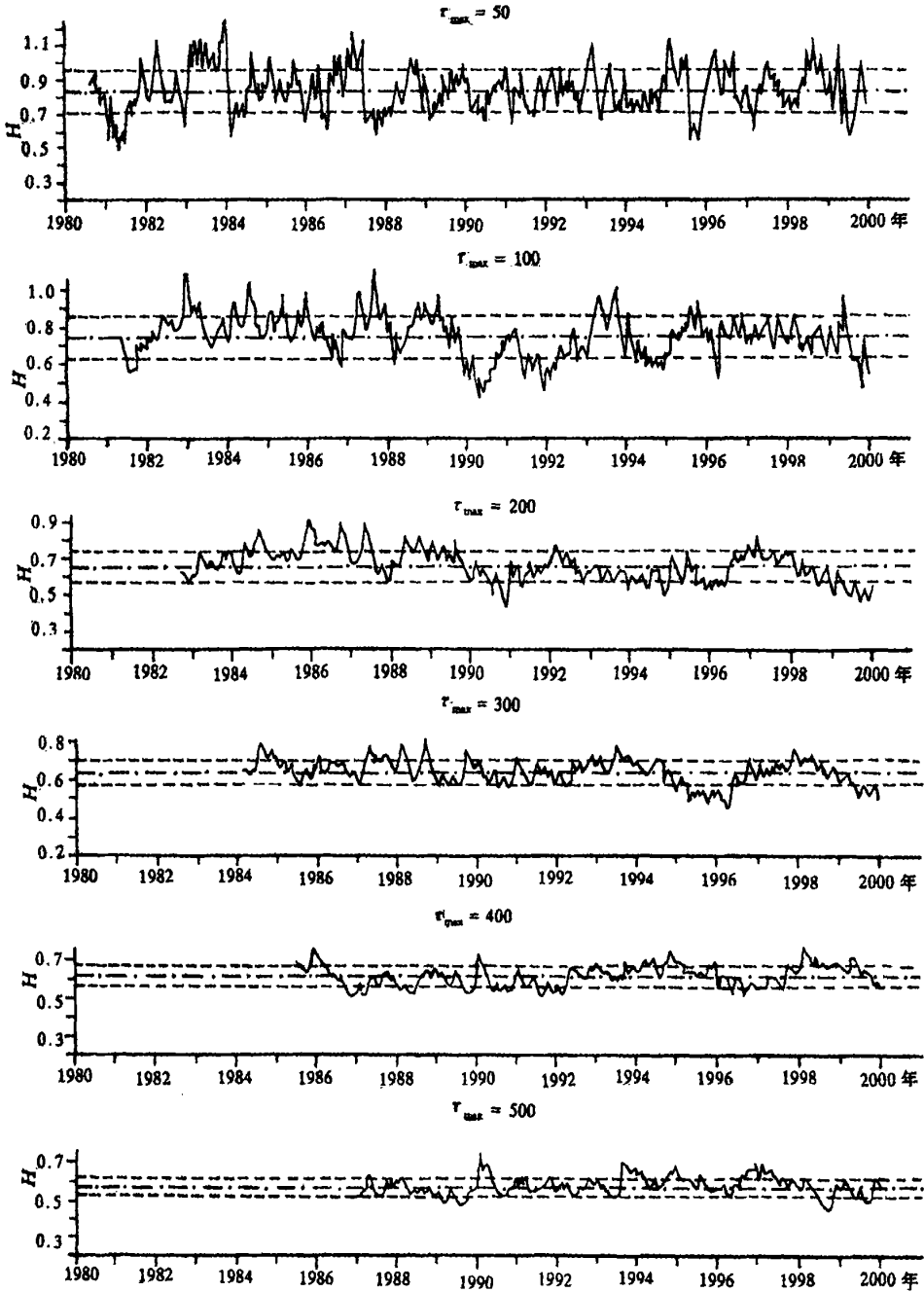


图 1 不同 τ_{max} 时模拟地震数据的 H 值滑动时间进程曲线

Fig. 1 Curves for sliding-time process of H values of simulated earthquakes with different τ_{max} .

(2) 随着样本数的增大, 相同窗长滑动计算的 H 值其精度也明显增大. 当 N 达到 4 000 条以上时, 不同窗长的 $R \geq 0.95$ 的比例已基本达 80% 以上. 当 $N \geq 5 000$ 条时不同窗长的 $R \geq 0.97$ 的比例达 50% 以上.

(3) 离散度 C_v 值随窗长及样本条件的增加而明显减小, 说明 H 值趋于稳定地在均值附近做平稳变化, 起伏幅度变小, 因而出现随机性异常变化的概率减小. 在各种样本条件下, 当 τ_{max}

达 300 时, C_v 值已基本小于 0.15.

(4) 随着拟合窗长的增大, 不但 H 值的时程越平稳, 其均值也系统性地减小, 显示出向其真值的逼近, 即与随机泊松时间过程($H = 0.5$) 偏离程度的减小.

(5) 即使是随机地震数据, 当窗长选择不合适时, H 值仍会随机出现较大幅度的异常变化. 而随着窗长的增加, H 值精度提高, 随机的异常起伏变化逐渐消失.

4 实际地震活动的震例分析

为了进一步说明上述结果的实际意义, 在此以 1996 年 2 月 3 日云南丽江地震为例讨论窗长的改变对计算结果可靠性、稳定性的影响. 对 1980 ~ 1997 年丽江地震研究区($26^\circ \sim 29^\circ \text{N}$, $99^\circ \sim 102^\circ \text{E}$) 内 2.5 级以上地震活动的频次序列进行 H 值的滑动计算. 频次的计算以 30 天为窗长, 5 天为滑动步长. 计算中同样选取上述 6 种不同的最大拟合窗长. 除 1980 年 2 月 22 日 6.1 级地震外, 丽江地震前研究区内未发生 $M_s 6.0$ 以上地震.

由图 2 可见, τ_{\max} 等于 50 和 100 时, H 值时间进程曲线的起伏变化具有较强的随机性; τ_{\max} 等于 200 时, 尽管震前已有一定的异常表现, 但不是很典型; τ_{\max} 等于 300 和 400 时, 1993 年以前 H 值一直在相对较为稳定的背景上变化, 1993 年以后其下降—低值异常过程已较为明显; $\tau_{\max} = 500$ 时, H 值在震前的异常也较为明显, 且震前出现有起伏的回升变化过程. 由表 2 可见, 随着 τ_{\max} 增大, C_v 值迅速减

表 2 丽江研究区不同拟合窗长时 H 值滑动计算结果统计

τ_{\max}	p_1	p_2	C_v	H
50	0.92	0.66	0.22	0.80
100	0.94	0.72	0.18	0.80
200	0.98	0.79	0.15	0.81
300	1.00	0.96	0.11	0.80
400	1.00	0.98	0.08	0.80
500	1.00	0.99	0.09	0.79

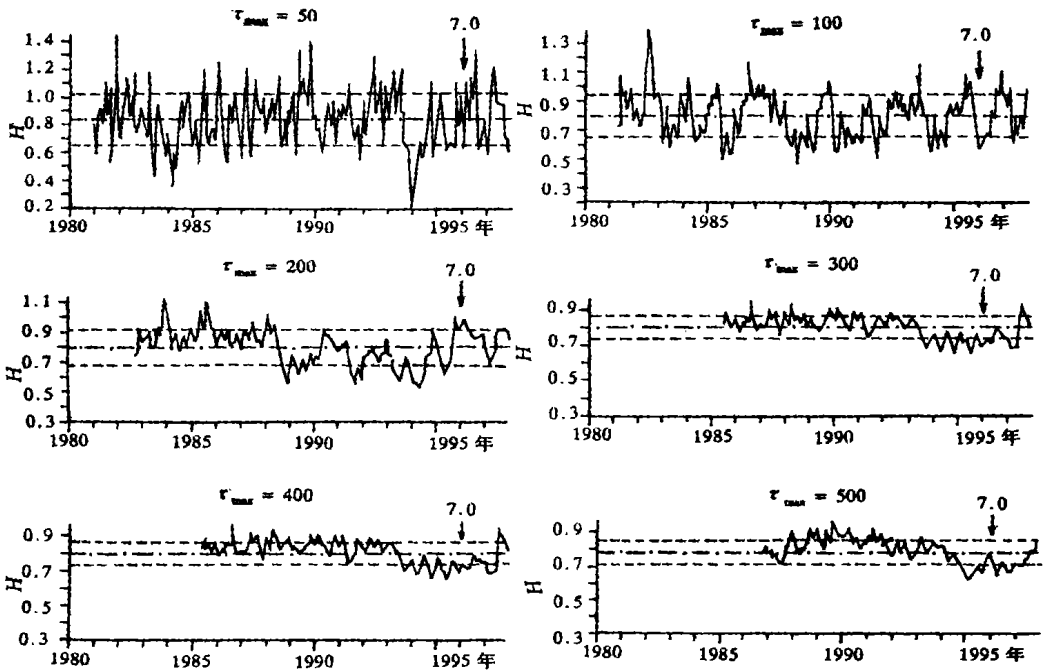


图 2 丽江地震前研究区 H 值滑动时间进程曲线

Fig. 2 Curves for sliding-time process of H value in the research region before Lijiang earthquake .

小, 时间进程曲线越来越平稳. p_1 、 p_2 在 $\tau_{\max} = 300$ 时已达 100%、96%, 说明系统的整体精度明显增加, 且 H 值已较为稳定.

丽江震例的计算结果在一定程度上说明, H 值拟合窗长 τ_{\max} 的不同取值对 H 值的结果具有显著的影响. 拟合窗长的合理选取可以抑制计算结果中随机性的异常变化, 从而突出强震前的异常过程, 有利于利用 H 值进行强震前的中短期预测.

5 结语与讨论

综上所述, 即使是随机地震时间数据, 当样本条件、窗长等计算条件选择不合适时, 仍会随机出现 H 值较大幅度的异常变化, H 值随机涨落频繁. 然而窗口过长也会压制异常信息的出现. 因此在实际应用 R/S 分析滑动计算 H 值时, 首要考虑 H 值的计算拟合窗长的合理选取.

本文在模拟实验的基础上认为, 当 $\tau_{\max} = 300$ 时, H 值显示出较为显著的稳定; 对实际地震时间序列的计算结果显示此时 H 值既具有较高的整体精度, 又能在稳定的背景上突出异常. 另外, 随着样本量的增大, H 值计算精度及稳定性也随之增加.

当然, 在利用实际地震数据进行 R/S 分析滑动计算 H 值时, 合理拟合窗长的选取可以在相应样本量随机地震数据模拟各种窗长的计算结果的基础上得到, 也可以在实际计算中尝试多种窗长以达到最佳的地震对应效果来确定.

朱令人研究员审阅了本文所做的工作, 在此表示感谢!

[]

- [1] , . R/S 分析与分式布朗运动[J]. , 1990, 13(8): 477—482.
- [2] , . [M]. : , 1998.
- [3] , , . [J]. , 1998, 41(): 244—253.
- [4] , , . R/S 分析[J]. , 1995, 17(3): 389—392.
- [5] , , . R/S 方法研究及其在地震预测中的应用[J], , 1995, 17(4): 528—532.
- [6] . H 值用于华北地区的预报方案研究及检验[J]. , 1996, 17(1): 76—80.
- [7] , , . [J]. , 1999, 19(3): 223—227.
- [8] , , . [J]. , 1980, 2(3): 181—293.
- [9] . Weibull 分布研究华北地区前兆震群的特征[J]. , 1992, 18: 13—22.
- [10] , . [M]. : , 1992. 17—115.
- [11] , . [J]. , 1998, 37(1): 48—59.

ANALYZING THE EFFECT ON HURST EXPONENT STABILITY FROM CALCULATIONAL CONDITIONS

ZHAO Cui-ping, WANG Hai-tao

(*Seismological Bureau of Xinjiang Uygur Autonomous Region, Urumqi 830011, China*)

Abstract: Based on random Poisson distributional model and Gutenberg-Richter law, we made 8 groups of simulated seismic catalogues with the sample size of 1 000 to 15 000 within 20 years so that we could discuss the effect on the H value of frequency sequence of seismicity caused by the calculational conditions of sample size and τ_{\max} of maximal fitting window length. The result shows that the whole precision and stability of H value improve with the sample size increasing. When τ_{\max} is equal to 300, it is specially obvious. Case calculations show that it is available to pick up well the process of anomaly change of H value before strong earthquake with selecting reasonable calculational conditions.

Key words: R/S analysis; Hurst exponent; H value; Sample condition; Maximal fitting window length; Simulated earthquake