

信息流合成方法在震情分析研究中的应用^{*}

薄万举

(国家地震局第一地形变监测中心, 天津 300180)

摘要 在八五地震短临预报攻关的研究中, 85-04-02-01 二级课题给出了一种新的信息合成方法——信息流合成方法. 作为 85-04-02 一级课题的继续及九五预研究工作的部分内容是使信息流合成方法进一步得到完善, 并在日常预报工作中进行了应用, 取得了初步的结果. 同时也为将科研与预报更紧密结合起来进行了有意义的探索.

关键词: 信息合成 统计分析 震情 首都圈

1 信息合成前的准备

1.1 前兆数据准备

进行信息合成之前, 要将待合成的多项资料数据按要求进行整理和准备. 九五计划将主攻一年尺度的中短期地震预报方面的科研课题, 本文特选定以每月一个采样值(每月一个观测值、月均值、内插成月均值均可)的数据为研究对象, 即要求单项资料数据序列为每月一个值. 要求这样的观测量具有明确的物理意义, 与孕震过程存在有机的联系, 从理论上认为该序列的变化中包含有地震孕育、发生的信息, 即可参加信息合成, 从而产生信息增益.

本文处理了京津唐地区的大部分流动跨断层的短水准、短基线资料, 共选用了相对观测时段较长的 30 份资料.

1.2 信息序列的提取及无量纲化

进行信息合成, 目的是将多项资料所可能含有的同源信息合成在一起, 以增强可识别信息的清晰度和可信度. 日常预报所用的资料中有水准、基线及其它各类前兆观测资料, 由于其量纲不同, 故首先需要无量纲化后才能进行信息合成.

在无量纲化的研究中, 已有一些比较好的方法, 如模糊从属度函数法等等. 本文给出另一类方法——多点组信息法. 该法用多个相邻的采样点构成一组, 构成一个曲线元, 在时轴上单值滑动, 可形成一个变化的曲线元序列, 包含有用信息的观测量所发生的异常变化必定反映在曲线元形态的变化之中, 而曲线元形态可用含有若干可变参数的曲线元模型来描述, 即可将观测物理量的变化转换成某种无量纲的曲线元参数的变化, 如曲线元的斜率、曲率等.

本文采用四个相邻的采样点构成一个曲线元, 用最小二乘法求出曲线元的线性斜率从而构成一个无量纲化的信息序列^[1].

^{*} 本文得到 85-04-02 课题及九五预研究课题联合资助

收稿日期: 1995-12-21

作者简介: 薄万举, 男, 1957 年 8 月生, 高级工程师, 现从事大地测量与地震预报研究工作.

1.3 信息序列的标准化^[2]

四点组斜率信息法尽管给出了无量纲信息,但不同的资料、地点和时段因存在气压、温度、降水等各种不同的系统性周期干扰因素,其斜率变化的幅度和规律是不一样的,若这一情况不加以考虑,必定会影响其信息合成效果.因此,不但要将不同资料进行无量纲化,还要从这些无量纲序列中去掉系统性干扰成分,并在量级尺度上归化到一个统一可比的尺度上去,我们称之为信息序列的标准化.

设 X_{ij} 表示 i 年 j 月的四点组斜率值,那么 i 年 j 月的四点组斜率标准化异常信息值为:

$$Y_{ij} = \begin{cases} \frac{|X_{ij} - X_j|}{Km(t)} & |X_{ij} - X_j| \leq |Km(t)| \\ 1 & |X_{ij} - X_j| > |Km(t)| \end{cases} \quad (1)$$

式中 K 为经验常数(一般取 2 ~ 3, 本文取 2.5), X_j 为用多年资料求得的 j 月份四点组斜率的期望估计值, 计算公式为:

$$X_j = \frac{1}{n} \sum_{i=1}^n X_{ij} \quad (2)$$

其中 n 为资料长度(a), $m(t)$ 为四点组斜率值标准差的时间函数,因不同季节有不同的噪声源,标准化计算中就要考虑到其随机误差的季节性差异,故 m 随着时间 t 而变化,为此,用下式表示 j 月的标准差值

$$m = \sqrt{\frac{\sum_{i=1}^n (X_{ij} - X_j)^2}{n}} \quad (3)$$

1.4 地震数据的准备

信息合成是有空间概念的,因此需确认某一特定区域的前兆资料对某一特定区域的地震存在有机的联系后才能对这些资料进行信息合成,从而得到对应区域内的孕震信息.基于上述认识,我们选用了京、津、唐地区 30 项流动跨断层短水准短基线资料分别进行了四点组斜率计算和标准化计算,同时还统计了首都圈有前兆观测资料时段内的 5 级以上地震和华北 7 级以上地震以用于信息评价.所统计的地震目录见表 1.

1.5 信息评价

为了证明信息流的有效性,必须有一个评价信息流的方法.对合成前的信息流进行评价,看其是否具有信息价值,而对合成后的信息流进行评价,看合成方法是否能产生信息增益,这部分研究成果见文献[3].

信息评价的具体作法是,首先利用计算机优化筛选出一个最佳警戒值 W ,将 T_0 时段内(本文取 $T_0 = 1$)超过 W 的全体信息值作为一个异常事件(即异常链),将异常事件开始至结束后的 T_0 时段内所有大于 5 级的地震作为一组地震事件(也称地震链),其余时段用 T_0 截取,有地震,则计为一组地震事件,有异常,则计为一次异常事件,这就可以在信息流占有时段内统计出地震链事件个数(l_1)、异常事件个数(l_2)、有异常后 T_0 时段内出现地震的事件个数(与二者对应的事件个数)(l_3).用 L 表示信息流总时段长度,则地震链的自然发生频率、异常链的自然发生频率和二者对应的自然频率(也称虚拟报准频率)分别为:

$$P_1 = l_1 / (L / T_0) \quad (4)$$

$$P_2 = l_2 / (L / T_0) \quad (5)$$

$$P_3 = l_3 / (L / T_0) \quad (6)$$

若 $P_3 > P_1 \circ P_2$, 证明二者实际对应频率大于其巧合概率. P_3 越大, 说明信息流中所含孕震信息越强. 反之, 说明信息价值越低, 或根本不含有该组地震的孕震信息. 为评价信息流的好坏, 我们将信息效益参数定义为:

$$G = \frac{I_3}{I_1 + I_2 + I_3} \circ (1 - P_1 \circ P_2) \quad (7)$$

显然, G 值较全面地反映了信息流价值的高低. 所谓最佳警戒值就是使 G 值最大的 W 值, 用计算机优化筛选即可报出最佳 W 值.

表 1 首都圈 5 级以上及华北 7 级以上地震的统计

地震序号	发震时间	经度(°E)	纬度(°N)	震级(M)
1	1966-03-22	115.03	37.32	7.2
2	1967-03-27	116.30	38.30	6.3
3	1969-07-18	119.24	38.12	7.4
4	1973-12-31	116.33	38.28	5.6
5	1975-02-04	122.42	40.42	7.3
6	1976-07-28	118.11	39.38	7.8
7	1976-11-15	117.50	39.24	7.1
8	1977-05-12	117.48	39.23	6.5
9	1977-11-27	118.01	39.12	5.8
10	1979-03-05	118.34	39.49	5.1
11	1979-09-02	118.23	39.44	5.1
12	1980-02-07	117.54	39.31	5.3
13	1981-08-13	113.25	40.30	5.8
14	1982-10-19	118.59	39.53	5.3
15	1983-04-03	114.47	40.45	5.1
16	1984-01-07	118.47	39.34	5.0
17	1985-04-22	118.46	39.45	5.0
18	1985-10-05	118.26	39.50	5.3
19	1988-08-03	118.39	39.39	5.1
20	1989-06-18	118.20	39.40	5.0
21	1989-10-18	113.50	39.57	5.7
22	1989-10-19	113.50	39.58	6.1
23	1989-12-25	118.48	40.05	5.2
24	1990-07-21	115.50	40.35	5.0
25	1991-03-26	113.50	39.52	6.1
26	1991-05-29	118.18	39.43	5.6
27	1991-05-30	118.18	39.43	5.6
28	1993-09-11	111.34	39.08	5.4
29	1995-10-06	118.50	39.80	5.0

用上面的方法对 30 项跨断层的流动短水准和短基线资料进行了多点组信息标准化处理后, 分别对其标准化信息流的效益进行了判断, 其结果列于表 2. 由表 2 可看出, 绝大部分资料都满足 $P_3 > P_1 \circ P_2$, 说明首都圈跨断层测量资料中是含有该区域地震前兆信息的. 以燕家台斜交水准为例, 其观测值曲线和标准化信息流的图象见图 1.

2 信息合成^[3]

在实践过程中, 信息合成模型也在不断地得到改进, 其模型为:

$$I_{0,t} = \frac{\sum_{i=1}^n \{ G_i \frac{\max\{U'_{i,k}, k \in [t, -T_0, t]\}}{W_i} \}}{\sum_{i=1}^n G_i} \quad (8)$$

其中

$$U'_{i,k} = \begin{cases} U_{i,k} & U_{i,k} < W_i \\ W_i & U_{i,k} \geq W_i \end{cases} \quad (9)$$

(8)、(9)两式中 $U_{i,k}$ 表示第 i 份资料 k 时刻的标准化信息值, W_i 表示第 i 份信息流的最佳警戒值, G_i 表示第 i 份信息流的最佳信息效益参数, n 表示参加信息合成的资料份数, $I_{0,t}$ 表示 t 时刻的信息合成值。

在表 2 的 30 份资料中, 信息效益参数 G 有高有低, 一般来说, G 值越高, 其信息效益越高。但是, 就统计理论而言, 这些参数都是由小子样中统计而得到的, 统计量本身具有一定的偏差, 其信度比较低, 况且各台点资料长度又不一样, 因此, 需要进行信息合成, 以增强其总体信息量的可信度。

分别以 $T_0 = 1.0a$ 和 $T_0 = 0.5a$, 按(8)、(9)两式对表 2 中的 30 份资料进行信息合成, 合成后信息流曲线见图 2。在表 2 中, 去掉一些信息量较低及认为有较明显干扰的资料后, 剩下施庄等 18 份资料, 又分别以 $T_0 = 1.0a$ 和 $T_0 = 0.5a$ 进行了信息合成, 其合成后的信息流曲线见图 3, 合成后的信息流参数 P_1 、 P_2 、 P_3 和 G 分别列入表 3。

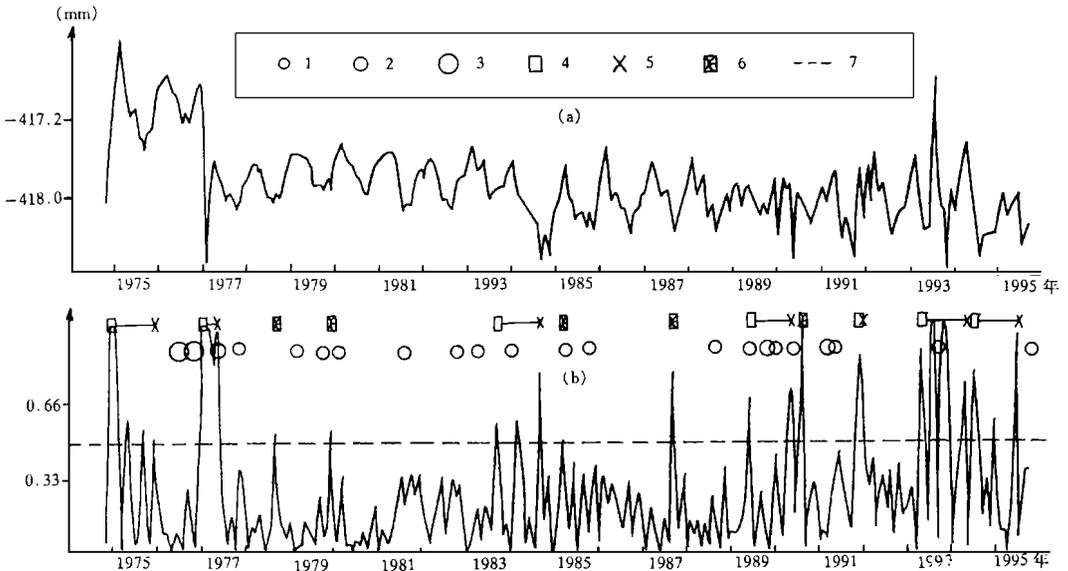


图 1 燕家台斜交水准资料

Fig. 1 Yanjaitai oblique leveling data.

(a) 观测值曲线; (b) 四点组斜率标准化信息流曲线;

- 1 $5.0 \leq M_S < 6.0$; 2 $6.0 \leq M_S < 7.0$; 3 $M_S \geq 7.0$; 4 异常链起点;
5 异常链终点; 6 长度短于两个月的异常链; 7 最佳警戒线

3 分析及讨论

由图 2 可看出, 1987 年以后, 异常出现的频率明显偏高, 而地震出现的频率比较低. 这与首都圈近年来形变异常比较明显而且没有对应地震发生的情况相吻合. 通过图 2(a) 的统计, 在异常链结束的 T_0 时段以内发生地震的事件有 16 次, 未发生地震的事件 21 次, 有地震链之前 T_0 时段内未出现异常的事件 17 次, 总时段长约 28a, 链长为 1a, 则可统计出地震的自然发生概率为 0.63, 出现异常后发生地震的条件概率为 0.76, 当前已发生一次 5 级地震, 通过统计, 认为未来 1a 内再次发生 5 级以上地震的条件概率为 0.50. 用同样方法对图 2 和图 3 的 4 种情况分别进行了统计, 统计结果见表 4.

表 2 信息流效益的评价结果

资料名	P_1	P_2	P_3	G
施庄基线	0.542	0.542	0.387	0.392
施庄水准	0.658	0.619	0.503	0.385
墙子路斜交基线	0.627	0.669	0.460	0.385
墙子路斜交水准	0.627	0.544	0.334	0.296
墙子路垂直基线	0.711	0.418	0.334	0.296
墙子路垂直水准	0.753	0.293	0.209	0.195
燕家台斜交基线	0.669	0.622	0.430	0.292
燕家台斜交水准	0.622	0.574	0.478	0.429
燕家台垂直基线	0.717	0.622	0.430	0.263
燕家台垂直水准	0.622	0.622	0.382	0.273
小水峪斜交基线	0.550	0.641	0.504	0.475
小水峪斜交水准	0.733	0.687	0.504	0.273
小水峪垂直基线	0.773	0.591	0.364	0.198
小水峪垂直水准	0.545	0.682	0.500	0.432
张家台斜交基线	0.706	0.581	0.498	0.372
张家台斜交水准	0.664	0.581	0.415	0.307
张家台垂直基线	0.581	0.623	0.457	0.390
张家台垂直水准	0.581	0.540	0.415	0.404
张山营基线	0.502	0.502	0.386	0.468
张山营水准	0.695	0.463	0.309	0.247
大灰厂斜交基线	0.621	0.583	0.388	0.304
大灰厂斜交水准	0.539	0.719	0.431	0.320
大灰厂垂直基线	0.541	0.469	0.468	0.422
大灰厂垂直水准	0.613	0.721	0.505	0.340
德胜口斜交基线	0.615	0.462	0.308	0.286
德胜口斜交水准	0.615	0.769	0.538	0.335
八宝山斜交基线	0.622	0.622	0.533	0.460
八宝山斜交水准	0.444	0.267	0.267	0.529
八宝山垂直基线	0.636	0.556	0.397	0.323
八宝山垂直水准	0.715	0.477	0.477	0.439

表 3 合成信息流效益参数的统计

项目名称	P_1	P_2	P_3	G	
30 项资料	$T_0 = 1.0a$	0.78	0.63	0.59	0.37
	$T_0 = 0.5a$	0.67	0.37	0.31	0.33
18 项资料	$T_0 = 1.0a$	0.68	0.72	0.64	0.43
	$T_0 = 0.5a$	0.39	0.44	0.26	0.41

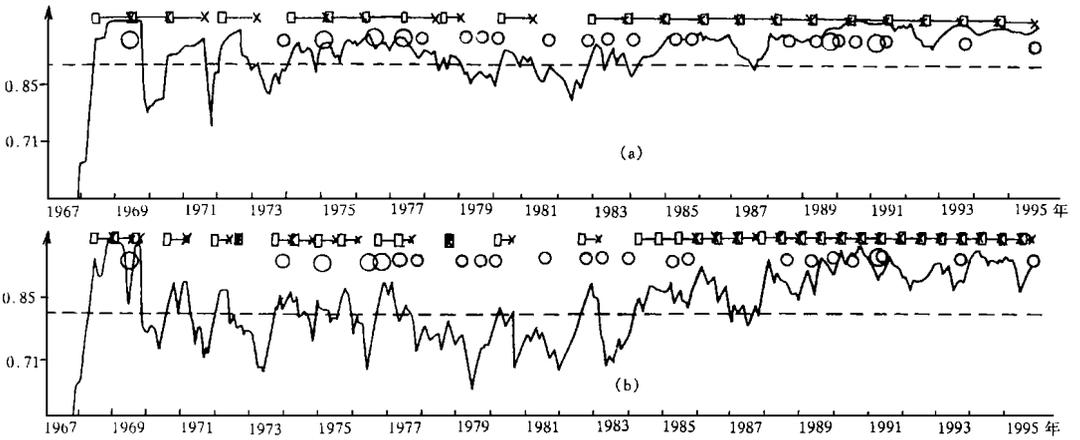


图 2 30 项资料的信息合成

Fig. 2 The information composition for 30 pieces of surveying data.

(a) $T_0 = 1.0a$; (b) $T_0 = 0.5a$; 其它图例符号意义同图 1

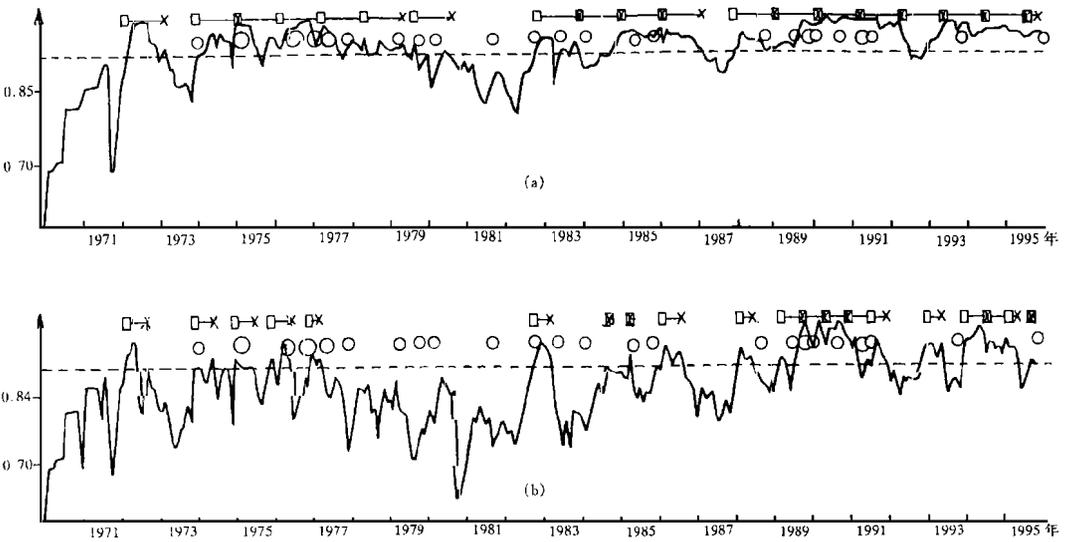


图 3 18 项资料的信息合成

Fig. 3 The information composition for 18 pieces of surveying data.

(a) $T_0 = 1.0a$; (b) $T_0 = 0.5a$; 其它图例符号意义同图 1

四种信息合成方案的结果均满足 $P_3 > P_1 \circ P_2$, 发生异常后出现地震的条件概率均高于地震的自然发生概率, 说明其均有一定的信息价值. 同时, 18 项资料给出的效益参数 G 稍高于 30 项资料的合成结果, 说明去掉效益稍低及干扰较明显的资料对信息合成结果有所改善, 但并不很明显. 这是由于信息合成模型中的权已经考虑了资料质量因素所致. 由表 4 不难看出, 未来 1a 内首都圈发生 5 级以上地震的概率较高. 由于 1995 年 10 月 6 日发生了一次 $M_s 5.0$ 地震, 故分别计算了未来 T_0 时段内再发生 5 级以上地震的条件概率. 显然, 未来 0.5a 内再次

表 4 首都圈 5 级以上地震发生概率计算

方法		地震的自然概率	有异常后 T_0 时段内发震的条件概率	已发生一次地震后, 未来 T_0 时段内再发生地震的概率
30 项资料的合成	$T_0=1.0a$	0.63	0.76	0.50
	$T_0=0.5a$	0.37	0.47	0.29
18 项资料的合成	$T_0=1.0a$	0.72	0.94	0.40
	$T_0=0.5a$	0.44	0.65	0.31

发生 5 级以上地震的概率较小, 在 0.31 以下, 但未来 1a 内再次发生 5 级以上地震的概率稍大一些, 约为 0.50. 这一理论研究结果对日常监测预报工作有一定的指导意义.

参考文献

- 1 薄万举, 王彦, 罗三明. 单项资料异常变化信息提取的一种新方法——多点组信息法. 山西地震, 1996, (1): 39~42.
- 2 薄万举, 周硕愚, 王彦. 四点组斜率信息标准化方法的进一步应用及其物理意义的探索. 四川地震, 1995, (1): 24~30.
- 3 薄万举, 吴翼麟. 信息合成方法及其应用. 地壳形变与地震, 1995, 15(3): 84~88.

APPLICATION OF INFORMATION COMPOSITION METHOD TO RESEARCHES ON SEISMIC REGIME ANALYSIS

BO Wanju

(The First Crustal Deformation Monitoring Center, SSB, Tianjin 300180)

Abstract

In the eighth Five-Year Plan period, a new information composition method called as Information Series Synthesis Method is given through the study of a second order project, 85-04-02-01, for short-impending earthquake prediction research. As a part of contents of the works for continuation of the first order project, 85-04-02, and pre-researches for project in the ninth Five-Year Plan period, the Information Series Synthesis Method is further perfected and applied to daily prediction work and the primary results have been obtained. This is a significant exploration for closely linking the earthquake study and prediction.

Key words: Information composition, Statistical analysis, Seismic regime, Capital Circle