

容量维 D_0 、信息维 D_1 、算法复杂性 $C(n)$ 与 b 值的相关性研究*

韩渭宾 易桂喜 辛 华

(中国成都 610041 四川省地震局)

关键词 相关性 地震时间分形 D_0 和 D_1 算法复杂性 $C(n)$ b 值

1 问题与思路

“八五”地震短临预报攻关号召大家在“新”和“实”字上下功夫,尤其是80年代末、90年代初,适逢地震界兴起非线性热,于是,引进、开发了许多地震学地震预报新指标和新方法,促进了地震预报研究的发展。可是,地震预报的实际水平并未因此发生质的飞跃。因此,在“九五”立项时,不少专家提出一个重要问题:这些新指标与老指标是否相互独立?能否筛选出相互独立,又有较好预报效能的指标和方法?

笔者在“八五”期间,引进、开发了容量维、信息维的变化率 ΔD_0 、 ΔD_1 ,以及算法复杂性 $C(n)$ 。本文着重研究 D_0 、 D_1 、 $C(n)$ 与 b 值相互之间的相关性问题。

研究两种预报指标之间的相关性,一般可从两方面入手:一是分析这两种指标的物理含义,看看它们有没有内在联系;二是计算两者的相关系数,看看能否通过统计相关检验。这两种方法虽然出发点很不相同,但都有科学道理。一般说来,结果应是一致的。笔者一向认为,统计是认识物理本质的敲门砖,只要样本量足够大,满足相关分析必要的条件,科学的统计相关检验得到的结果一定有其物理意义。可是,在地震预报研究中,常常遇到小样本等条件不充分满足的情况。这样,统计相关检验的结果可能与物理分析的结论不一致。为此,本文采用物理分析与统计相关检验双管齐下的办法,研究上述指标之间的相关性问题。

2 容量维 D_0 与信息维 D_1

容量维 D_0 和信息维 D_1 是两种不同的分维数,它们有不同的计算公式

$$D_0 = - \lim_{\epsilon \rightarrow 0} \frac{\lg N(\epsilon)}{\lg \epsilon} \quad (1)$$

$$D_1 = - \lim_{\epsilon \rightarrow 0} \frac{\sum_i p_i(\epsilon) \lg p_i(\epsilon)}{\lg \epsilon} \quad (2)$$

其基本区别是,用“标度变换法”计算 D_0 时,只统计有多少尺度为 ϵ 的格子有震,而不管每一格中有几次地震;计算 D_1 时,却特别注重地震落入尺度为 ϵ 的格子中的概率,这个概率实际上是通过统计落入尺度为 ϵ 的格子中的地震频数来计算的。这就决定了 D_0 和 D_1 一般是不相等的。

可是,笔者(韩渭宾,王维恩,1994)对川滇地区11个地震带(区)计算 D_0 和 D_1 随时间变化曲线都几

* “九五”地震预报攻关课题 95-04-01-01-11。

1997-10-05 收到初稿,1998-02-17 收到修改稿并决定采用。

乎重合. 详见上述文献图 2~图 12. 这可能是因为, 为了缩短预报期, 向短临预报过渡, 我们以 1 年为时间窗, 半年为滑动步长, 计算 D_0 和 D_1 随时间变化曲线. 这样, 每一个计算 D_0 和 D_1 的样本的地震数不大, 一般为数十个. 在这种条件下, 用“标度变换法”计算时, 同一时间格中落入 2 次以上地震的机会较小, 就可能使 D_0 和 D_1 几乎相等.

朱传镇和王琳瑛(1996)对于唐山、海城、黄海、江苏、和林格尔、首都圈中区和首都圈西区等 7 个地区分别计算了地震活动空间和时间多分形谱, 发现空间分形不均匀比较明显, 而这 7 个地区的地震时间 D_q 谱不均匀性不很明显, 即 D_0 与 D_1 差别也不大. 这与我们的上述结果也是一致的.

由于 D_0 和 D_1 曲线几乎相重, 由此演算出来的 ΔD_0 和 ΔD_1 也很接近. 造成的根源如前所述. 后面讨论与 b 值、 $C(n)$ 的相关性, 就只用 D_0 作代表了.

3 D_0 与 b 值

80 年代末, 地震界开始引入分形概念时, 国内外都有专家认为, 描述震级-频度关系的古登堡-里克特公式反映了地震活动在强度域里存在自相似性. 陈颙(1988)和洪时中(1990)分别导出了有关地震断层长度的自相似系统的分维数 d 和地震能量的容量维 D_0 与 b 的关系

$$d = 2b \tag{3}$$

$$D_0 = b/1.5 \tag{4}$$

这样, 近年一直有人持有这样的疑问, 既然已导出分维数 D_0 与 b 值的定量关系, 为什么还把 D_0 作为一种新指标, 花这么大的力气去研究? 问题是, 除个别人外, 人们大多研究地震活动的时空分维在地震预报中的应用. 这与地震强度域的分形不是一回事. 没有人, 也无法导出地震活动时空分维, 包括地震时空容量维 D_0 与 b 值的关系.

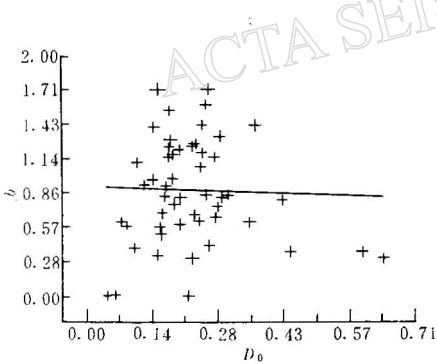


图 1 鲜水河地震带 D_0 与 b 值相关图

表 1 D_0 、 D_1 、 $C(n)$ 与 b 值的相关系数

预报 指标	鲜水河	松潘— 龙门山	名山—马边 —昭通	安宁河 —则木河	九龙— 永胜
D_0 与 b	-0.036	-0.257	0.107	-0.120	-0.278
D_1 与 b	-0.093	-0.251	-0.037	-0.187	-0.279
$C(n)$ 与 b	-0.404	-0.717	-0.483	-0.507	-0.654
D_0 与 $C(n)$	0.638	0.488	0.486	0.153	0.485
D_1 与 $C(n)$	0.637	0.493	0.510	0.166	0.435

四川地区 5 个主要地震带的地震时间容量维 D_0 、信息维 D_1 与 b 值等的相关系数见表 1. 除九龙—永胜地震带外, D_0 、 D_1 与 b 值的相关系数都低于置信度为 95% 的临界相关系数 $r_{0.05}$. 在以往的文献里, 我们曾指出, 也许由于九龙—永胜地震带的成带性不强, 划带根据不充分, 这个地震带的 ΔD_0 、 ΔD_1 的预报评分值很低. b 值的 R 评分也只有 0.33. 这样, 我们可以说, 经统计相关检验, D_0 、 D_1 与 b 值是相互独立的. 作为一个例子, 图 1 给出了鲜水河地震带 D_0 与 b 值的相关图. 图中清楚地显示出点子散乱, 明显地偏离回归直线.

4 $C(n)$ 与 b 值

$C(n)$ 与 b 值, 从算法所反映的物理实质来看, 也是不相同的. b 值反映大小地震的比例关系; 从一定程度上看, $C(n)$ 描述一串地震先后大小的变化. 前者没有先后次序概念, 后者研究的就是这串地震的先后顺序的复杂性.

现在遇到的矛盾是, 从表 1 给出的四川地区 5 个主要地震带 $C(n)$ 与 b 值的相关系数看, 不仅都大于

置信度 97.5% 的临界相关系数 $r_0(97.5\%)=0.273$, 甚而超过了置信度为 99% 的临界值 $r_0(99\%)=0.354$. 这就是说, 从这 5 个地震带的资料计算, $C(n)$ 与 b 值通过了相关检验.

为什么从物理内涵分析, 认为两个独立的预报指标, 在实际应用中又能通过相关检验, 被认为相互不独立呢? 我们分析, 最可能的原因是, 这两种指标或其中之一在实际计算中遇到了什么困难, 出现了局限性, 使得它或它们的物理内涵无法充分反映. 为此, 我们做了以下工作.

4.1 增加计算 $C(n)$ 的震级分档

我们首先想到的是, 计算 $C(n)$ 时, 震级分为 3 档: $2.5 \leq M < 3.5$, $3.5 \leq M < 4.5$ 和 $M \geq 4.5$, 而 $M \geq 4.5$ 的地震又太少, 不少地震带一年也没有 1 次. 会不会因为所分震级档太少, 算法复杂性不能充分被刻划?

为此, 我们试分为 5 档: $2.5 \leq M < 3.0$, $3.0 \leq M < 3.5$, $3.5 \leq M < 4.0$, $4.0 \leq M < 4.5$ 和 $M \geq 4.5$. 还是以 1 年为时间窗, 半年为滑动步长, 计算这 5 个地震带 $C(n)$ 随时间变化曲线, 进而计算 $C(n)$ 与 b 值、 D_0 、 D_1 值的相关系数. 其结果列于表 2.

结果表明, 尽管总的来说, $C(n)$ 与 b 值的相关系数还比较大, 但与分为 3 档的情况(见表 1)相比较, 5 个带里有 4 个带的相关系数有所下降. 这也许说明, 在目前资料情况下, 增加分档数, 有可能更好地刻划算法复杂性.

表 2 震级分为 5 档的 $C(n)$ 与 b 值、 D_0 、 D_1 的相关系数

预报指标	鲜水河	松潘— 龙门山	名山—马边 —昭通	安宁河 —则木河	九龙— 永胜
$C(n)$ 与 b	-0.387	-0.583	-0.273	-0.387	-0.546
$C(n)$ 与 D_0	0.659	0.451	0.664	0.168	0.570
$C(n)$ 与 D_1	0.658	0.449	0.671	0.165	0.514

4.2 增大时间窗的试验

为扩大每一个计算样本的地震数, 我们开展增大时间窗的试验. 作为试算, 只选了地震数相对较多的松潘—龙门山地震带和安宁河—则木河地震带为例. 先将时间窗增为 2 年, 滑动步长仍为半年, 计算 $C(n)$ 、 b 及 D_0 曲线, 再计算两两之间的相关系数. 结果列于表 3. 从表 1 和表 2 可见, 在各种条件下, D_0 、 D_1 与 b 值、 $C(n)$ 值的相关系数都很接近. 在做扩大时间窗的试验时, 只选用了 D_0 , 没有再用 D_1 .

由表 3 可以看到, 与表 1 相比, 增大时间窗, 这两个地震带的 $C(n)$ 值与 D_0 值的相关系数也有所下降, 但降幅不大, 仍能通过统计相关检验. 这也许说明, 样本量大小对计算这些预报指标都是有

表 3 时间窗为 2 年的 $C(n)$ 、 b 、 D_0 间的相关系数

预报指标	松潘—龙门山地震带	安宁河—则木河地震带
$C(n)$ 与 b	-0.459	-0.454
$C(n)$ 与 D_0	0.262	0.814

影响的. 增大样本量有可能改善 $C(n)$ 与 b 值的相互独立性. 但由于将时间窗由 1 年扩大到 2 年, 对于多数样本只是将地震数由数十增加到一百左右, 有些样本还只有数十, 样本仍不大. $C(n)$ 与 b 值仍能通过相关检验, 不能证明为相互独立可能仍与此有关. 时间窗再扩大, 或将统计的空间范围增大, 则又可能使预报期增大、预报地域范围扩大, 损失预报信息的精度.

5 $C(n)$ 与 D_0 、 D_1

从前面的讨论中已可看到, $C(n)$ 与 D_0 、 D_1 的物理概念不同. 问题是, 表 1、表 2 和表 3 都表明, 除安宁河—则木河地震带外, 有关 $C(n)$ 和 D_0 、 D_1 的相关系数都较大. 增加震级分档的 $C(n)$, 将时间窗由 1 年增到 2 年, 效果也不明显. 关于实际计算条件的局限性问题还有许多值得研究的问题.

6 结论与问题

(1) D_0 与 D_1 是两种物理意义不同的分维数, 但当样本量不够大时, 实际计算的 D_0 和 D_1 十分接近, 故不必重复计算, 可只算 D_0 . 显然, ΔD_0 与 ΔD_1 也同样不必重复计算. 就川滇地区 11 个地震带(区)的具

体情况,可只算 ΔD_0 .

(2) 物理意义分析和统计相关检验都表明,地震时间分形 D_0 、 D_1 与 b 值是相互独立的. 这两种指标的预报评分也较高,可以继续独立使用.

(3) 从计算揭示的物理本质看, $C(n)$ 与 b 值、 D_0 (D_1) 值是不同的,但在样本量不够大,震级分档少的情况下,计算的 $C(n)$ 值有可能与 b 值、 D_0 (D_1) 值通过相关检验. 又考虑到其预报评分相对较低(韩渭宾等, 1996), 建议在日常预报实践中暂不必将 $C(n)$ 作为必做指标.

(4) 从增加分档数和扩大时间窗的试验来看,改善计算 $C(n)$ 的条件,提高其预报效能,增大它与 b 值、 D_0 的独立性还是有可能的. 算法复杂性还刚刚被引入地震预报研究,还有许多问题值得深入研究.

关于“八五”攻关引入的新指标和一些老指标的相关性研究,本文只是作了一些初步探讨,希望能引出更广泛的讨论和更深入的研究.

研究过程中与洪时中同志作过不少有益的讨论,特此感谢.

参 考 文 献

- 陈颙, 1988. 分形与分维在地球科学中的应用. 北京: 学术期刊出版社, 37.
- 韩渭宾, 王维恩, 1994. 川滇地区各地震带容量维、信息维的变化率与地震短期预报的研究. 中国地震, 10(增刊): 122~132.
- 韩渭宾, 江道崇, 邓建平, 等, 1996. 算法复杂性与川滇地区 11 个地震带(区)地震预报的研究. 地震, 16(3): 225~231.
- 洪时中, 1990. 分维与地震科学. 科学, 42(2): 105~108.
- 朱传镇, 王琳瑛, 1996. 前震活动特征及其识别的研究(1). 地球物理学报, 39(1): 80~87.