

分形法预测太阳事件等数据

付昱华¹ 付安捷² 赵戈³

(1 中国海洋石油研究中心, 北京, 100027, E-mail: fuyh1945@sina.com)

(2 索尼(中国)有限公司, 北京, 100027)

(3 微软(中国)有限公司, 北京, 100027)

摘要 应用分形方法(包括重标极差(R/S)方法), 预测太阳事件等数据。预测方式分为两类: 单值预测(根据已知数据仅仅预测下一个数据)和多值预测(根据已知数据预测下几个数据)。预测实例包括太阳黑子数目, 宇宙射线指标, 2800MHz 太阳通量, 台风路径, 以及股票市场的走势等。其中预测宇宙射线指标时, 给出了“预警区间”的概念, 用以补充“置信区间”的概念。最后, 简述变维分形形式的引力公式计算(预测)太阳系内部的引力。

关键词 分形, 变维分形, 重标极差(R/S)方法, 预测, 太阳事件数据

1 前言

太阳事件数据的预测是一个非常重要的问题, 本文以作者已经发表的几篇论文为基础, 讨论应用分形方法(包括重标极差(R/S)方法), 预测太阳事件等数据。

2 常用的常维和变维分形预测方法及其应用实例

在最近十几年间, 分形方法已在一些领域获得成功应用, 它被用来揭示复杂现象中深藏的有组织结构。反映有组织结构特征的量称为分维数, 用 D 值来表示。在目前一般应用的分形方法中, 分维数 D 为常数, 例如不同地段海岸线的分维数 D 值可以取为 1.02、1.25 等。分形分布可用如下幂指数分布定义^[1]

$$N = \frac{C}{r^D} \quad (1)$$

式中: r 为特征线度, 如时间, 长度, 坐标等; N 为与 r 有关的物体数目或量值, 如太阳黑子数目, 宇宙射线指标, 价格, 温度等; C 为待定常数, D 为分维数。

在目前一般应用的分形方法中, D 为常数, 这种分形可称为常维分形。它在双对数坐标上是一条直线。根据该直线上的任意两个数据点 (N_i, r_i) 和 (N_j, r_j) , 可以确定该段直线的分形参数, 亦即分维数 D_{ij} 和常数 C_{ij} , 实际上将两个数据点的坐标代入(1)式后可解出

$$D_{ij} = \frac{\ln(N_i/N_j)}{\ln(r_j/r_i)} \quad (2)$$

$$C_{ij} = N_i r_i^{D_{ij}} \quad (3)$$

对于双对数坐标上的直线函数关系, 就可以用这种常维分形的方法进行预测和计算。

但是对于双对数坐标上非直线的函数关系, 常维分形就无法处理。而许多问题正属于这种情况。为克服这一困难, 我们在参考文献[2]~[4]中引入变维分形的概念, 即分维数 D 是特征线度 r 的函数。

$$D = F(r) \quad (4)$$

下面给出一种最常用的变维分形预测方法。

为方便计，将 r 取为时间的编号，例如规定某一年为第一年，则有 $r_1 = 1$ ，第二年的 $r_2 = 2$ ，等。 N 为待预测数据之值，如可将 N_1 取为第一年的待预测数据之值，将 N_2 取为第二年的待预测数据之值等。

现在假设，已知 n 个数据点，即第 1 年到第 n 年的待预测数据之值为已知，于是问题变为如何预测第 $n+1$ 年，第 $n+2$ 年等等的待预测数据之值。

由于第 n 个数据点，即第 n 年的 N_n 和 r_n 之值是已知的 ($r_n = n$)，第 $n+1$ 年的 r_{n+1} 之值也是已知的 ($r_{n+1} = n+1$)，如果再知道第 n 个数据点和第 $n+1$ 个数据点确定的常维数形的分维数 $D_{n,n+1}$ ，则可以从 (2) 式解出第 $n+1$ 年的待预测数据之值

$$N_{n+1} = N_n \left(\frac{r_n}{r_{n+1}} \right)^{D_{n,n+1}} \quad (5)$$

以此类推，可以求出第 $n+2$ 年等等的待预测数据之值。

至于如何确定分维数 $D_{n,n+1}$ ，需要由已知的第一个和第二个数据点确定的分维数 D_{12} ，以及其他已知数据点确定的 $D_{23} \cdots D_{n-1,n}$ 来求出。但是在一般情况下这些分维数的变化规律难以发现。

在这种情况下，我们不能直接应用上述方法。需要先将已知的数据进行累计和的变换，然后再应用上述方法预测第 $n+1$ 年，第 $n+2$ 年等等的累计和之值。最后再由累计和之值求出待预测数据之值。

应用累计和的优点是，可以将有升有降的序列变为单调上升的序列。

该法可简介如下。

第一步，将原始数据点 $(N_i, r_i) (i = 1 \sim n)$ 绘于双对数坐标上，一般情况下它们不能与常维分形分布模型符合良好，于是可将 N_i 排成一个基本序列，即有：

$$\{N_i\} = \{N_1, N_2, N_3, \dots\} \quad (i = 1, 2, \dots, n)$$

其他的序列均可以根据基本序列构造。例如构造一阶累计和序列 $S^{(1)}$ ，其中 $S_1^{(1)} = N_1$ ， $S_2^{(1)} = N_1 + N_2$ ， $S_3^{(1)} = N_1 + N_2 + N_3$ ，等等，依类推可构造二阶、三阶累计和等等，即有：

$$\{S_i^{(1)}\} = \{N_1, N_1 + N_2, N_1 + N_2 + N_3, \dots\} \quad (i = 1, 2, \dots, n) \quad (6)$$

$$\{S_i^{(2)}\} = \{S_1^{(1)}, S_1^{(1)} + S_2^{(1)}, S_1^{(1)} + S_2^{(1)} + S_3^{(1)}, \dots\} \quad (i = 1, 2, \dots, n) \quad (7)$$

$$\{S_i^{(3)}\} = \{S_1^{(2)}, S_1^{(2)} + S_2^{(2)}, S_1^{(2)} + S_2^{(2)} + S_3^{(2)}, \dots\} \quad (i = 1, 2, \dots, n)$$

$$\{S_i^{(4)}\} = \{S_1^{(3)}, S_1^{(3)} + S_2^{(3)}, S_1^{(3)} + S_2^{(3)} + S_3^{(3)} \dots\} \quad (i = 1, 2 \dots n)$$

需要指出的是, $S_i^{(2)}$ 表示二阶累计和, 而不表示 S_i 的二次方。 $S_i^{(3)}$ 等需要同样理解。

第二步, 建立各阶累计和的分形模型。

以二阶累计和为例。将数据点 $(S_i^{(2)}, r_i)(i = 1 \sim n)$ 绘于双对数坐标上, 将这些点依次连接起来, 可得分段的常维分形模型。例如根据 n 个数据点, 可以得到由 $n-1$ 条直线组成的分段常维分形模型 (对于不同的直线, 其分维数也不同, 这也是最简单的变维分形模型), 各条直线的分形参数 $D_{ij}^{(2)}, (i = 1 \sim n-1, j = i+1)$ 和 $C_{ij}^{(2)}$, 可根据式 (2) 和 (3) 计算 (其中的 N_i 值用 $S_i^{(2)}$ 值代替)。

$$D_{ij}^{(2)} = \ln(S_i^{(2)} / S_j^{(2)}) / \ln(r_j / r_i) \quad (8)$$

$$C_{ij}^{(2)} = S_i^{(2)} r_i^{D_{ij}^{(2)}} \quad (9)$$

第三步, 选择效果最好的变换并确定其相应的分形参数。

将各阶累计和构成的数据点分别绘于双对数坐标上之后, 即可选择效果最好 (其分维数之值均匀增加或均匀减少) 的变换并确定其相应的分形参数。

由于一般情况下使用二阶累计和效果最好, 下面仅讨论二阶累计和的情况。

当分形分布模型选定之后, 即可先用适当方法确定分维数 $D_{n,n+1}^{(2)}$, 再用经过改造的 (5)

式进行累计和的预测。由于分维数之值均匀增加或均匀减少, 分维数 $D_{n,n+1}^{(2)}$ 可用下面的线性插值公式求出

$$D_{n,n+1}^{(2)} = 2D_{n-1,n}^{(2)} - D_{n-2,n-1}^{(2)} \quad (10)$$

对于二阶累计和, (5) 式变为

$$S_{n+1}^{(2)} = S_n^{(2)} \left(\frac{r_n}{r_{n+1}} \right)^{D_{n,n+1}^{(2)}} \quad (11)$$

由于 $S_1^{(1)} \sim S_n^{(1)}$, $S_1^{(2)} \sim S_n^{(2)}$ 都已经计算出, 由预测的二阶累计和可以得到预测的一阶累计和

$$S_{n+1}^{(1)} = S_{n+1}^{(2)} - S_n^{(2)} \quad (12)$$

由预测的一阶累计和可以得到待预测数据之值

$$N_{n+1} = S_{n+1}^{(1)} - S_n^{(1)} \quad (13)$$

以此类推可以得到 N_{n+2} , N_{n+3} , 等等。

由于这类预测方式是根据 n 个已知数据 $(N_1, N_2, N_3 \dots N_n)$ 预测下几个数据 $(N_{n+1},$

N_{n+2} , N_{n+3} , 等), 所以称为多值预测。

根据观察, 多值预测仅仅是用于数据是单调上升或单调下降的情况, 否则的话就要采用其他方式进行预测。

本文提出的另外一类方式是单值预测: 开始时根据 n 个已知数据 ($N_1, N_2, N_3 \cdots N_n$) 仅仅预测下一个数据 N_{n+1} , 等到实际的 N_{n+1} 出现之后, 再根据 $n+1$ 个已知数据 ($N_1, N_2, N_3 \cdots N_{n+1}$) 来下一个数据 N_{n+2} , 以此类推可以得到 N_{n+3} 。

为了与其他方法的预测结果进行对比, 我们选择了如下的线性插值公式进行对比。

$$N_{n+1} = 2N_n - N_{n-1} \quad (14)$$

3 置信区间的计算

假设根据单值预测或多值预测方式已经预测出 n 个数据 ($Y_1', Y_2', Y_3' \cdots Y_n'$), 其相应的实际值为 ($Y_1, Y_2, Y_3 \cdots Y_n$), 则第 $n+1$ 个预测值 Y_{n+1}' 的 $(1-\alpha) \times 100\%$ 置信区间为:

$$Y_{n+1}' \pm t_{\alpha} S_Y \quad (15)$$

式中: $S_Y = \sqrt{\frac{\sum (Y - Y')^2}{n-2}}$

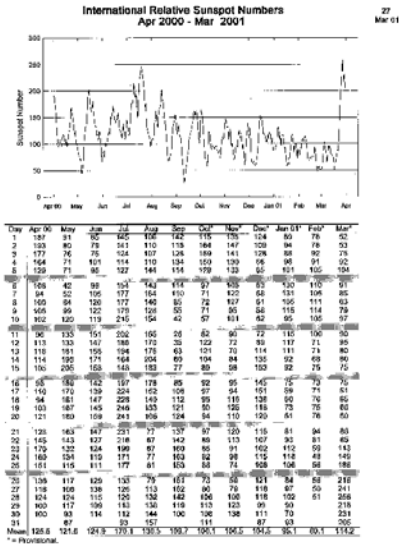
在本文中, $\alpha = 0.05$ 。

下面给出几个应用实例。

4 单值预测方式的应用实例

例 1, 根据 2000 年 4 月 1 日到 4 月 20 日的太阳黑子数目, 用单值预测方式预测 4 月 21 日到 6 月 9 日的太阳黑子数目。

已知数据如下 (下载地址: <http://www.ngdc.noaa.gov/ngdc.html>, <http://sgd.ngdc.noaa.gov/sgd/jsp/solarfront.jsp>)。



取2000年4月1日为第一天, $r_1 = 1$, 该日太阳黑子数目为 N_1 , 从表中查得: $N_1 = 187$ 。

以此类推可得4月20日的 $r_{20} = 20$, $N_{20} = 121$ 。根据4月1日到4月20日的太阳黑子数目以及公式(10) - (13), 用单值预测方式预测出4月21日的太阳黑子数目为124, 而实际值为128。根据4月1日到4月21日的太阳黑子数目预测出4月22日的太阳黑子数目为132, 而实际值为145。注意, 预测4月22日的太阳黑子数目时, N_{21} 取为实际值128, 而不是预测值124。全部预测结果如下表所示。

表1, 2000年4月21日到6月9日太阳黑子数目预测

编号	月 日	实际值	分形预测值	分形误差	线插预测值	线插误差
1	4 21	128	124	-4	139	11
2	22	145	132	-13	135	-10
3	23	170	151	-19	162	-8
4	24	160	180	20	195	35
5	25	151	167	16	150	-1
6	26	136	151	20	142	6
7	27	118	138	20	121	3
8	28	124	117	-7	100	-24
9	29	100	124	24	130	30
10	30	100	97	-3	76	-24
11	5 1	91	97	6	100	9
12	2	80	88	8	82	2
13	3	76	76	0	69	-7
14	4	71	72	1	72	1
15	5	71	67	-4	66	-5
16	6	42	67	25	71	29
17	7	52	36	-16	13	-39

18	8	64	48	-16	62	2	
19	9	99	61	-38	76	-23	
20	10	120	99	-21	134	14	
21	11	133	122	-11	141	8	
22	12	133	136	3	146	13	
23	13	161	136	-25	133	-28	
24	14	193	160	-27	189	-4	
25	15	205	200	-5	225	20	
26	16	189	213	24	217	28	
27	17	170	195	25	179	3	
28	18	161	174	13	151	-10	
29	19	167	164	-3	152	-15	
30	20	180	171	-9	173	-7	
31	21	163	185	22	193	30	
32	22	143	166	23	146	3	
33	23	132	144	12	123	-9	
34	24	134	132	-2	121	-13	
35	25	115	134	19	136	21	
36	26	117	114	-3	96	-21	
37	27	106	116	10	119	13	
38	28	124	104	-20	95	-29	
39	29	117	124	7	142	25	
40	30	93	116	23	110	17	
41	31	67	91	24	69	2	
42	6	1	85	63	-22	41	-44
43	2	79	83	4	103	24	
44	3	75	76	1	73	-2	
45	4	101	72	-29	71	-30	
46	5	95	100	5	127	30	
47	6	99	94	-5	89	-10	
48	7	105	98	-7	103	-2	
49	8	120	104	-16	111	-9	
50	9	122	120	-2	135	13	

从上表中可以看出，在分形方法预测的 50 个数据中，误差小于等于 10 的共有 22 个，占 44%；误差大于等于 11、小于等于 20 的共有 14 个，占 28%；误差大于等于 21、小于等于 30 的共有 13 个，占 26%；误差大于等于 31 的共有 1 个，占 2%。其中最大误差为 38。

而线性插值方法预测的结果为，误差小于等于 10 的共有 23 个，占 46%；误差大于等于 11、小于等于 20 的共有 9 个，占 18%；误差大于等于 21、小于等于 30 的共有 15 个，占 30%；误差大于等于 31 的共有 3 个，占 6%。其中最大误差为 44。

下面给出本例分形法置信区间的计算。

对于 4 月 30 日的预测值 97 (实际值也为 97), $S_Y = 19.26$, $t_{0.05}(7) = 2.365$, 根据公式 (15) 可得

置信区间为: 97 ± 45.5 。

对于 5 月 10 日的预测值 99 (实际值为 120), $S_Y = 17.71$, $t_{0.05}(17) = 2.110$, 根据公式 (15) 可得

置信区间为: 99 ± 37.0 。

对于 5 月 20 日的预测值 171 (实际值为 180), $S_Y = 17.94$, $t_{0.05}(27) = 2.052$, 根据公式 (15) 可得

置信区间为: 171 ± 36.8 。

对于 5 月 30 日的预测值 116 (实际值为 93), $S_Y = 17.12$, $t_{0.05}(37) = 2.028$, 根据公式 (15) 可得

置信区间为: 116 ± 34.7 。

对于 6 月 9 日的预测值 120 (实际值为 122), $S_Y = 17.01$, $t_{0.05}(47) = 2.014$, 根据公式 (15) 可得

置信区间为: 120 ± 34.3 。

需要指出的是, 在本例中的时间间隔取为一天, 因此只能够预测下一天的太阳黑子数目。如果时间间隔取为一周, 一月, 一年等, 则可以预测下一周, 下一月, 下一年等的太阳黑子数目。

例如时间间隔取为一个月, 如果已知数据按下面的方式选取

1 月总数, 2 月总数, ...11 月总数

则可以预测 12 月的总数。

然而, 如果已知数据按下面的方式选取

1 月 15 日数目, 2 月 15 日数目, ...11 月 15 日数目

则可以提前一个月预测 12 月 15 日的数目。

例 2, 根据 2000 年 4 月 1 日到 4 月 20 日的 2800MHz 太阳通量值, 用单值预测方式预测 4 月 21 日到 6 月 9 日的 2800MHz 太阳通量值。

已知数据如下 (下载地址: <http://www.ngdc.noaa.gov/ngdc.html>)。

应用与例 1 相同的方法, 得到的全部预测结果如下表所示。

表 2, 2000 年 4 月 21 日到 6 月 9 日 2800MHz 太阳通量预测

编号	月	日	实际值	分形预测值	分形误差
1	4	21	189.2	185.7	-3.5

2		22	204.1	193.3	-10.8
3		23	208.4	210.1	1.7
4		24	208.1	214.4	-6.3
5		25	205.1	213.3	-8.2
6		26	192.4	209.3	16.9
7		27	186.0	194.2	8.2
8		28	186.0	186.7	0.7
9		29	177.5	186.6	9.1
10		30	172.0	177.0	5.0
11	5	1	160.1	170.8	10.7
12		2	155.3	157.6	2.3
13		3	139.6	152.5	12.9
14		4	136.8	135.4	-1.4
15		5	132.1	132.7	0.6
16		6	129.1	128.6	-1.1
17		7	133.4	125.1	-8.3
18		8	139.6	130.3	-9.3
19		9	152.5	137.3	-15.1
20		10	182.8	151.6	-31.2
21		11	181.3	184.8	3.5
22		12	194.4	183.1	-11.3
23		13	222.0	197.1	-24.9
24		14	237.6	227.0	-10.6
25		15	249.9	243.7	-6.2
26		16	264.5	256.7	-7.8
27		17	268.1	272.1	4.0
28		18	258.8	275.6	16.8
29		19	260.4	265.0	4.6
30		20	251.6	266.3	14.7
31		21	238.0	256.5	18.5
32		22	220.3	241.5	21.2
33		23	209.5	222.1	12.6
34		24	194.3	210.4	16.1
35		25	177.4	193.9	16.5
36		26	172.4	175.7	3.3
37		27	166.2	170.6	4.4
38		28	160.2	163.8	3.6
39		29	153.1	157.8	4.7
40		30	150.5	150.3	-0.2
41		31	158.7	147.7	-11.0
42	6	1	152.3	156.6	4.3
43		2	192.7	150.0	-42.7
44		3	170.7	193.0	22.3
45		4	174.7	169.7	-5.0

46	5	176.1	174.1	-2.0
47	6	192.0	175.4	-16.6
48	7	185.8	193.3	6.5
49	8	179.9	185.8	5.9
50	9	174.1	179.6	5.5

从上表中可以看出，在分形方法预测的 50 个数据中，误差小于等于 10 的共有 30 个，占 60%；误差大于 10、小于等于 20 的共有个 15，占 30%；误差大于 20、小于等于 30 的共有 3 个，占 6%；误差大于 30 的共有 2 个，占 4%。其中最大误差为 42.7。

而线性插值方法预测结果的最大误差为 62.4。

下面给出本例分形法置信区间的计算。

对于 4 月 30 日的预测值 177.0（实际值为 172.0）， $S_Y = 9.82$ ， $t_{0.05}(7) = 2.365$ ，根据公式（15）可得

置信区间为： 177.0 ± 23.2 。

对于 5 月 10 日的预测值 151.6（实际值为 182.8）， $S_Y = 9.00$ ， $t_{0.05}(17) = 2.110$ ，根据公式（15）可得

置信区间为： 151.6 ± 19.0 。

对于 5 月 20 日的预测值 266.3（实际值为 251.6）， $S_Y = 11.60$ ， $t_{0.05}(27) = 2.052$ ，根据公式（15）可得

置信区间为： 266.3 ± 23.8 。

对于 5 月 30 日的预测值 150.3（实际值为 150.5）， $S_Y = 12.08$ ， $t_{0.05}(37) = 2.028$ ，根据公式（15）可得

置信区间为： 150.3 ± 24.5 。

对于 6 月 9 日的预测值 179.6（实际值为 174.1）， $S_Y = 13.24$ ， $t_{0.05}(47) = 2.014$ ，根据公式（15）可得

置信区间为： 179.6 ± 26.7 。

预警区间的概念如下。

置信区间的概念已经获得了广泛的应用。然而美中不足的是，没有发挥已有的最大预测误差的作用。为此，我们引入预警区间的概念，用以补充“置信区间”的概念，并且希望其上下限能够更接近于实际值。

假设第 $n+1$ 个预测值为 N_{n+1} ，在此之前的 n 个预测值中，最大的预测误差为 E_{\max} ，

则第 $n + 1$ 个预测值的预警区间为:

$$N_{n+1} \pm E_{\max}$$

例 3, 根据 2001 年 3 月 1 日到 19 日的 THULE 宇宙射线指标, 用单值预测方式预测 3 月 20 日到 31 日的宇宙射线指标, 并给出相应的预警区间。

已知数据如下 (下载地址: <http://www.ngdc.noaa.gov/ngdc.html>)。

应用与例 1 相同的方法, 得到的全部预测结果如下表所示。

表 3, 2001 年 3 月 20 日到 31 日宇宙射线指标预测

编号	月	日	实际值	分形预测值	分形法预测误差	预警区间	置信区间
1	3	20	4079.7	4172.9	93.2		
2		21	4094.7	4069.0	-25.7	4069.0 ± 93.2	
3		22	4136.1	4089.4	-46.7	4089.4 ± 93.2	
4		23	4146.5	4139.3	-7.2	4139.3 ± 93.2	4139.3 ± 1304.9
5		24	4140.0	4151.7	1.7	4151.7 ± 93.2	4151.7 ± 314.6
6		25	4147.1	4143.8	-3.3	4143.8 ± 93.2	4143.8 ± 191.8
7		26	4156.5	4152.5	-4.0	4152.5 ± 93.2	4152.5 ± 145.4
8		27	4058.4	4162.7	104.3	4162.7 ± 93.2	4162.7 ± 120.5
9		28	4001.3	4050.0	48.7	4050.0 ± 104.3	4050.0 ± 148.8
10		29	4003.7	3986.9	-16.8	3986.9 ± 104.3	3986.9 ± 133.2
11		30	4066.0	3991.3	-74.7	3991.3 ± 104.3	3991.3 ± 125.4
12		31	3975.2	4063.7	88.5	4063.7 ± 104.3	4063.7 ± 126.6

从上表中可以看出, 在分形方法预测的 12 个数据中, 误差小于等于 10 的共有 4 个, 占 33%; 误差大于 10、小于等于 20 的共有 1 个, 占 8%; 误差大于 20、小于等于 30 的共有 1 个, 占 8%; 误差大于 30 的共有 5 个, 占 42%。其中最大误差为 104.3。

而线性插值方法预测结果的最大误差为 153.1。

从表中的数据可以看出, 对于本例, 所有预警区间的上限 (或下限) 均比置信区间相应的上限 (或下限) 更接近于实际值。

例 4, 应用单值预测方式预测股票指数。

自 2000 年 11 月 1 日至 2001 年 2 月 7 日, 北京有线电视台《每日财经》节目举办了股指竞猜活动, 要求参赛者于每日 13 时以前将预测的当日收盘指数用电话传给电视台参加竞赛, 每日设一等奖 2 名 (沪市深市各 1 名), 二等奖 8 名, 三等奖 10 名。我们是在 11 月 17 日得到消息并开始用变维分形的预测结果参加竞赛的, 到 2 月 7 日竞赛因故停止, 共获得一等奖 2 次 (沪市深市各 1 次), 二等奖 2 次, 三等奖 7 次。这是一次实战应用, 其效果如何, 读者可以做出自己的判断。下面是上证指数的全部实际预测结果。

表 4, 上证指数 (1A0001) 的预测结果

编号	日期	预测值	实际值	误差	获奖情况
1	2000 年 11 月 16 日	2087.87	2095.98	-8.11	
2	2000 年 11 月 17 日	2104.99	2093.23	11.76	

3	2000年11月20日	2103.48	2101.38	2.10	获当日三等奖
4	2000年11月21日	2115.14	2097.98	17.16	
5	2000年11月22日	2109.29	2113.30	-4.01	
6	2000年11月23日	2125.61	2119.43	6.18	
7	2000年11月24日	2131.43	2053.37	78.06	
8	2000年11月27日	2048.50	2049.67	-1.17	获当日三等奖
9	2000年11月28日	2071.23	2079.39	-8.16	
10	2000年11月29日	2082.63	2067.49	15.14	
11	2000年11月30日	2063.54	2070.61	-7.07	
12	2000年12月1日	2082.96	2081.84	1.12	
13	2000年12月4日	2092.32	2092.13	0.19	获当日二等奖
14	2000年12月5日	2099.49	2091.66	7.83	
15	2000年12月6日	2095.93	2075.62	20.31	
16	2000年12月7日	2065.51	2075.04	-9.53	
17	2000年12月8日	2085.09	2073.16	11.93	
18	2000年12月11日	2044.21	2046.07	-1.86	获当日三等奖
19	2000年12月12日	2047.74	2059.05	-11.31	
20	2000年12月13日	2057.62	2056.12	1.50	
21	2000年12月14日	2055.93	2051.07	4.86	
22	2000年12月15日	2041.31	2039.36	1.95	
23	2000年12月18日	2026.44	2044.54	-18.10	
24	2000年12月19日	2052.43	2049.03	3.40	
25	2000年12月20日	2058.43	2071.26	-12.83	
26	2000年12月21日	2084.98	2076.89	8.09	
27	2000年12月22日	2079.10	2069.77	9.33	
28	2000年12月25日	2071.63	2068.17	3.46	
29	2000年12月26日	2075.03	2076.26	-1.23	
30	2000年12月27日	2070.66	2058.24	12.42	
31	2000年12月28日	2057.65	2053.70	3.95	
32	2000年12月29日	2070.41	2073.47	-3.06	
33	2001年1月2日	2095.00	2103.46	-8.46	
34	2001年1月3日	2121.09	2123.89	-2.80	
35	2001年1月4日	2123.90	2117.40	6.50	
36	2001年1月5日	2125.34	2125.30	0.04	获当日一等奖
37	2001年1月8日	2108.06	2102.06	6.00	
38	2001年1月9日	2098.75	2101.13	-2.38	
39	2001年1月10日	2120.91	2125.61	-4.70	
40	2001年1月11日	2132.74	2119.14	13.60	
41	2001年1月12日	2106.41	2104.74	1.67	
42	2001年1月15日	2054.82	2032.44	22.38	
43	2001年1月16日	2003.01	2006.88	-3.87	
44	2001年1月17日	2035.48	2034.58	0.90	
45	2001年1月18日	2043.70	2043.10	0.60	获当日三等奖
46	2001年1月19日	2063.47	2065.60	-2.13	

47	2001年2月5日	2036.62	2008.03	28.59	
48	2001年2月6日	1960.85	1995.31	-34.46	
49	2001年2月7日	1979.34	1979.93	-0.59	获当日二等奖

在上述连续的 49 次实际预测中，误差在 0.5 点以下的共有 2 次，占 4%。误差在 1 点以下的共有 5 次，占 10%。误差在 2 点以下的共有 12 次，占 24%。误差在 5 点以下的共有 24 次，占 49%。误差在 10 点以下的共有 35 次，占 71%。误差在 10 点以上的共有 14 次，占 29%。

显然，这种方法也可以试用于股票价格的预测。

4 多值预测方式的应用实例

例 5，已知到 1980 年 7 月 20 日 08 时为止，8007 号强台风 (JOE) 中心的路径如表 5 所示，试用多值预测方式求该台风中心未来的预报位置。本例取自参考文献[5]。

表 5，8007 号台风中心的已知路径

编号	月	日	时	北纬 ϕ (度)	东经 λ (度)
1	7	16	14	10.0	147.0
2			20	11.0	146.0
3		17	02	12.0	145.0
4			08	12.7	143.8
5			14	12.3	143.2
6			20	12.5	142.0
7		18	02	13.1	140.2
8			08	13.5	138.9
9			14	14.0	137.5
10			20	14.2	136.2
11		19	02	14.3	134.7
12			08	14.7	133.1
13			14	15.0	131.7
14			20	15.2	130.1
15		20	02	15.7	128.1
16			08	16.1	126.7

台风路径本来是一个二维问题，然而应用多值预测方式，可以简化为一维问题，亦即分别得出纬度和经度的预测值。

全部预测结果，及其与实际值和参考文献[6]的预测值的对比见表 6 和表 7。需要说明的是，应用本文所述方法，无法预先确定台风中心将在编号 30 以后移出，不过为了进行对比，我们计算到编号 30 处结束。

表 6，8007 号台风中心纬度的预测值

编号	月	日	时	实际值	本文预测值	文[6]预测值
17	7	20	14	16.3	16.4	
18			20	16.4	16.7	17.0
19		21	02	17.1	17.1	

20	08	17.4	17.4	18.2
21	14	18.1	17.8	
22	20	18.7	18.1	19.2
23	22 02	19.1	18.5	
24	08	19.5	18.8	20.0
25	14	20.1	19.2	
26	20	20.2	19.5	
27	23 02	20.4	19.9	
28	08	20.9	20.3	
29	14	20.9	20.6	
30	20	20.5	21.0	

表 7, 8007 号台风中心经度的预测值

编号	月 日 时	实际值	本文预测值	文[6]预测值
17	7 20 14	125.3	125.3	
18	20	123.8	123.8	123.6
19	21 02	122.1	122.4	
20	08	120.8	121.0	120.9
21	14	119.0	119.6	
22	20	117.2	118.2	118.4
23	22 02	115.3	116.8	
24	08	113.6	115.5	116.0
25	14	112.2	114.1	
26	20	110.3	112.8	
27	23 02	108.4	111.4	
28	08	106.7	110.1	
29	14	105.3	108.9	
30	20	103.0	107.5	

从表 6 和表 7 可以看出, 虽然经度的结果稍差, 但是总的来说本文的预测结果是令人满意的。

5 改进的重标极差 (R/S) 预测方法

火灾和地震等自然灾害所释放的能量是巨大的。如果这些能量能够有效利用, 哪怕是一小部分, 人类也将获得丰富的能源。在需要对自然灾害的情况做出预测的情况下, 参考文献[7]中改进的重标极差 (R/S) 方法是非常有效的。

赫斯特(H. E. Hurst)是著名的水文学家。他在参考文献[8]中提出了 R/S 分析方法或重标极差方法。其中最重要的工作是计算赫斯特指数 H。

我们在参考文献[7]中采取两种新的分区方法计算赫斯特指数 H。

第一种方法: 区间所含数据的数目是依次递增的。例如可以令第一区间含 1950~1955 年的数据 (共有 6 个数据), 第二区间含 1950~1956 年的数据 (共有 7 个数据), 依此类推, 最后一个区间含 1950~2000 年的数据 (共有 51 个数据)。需要说明的是, 第一区间所含数据计算的赫斯特指数作为 1955 年的赫斯特指数, 第二区间所含数据计算的赫斯特指数作

为 1956 年的赫斯特指数，依此类推，最后一个区间所含数据计算的赫斯特指数作为 2000 年的赫斯特指数。

第二种方法：区间所含数据的数目是固定的。例如可以令第一区间含 1950~1955 年的数据（共有 6 个数据），第二区间含 1951~1956 年的数据，依此类推，最后一个区间含 1995~2000 年的数据（共有 6 个数据）。与第一种方法一样，每一个区间所含数据计算的赫斯特指数作为该区间最后一年的赫斯特指数。

下面给出第一种方法的应用实例。

例 6，根据 1950~1999 年的中国全国火灾起数预测 2000 年的全国火灾起数。

首先计算出 1955~1999 年的赫斯特指数 H 。然后根据这些赫斯特指数用分形方法计算 2000 年的赫斯特指数 H_{2000} ，则可以得到

$$H_{2000} = 0.7750$$

根据计算的 H_{2000} ，就可以用打靶法（试射法）确定 2000 年的全国火灾起数。最后得到，当 2000 年的全国火灾起数为 199960 时 $H_{2000} = 0.7750$ 。即预测的 2000 年全国火灾起数为 199960，而实际值为 189185，预测误差为 5.7%。应该说预测效果还是不错的。

为了更有效地对统计数据进行分析预测，参考文献[7]中还引入高阶赫斯特指数的概念。对于已经计算出的所有赫斯特指数 H ，将其视为一组已知的统计数据。于是可以对这一组统计数据再进行 R/S 分析，从而得到一组新的赫斯特指数，将其记为 H_1 ，即赫斯特指数的赫斯特指数。依此类推可以得到更高阶的赫斯特指数 H_2 ， H_3 等。应用赫斯特指数和高阶赫斯特指数，可以判断下一年的火灾数目是否会激增。

6 变维分形形式的引力公式计算（预测）太阳系的引力

在参考文献[9]中，给出了如下的改进的万有引力公式

$$F = -\frac{GMm}{r^2} - \frac{3G^2M^2mp}{c^2r^4}$$

式中： G 为引力常数； M 和 m 为两物体的质量； r 为两物体间的距离； c 为光速； p 为质量为 m 的物体在质量为 M 的物体的引力场中沿圆锥曲线运动时的半正焦弦，而且有： $p = a(1-e^2)$ ，对于椭圆； $p = a(e^2-1)$ ，对于双曲线； $p = y^2/2x$ ，对于抛物线。

该公式对水星近日点进动问题和光线近日偏折问题均能给出与广义相对论一样的正确解。当两物体做中心对中心运动（含相对静止）时，可视为半正焦弦 $p = 0$ 的情况，此时改进的万有引力公式与原有的万有引力公式相同。

可以将改进的万有引力定律写为变维分形的形式：

$$F = -\frac{GMm}{r^D}$$

根据改进的万有引力公式可得

$$D = 2 - \ln\left(1 + \frac{3GMp}{c^2r^2}\right) / \ln r$$

D 的取值随问题的不同而不同。对于光线近日偏折问题， $1.954997 \leq D \leq 2$ 。对于水星与太阳之间的引力问题，设 $D = 2 - \epsilon$ ，则 $2.018165 \times 10^{-9} \leq \epsilon \leq 4.935239 \times 10^{-9}$ 。

7 结论

本文讨论应用分形方法（包括重标极差（R/S）方法），预测太阳事件等数据。文中讨

论的实例表明, 这些预测和计算方法在预测太阳事件等数据时将具有良好的应用前景。

参考文献

- 1 Turcotte, D L. Fractals and chaos in geology and geophysics[M]. Cambridge: Cambridge University Press, 1992
- 2 付昱华. 流体力学中的变维分形[A]. 第二届全国计算水力学会议论文集[C]. 武汉: 中国水利学会水力学专业委员会, 1993, 202~207
- 3 付昱华. 分形技术在油气勘探应用中的改进 — 变维分形技术[J]. 中国海上油气(地质), 1994, (3): 210~214
- 4 Fu Yuhua. Analysed and fractal single point method for solving hydraulic problems in ocean engineering[R](i.e., SPE(Society of Petroleum Engineers) 29986). International meeting on petroleum engineering, Beijing, Nov 1995. 347~356
- 5 用分形分布模型预测台风路径[J]. 中国海上油气(工程), 1999, (1): 36~39
- 6 牟少杰, 陆静珍. 台风路径预报查算表 [M]. 北京: 海洋出版社, 1982
- 7 付昱华, 付安捷. 改进的R/S方法与中国火灾数据的分析预测[J]. 中国工程科学, 2004, 6 (5): 39~44
- 8 Hurst H E. The Long-Term Storage Capacity of Reservoirs. Transactions of the American Society of Civil Engineers, 116, 1951
- 9 付昱华. 改进的牛顿万有引力公式. 自然杂志, 2001, (1):58~59

Fu Yuhua¹, Fu Anjie², Zhao Ge³

(1 China Offshore Oil Research Center, Beijing, 100027, China)

(1 E-mail: fuyh1945@sina.com)

(2 Sony (China) Limited, Beijing, 100027, China)

(3 Microsoft (China) Co., Ltd, Beijing, 100027, China)

Abstract: With the fractal method (including the rescaled range analysis (R/S analysis) method), some important questions for forecasting solar events data and the like are discussed. Two forecasting manners are proposed: single-value-forecasting (according to the given data to forecast the next one datum only), and multiple-value-forecasting (according to the given data to forecast the next several data). The examples for forecasting sunspot number, 2800MHz solar flux, cosmic ray index, typhoon track, and stock index are presented. As forecasting the cosmic ray index, the concept of “warning interval” is presented to cooperate with the concept of “confidence limit”. Finally the improved Newton’s formula of universal gravitation with the form of variable dimension fractal is briefly introduced to calculate (forecast) the gravitation in solar system.

Key words: Fractal method, variable dimension fractal, rescaled range analysis (R/S analysis), forecasting, solar events data