

相似区间叠集成法在降水预报中的应用

朱德生 方 茸

(安徽省气象台, 合肥 230061)

提 要

在用相似预报方法制作安徽省汛期降水预报中, 同时用相似系数和距离系数表征预报因子场的相似程度。并采用相似区间叠集成法对找出的数十个相似年进行集成, 客观地确定出预报量可能出现的范围, 此方法预报效果良好。

关键词: 预测 相似 叠合 集成法

引 言

相似预报法是目前短期气候预测业务中广泛使用的一种统计方法。其基本原理是:把当前的大气状况和过程与历史上曾经出现过的大气状况和过程相比较, 从中找出一个或几个与当前大气状况和过程最相似的个例, 假定相似原因将产生相似的结果, 那么就可按照历史上出现过的最相似个例来类推预测今后将要出现的天气过程和天气现象。这是一种非线性预报方法。因为外推是按照大气本身的演变规律进行的, 它包含了大气本身所有的非线性变化^[1]。

1 方法和步骤

统计预报中通常采用相似系数和距离系数来表征两个变量场的相似程度。这两个系数意义是不同的, 相似系数着眼于两个变量场的分布规律是否一致, 即形态是否相似, 而不具体考查每个变量的数值差异大小; 而距离系数则是主要考查变量场中每个变量的数值差异大小。

为了找出的相似样本不仅形态相似, 而且数值差异也较小, 所以本文采用相似系数和距离系数综合分析, 以期提高预测效果, 具体步骤如下。

1.1 筛选因子

长江中下游地区汛期降水与 100hPa、500hPa 和海温场之间存在着明显的隔季相关已是公认的事实,但是并非场中每一个格点的数值都与汛期降水存在明显的相关,为此在相似分析之前,必须对备选因子场进行筛选,剔除与预报对象关系不好的因子。本文采用常用的相关普查法来筛选预报因子。

求出备选因子场中每一个格点数值与预报对象之间的单相关系数 $\tau_{x(k)y}$, 给定相关系数临界值 τ_0 , 选取 $|\tau_{x(k)y}| \geq \tau_0$ 的因子作为正式预报因子, 设共取得 m 个, 构成预报因子集 $X(x_1, x_2, \dots, x_m)$ 。

1.2 因子预处理

为了消除因子量纲(量级)不同造成的影响, 须对所有预报因子进行标准化处理。

1.3 样本相似程度

用实时资料的预报因子集 X_{n+1} 和 n 个历史样本的预报因子集 $X_i (i=1, 2, \dots, n)$ 逐一按式(1)和(2)分别计算相似系统和距离系数。

相似系数:

$$r_{n+1,i} = \frac{\sum_{k=1}^m (x_{ik} - \bar{x}_i)(x_{n+1,k} - \bar{x}_{n+1})}{\left[\sum_{k=1}^m (x_{ik} - \bar{x}_i)^2 (x_{n+1,k} - \bar{x}_{n+1})^2 \right]^{\frac{1}{2}}} \quad (1)$$

距离系数:

$$d_{n+1,i} = \left(\sum_{k=1}^m (x_{ik} - x_{n+1,k})^2 \right)^{\frac{1}{2}} \quad (2)$$

1.4 相似区间叠合集成预报

对一个因子集, 把求出的 n 个 $r_{n+1,i}$ 和 n 个 $d_{n+1,i} (i=1, 2, \dots, n)$ 各自按照其表征相似程度由大到小的顺序排列, 即 r 值由大到小, d 值由小到大排列。

选取前 L 个 r 和 d , 则这 $2L$ 个历史样本的预报因子集是与实时资料的预报因子集最相似的。按照相似原因产生相似结果的原则, 这前 $2L$ 个最相似的样本对应的预报对象的实况值, 就是 $n+1$ 时刻预报对象最可能出

现的值。

对一个因子场, 有 $2L$ 个预报对象的预报值, 如有 P 个因子场, 则有 $2L \times P$ 个预报值, 这些值是不尽相同的, 为此, 本文提出用相似区间叠合集成法来进行预报集成, 具体过程如下:

对一个因子集, 在相似系数 r 确定的 L 个最相似样本对应的预报值中, 取最接近的两个值作为预报量区间 A ; 同理, 距离系数确定的相应 L 个预报值中, 取最接近的两个值为预报区间 B , P 个因子集, 构成 $2P$ 个预报量区间 $A_i, B_i (i=1, 2, \dots, P)$, 然后把 $2P$ 区间在预报量数轴上进行叠合, 重复次数最多的区间作为预报对象的最后预报值范围。若有 2 个因子集其示意图如图 1 所示。

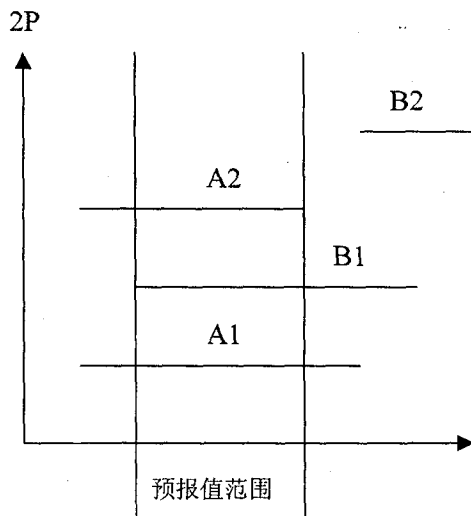


图1 相似区间叠合集成示意图

2 1999 年预报试验

①用 1~4 月北半球 500hPa、100hPa 月平均位势高度场和北太平洋月平均海表温度场作为预选因子场, 预报 1999 年安徽省沿淮、淮北、江淮之间、沿江、江南 4 片汛期 6~8 月总降水量。样本长度为 1951~1998 年, 共 48 年 (100hPa 资料为 1956~1998 年)。

分别计算 500hPa、100hPa、海温场各个格点值与 4 片汛期降水量的单相关系数, 取 $\tau_0 = 0.29$ (即 $\alpha = 0.05$)。把每个场中相关系数 $|\tau| \geq \tau_0$ 的格点值 (设有 m 个格点) 构成预报因子集 X_i ($i=1, 2, \dots, P$), 并且设定当 $m < M \times 5\%$ 时 (M 为预选因子场总格点数), 则认为此预报因子场与降水量相关不明显, 舍去。

②对因子集中每个因子进行标准化处

表 1 淮北片 6~8 月相似年降水量/mm

月份	500hPa 因子				100hPa 因子				北太平洋海温			
	r		d		r		d		r		d	
	相似年	降水量	相似年	降水量	相似年	降水量	相似年	降水量	相似年	降水量	相似年	降水量
1	1988	257	1967	415	1983	452	1986	351	1988	257	1989	615
	1987	478	1951	405	1988	257	1987	478	1977	375	1985	315
	1966	236	1989	615	1993	408	1966	236	1961	348	1982	520
2	1983	452	1973	322	1998	627	1988	257	1994	306	1954	873
	1970	310	1953	349	1987	478	1981	353	1986	351	1961	349
	1996	653	1952	466	1983	452	1987	478	1988	257	1955	461
3	1973	322	1953	349	1998	627	1961	348	1988	257	1953	349
	1994	306	1971	466	1995	522	1977	375	1993	408	1960	436
	1977	375	1966	236	1988	257	1988	257	1977	375	1985	315
4	1988	257	1975	469					1988	257	1953	349
	1970	370	1979	537					1960	436	1973	322
	1998	627	1981	353					1992	310	1976	344

注: 空白处说明相关不明显。

③从相似年中按相似区间叠集成法的原则, 在 3 个相似年中取预报量最相近的两年作为一个预报区间, 共 22 个区间, 然后将 22 个区间分别描绘在预报量的数轴上, 重叠次数最多的区间为 312~349mm。同时考虑相似年出现最多的年份为 1988 年, 其年降水量为 257mm。1999 年淮北降水可能出现的范围为 257~349mm。同样, 计算出江淮之间、沿江、江南三片的 1999 年预报区间结果列于表 2。

表 2 1999 年 6~8 月降水量预报和实况值/mm

	预报区间	实况	多年平均
淮北	312~349	246	441
江淮之间	431~483	599	456
沿江	745~760	1355	544
江南	628~694	1431	577

理。按式 (1) 和 (2) 计算每个因子集历史样本与 1999 年实时资料的相似系数 $r_{n+1,i}$ 和距离系数 $d_{n+1,i}$ ($i=1, 2, \dots, P$), 取前 3 个为最相似年。对一片降水量而言, 1~4 月 500hPa、100hPa 和海温共 12 个因子场, 最多可取 72 个最相似年, 现将三类因子场和淮北片汛期 6~8 月降水量的相似计算结果列于表 1。

从表 2 中可以看到四片的趋势预报和实况是一致的。但南部两片数值差异较大。

3 讨论

①相似预报是从历史上已经发生过的一系列事件中寻找最相似的个例, 它不可能预报出历史样本以外的值, 因此, 只有样本足够大, 涵盖因子所有的可能组合情况, 并且只有当实时资料也在样本出现过的范围内, 才能取得较好的预报效果。

②相似系数和距离系数都是表征相似程度的系数, 但两者的相似含义不同, 同时采用两种系数, 即考虑到场与场之间形态分布的相似, 又考虑到两个场之间变量数值上相近, 有利于预报准确性的提高。

③相似年集成的方法有多种, 最简单, 也

是最常用的,是所谓“举手表决法”。即把所找出的相似年中,重复出现次数最多的那一年作为最相似年,把其所对应的预报量的实况值作为预报年的预报值。在业务实践中,往往找出的相似年并不重复,或重复次数相同的年份不只一个,如果对应的预报量趋势相反则增加了判别的难度。本文提出的相似区间叠合集成法是根据“集合论”的原理,在若干个相似区间中(每一个区间可看作是一个集合),找出其中多数区间所包含的公共子区间,它是所有找出的相似年中,预报量分布最集中的区域,即未来预报量最可能出现的范围。为了检查举手表决法与叠合集成法的效果,作者用1989~1998年共10年资料进行预报试验,结果列于表3。

从表3中可以看出用相似区间叠合集成法正负趋势准确率为80%,举手表决法为60%,说明叠合法优于举手表决法。

④相似预报法是一种非线性方法,它可以避免某些统计方法(如回归)由于线性化所带来的预报值趋向平均值的弊端。相似区间叠合集成法是把多个相似区间的公共子区间找出来,因此它有平滑作用,就降水量

表3 淮北片1989~1998年6~8月降雨量距百分率预测和实况对比

年份	实况	叠合法预报值	相似年	及雨量
1989	+39	-6~-1	1988	-42
1990	+7	+6~+8	1988	-42
1991	+29	+6~+7	1990	+7
1992	-30	+2~+8	1987	+18
1993	-7	-27~-20	1986	-20
1994	-31	-42~-34	1993	-7
1995	+18	+2~+11	1987	+8
1996	+48	+1~+2	1986	-20
1997	+4	+2~+7	1983	+2
1998	+42	+2~+7	1983	+2

注:淮北6~8月降雨量多年平均值为441mm

而言,异常偏多或异常偏少的样本往往会平滑掉。从表2、表3可以看出,相似法可以报出1999年6~8月淮北降水量异常偏少的趋势;叠合集成法对1989~1998年淮北6~8月降水预报效果的检验中,预报偏少比预报偏多的效果好,这可能是降水异常偏少是大尺度天气系统长期稳定影响的结果,而降水异常偏多除上述原因外,往往受中小尺度天气系统影响更大。

参考文献

- 1 严华生等.非线性统计预报方法及其应用.昆明:云南科技出版社,1998:103.

The Application of Consensus Method Superposed Analogue Range to Precipitation Forecast

Zhu Desheng Fang Rong

(Anhui Meteorological Observatory, Hefei 230061)

Abstract

A superposed analogue range consensus forecast method was presented. The method was applied to the precipitation forecast in flood season. The result shows that calculation is in good consistent with observations.

Key Words: forecast analogue superpose consensus method