

最优气候相似法及其在降水预报中的应用

刘 兵

(湖南省张家界市气象局, 427000)

提 要

介绍了一种短期气候预测方法——最优气候相似法,并应用于张家界地区 1999~2002 年 5~7 月总降水预报中,结果显示最优气候相似法能够有效地提高短期气候的预测能力,特别对气候异常具有良好的反应能力,适合于短期气候业务预报。

关键词: 最优 相似 气候预测 气候异常

引 言

由于影响气候变化的因素极为复杂,目前人们对短期气候各要素及其变化和短期气候过程的预测仍处于比较低的水平。为了提高短期气候的预测能力,更有效、更切实地做好防灾减灾工作,气象学者们研究了许多科学的预测气候的方法,并仍在不断地补充、更新和完善,开拓新的预报思路。在实际应用中,笔者运用了一种新的短期气候预报方法,比照美国气候预报中心研制的制作气候上温度预报的最优气候均态法(optimal climate normals)的这一提法,称之为最优气候相似法(optimal climate analogues)。最优气

候均态法实际上是在众多样本记录中选用一个最优的最近 k 年的要素平均,作为下一年该要素的预报^[1],在常规气候要素中仅对温度预报有一定的预报能力,但对气候异常的反应能力很差,实际应用中的效果不甚理想。最优气候相似法不同于短期气候预测中一般的相似预报法^[2],而是提供了一种新的预报思路,在预测过程中力求反映气候要素变化的内在规律,并突出对气候异常的反应能力,在实际应用中对各种气候要素的短期预测具有良好的效果,适合于业务预报。

1 基本原理

最优气候相似法实际上借鉴了最优气候

均态法和一般相似预报法的部分思想,又有其独有的特色,提出的基本原理是:把预报年最近 k 年的气候要素统计特征与历史上曾经出现过的任意连续 k 年的该气候要素统计特征作相似比较,从中找出一个或几个与最近 k 年气候统计特征及变化趋势最相似的个例,称为气候相似区间,形成一个气候相似区间序列,假定所有气候相似区间的下一年也相互具有相似的气候特征,并得到一个新的随机抽样气候相似要素序列。通过对 k 取不同的值(如:1,2... n)可以得到 n 个具有不同程度相似气候特征的抽样相似序列,取其中相似最优的一个或几个抽样气候相似序列的平均值来类推该预报年的要素预报值(图略)。这是一种非线性预报方法。因为外推是按照气候要素特征本身的演变规律进行的,根据非线性统计预报方法理论^[3],它包含了气候要素本身所有的非线性变化。

2 方法和步骤

为了反映单要素不同时段的气候统计特征的相似程度,本文设计了两个相似系数:气候均值相似系数和气候波动相似系数。气候均值相似系数着眼于两个时段要素值的总体数值差异大小,而气候波动相似系数是在此基础上的加强条件,主要考查这两个时段的要素值变化的相似程度。两者综合考虑有利于提高预测效果,具体步骤如下。

2.1 研究对象

要研究的对象为某种气候要素 X 。设某站该要素年、季或月气候要素值构成序列 $\{X_i\}$, $i = 1, 2, \dots, N, N = EY - BY + 1$, 其中 EY 为预报年, BY 为起始年。求预报年该气候要素值水平。要求 N 值越大越好,有利于得到更好的预报结果。

2.2 构造气候特征值序列并计算气候特征值

气候统计特征有多种,它们从不同方面描述气候的特征,可以选取一个或多个来构

造气候特征值序列,这里我们选取要素平均这一最基本的统计特征来构造序列:

$$\overline{x_{i,k}} = \frac{1}{k} \sum_{j=0}^{k-1} x_{i+j}$$

其中 $i = 1, 2, \dots, N - k, k = 1, 2, \dots, n$ (一般取 $n = 5$ 就足够了), k 为相似时段长度,则 $x_{i,k}$ ($i \leq N - k - 1$) 表示观测样本中第 i 个样本记录年开始连续 k 年的要素值平均,称为比较气候特征值, $\overline{x_{N-k,k}}$ 表示最近 k 年的要素值平均,称为最近 k 年的参考气候特征值。

2.3 定义并计算气候均值相似系数 $\beta_{k,i}$

由于是一维变量,现有的描述相似系数的公式在这里已不适用,为了更准确的描述气候要素不同阶段的相似程度,笔者研究了多种气候要素的分布和变化特点及相关的统计特征,精心设计了如下适合求各种常规要素的气候均值相似系数公式:

$$\beta_{k,i} = 1 - \frac{|\overline{x_{i,k}} - \overline{x_{N-k,k}}|}{\frac{1}{2}(x_{\max} - x_{\min}) + |\overline{x_{N-k,k}} - \overline{x}|}, \quad \beta_{k,i} \leq 1$$

其中 $i = 1, 2, \dots, N - k, x_{\max}, x_{\min}$ 分别为历史最大与最小值, \overline{x} 为气候均值(世界气象组织推荐最近 30 年的平均作为气候均值)。

2.4 定义并计算气候波动相似系数 $\zeta_{k,i}$

与气候均值相似系数相仿,定义气候波动相似系数为:

$$\zeta_{k,i} = 1 - \frac{\frac{1}{k} \sum_{j=0}^{k-1} |x_{i+j} - x_{N-k+j}|}{\frac{1}{2}(x_{\max} - x_{\min}) + |\overline{x_{N-k,k}} - \overline{x}|}, \quad \zeta_{k,i} \leq 1$$

其中 $i = 1, 2, \dots, N - k$ 。

2.5 构造气候相似区间序列

上述两公式除了能较好的描述气候要素各阶段的相似程度外,还有两个显著的优点:一是对于给定的常规气候资料,无须任何预处理即可使用,因为公式已经包含了某种标准化处理,消除了气候要素本身带来的量级差异;二是对于服从正态分布的气候要素,

在最近 k 年出现异常偏高或异常偏低时有较强的反应能力。给定均值相似水平 β_0 ($\beta_0 \leq 1$) 和波动相似水平 ζ_0 ($\zeta_0 \leq 1$), 当 $\beta_{k,i} \geq \beta_0$ 与 $\zeta_{k,i} \geq \zeta_0$ 同时满足时, 即认为从第 i 个记录年开始连续 k 年期间的气候特征与最近 k 年总体上是相似的, 此连续 k 年作为一个历史相似个例入选气候相似区间序列 $\{\text{range}_k(p)\}$, $p = 1, 2, \dots, q, q \leq N - k - 1$ 。 β_0 与样本容量的大小有关, 样本容量越大取值可越大, 40 年以上可取 $\beta_0 \geq 0.9$, ζ_0 的值与最近 k 年的气候波动情况有关, 波动越大取值越大。

2.6 构造抽样气候相似要素序列并计算序列内要素平均 μ_k 、均方差 σ_k

依顺序选取气候相似区间序列 $\{\text{range}_k(p)\}$ ($p = 1, 2, \dots, q, q \leq N - k - 1$) 内各相似区间下一年的气候要素值构造一个新的序列 $\{y_k(p)\}$, $p = 1, 2, \dots, q, q \leq N - k - 1$, 称为抽样气候相似要素序列。按下列公式计算序列内要素平均 μ_k 与均方差 σ_k :

$$\mu_k = \frac{1}{q} \sum_{p=1}^q y_k(p),$$

$$\sigma_k = \sqrt{\frac{1}{q} \left(\sum_{p=1}^q y_k(p) - \mu_k \right)^2}$$

2.7 计算抽样序列内各要素值的总体相似系数 η_k

仿照气候相似系数公式, 可以写出抽样序列内要素的总体相似系数公式:

$$\eta_k = 1 - \frac{\sigma_k}{\frac{1}{2}(x_{\max} - x_{\min}) + 2|\mu_k - \bar{x}|},$$

$$\eta_k \leq 1$$

该公式同样考虑到了当 μ_k 异常偏大或异常偏小时, 对于准正态分布的气候要素观测值较稀疏(特别是原始样本容量不够大时), 因而均方差也较大的事实, 因此在计算总体相似系数时有意识的适当增大了当 μ_k 异常偏大或异常偏小时对均方差的容忍度。由公式知序列内均方差越小 η_k 越大。

2.8 选取一个或几个最优抽样气候相似要素序列并得出预报年的气候要素预报值

分别求出当 $k = 1, 2, \dots, n$ 时 η_k 的值。给定总体相似水平 η_0 , 如果有两个或两个以上 $\eta_k \geq \eta_0$, 选取对应几个序列为最优抽样气候相似要素序列, 对应 μ_k 值的平均确定预报年的气候要素预报值; 否则选取其中 η_k 最大的序列为最优抽样气候相似要素序列, 取对应的 μ_k 值为预报年的气候要素预报值。当 $\eta_k = 1$ 时, 建议舍掉, 因为这时只从历史资料中选出一个相似样本, 有时并不稳定。在实际应用中, 由于 ζ_0 对最近 k 年的波动特征是否明显比较敏感, 特别对于样本偏少时要经过试报反复调试, 以得到最优的气候相似区间序列, 提高抽样相似要素序列的稳定度。所得预报值加减所在最优气候相似要素序列的平均标准差可做为预报值的参考范围。

3 1999~2002 年预报实验

根据本地气候特点, 5~7 月(主汛期内)降水最为充沛, 实验通过预报 1999~2002 年逐年张家界市市区、桑植县、慈利县三地汛期 5~7 月总降水量, 尝试它在降水预报中的应用。所用资料为张家界市以上三站 1957~2001 年 5~7 月总降水量资料, 最大原始样本长度为 45 年。若预报本站 Y 年 5~7 月总降水量, 则实际用到的只是本站自 1957 年至 $Y-1$ 年 5~7 月的月降水资料。

①取最长相似时段长度为 5 年, 按上述方法分别找出当 $k = 1, 2, 3, 4, 5$ 时各地逐年最近 k 年的气候相似区间序列。

②调试 β_0 和 ζ_0 的值(见表 1), 得到最优气候相似要素序列(见表 2)。表 1 中由于慈利最近 k 年降水量波动相对较小, 故 ζ_0 取值较小, 这时要增大 β_0 以找到最优气候相似要素序列。

表 1 1999~2002 年试报期给定的最优相似系数

站名	β_0 值	ζ_0 值
张家界	0.90	0.820
桑植	0.90	0.693
慈利	0.99	0.400

表2 张家界地区各地试报年最优气候相似要素序列最优k值及相似年降水量/mm

预报年份	张家界			桑植			慈利			
	最优k值	相似年	降水量	最优k值	相似年	降水量	最优k值	相似年	降水量	
1999	2	1970	674.1	2	1968	414.8	5	1968	390.3	
		1996	982.1		1971	638.4		1983	1014.0	
	3	1970	674.1	3	1980	1135.6	3	1987	621.5	
		1996	982.1		1994	518.6		1996	1226.0	
	2000	4	1971	468.1	2	1965	270.4	3	1960	593.3
			1997	502.7		1981	428.4		1985	493.1
2001	3	1960	609.2	2	1966	537.2	1*	1995	560.5	
		1966	495.6		1979	776.2		3	1965	319.3
	2	1967	1089.6	3	1962	669.1	4	1966	633.5	
			1986		728.1	1963		862.1	1970	622.5
2002	5	1990	682.5	5	1990	566.8	4	1971	471.2	
		1970	844.7		1986	753.6				
	2	1967	1089.6	3	1962	669.1	4	1966	633.5	
			1986		728.1	1963		862.1	1970	622.5
3	1966	495.6	2	1966	537.2	1*	1995	560.5		
		1960		609.2	1979		776.2	3	1965	319.3
4	1971	468.1	2	1965	270.4	3	1960	593.3		
		1997		502.7	1981		428.4	1985	493.1	
5	1962	669.1	3	1962	669.1	4	1966	633.5		
		1986		728.1	1963		862.1	1970	622.5	
6	1968	414.8	2	1968	414.8	5	1968	390.3		
		1971		638.4	1983		1014.0			
7	1970	674.1	3	1980	1135.6	3	1987	621.5		
		1996		982.1	1994		518.6	1996	1226.0	
8	1965	270.4	2	1965	270.4	3	1960	593.3		
		1981		428.4	1985		493.1			
9	1966	537.2	2	1966	537.2	1*	1995	560.5		
		1979		776.2	3		1965	319.3		
10	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
11	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
12	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
13	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
14	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
15	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
16	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
17	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
18	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
19	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
20	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
21	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
22	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
23	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
24	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
25	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
26	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
27	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
28	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
29	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7
30	1966	495.6	2	1966	495.6	3	1965	319.3		
		1960		609.2	1979		776.2	4	1997	397.7

注：*处各最优k值均只有一个相似年，计算预报值时取其平均。

③由表2可以算出张家界地区三地1999~2002年5~7月总降水量，将其结果列于表3。

表3 张家界地区1999~2002年5~7月(主汛期)总降水量试报情况/mm

预报年份	张家界			桑植			慈利		
	实况	预报	误差	实况	预报	误差	实况	预报	误差
1999	862.2	828.1	34.1	831.6	786.7	44.9	645.0	675.3	-30.3
2000	361.4	485.4	-124.0	304.7	349.7	-45.0	548.3	543.2	5.1
2001	329.4	551.9	-222.5	313.7	568.2	-254.5	321.8	456.3	-134.5
2002	946.0	833.4	112.6	744.6	707.1	37.5	680.6	620.2	60.4

从表3可以看出，用最优气候相似法对1999~2002年张家界三地均做出了相对准确的预报，其中1999年最好，2000年和2002年次之，2001年最差。这是由于2001年连续第二年出现了异常偏少的现象，这在历史资料中还是第一次，因此在数值上出现了较大的偏差，且均为预报值偏大。这也反映出样本资料偏少时对预报准确率的影响。但对近30年5~7月总降水量统计量分析(见表4)可见，对1999年张家界和桑植明显偏多，2000年张家界和桑植异常偏少，及2002

年张家界异常偏多都有良好的表现，证明最优气候相似法对气候异常具有很好的反应能力。

表4 张家界地区3测站近30年(1971~2000)5~7月总降水量统计量/mm

站点	统计年数	最小值	最大值	<400mm年数	>810mm年数	平均值	平均差	标准差
张家界	30	352.0	1153.4	3	4	621.7	162.3	198.0
桑植	30	304.7	1419.0	3	8	680.6	206.9	267.5
慈利	30	286.7	1292.4	4	4	606.8	165.6	219.6

表5 张家界地区3测站1999~2002年5~7月(主汛期)总降水量预报趋势效果检验

站名	张家界	桑植	慈利
1999	实况	+	+
	预报	+	+
2000	实况	-	-
	预报	-	-
2001	实况	-	-
	预报	-	-
2002	实况	+	+
	预报	+	+

表5列出了试报期总降水量距平趋势预

报情况,可以看出最优气候相似法对张家界各地逐年5~7月总降水量趋势均做出了准确的预报,比预想的要好。这说明最优气候相似法在短期气候趋势预测中具有良好的效果。

4 讨论

用最优气候相似法作短期气候预测,能够有效的提高短期气候的预测能力,提高业务预报水平,适合于短期气候的业务预报。其主要优点表现在:

①气候均值相似系数和气候波动相似系数都是表征相似程度的系数,但两者的相似含义不同,两者既有联系又有区别,既相互制约又相互加强,同时采用两种系数有利于预报准确性的提高。

②利用加强的相似系数力求反映气候要素变化的内在规律,利用抽样气候相似要素序列的随机性突出对气候异常的反应能力,在实际应用中,特别是对气候异常的短期预测中有优异的表现。

③最优气候相似法是一种非线性预报方法,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。同时由于它本身包含了气候要素本身所有的非线性变化,可以不需考虑诸如青藏高原、本地地形、厄尔尼诺、拉尼娜、冰雪及陆地表面状况等复杂因素的影响。

④所需的计算资源容易满足,同一套公式适用于各种气候要素,易于编程实现。应用于月预报可提前一年作出预报,季预报可提前9个月,制作简单而实用。

其主要局限性在于:

①相似预报是从历史上已经发生过的一

系列事件中寻找若干个最相似的个例取其平均,它不可能预报出历史样本以外的值。40多年的历史样本只是反映气候要素变化的一小部分,而且气候变化还有新的表现,因此只有样本足够的大,尽可能多的涵盖气候要素本身可能变化的情况,才能取得更好的预报结果。

②对于新出现的气候波动变化现象,最优气候相似法有一个适应的过程,并不能马上作出很好的反应。

在实际应用中应注意和有待探讨的是:

①对于气候均值相似系数水平值 β_0 和气候波动相似系数水平值 ζ_0 ,在实际应用中应根据本地气候要素的变化特征和历史资料的多少进行调整,如果气候要素波动特征明显, ζ_0 值可取大些,否则应将 ζ_0 值取小,并增加 β_0 值。它们共同决定了历史上相对于最近 k 年的气候相似区间的取舍。对于同一地点同一要素,分别作连续年的预报时 β_0 值与 ζ_0 值一般是稳定的。

②最优气候相似法提供了一种新的短期气候预报思路,通过其在降水预报中的初步应用得到了部分检验,并取得了良好的效果,但作为一种新的短期气候预测方法,还有待进一步完善,有待在不同的地理区域作更进一步的检验与探讨。

参考文献

- 1 罗晓玲. 最优气候均态法作广东省月平均气温预报, 广东气象, 1998. 4(增2): 31~32.
- 2 朱乾根等. 天气学原理和方法(修订本). 北京: 气象出版社, 1992: 338.
- 3 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.

Optimal Climate Analogues and Application to Precipitation Forecast

Liu Bing

(Zhangjiajie Meteorological Office, Hunan Province 427000)

Abstract

A climatic forecast method, optimal climate analogues, is presented. The method was applied to the precipitation forecast in flood season. The result shows that the calculation is consistent with observations, specially in climatic anomalies.

Key Words: optimal analogue climatic forecast climatic anomalies