

大气环流异常分析中 使用合成方法的一个问题

施 能

古文保*

(南京气象学院, 210044)

(广西河池地区气象局)

提 要

通过一个实例的研究,指出,如果仅根据差值的绝对值大小来挑选关键区,其结果可能不可靠。它可能导致虚假的关键区或遗漏掉重要的关键区(在低纬度)。提出了解决问题的方法。

关键词: 大气环流 合成分析 关键区 统计检验

1 问题的提出

在气象研究中,为了分析不同或相反类的气象现象的同期或前期特征及差异时,经常使用合成图、差值图。例如,对旱年、涝年分别求同期或前期某环流场(例如 500hPa)的平均场或差值场。差值场的中心(绝对值)往往被认为影响旱涝的关键区。这种方法简单,应用较多。但是,所分析的环流场各格点的均方差变化比较大,很不均匀时,差值图所得的关键区可能不可靠:它有可能将均方差较小的关键区漏掉;另一方面,又可能将均方差较大的区域误认为影响旱涝的关键区。这是因为均方差大的地区、高度值变化大,随机样本之间的差值也大。即使不针对旱、涝年作差值图,差值的中心也仍可能出现在这些地方,这样,也就很难认为差值绝对值大的地区就是影响旱、涝的关键区了。为了避免这种情况,得到较可靠的结果,方法之一是对差值做差异显著性检验。将经过统计检验的显著区定为关键区。检验方法可用

$$t_i = \frac{\overline{x_i(1)} - \overline{x_i(2)}}{\sqrt{\frac{n_1\sigma_{1i}^2 + n_2\sigma_{2i}^2}{n_1 + n_2} \left(\frac{1}{n_1} + \frac{1}{n_2} \right)}} = \frac{\overline{x_i(1)} - \overline{x_i(2)}}{\sigma_i \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \quad (1)$$

式(1)中 n_1, n_2 分别为旱年、涝年数。 $\overline{x_i(1)}, \overline{x_i(2)}$ 就是第 i 格点,旱年、涝年的平均值。 $\sigma_i, \sigma_{1i}, \sigma_{2i}$ 分别是第 i 格点均方差,旱年均方差、涝年均方差。或者用 Mahalanobis 平方距离

$$Maha_i = \left[\frac{\overline{x_i(1)} - \overline{x_i(2)}}{\sigma_i} \right]^2 \quad (2)$$

这两种方法本质相同,因为

$$Maha_i = t_i^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \quad (3)$$

计算出 $|t_i|$ 或 $Maha_i$ 后,将它们的值与自由度为 $n_1 + n_2 - 2$ 的临界值比较,大于信度标准下的格点 i 组成的区定为关键区。由式(3)可知, Mahalanobis 平方距离的临界值

$$Maha_c = t_c^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) = t_c^2 \left(\frac{n_1 + n_2}{n_1 n_2} \right) \quad (4)$$

* 南京气象学院 1993 年硕士生

所以,综上所述,合成分析中,将 $|\overline{x_i(1)} - \overline{x_i(2)}|$ 的大值区作为关键区,结果可能不可靠。而式(1)、式(2)中除了 σ_i ,消除了 σ_i 不均匀造成的后果。

需要指出的是,如果不消除 σ_i 的不均匀性,将关键区高度作为统计预报模型中的预报因子来使用,则会影响到统计预测效果,使预报不可靠。

2 实例

广西早春旱前期1月份500hPa环流异常形势研究。取旱、涝年如下:

旱年:1954,1962,1966,1969,1971,1974,1975,1976,1977,1984

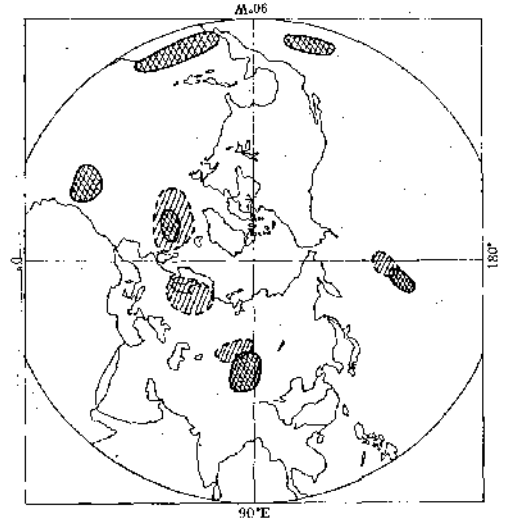
涝年:1952,1953,1956,1957,1959,1963,1981,1983,1985

$n_1 = 10, n_2 = 9$ 。500hPa资料用1951—1985年,10—85°N,5°纬度×10°经度,共576个格点。格点序号同《气候监测公报》。附表是计算结果。

因为自由度为17,取0.05信度, $t_{\alpha} = 2110$,所以 $Maha_{\alpha} = t_{\alpha}^2 \cdot (\frac{19}{90}) = 0.94$,这样Mahalanobis平方距离 ≥ 0.94 时,就达到0.05信度。由附表右半部看出,根据Mahalanobis平方距离从大到小挑选共有22个格点达到0.05信度。这22个格点中除了有个别2个点不成片外,共组成6个区,见附图中黑影区域。如果根据差值的绝对值从大到小也选出22个点(需 $\geq 5.83 \times 10\text{gpm}$,附图中斜线区)。可以看出,两种方法所得的结果有较明显的区别,表现在

2.1 差值图上显示不出20°N以南信度达0.05的关键区,实际上是给遗漏掉了。

2.2 在中高纬度,差值图所显示的关键区域范围偏大,并且位置也有所变动。



附图 广西早春旱前期1月份500hPa关键区
网影区:信度达0.05,斜线区:据差值绝对值选取(未检验)

附表 广西早春旱前期1月500hPa关键区

序号	$\overline{x_i(1)} - \overline{x_i(2)}$		Maha值		
	值/10gpm	格点序	Maha值	值	格点序
1	-9.90	250	1.43	2.25	570
2	-9.30	285	1.12	1.50	333
3	-9.20	286	1.23	1.47	569
4	-9.12	251	1.41	1.45	565
5	-7.39	249	0.79	1.43	250
6	7.36	183	1.06	1.41	529
7	-7.34	214	0.57	1.41	251
8	-7.17	284	0.66	1.29	378
9	-6.96	287	0.87	1.29	466
10	-6.81	213	0.44	1.25	297
11	6.54	220	0.85	1.23	286
12	6.47	182	0.77	1.23	502
13	6.22	148	0.86	1.15	568
14	6.15	184	0.80	1.12	501
15	-6.14	260	1.06	1.12	285
16	6.09	147	0.74	1.06	183
17	6.03	378	1.29	1.06	260
18	6.03	342	0.53	1.03	332
19	6.02	219	0.72	1.02	339
20	-5.94	225	0.85	0.98	415
21	-5.91	259	0.89	0.97	566
22	-5.83	321	0.42	0.94	261

2.3 在中高纬度,差值图上出现一个虚假的关键区,格点序号为 183,148,147,182,184,219,220(55—65°N,20—40°E)。这个区域实际上只有一个孤立点信度达到 0.05,其余的差异均不显著,不能认为是关键区。

2.4 由附表看出,根据差值绝对值最大挑选的 22 个点中,有 15 个点 Maha 值 < 0.94 ,达不到 0.05 信度。而根据 Maha 值,差异最显著的点是在低纬度,格点序号为 570,333,569,565,这个最显著的区域在差值图上给漏掉了。

需要指出,本文给出的合成分析中的个例数并不很少($n_1 = 10, n_2 = 9$),但已经出现了不可靠的结果。如果计算平均场或差值场时的个例数非常少(< 5),那么产生虚假或遗漏的可能性就更大了。这可能就是某些合成分析结果的天气意义不好解释或预测错误的一个重要原因。

3 结论和讨论

3.1 在合成分析中,如果仅用差值绝对值大

值中心来表示关键区,则可能会遗漏掉重要的关键区或将一些不重要的区域误认为关键区。差异显著性统计检验是解决方法之一。

3.2 如果合成分析不是在空间点上进行,而是在时间点上进行,那么均方差随时间点不均匀时,也会发生类似的问题,解决方法类似。

3.3 差异显著性检验仅是一种常用的处理方法。还有如下两种处理方法。第一,将格点资料作归一化处理,各格点除以该格点的均方差以后,再制作差值图。然后将绝对值大的区域定为关键区。第二,直接计算相关系数。将格点的历史序列资料与所分析的对象的历史序列资料(例如,降水量)求相关系数。这时,经统计检验,绝对值大的区自然就是对降水量关系密切的关键区了。因为,对格点资料的任何线性变换后,相关系数绝对值是不变的。所以,对格点资料不必进行归一化或标准化处理。但是,某些天气现象不是能用定量数据表示的(例如,有无厄尔尼诺事件发生),这时,就不能用这种方法了。

A Note on Composite Analysis of Atmospheric Circulation Anomalies

Shi Neng

Gu Wenbao

(Nanjing Institute of Meteorology, 210044)

(Hechi Meteorological Bureau of Guangxi Province)

Abstract

It is pointed out, that with a real study example, the key regions may be unreliable if only according to the absolute values of difference of the mean values to pick out the key regions. That may result in false regions and miss the important ones. The methods of solving this problem are presented.

Key Words: atmospheric circulation composite analysis statistical test