第 20 卷第 3 期
1997 年 9 月

南 京 气 象 学 院 学 报
Journal of Nanjing Institute of Meteorology

Vol. 20 No. 3
Sep. 1997

气象场相关分析及合成分析中蒙特卡洛检验方法及应用*

施 能[1)] 魏风英[2)] 封国林[3)] 沈桐立[1)]

(1) 南京气象学院气象学系, 南京 210044; 2) 中国气象科学研究院, 北京 100081;
3) 扬州大学基础科学部, 扬州 225009)

摘 要 指出了气象要素及气象要素场显著性检验中的问题, 介绍了解决这种问题的蒙特卡洛模拟试验方法。若干要素及气象场相关场检验的蒙特卡洛试验表明, 所得结果可以在实际工作中参考使用。

关键词 相关分析, 合成分析, 蒙特卡洛检验
分类号 O212

在气象研究与预报中经常要进行相关分析及合成分析, 根据相关系数的大小指出密切的相关区及相关变量, 用合成分析方法确定前期(或同期) 大气环流在不同的天气、气候状态(某要素正、负距平, 强、弱季风, ……) 下, 后期(或同时) 另一要素场或环流场有无明显差异, 以确定不同气候态造成的影响程度。这些方法能否成功的关键在于所得的结果是否具有统计学意义及物理解释。对于第二个问题, 需要对研究对象做深入细致的物理成因、经验等多方面的研究, 往往不是统计方法本身所能解决的。对于第一个问题, 属于统计检验问题, 它是研究第二个问题的基础, 也是一个重要的问题。除非客观条件不允许进行这项工作, 否则, 当所得的结果不具有统计学意义时, 就只能认为是一个随机的不可信的结果, 从而无法进行物理本质的探索, 而且有可能造成误导, 使研究与预报走向歪路。

长期以来, 对单相关系数的统计检验主要用 t 检验方法。对于高度自相关的变量之间的单相关系数的检验, 可以用 Davis(1976) 提出的用有效自由度调整 t 检验自由度的方法[1,2]。对于合成分析, 我们可以用样本平均值差异的 t 检验方法。然而, 这些方法都需在变量服从正态分布条件下使用。而我们在相关系数检验时, 往往不对是否服从正态分布进行检验。此外, 这些检验方法并不适用于大量、重复的试验。当我们将一个气象要素序列与成千上万个与气象毫不相干的数据序列求相关时, 肯定能找出 "高相关 ", 并且能通过相关系数 t 检验的标准。高速电子计算机又为我们进行这种相关计算提供了可能。如何防止或检测这种虚假的高相关呢?

另一个更为重要的问题是气象场的相关问题。我们经常需要研究一个气象变量(例如: 季、月降水量, 季、月气温等) 与一个气象场(例如: 前期或同期的海温场, 500 hPa 高度场, ……) 是否相关。因为气象场有许多格点或测站组成, 相关系数计算类似于大量、重复试验。有多少个格点或测站通过显著性检验才能认为要素与气象场之间有相关呢? Livezey[3] 与 Gordon[4] 早

收稿日期: 1996- 10- 28; 改回日期: 1997- 01- 10
* 自然科学基金 49675260 资助项目
第一作者简介: 施 能, 男, 1942 年 5 月生, 教授

已提出过这个问题, 并给出过解决方法。显然, 如果气象场中格点值是相互独立的, 则可以用二项分布计算出达到显著性标准的格点数的下限。但事实上, 气象场的空间格点有相关性, 而且不同的气象场有不同紧密程度的空间相关性。这样就不可能有统一的模式来解决变量与气象场相关的显著性检验问题。Livezey[3]曾用统计模拟(Monte Carlo)方法来做实际相关场的检验, 这种检验需要大量的模拟计算, 但是, 在计算机高速发展的今天, 已不难实现。

本文给出了单点及气象场相关分析及合成分析中使用蒙特卡洛检验的具体方法, 并用 44 年资料对北半球 500 hPa 高度场与我国的气温、降水场作了若干模拟试验, 给出若干定性标准。

1 单相关系数及均值显著性差异的蒙特卡洛检验

1.1 相关系数的非参数检验——蒙特卡洛检验

假定有任意概率分布的气象变量 x 的 n 次观测(n 为样本容量), 与 x 的相关系数(绝对值)至少多大才能认为与变量 x 有相关呢? 蒙特卡洛方法是用随机数序列与 x 变量序列求相关系数, 长度为 n 的随机数序列由机器程序大量产生(例如 1 000～10 000 次)。假如计算了 1 000 个模拟的相关系数, 则将 1 000 个模拟相关系数的绝对值从大到小排序, 最大的序号为 1。称序号为 2、11、51 的模拟相关系数为 $r_{0.001}$、$r_{0.01}$、$r_{0.05}$。当实际计算的与变量 x 的相关系数 $|r|$ 大于上述模拟值, 才能认为有相关关系存在(信度分别为 0.001、0.01、0.05)。经我们试验, 模拟 1 000 次以上所得的临界相关系数已足够精确, 而微机计算仅要 1～2 分钟。

1.2 合成分析中样本差值显著性的蒙特卡洛检验

假定在 A、B 两种气候状态下(例如: 强, 弱南方涛动) 某气象变量 x 的样本值为

A: 94、197、16、38、99、141、23 $n_A = 7$

B: 52、104、146、10、50、31、40、27、46 $n_B = 9$

A、B 两种气候态下, 气象变量 x 有无明显不同呢? 或 A、B 两种气候态对变量 x 有无显著的影响? 通常的参数检验方法就是看总体的均值是否可信相等的 t 检验方法, 但是它需要样本来自正态总体以及方差 σ_A^2 与方差 σ_B^2 相等的假定。蒙特卡洛方法是将上述气候态下的样本值看为 $n_A + n_B$ 中任意取出的一种取法。两种气候态下有 16 个样本($n_A + n_B$), 16 个样本中任取 7 个(或 9 个)共有 $C_{16}^7 = \frac{16!}{7!\ 9!} = 11\ 440$ 种取法(一般情况下有 $\frac{(n_A + n_B)!}{n_A!\ n_B!}$ 种取法), 每一种取法可计算出一个变量 x 的样本均值差 $\bar{x}_A - \bar{x}_B$ 的绝对值 $|\bar{x}_A - \bar{x}_B|$。将所有可能的($\bar{x}_A - \bar{x}_B$)从大到小排序, 如果气候态所对应的那种取法的 $|\bar{x}_A - \bar{x}_B|$ 落入最大的 1%(5%)范围之内, 就认为 A、B 两种气候态对气象变量 x 有明显的影响, 这种推断的可靠性分别为 99%(95%), 信度为 0.01(0.05)。

从上所述, 无论是相关系数的统计检验还是合成分析中均值差异的显著性检验, 都可以用蒙特卡洛模拟方法解决。这种方法摆脱了正态分布和其他各种假定, 但计算量大。高速计算机向统计检验理论的应用价值提出了强有力的挑战。但是蒙特卡洛方法只能对一个具体的问题给出检验标准。下面我们用 1951～1994 年 44 年的资料, 进行若干气象问题的蒙特卡洛试验。与传统的统计检验结果相比较, 以得出可供借鉴的定性结果及标准。

2 气象变量相关系数的蒙特卡洛检验

取乌鲁木齐、广州、北京、成都、南京的 1951～1994 年 1 月份、8 月份的降水量与平均气温

为气象变量, 将它们分别与随机序列计算 1 000 个相关系数, 用前述的蒙特卡洛方法定出蒙特卡洛模拟标准, 见表 1。表 1 中的末行是用 t 检验方法时的临界相关系数。

从表 1 中可看出:

1) 蒙特卡洛方法的临界相关系数与 t 检验方法的结果不一致, 当信度标准比较低时(0.05 或 0.10) 蒙特卡洛方法的临界相关系数绝大多数不低于 t 检验的临界相关系数(特别是气温), 但是, 当信度高达 0.001 标准时, t 检验的临界值总大于蒙特卡洛方法。由此看来, 使用 t 检验时, 应取 0.01 以上信度标准, 0.05 以下信度的 t 检验标准有点偏低, 宜慎用。

2) 取决于要素本身的结构特征, 蒙特卡洛的临界相关系数随月份、地点、要素改变。从表 1 的结果看, 无明显规律可循。但尚可看出, 蒙特卡洛方法对 8 月份降水的相关临界值大多数(约占 16/20) 大于 1 月份降水的临界值。所以, 与 8 月份降水的相关变量宜取更高的标准。

表 **1**　降水、气温相关系数的蒙特卡洛检验标准及 ***t*** 检验标准(序列长: **44**)

Table 1　Standars of Monte Carlo and *t* tests for correlativity of rainfall and temperature (n= 44)

时间	站名	降水的检验标准				气温的检验标准			
		0.001	0.01	0.05	0.10	0.001	0.01	0.05	0.10
8 月	乌鲁木齐	0.42	0.40	0.31	0.26	0.40	0.37	0.32	0.26
	广州	0.40	0.39	0.33	0.30	0.40	0.39	0.30	0.27
	北京	0.44	0.37	0.30	0.26	0.42	0.39	0.30	0.26
	成都	0.43	0.41	0.30	0.25	0.39	0.35	0.29	0.25
	南京	0.40	0.36	0.29	0.25	0.47	0.45	0.29	0.26
1 月	乌鲁木齐	0.39	0.35	0.28	0.25	0.39	0.37	0.32	0.27
	广州	0.40	0.37	0.30	0.26	0.38	0.36	0.30	0.27
	北京	0.39	0.35	0.29	0.25	0.45	0.37	0.31	0.26
	成都	0.39	0.37	0.34	0.28	0.40	0.35	0.31	0.27
	南京	0.38	0.35	0.28	0.26	0.43	0.41	0.35	0.31
	t 检验	0.47	0.38	0.29	0.25	0.47	0.38	0.29	0.25

3　单要素与气象场的相关

由于气象场往往有许多格点或测站组成。当某一个变量与气象场求相关时, 往往需要计算很大数量的相关系数, 而其中必有一些相关系数达到较大的数值并且通过相关系数的统计检验。但这并不一定表示气象变量与该气象场存在相关。蒙特卡洛方法提供了检验单要素与气象场相关的检验方法。

首先, 我们产生一个样本容量与气象变量相同的随机序列, 将这个随机序列与气象场各格点序列求相关, 统计出在全部相关系数中, 正负相关的格点数, 达 0.05、0.01、0.001 的点数。这样的试验重复 100 次或 1 000 次。如取 1 000 次, 将 1 000 次试验所得的格点数进行从大到小排序, 序号 51 所对应格点数为 0.05 的随机试验标准。气象要素与该气象场的相关系数中, 达到统计检验 0.05 信度格点数应大于这个标准, 才能有 95% 的把握认为气象要素与该气象场存在相关。

表 2 是对序列长度为 44 时, 进行 1 000 次随机蒙特卡洛试验的结果。

表 2 单变量与若干气象场相关达到信度标准的点数的蒙特卡洛检验标准(序列长: 44)

Table 2 Number of points reaching significance that meet Monte Carlo testing standards for a meteorological variable (e. g., rainfall and temperature) in correlation with 500 hPa height in summer and winter months (n= 44)

气象要素	时间	正或负	0.05	0.01	0.001	总数
H_{500}	夏季月	420(72.9%)	58(10.1%)	18(3.1%)	2(0.35%)	576 点
	冬季月	390(67.7%)	50(8.7%)	15(2.6%)	2(0.35%)	
降水	夏季月	108(67.5%)	12(7.5%)	4(2.5%)	2(0.63%)	160 站
	冬季月	120(75.0%)	15(9.4%)	7(4.4%)	2(0.63%)	
气温	夏季月	125(84.4%)	25(15.6%)	6(6.3%)	2(0.63%)	160 站
	冬季月	135(84.4%)	25(15.6%)	10(6.3%)	2(0.63%)	

注: 括号内是指在576 个格内中的百分比

表 2 说明, 某气象变量与 576 个格点资料组成的 H_{500}场求相关时, 对夏季月份, 需至少 420 个格点(占总数 72.9%)为正(或负)相关; 58 个格点达 0.05 信度, 18 个格点达 0.01 信度; 2 个格点达 0.001 信度时才能认为气象变量与该 H_{500}场有相关。当计算结果全部达不到上述标准时, 变量与场的相关不可靠, 很可能是随机的。计算经验表明, 上述标准往往是几乎同时满足(或几乎同时达不到)。当然, 对具体的问题, 需对相关场的分布结合可能的物理解释作仔细的分析。

从表 2 还可以看出, 单变量与气象场的蒙特卡洛检验标准, 对不同的要素场, 不同的季、月标准是不完全相同的。H_{500}场夏季月份的标准要比冬季月份高些, 与气温场的相关标准要比降水场高些, 冬季月份又要比夏季月份高些。这些都是由气象场本身的结构决定的。

当某气象变量与另一气象场存在相关时, 对该气象变量的不同气候态(气象变量高值、低值)进行气象场的合成分析能得出可靠的结果, 否则合成分析的结果不会可靠。所以, 上述方法原则上适用于气象场的合成分析的统计检验。

3 结论与总结

1) 为了得到真正可靠的相关分析结果, 使用单纯的参数统计检验方法是不够的。蒙特卡洛模拟试验为相关分析提供了 "去伪存真" 的方法。这种方法需大量计算, 但是在计算机高速发展的今天, 已经完全可能实现了。

2) 从蒙特卡洛模拟试验结果看, 相关系数的 t 检验方法的信度取 0.05 时, 有可能引入虚假相关。蒙特卡洛试验结果推存 0.01 以上的相关系数参数检验。

3) 单变量与气象场相关的蒙特卡洛检验标准已经得出, 可以参考使用。但是, 蒙特卡洛方法对各类具体问题(变量、时段)进行模拟试验仍是必要的。

4) 近来, 蒙特卡洛模拟试验已广泛深入气象诊断研究各领域。例如气象场自然正交展开(EOF)时, 对特征值与模态的统计检验; 用于两个相关场的奇异值分解(SVD)模态的显著性检验[5~7]。我们很有必要在工作中熟悉并使用这种方法。

参　考　文　献

1 Davis R E. Predictability of sea-surface temperature and sea-level pressure anomalies over the North Pacific Ocean. J Phys Oceanogr, 1976, 6: 249～266

2 施　能, 气象科研与预报中的多元分析方法。北京: 气象出版社, 1995, 55～57

3 Livezey R E, Chen W Y. Statistical fieed significance and its determination by Mante Carlo techniques. Mon Wea Rev, 1983, 111(1): 46～59

4 Gordon N D. The Southern Oscillation and New Zealand weather. Mon Wea Rev, 1985, 113(1): 46～59

5 Shen S, Lau K M. Biennial Oscillation Associated with the East Asian Summer monsoon and tropical sea surfoce temperature. J Meteor Soc Japan, 1995, 73(1): 105～124

6 Iwasaka N, Wallace N M. Large scale air sea interaction in the Northern Hemisphere from a view point of variation of surface heat flux by SVD analysis. J Meteor Soc Japan, 1995, 73(4): 780～793

7 施　能. SVD 模态的显著性检验. 气象科技, 1996, (4): 5～6

MONTE CARLO TEST USED IN CORRELATION AND COMPOSITE ANALYSIS OF METEOROLOGICAL FIELDS

Shi Neng
(Department of Meteorology, NIM, Nanjing　210044)

Wei Fengying
(Academy of Meteorological Science of China, Beijing　100081)

Feng Guolin
(Department of Basic Courses, Yangzhou University, Yangzhou　225009)

Sheng Tongli
(Department of Meteorologicy, NIM, Nanjing 210044)

Abstract　This article is concerned with the problem of significance test of meteorological elements and their fields and introduces a Monte Carlo simulation experiment scheme with the results of usefulness in practice.

Keywords　correlation analysis, composite analysis, Monte-carlo test