

显著性检验的统计模拟^①

朱平盛 张苏平

(山东省气象科学研究所, 济南 250031)

提 要

对天气预报,特别是气候分析和预测中经常用到的几个气象要素场,应用相关场显著性检验的统计模拟方法,计算出了某一单点或区域平均气象要素与全球 OLR、北太平洋 SST 和北半球 500hPa 高度的相关场的显著性检验判据,并对气象统计预报中相关场显著性检验的问题进行了讨论。

关键词: 气象要素相关场 显著性检验 统计 模拟

引 言

统计方法是气象预报中应用最为广泛的一种数学方法。相关分析是气象统计预报的基础,而这些相关关系必须通过统计检验才能确认。对于某一单点(或区域平均)气象要素与另一气象要素场构成的相关场,目前多通过对相关场中点的显著性检验,找出相关显著的区域,但这并不意味着该气象要素场是显著相关的。那么,多少个点通过检验才能判断该场显著呢?对于气象要素相关场显著信度的检验,Livezey 等^[1,2]提出了采用 Monte-carlo 试验的统计模拟方法。首先选取数百个乃至上千个遵从已知分布(如正态分布或均匀分布)的随机数序列,然后计算各随机数序列与实际要素场各格点的相关系数,对所产生的不同相关场计算出其显著相关区域的百分比,由此得到一个显著区域比例的经验概率分布,从而定出某一信度下场的显著性判据。王绍武等^[3]则选取 10 组统计性质良好

(无严格周期和强的准周期)的随机数序列,分别与构成实际场的各格点求相关,再由各相关场中在某一信度下的显著相关点数的概率分布,确定场的显著性判据。这种方法较 Livezey 等提出的 Monte-carlo 检验方案更为简单,计算量大大减少。

目前,地-气系统射出长波辐射(OLR)、海表温度(SST)和 500hPa 高度场资料在气候分析预测和长期天气预报中越来越被广泛应用。本文根据实际需要,采用王绍武等提出的气象要素相关场显著性检验的统计模拟方法,计算得出全球月平均高度相关场的显著性检验判据。以便气象台站制作预报工具时使用。

1 气象要素相关场显著性检验的统计模拟

1.1 资料与方法

本文所用气象要素场资料为 20 年(1975~1994 年,缺 1978 年 3~12 月)全球(范围:40°S~40°N,格距:5°×5°)网格点的月平均

① 本文受山东省气象局青年气象科学基金资助。

OLR 场、45 年(1951~1995 年)北太平洋 286 网格点(格距:5°×5°)月平均 SST 场和 46 年(1951~1996 年)北半球(范围:5~85° N, 格距:10°×5°)500hPa 月平均高度场;降水资料为 46 年(1951~1996 年)山东全省平均季降水量。

首先,选取 5 组(每组 1000 个)随机数,分别对其进行功率谱分析,将其中周期性表现最不明显的一组,再等分成 10 组。然后,计算各组随机数与 1~12 月各网格点上月平均要素场资料的相关系数,由此构成 120 个相关场。最后,统计每个相关场中达到给定信度的点数,按照显著相关点数的频数分布,确定相关场的显著性判据。

1.2 全球月平均 OLR 相关场的显著性检验

表 1 是 10 组随机数与 20 年 1~12 月全球 1224 个网格点的月平均 OLR 场显著相关($\alpha=0.05$)格点数。根据表 1 统计出显著相关格点数的频数分布,并绘出频数分布图(图 1)。对于表 1 中的 120 个数据,假若其服从均匀分布,那么,有 6 个数落在单边 5%的信度区域内(图 1 上的阴影部分),这 6 个数中最小值是 113。也就是说,当信度达到 0.05 的

格点数 $m \geq 113$ 时,则落在图 1 中 $\alpha=0.05$ 的否定域内,即说明该场是一个信度为 0.05 的显著相关场。

计算山东省平均历年季降水量与全球月平均 OLR 场各网格点的相关系数,每张图(共 48 张)上相关信度达到 0.05 的格点数列于表 2。分析表 2 可以看出,山东省春季降水量与前期上年 1 月 OLR 场显著相关;夏季降水量与上年 3 月和当年 1 月的 OLR 场显著相关;秋季与上年 6 月,冬季与上年 5 月及当年 2 月的 OLR 场相关显著。

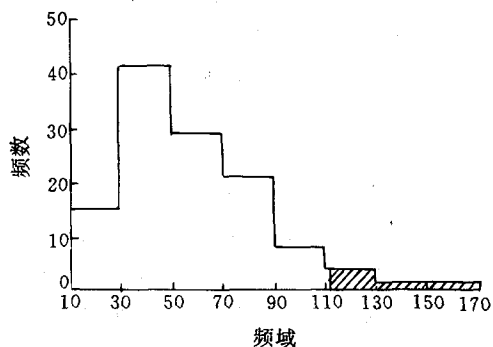


图 1 表 1 中数据的频数分布

表 1 10 组随机数与 1~12 月全球 OLR 场显著相关格点数($\alpha=0.05$)

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	平均
1	36	50	102	46	52	41	51	48	32	108	28	87	57
2	42	21	33	28	34	29	86	152	61	53	52	50	53
3	43	133	50	40	69	78	75	43	62	48	91	38	64
4	20	90	84	92	31	52	42	62	37	64	36	82	58
5	108	76	20	70	56	31	49	73	19	30	69	85	57
6	68	65	54	89	105	16	37	72	25	80	34	66	59
7	51	30	71	114	66	124	70	39	74	52	46	37	64
8	22	19	100	74	16	38	57	75	43	31	61	71	51
9	77	79	50	37	48	40	127	56	47	101	39	67	64
10	69	63	39	43	61	62	36	48	113	22	74	44	56
平均	54	63	60	63	54	51	63	67	51	59	53	63	58

表2 山东季降水量与各月 OLR 场显著相关格点数 ($\alpha = 0.05$)

月份	-1	-2	-3	-4	-5	-6	-7	-8	-9	-10	-11	-12
春	114	33	73	66	43	47	27	32	80	60	52	94
夏	68	58	118	64	64	47	50	47	58	56	93	75
秋	46	41	41	48	70	114	67	77	59	81	40	24
冬	28	63	70	85	131	50	47	51	52	58	49	80
月份	1	2	3	4	5	6	7	8	9	10	11	12
春	16	34	76	70	31	41	43	67	47	125	64	89
夏	120	45	60	61	39	44	26	28	45	54	72	95
秋	42	36	48	64	68	42	48	66	43	43	72	81
冬	44	113	49	78	73	50	63	71	52	60	31	56

注:负数表示前一年

1.3 北太平洋月平均 SST、北半球 500hPa 月平均高度相关场的显著性检验

用类似上述的方法,对 286 个网格点的北太平洋月平均 SST 和 576 个网格点的 500hPa 月平均高度场作相关显著性检验的统计模拟。每个相关场中信度达到 0.05 的格点数分别列于表 3 和表 4,图 2 和图 3 分别是表 3 和表 4 中数据的频数分布图。由图 2 可以看出,在 0.05 信度域内的 6 个数据中的最小值是 26。由图 3 可以看出,在 0.05 信度域内的 6 个数据中的最小值是 79。也就是说,对于样本长度 $n = 45$,由 286 个网格点构成的北太平洋月平均 SST 场,当其中相关信度达到 0.05 的格点数 $m \geq 26$ 时,则该场是显著相关($\alpha = 0.05$);对于样本长度 $n = 46$,

由于 576 个格点构成的北半球 500hPa 月平均高度场,当其中相关信度达到 0.05 的格点数 $m \geq 79$ 时,则该场是显著的。

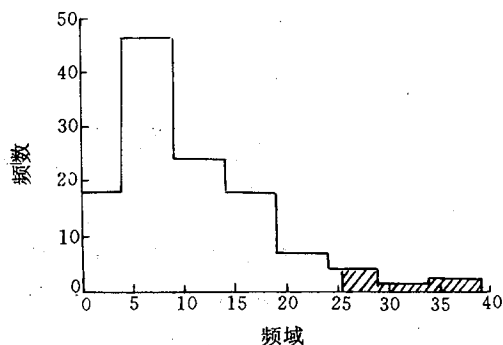


图2 表3中数据的频数分布

表3 10组随机数与1~12月北太平洋 SST 场显著相关格点数 ($\alpha = 0.05$)

	1月	2月	3月	4月	5月	6月	7月	8月	9月	10月	11月	12月	平均
1	12	11	2	7	5	0	6	22	18	13	4	4	9
2	9	6	17	17	4	2	11	11	11	6	8	8	9
3	27	39	30	16	10	23	9	8	20	7	3	13	17
4	9	8	2	5	5	19	26	7	12	26	38	28	16
5	18	9	3	5	9	14	9	3	7	16	18	19	11
6	16	4	9	2	12	8	7	2	9	16	4	6	8
7	8	8	5	19	14	11	3	5	13	12	24	9	11
8	5	4	5	3	8	9	15	20	10	6	18	13	10
9	2	8	4	10	9	6	5	11	14	8	7	17	8
10	14	15	16	12	14	5	10	15	20	23	5	5	13
平均	12	11	9	10	9	10	10	10	13	13	13	12	11

表4 10组随机数与1~12月北半球500hPa高度场显著相关格点数($\alpha=0.05$)

	1	2	3	4	5	6	7	8	9	10	11	12	平均
1	99	10	22	33	23	46	90	26	12	25	21	1	34
2	13	18	10	14	38	7	11	26	29	15	44	3	19
3	40	31	33	45	12	22	17	6	45	20	81	55	34
4	12	38	79	44	36	50	41	37	30	44	79	115	50
5	18	53	7	33	24	15	52	38	35	11	63	3	29
6	18	5	5	41	54	11	18	9	12	9	16	4	17
7	20	3	21	5	6	13	11	15	37	22	26	0	15
8	27	0	51	29	27	87	24	12	32	22	9	7	27
9	10	17	4	0	13	32	18	19	29	17	24	42	19
10	26	39	6	16	23	9	13	19	44	14	23	3	20
平均	28	21	23	26	26	29	30	21	31	20	39	23	26

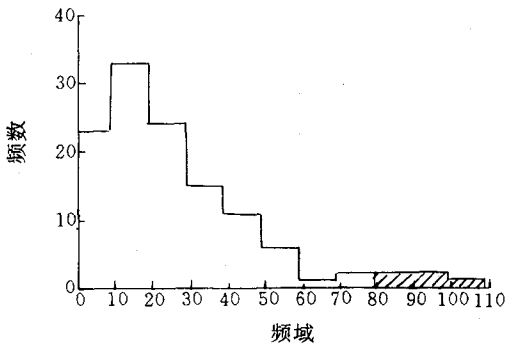


图3 表4中数据的频数分布

我们还计算了山东季降水量与上年和当年各月北太平洋 SST、北半球 500hPa 高度场各网格点的相关系数,统计了各相关场中显著相关($\alpha=0.05$)格点数(表略)。分析山东各季降水量与前期和同期北太平洋月平均 SST 场的相关性可以看出,春季降水量与上年 1、2、4 月的 SST 场显著相关;夏季与当年 3、4、7、8 月显著相关;秋季与上年 5 月和当年 9、11 月显著相关;冬季与上年 2、3 月显著相关。分析山东各季降水量与前期和同期北半球 500hPa 月平均高度场的相关性可以看出,春、夏、秋、冬四季降水量分别与当年 3、7、9 和 11 月显著相关(表略)。

2 结论与讨论

本文采用统计模拟的方法,计算出了长期天气预报和气候分析预测中常用的 3 个气

象要素相关场的显著性检验判据。某一气象要素与 20 年全球(40°N~40°S)1224 个格点月平均 OLR 的相关场中,当有 113 个以上格点的单点相关信度达到 0.05 时,则此相关场在 0.05 的信度水平下是显著的。某一气象要素与 45 年北太平洋月平均 SST 的相关场(286 个格点)中,当单点相关信度达到 0.05 的格点数超过 26 个,则此相关场在 0.05 的信度水平下是显著的。某一气象要素与 46 年北半球 500hPa 月平均高度的相关场(576 个格点)中,如果单点相关信度达到 0.05 的格点数超过 79 个时,则此相关场在 0.05 的信度水平下是显著的。

一般来说,气象要素场既存在时间上的持续性,也存在空间上的相互依赖性。例如, SST 有较强的定常性;大气环流等不仅在很大范围内的变化位相是一致的,而且在空间上存在很强的遥相关。因此,对于其相关场的检验,即使用有效自由度来消除单点要素的持续性,也不满足样本相互独立的假设检验条件。特别是为了消除资料噪音,对要素场进行时间平滑或空间滤波处理,会使资料的时空相关性明显提高。这种虚假的相关性,必将严重影响预报因子筛选的准确性和统计关系及其预报准确率的稳定性。分析上述 3 个要素与各组随机数显著相关格点的分布(图略),在许多图上点的分布并不是杂乱无章的,这种有规律的分布说明了要素场的空间

连续性。因此,避免由于气象要素的时间持续性和空间非独立性所造成的相关信度估计偏高,对于有效提高统计预报准确率和稳定性有很大帮助。

本文只给出了上述3个气象要素相关场,在0.05信度下的显著性判据。当构成场的格点数或资料序列长度不同时,其显著判据会有相应的变化。

参考文献

- 1 Livezey, R. E. and W. Y. Chen. Statistical field significance and its determination by Monte Carlo techniques, *Mon. Wea. Rev.*, 1983, 111: 46~59.
- 2 黄嘉佑. 气象要素场的显著性检验. *气象*, 1989, 15(4): 3~8.
- 3 王绍武, 胡增臻. 气象要素相关场显著性检验的统计模拟. 长江黄河旱涝灾害发生规律及其经济影响的诊断研究. 北京: 气象出版社, 1993: 215~221.

Statistical Simulation of Significance Test of correlative Fields for Several Common Weather Factors

Zhu Pingsheng Zhang Suping

(Shandong Research Institute of Meteorological Science, Jinan, 250031)

Abstract

The statistical simulation of significance test of correlative fields was used to deal with several weather climatic analysis and prediction. The significance test criteria were calculated for the correlative fields between single point or area mean values of weather factors and global OLR, north Pacific SST and north hemisphere geopotential height of 500hPa. Some problems about significance test of correlative fields in statistical forecast were also discussed.

Key Words: correlative field for weather factors significance test statistical simulation