

一种预测高温危害天气的统计模型

王春乙 潘亚茹

(国家气象局气象科学研究所)

提 要

本文使用自然正交分解方法,分析了华北及渭河流域冬麦区冬小麦灌浆乳熟期高温危害天气的时空分布特征。在此基础上,利用主分量分析倒筛选和调整复相关系数相结合的方法建立了预测高温危害天气的最佳回归模型。

一、引 言

冬小麦灌浆乳熟期的高温天气是北方冬麦区冬小麦后期产量形成的主要农业气象灾害之一。高温危害重的年份可减产10—20%,轻的年份可减产5—10%。因此,搞清高温天气的发生规律及危害,并且提出行之有效的预防措施,是确保冬小麦稳产高产一项不可缺少的技术环节。本文采用自然正交分解方法,利用旬平均气温资料来分析华北和渭河流域5月下旬至6月上旬高温天气各个典型场的时空分布特征,并根据各个典型场时间系数的演变,将高温天气划成重、轻、无三种类型,再利用大气环流因子来组建各时间系数预测模式。

二、华北和渭河流域冬麦区气温场的时空分布特征

1. 资料

本文使用1956—1987年5月下旬和6月上旬的旬平均气温资料,采用这两旬温度的平均值来描述冬小麦灌浆乳熟期高温危害的时空分布特征。在华北和渭河流域共选12个代表站,它们是北京、石家庄、沧州、介休、运城、西安、安阳、郑州、商丘、德州、泰安和宿县。

2. 计算方法

首先按上述12个代表站的平均温度构成初始矩阵 F_{ij} ($i=1, 2, \dots, m; j=1, 2, \dots, n$) $m=12, n=32; m$ 为站点, n 为年代。用下式对 F 矩阵进行中心化处理:

$$X_{ij} = F_{ij} - \bar{F}_i \quad (1)$$

$$\bar{F}_i = \sum_{j=1}^n F_{ij} / N$$

相应的协方差矩阵 S 为:

$$S = XX^T \quad (2)$$

$$S_{ij} = \sum_{v=1}^n x_{iv} \cdot x_{vj}$$

其中, X^T 为 X 的转置矩阵。利用雅可比算法求其 S 矩阵的全部特征值 λ_k 及特征向量矩阵 U , 则主分量矩阵 Y 为:

$$Y = UX \quad (3)$$

3. 气温场的空间分布特征

由协方差矩阵 S 求得全部特征值 λ_k 及累积方差见表1(取其前6项)。

表1 矩阵 S 的特征值及累积方差

序号	1	2	3	4	5	6
特征值	412.43	87.34	47.80	18.09	9.45	6.96
累积方差	69.06	83.68	91.68	94.71	96.29	97.46

从表1中可以看出,5月下旬至6月上旬平均气温场的自然正交分解的结果收敛很快,前两项特征值之和占总体方差已达83.7%,第7项以后各项,特征值均小于1,具有良好的精度。

第1特征向量场如图1所示。华北及渭河流域5月下旬至6月上旬平均气温场的空间分布呈明显的全区一致性。它表明,若是气温偏高,则是全区偏高,反之亦然。

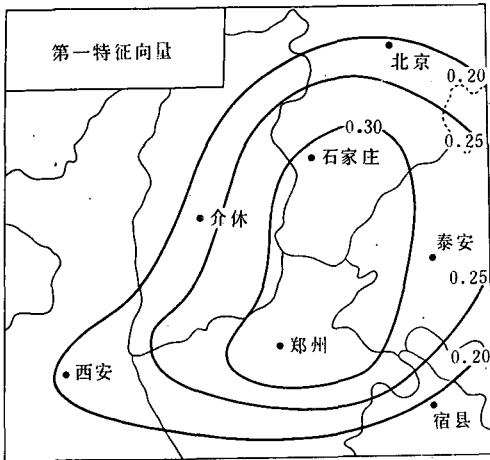


图1 第一特征向量空间分布图

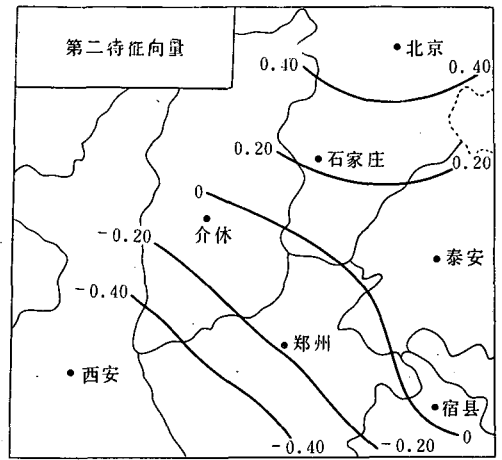


图2 第二特征向量空间分布图

第2特征向量场如图2所示。华北及渭河流域5月下旬至6月上旬平均气温场的空间分布呈东北、西南相反型。它表明,若是东北区气温偏高,则西南区气温偏低;若是东北区气温偏低,则西南区气温偏高。

第3特征值以后各特征值占总体方差均较小,对应的空间分布较零乱,实际意义不大。

4. 气温场的时间分布特征

根据第1特征值所对应的时间系数演变和历史上高温天气的出现情况^[1],可将第一典型场分为3类,当时间系数 $\alpha \geq 3.0$ 时,为全区重高温危害年,如1960、1961、1962、1967、1968、1982、1986年,所占机率为7/32,平均约每4至5年一遇。当时间系数 $1.0 \leq \alpha < 3.0$ 时,为全区轻高温危害年型,如1958、1965、1975、1978、1979、1981、1983年,所占机率为7/32,平均约4至5年一遇。当时间系数 $\alpha < 1.0$ 时,为全区无高温危害年型,所占机率为18/32,平均约2年一遇。同样,根据第2特征值所对应的时间系数演变和历史上高温天气的出现情况^[2],也可将此典型场按时间系数分为3类。当时间系数 $\alpha \geq 2.0$ 时,东北区为轻至重高温危害年型,西南区为无高温危害年型,如1964、1983、1986年。当时间系数 $\alpha \leq -2.0$ 时,东北区为无高温危害年,西南区为轻至重高温危害年,如1969、1973、1974、1978年。当时间系数 $-2.0 < \alpha < 2.0$ 时,为正常年份,即东北、西南区气温分布无明显差异。

三、华北及渭河流域冬麦区高温危害年的统计预报

本文利用某些大气环流因子,采用多元回归分析方法来建立第一时间系数的统计预报模式。

1. 因子的粗选及膨化处理

利用各大气环流因子同第一时间系数进行单相关普查,选出4个相关显著的因子,它们是当年2月副高脊线平均位置(x_1),当年1月副高脊线平均位置(x_2),当年2月冷空气次数(x_3),上年12月东亚槽强度(x_4)。

为了避免预报因子的阶段性和稳定性对回归模式的影响^[2],本文对粗选因子进行膨化处理,即附加预报量和预报因子的周期分量,使方程同时考虑预报量、预报因子的周期分量及预报因子本身对回归模式的贡献。预报量和预报因子的周期分量生成参看文献[3]。

各周期分量和上述4个环流因子同第一时间系数的单相关系数见表2(显著者):

表2 预报因子及周期分量同第一时间系数的相关系数

序号	因子名称	相关系数	序号	因子名称	相关系数
x_1	上年12月副高脊线平均位置	0.3362	x_7	x_3 的14年周期分量	-0.4325
x_2	当年1月副高脊线平均位置	0.3288	x_8	x_4 的6年周期分量	-0.5508
x_3	当年2月冷空气次数	-0.4057	x_9	x_4 的13年周期分量	-0.5464
x_4	当年1月东亚槽强度	-0.3479	x_{10}	预报量的5年周期分量	0.3513
x_5	x_2 的14年周期分量	0.4388	x_{11}	预报量的7年周期分量	0.4649
x_6	x_3 的13年周期分量	-0.5443	x_{12}	预报量的11年周期分量	0.5456

2. 回归模式的建立

目前应用较为广泛的选择变量组合的方法较多,本文采用主分量分析倒筛选法结合调整复相关系数准则来解决这个问题。该方法是以各因子对主分量的负荷量为依据来判断取舍因子。因子 i 对第 j 个主分量的负荷量为:

$$L_{ij} = \sqrt{\lambda_j} U_{ji} \quad i, j = 1, 2, \dots, p \quad (4)$$

p 为因子数。删除因子的原则为,首先删除对最小主分量贡献最大的因子。当删除一因子后,利用剩下变量,再进行主分量分析,依上述原则删除第2个因子,依次进行下去,直到最后一个主分量的特征值不是很小为止。利用表2的12个因子进行主分量分析,求得各特征值及累积方差见表3:

表3 12个因子的特征值及累积方差

序号	1	2	3	4	5	6	7	8	9	10	11	12
特征值	308.4	169.9	125.9	78.0	69.7	30.6	25.2	16.4	11.8	6.6	2.6	1.9
累积方差	36.4	56.5	71.3	80.5	88.8	92.4	95.4	97.3	98.7	99.5	99.8	100.0

第12个特征值为1.9,该特征值所对应的负荷量为:(-0.1545 0.0211 -0.4576 -0.1519 0.3839 -0.1909 1.1741 0.2170 0.0776 -0.0622 -0.0508 0.0701),其中第7个因子所占的权重最大,首先将它删除。经过7次主分量分析后,共删除7个变量,余下的5个变量为 $x_1, x_4, x_9, x_{11}, x_{12}$ 。

在回归模型中增加新的变量时,复相关系数也随之增大,而使复相关系数增大的代价是剩余自由度减小,它意味着估计和预测的可靠性降低。考虑到拟合优度和可靠度之间的关系,本文引入自由度调整复相关系数 R^* ,作为选择变量组合的准则:

$$R^{*2} = 1 - \frac{n-1}{n-p-1} (1 - R^2) \quad (5)$$

R^* 称为自由度调整复相关系数。使用多元回归分析方法,组建剩下5个变量不同组合的回归模式。结果表明,模式

$$\hat{Y} = 10.63 + 0.49x_1 + 0.04x_4 - 1.02x_9 + 0.64x_{11} + 0.56x_{12} \quad (6)$$

为较佳的回归模型,它与 Y 的复相关系数达0.9578。但从自由度调整复相关系数可知

$$Y = 10.73 + 0.49x_1 - 0.98x_9 + 0.64x_{11} + 0.56x_{12} \quad (7)$$

为最佳稳健的回归模型, R^* 为0.9509。另外,从变量之间的共线性也可证明这一点。5个变量彼此之间的相关系数中 x_4 和 x_9 之间相关显著,已达0.01水平,所以应将因子 x_4 删除,剩下4个变量构成的彼此独立的最优回归子集。

利用(7)式在1988年3月对当年冬小麦灌浆乳熟期的高温危害天气进行预报,结果 Y 为-1.10。实况该年为无高温危害,预报准确。历史拟合(虚线)如图3所示,结果较为理想。

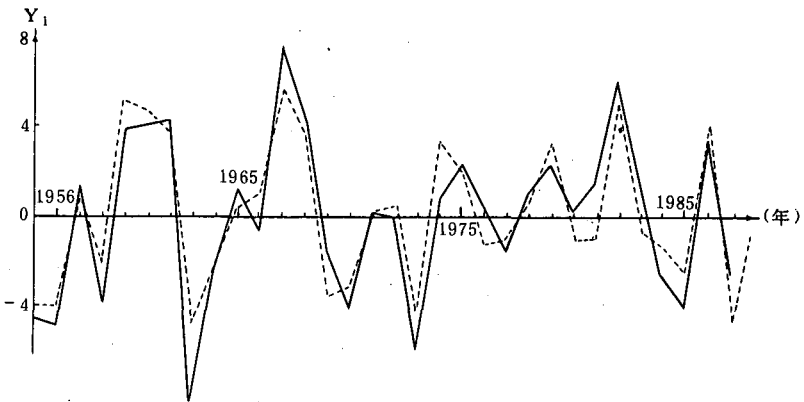


图3 第一时间系数逐年变化图(实线为实际值,虚线为计算值)

四、结 论

1. 华北及渭河流域冬麦区冬小麦灌浆乳熟期高温危害天气的空间分布特征是以全区性为主导,其次是东北、西南相反型。时间分布特征是全区性干热风年型历史上共出现14次,约2—3年一遇,东北、西南相反型共出现8次,约8年一遇。

2. 运用主分量分析倒筛选法结合调整复相关系数准则,可以使工作量大大减小,而又能准确地找到最佳变量组合的回归模型。但本文只作了一年预报,还需经多年预报实践来检验和完善。

参 考 文 献

- [1] 余优森等,北方小麦干热风气候区划,气象,第5期,1981。
- [2] 朱盛明,相关系数稳定性分析方法及其应用,气象学报,4,497—501,1982。
- [3] 王春乙、潘亚茹,我国北方主要产麦区冬小麦产量海温业务预报模式,数学的实践与认识,第1期,1989。

A STATISTICAL FORECASTING MODEL FOR THE HIGH TEMPERATURE DISASTROUS WEATHER

Wang Chunyi Pan Yaru

(Academy of Meteorological Science, SMA)

Abstract

In this paper, using empirical orthogonal function (EOF) expansion method, the temporal and spatial distribution features of high temperature weather causing disaster during the winter wheat milk-filling and milking maturity phases in North China and Weihe River reaches are analysed. Based on this analysis, the optimal regression model for high temperature weather forecasting have been developed by using the principal component reverse selecting method combined with multiple correlation coefficient criteria.