

文章编号: 100F 1498(2004) 03 0300 05

沙尘暴天气统计预报模型的研究

赵峰, 鞠洪波, 陈永富

(中国林业科学研究院资源信息研究所, 北京 100091)

摘要: 选用气象因子和地表植被因子作为预报因子, 其中以归一化植被指数作为植被因子, 采用逐步判别分析法对浑善达克地区 1988—2000 年 3—5 月份 301 个大风天气和沙尘天气样本资料进行了统计和分析, 建立了沙尘暴天气的多级预报模型, 并用 2001 年 3—5 月的 52 个试报样本进行检验, 结果表明试报准确率平均为 79.4%。

关键词: 浑善达克地区; 沙尘暴; 判别分析; 归一化植被指数

中图分类号: S165+.25 文献标识码: A

沙尘暴是一个重要的环境问题^[1], 是沙漠化的主要过程(植被退化、风蚀、水蚀、盐渍化、土壤板结、有机物减少和毒物积累)之一, 也是在沙漠及其边缘等特殊下垫面条件下产生的一种灾害性天气。它给人们的生活生产、交通运输及国民经济都带来很大的危害, 而且沙尘天气造成的空气污染严重影响人们的身体健康, 因此, 及时准确地对沙尘天气作出预报是目前十分重要的任务。

国内外对沙尘暴天气的研究已逐步系统化, 从沙尘暴的天气气候学特征、分析预报、遥感监测、沙尘暴成分分析、数值模拟、灾害预报等方面都作了深入探讨^[2]。以往沙尘暴天气的预报多从气象方面进行预报, 本文采用归一化植被指数作为植被因子, 将气象因素与地表因素相结合, 以浑善达克地区的沙尘天气为例建立沙尘暴天气的统计预报模型。

1 试验区概况

浑善达克沙地位于内蒙古东南部, $41^{\circ}56' \sim 44^{\circ}24' N$, $112^{\circ}22' \sim 117^{\circ}57' E$, 海拔多在 1 000~1 400 m 之间, 南部为阴山山地隆起区, 东部为大兴安岭低山丘陵, 沙地间具有较广阔的丘间低地, 并分布有 100 多个面积大小不等的湖泊。浑善达克沙地地处中纬度西风带, 属温带大陆性气候。降水不均, 干旱少雨, 风大沙多, 是该沙区的主要气候特征。冬春风强而多, 4—5 月风速较大, 可高达 12 级, 为沙尘天气的产生提供了有利的动力条件。浑善达克沙地东端位于大兴安岭南段西麓的草甸草原地带, 西端属于荒漠草原区, 中间广大沙区处于半干旱草原带, 植物种类繁多, 植被类型丰富, 以草原植被为主, 针阔叶乔木等超地带性植被明显。

气象站点的位置分布及其周围的沙化土地类型和植被类型见表 1。

收稿日期: 2003 06 24

基金项目: 中央级科研院所社会公益研究专项基金项目“沙尘暴监测技术”(17)部分研究内容

作者简介: 赵峰(1976—), 女, 山东德州人, 研究实习员。

表 1 试验区各气象站点的地理位置、沙化土地类型与植被类型

站号	站名	海拔/m	纬度/N	经度/E	沙化土地类型	植被类型
53068	二连浩特	965.96	43° 39'	111° 58'	潜在沙漠化土地、风蚀化沙地、固定沙地	草地
53195	苏尼特左旗	1 036.76	43° 52'	113° 38'	潜在沙漠化土地、风蚀化沙地、固定沙地	草地
53276	朱日和	1 151.96	42° 24'	112° 54'	潜在沙漠化土地、风蚀化沙地、固定沙地	草地
53391	化德	1 484.66	41° 54'	114° 00'	风蚀化沙地、潜在沙漠化土地	草地
54102	锡林浩特	990.86	43° 57'	116° 04'	固定沙地、潜在沙漠化土地	草地
54208	多伦	1 247.06	42° 11'	116° 28'	固定沙地、半固定沙地、潜在沙漠化土地 流动沙地	草地 未利用地

2 试验数据及处理

2.1 试验数据

2.1.1 气象数据 利用浑善达克地区气象站观测的 1988—2000 年 3—5 月份 301 个大风天气和沙尘天气样本资料建模, 其中有 127 个大风天气样本、86 个扬沙或浮尘天气样本以及 88 个沙尘暴天气样本。然后用 2001 年 3—5 月的 52 个大风天气和沙尘天气样本检验, 其中大风天气样本数为 26, 扬沙或浮尘天气样本数为 16, 沙尘暴天气样本数为 10。

2.1.2 植被数据 采用归一化植被指数作为植被因子。该因子通过对 NOAA/AVHRR 卫星接收数据进行图像处理来获取, 数据以每日 4 6 轨白天影像为信息源(N-14, N-15), 数据的时间序列是 1988—2001 年 3—5 月的影像。

2.2 数据处理

2.2.1 计算归一化植被指数(NDVI) 植被指数是根据植被反射波段的特性计算出来的, 是反映地表植被种类情况的间接指标, 也是反映植物生长状况及植被空间分布密度的最佳指示因子^[3,4]。

NOAA 影像的时间分辨率高, 每天 2~4 次, 有利于捕捉地面动态信息, 实现同步监测。从 NOAA/AVHRR 传感器 CH1 和 CH2 数据资料中, 可以获得多种形式的植被指数, 其中归一化植被指数被广泛应用于植被覆盖的定量研究。由于该指标一定程度上有助于减少外界因素, 如太阳高度角、大气状态和非零下点观测带来的误差。因此, 本文采用归一化植被指数作为反映地表植被信息的因子。

其计算公式如下:

$$NDVI = (CH2 - CH1) / (CH2 + CH1) \quad (1)$$

上式中, CH1、CH2 代表第一、第二通道的光谱反射比。

其计算过程是由极轨卫星资料接收与处理系统和 ERDAS 图像处理系统来实现的。影像处理主要包括数据格式转换、地理校正和辐射校正, 以及归一化植被指数计算。

2.2.2 数据分组 采用最低能见度的沙尘暴强度分级标准(见表 2)将预报量分为 4 个类别, 然后对各预报因子也相应地划分为 4 组。

2.2.3 数据标准化 统计天气预报的一个特点是不管物理量的量纲怎样, 变化多大, 都可以放在同一个预报式中。如果不做一些处理, 有可能突出变异较大资料的作用。因此, 需对观测数据进行无量纲化, 使各因子的取值范围变得大致相同^[5]。

3 研究方法 with 建模

3.1 判别分析法

本文采用逐步判别分析法进行预报^[6-8]。因为在逐步判别过程中,对预报因子进行逐步检验,保证了每个因子对预报量都有较好的相关性,而且几个因子组合对预报量也有较好的相关作用。因子检验方法如下:计算各因子的判别能力,在给定的 F_α 判据值下,进行显著性检验,确定剔除或引进的因子,直到因子无剔除和无引进为止。通常情况下, F_α 的取值为 2。判别准则选用 Bayes 定理,即将已出现的 p 个因子的样品, $x_{10}, x_{20}, \dots, x_{p0}$, 代入到判别函数中,进一步计算其值出现的条件下,第 g 类出现的概率,从 $g = 1, 2, \dots, G$ 中选择出现最大概率的类别作为预报量的类别。沙尘天气的建模过程主要是由统计软件 SPSS 来完成的。

3.2 预报量和预报因子的确定

表 2 最低能见度为标准的预报量分级

3.2.1 预报量 本文参照已有的沙尘天气分级标准^[9],采用最低能见度的沙尘暴强度分级标准(见表 2)将预报量分成 4 个类别,即大风天气、扬沙天气、浮尘天气和沙尘暴天气。

预报量	最低能见度 VV/km	风速 v ($m \cdot s^{-1}$)
大风天气	$10 \leq VV < 30$	
浮尘天气	$1 \leq VV < 10$	$v < 4$
扬沙天气	$1 \leq VV < 10$	$v \geq 4$
沙尘暴	$0 \leq VV < 1$	

在进行预报时,采用多次二级判别的方法。

先将预报量划分为 4 级,每次仅预报两级,分多次预报。即先预报是否产生沙尘天气,然后在预报是沙尘天气的情况下,再预报是浮尘或扬沙天气还是沙尘暴天气。如果是浮尘或扬沙天气,可通过风速因子(以风速 $4 m \cdot s^{-1}$ 为临界值)进一步预报是浮尘还是扬沙天气。

3.2.2 预报因子 预报量和预报因子取不同状态时,通常选用不同的因子筛选方法。本文中预报量取离散状态,而预报因子为连续变量,对于这种情况,常采用判别式法来选取预报因子。

决定因子 m 是否入选的判别式为

$$F = \frac{\sum_{g=1}^G N_g (\bar{x}_g - \bar{x})^2 / G - 1}{\sum_{g=1}^G \sum_{k=1}^g (x_{gk} - \bar{x}_g)^2 / N - G} \quad (2)$$

N_g —— 第 g 类气象要素的样本数; \bar{x}_g —— 第 g 类气象要素值的平均数; G —— 气象要素的类别数; N —— 气象要素的样本数

表 3 沙尘暴预报因子

其分子表示类间差异,分母表示类内差异。 F 值越大,即类间差异越大,类内差异越小,分类的性能越好,因此,可将 F 值超过某一数值作为前兆因子入选标准。取某一显著水平 α ,查 F 分布表,其中分子自由度为 $G-1$,分母自由度为 $N-G$,得到的 F_α 即判断的临界值。本文中显著水平采用 $\alpha = 0.1$,临界值 $F_{0.1} = 2.32$ 。

编号	预报因子	物理意义
1	m_1	气压
2	m_2	气温
3	m_3	相对湿度
4	m_4	风速
5	m_5	最大风速
6	m_6	地面温度
7	m_7	植被指数

根据判别式(2)对样本数据进行计算,最终筛选后的因子见表 3。

3.3 建立沙尘暴天气预报模型

通过对 1988—2000 年 3—5 月 301 个大风天气和沙尘天气样本资料统计计算得出,沙尘天

气各类预报方程分别为:

(1) 大风天气与沙尘天气的预报方程

$$\text{大风天气: } f_1 = 0.3543 m_1 - 0.0818 m_4 + 481.5019 m_7 - 1570.1564 \quad (3)$$

$$\text{沙尘天气: } f_2 = 0.3513 m_1 - 0.0350 m_4 + 473.3017 m_7 - 1548.3110 \quad (4)$$

(2) 扬沙或浮尘天气与沙尘暴天气的预报方程

$$\text{扬沙或浮尘天气: } f_3 = 0.1406 m_4 - 8.9992 \quad (5)$$

$$\text{沙尘暴天气: } f_4 = 0.1752 m_4 - 13.5578 \quad (6)$$

从上面的判别方程可以看出, 风速、植被指数和气压因子对沙尘天气的预报起重要作用。

利用判别方程进行预报的基本步骤: 首先将检验样本个例的因子值分别代入(3)、(4)式, 根据 Bayes 定理, 比较 f_1, f_2 值大小。若 $f_1 > f_2$, 则预报为无沙尘天气; 否则, 预报为沙尘天气, 并且进一步将检验样本个例的因子值分别代入(5)、(6)式。然后再比较 f_3, f_4 值大小。若 $f_3 > f_4$, 则预报为扬沙或浮尘天气, 然后根据风速的大小进一步判定是扬沙还是浮尘天气(风速大于或等于 $4 \text{ m} \cdot \text{s}^{-1}$ 为扬沙天气, 否则为浮尘天气); 若 $f_3 < f_4$, 则预报为沙尘暴天气。

根据上述方程及判别步骤, 用 2001 年 3—5 月的 52 个大风天气和沙尘天气样本资料进行检验, 检验结果见表 4、表 5。

表 4 建模样本资料和检验样本资料情况

天气类别	建模样本资料拟合(1988—2000)					检验样本预报分类(2001)				
	大风天气	沙尘天气	扬沙或浮尘天气	沙尘暴天气	总计/次	大风天气	沙尘天气	扬沙或浮尘天气	沙尘暴天气	总计/次
大风天气	90	37			127	21	5			26
沙尘天气	21	150			171	7	19			26
扬沙或浮尘天气			59	27	86			15	1	16
沙尘暴天气			24	64	88			3	7	10

表 5 判别结果的效果检验

项 目	大风天气	沙尘天气	扬沙或浮尘天气	沙尘暴天气	平均
1988—2000 年建模样本数据的拟合率/%	70.9	87.7	69.1	72.7	75.1
2001 年检验样本数据的试报准确率/%	80.8	73.1	93.8	70.0	79.4

从表 4、5 可以看出, 根据判别分析法所建立的沙尘天气的预报方程对 2001 年 3—5 月的检验样本数据的试报准确率为 70.0%~93.8%, 平均准确率为 79.4%。

4 小结及存在问题

(1) 天气预报中, 通常将预报量划分为多级或多个类别, 例如降水量的预报可分为: 暴雨、大雨、中雨、小雨和无雨等五级。这种情况下多采用多级预报的方法进行预报。然而, 由于沙尘暴天气属小概率事件, 为了使小概率事件的比重增大, 从而使其统计规律性更明显, 本文在用逐步判别分析法对沙尘天气进行分级预报时, 将对预报量的多级预报划分为多次二级预报。

(2) 研究表明, 沙尘暴发生的三个必要条件包括沙源、大风和低层大气层结构的不稳定状态。强风是沙尘暴发生的不可缺少的动力条件, 不稳定大气有利于强对流的发生和发展, 从而加强对流性天气过程, 而丰富的沙源则是沙尘暴发生的物质基础。由于地表覆被的好坏直接

影响到沙尘的物质源, 较低的植被覆盖率或裸地也将为沙尘暴的发生提供物质条件, 因此, 植被覆盖率的高低也是沙尘暴发生与否的重要条件之一。所以本文综合考虑了气象因素和地表植被因素, 并选用归一化植被指数因子作为预报因子。

(3) 逐步判别分析法是根据各预报因子对预报方程贡献大小来进一步筛选因子的, 因此, 从文中所建立的四个预报方程可以看出, 气压、风速和植被指数因子对预报方程的贡献较大, 是预报沙尘天气发生与否的主要因子。

(4) 选用检验样本数据对判别函数进行效果检验时发现, 用3、4月份的样本数据试报准确率高于其它月份, 这与3、4月份是沙尘暴天气发生的多发季节有关。

(5) 气象因子是沙尘暴发生的动力因子, 因此在对沙尘暴天气进行预测预报时, 须结合典型的天气形势, 根据不同机制下产生沙尘暴的天气形势对预报因子进行分组, 以保证预报因子确切反映预报量的信息, 从而提高预报准确率。但是, 由于本文在数据分析过程中没有其它的资料参考, 不能对沙尘天气过程进行天气形势的分析, 因此难免会影响预报效果。

(6) 本文采用统计预报的方法对沙尘暴天气进行预报。统计预报方法简单易操作, 并且根据统计预报方程也可做出比较客观的定量或定性预报。但是由于其往往把复杂问题简单化或模式化, 与天气形势复杂变化的实际情况有一定距离从而影响了预报效果。建议在以后的预报中, 对于同一个预报问题, 可以采用不同的预报方法并把各种预报意见进行“集成”, 得出一个统一的预报结论, 以提高预报准确率。

参考文献:

- [1] Jauregui E. The dust storms of Mexico city[J]. International Journal of Climatology, 1989, 9(2): 169~ 180
- [2] 史培军, 严平, 高尚玉, 等. 我国沙尘暴灾害及其研究进展与展望[J]. 自然灾害学报, 2000, 9(3): 72~ 77
- [3] 吴传钧, 郭焕成. 中国土地利用[M]. 北京: 科学出版社, 1994
- [4] 叶笃正. 中国的全球变化预研究(总论)[M]. 北京: 气象出版社, 1992. 23 ~ 54
- [5] 朱伯承. 统计天气预报[M]. 上海: 上海科学技术出版社, 1981
- [6] 李祚泳, 刘少依, 邓新民. 基于 Fisher 判别分析法的霾云识别模式及其效果检验[J]. 成都气象学院学报, 1994, 28(1): 64~ 69
- [7] 邢纪元, 杨瑞峰, 王建波. 泰安冬季降水概率预报判别分析[J]. 山东气象, 2001, 21(1): 16~ 18
- [8] 黄嘉佑. 气象统计分析与预报方法[M]. 北京: 气象出版社, 2000
- [9] 方宗义, 朱福康, 江吉喜, 等. 中国沙尘暴研究[C]. 北京: 气象出版社, 1997

Research on Statistic Model for Sand and Dust Storm Forecasting

ZHAO Feng, JUH ong-bo, CHEN Yong-fu

(Research Institute of Forest Resource Information Techniques, CAF, Beijing 100091, China)

Abstract: As a meteor, sand and dust storm results from a special atmospheric movement. But weather was not the only factor, the earth's surface had made a significant effect on its forming and development. It's qualitative to predict sand and dust storm, so Fisher's discriminant was used to analyze historical data of wind and sandstorm in the Hunshandake area. Based on the available high quality surface and meteorological observation data from 1988 to 2001, the predictors were analyzed by stepwise method to build linear equations of different groups. The testing ratio of the equations were 79.4%.

Key words: Hunshandake; sand and dust storm; discriminant; NDVI