

报名号#1102

2008 年 第一届“数学中国杯”

数学建模网络挑战赛

承 诺 书

我们仔细阅读了首届“数学中国杯”数学建模网络挑战赛的竞赛规则。

我们完全明白，在竞赛开始后参赛队员不能以任何方式（包括电话、电子邮件、网上咨询等）与队外的任何人（包括指导教师）研究、讨论与赛题有关的问题。

我们知道，抄袭别人的成果是违反竞赛规则的，如果引用别人的成果或其他公开的资料（包括网上查到的资料），必须按照规定的参考文献的表述方式在正文引用处和参考文献中明确列出。

我们郑重承诺，严格遵守竞赛规则，以保证竞赛的公正、公平性。如有违反竞赛规则的行为，我们将受到严肃处理。

我们允许数学中国网站(www.madio.net)公布论文，以供网友之间学习交流，数学中国网站以非商业目的的论文交流不需要提前取得我们的同意。

我们的参赛报名号为：1102

参赛队员（签名）：

队员 1：肖正武

队员 2：梁乾

队员 3：毛维夫

参赛队教练员（签名）：郭刚

参赛队伍组别：大学组

报名号#1102

2008 年 第一届“数学中国杯”

数学建模网络挑战赛

编 号 专 用 页

参赛队伍的参赛号码：（请各个参赛队提前填写好）：

1102

竞赛统一编号（由竞赛组委会送至评委团前编号）：

竞赛评阅编号（由竞赛评委团评阅前进行编号）：

报名号#1102

2008 年 第一届“数学中国杯” 数学建模网络挑战赛

题 目 沙尘暴与各项气象指标的关系分析关 键 词 基于改进遗传思想的 K-Means 算法 数据挖掘
聚类分析 遥感反演法 起沙机制

摘 要：

沙尘暴日益成为影响我国西北发展的突出问题，本文仅就 45 年来西北地区一些气象观测站点的气象数据进行建模分析，探讨影响沙尘暴产生的各个气候因素之间的关系及其预测问题。

对于问题一，该问题为一统计分析问题，本文对随机选取的一个观测站点的观测数据，通过宏观分析，得出该地区的年沙尘暴发生天数的整体趋势，及各个气候因素与它的定性关系。再对其进行微观分析，进行 SPSS 软件聚类分析，并采用非线性映射分析法来优化算法，其对数据的分对率达到 82.6%。为继续提高精度，本文采用基于改进遗传算法的 K-Means 聚类分析，使分对率提高到 93.67%，表明该算法具有很高的精度，可信度较高。最后本文对结果作了分析和合理性说明。问题中采用基于改进遗传算法的 K-Means 聚类分析，较大提高估计的准确率是本文的一大特色。

对于问题二，本文将沙尘天气的气象等级转化为能见度等级。针对不同的预测周期，将其分为长期与短期预测，分别采用不同原理建模。对于长期预测，通过太阳辐射量、气溶胶透射比和气溶胶厚度的关系，建立模型，得到能见度与太阳辐射的关系，以此来评估沙尘天气气象等级；对于短期预测，采用遥感反演法，引入沙尘指数，通过沙尘指数与能见度的等级对应表，判断气象数据对应的沙尘天气气象等级。两个模型通过互补，能较好地沙尘暴天气进行预测，并估计沙尘天气的等级。

对于问题三，本文首先对一个省各个站点沙尘暴 25 年总发生天数进行统计分析，推断出沙源地，再通过各个站点月沙尘暴发生天数的相关性判断沙尘暴的运动路线。最后建立基于起沙机制的模型，通过沙粒的漂移沉淀机制，判断沙尘暴的影响范围。该模型的可推广性较强，能适用于其他省份不同站点所在地区。

该文重点论述自然因素对沙尘暴的影响，拟从七个方面对沙尘暴的形成及研究现状进行评述，即沙源分布与移动路径、时间变化、气候变化的影响、产生沙尘暴的天气系统、沙尘的物理化学性质及起沙机制、大气辐射特性及卫星遥感监测、沙尘暴的数值模拟。文章还介绍了国内外在沙尘暴方面的主要研究成果。

最后本文还对模型的优缺点进行了客观的评价，并对沙尘暴的防治作了拓展讨论。

参赛队号 1102所选题目 C参赛密码 _____
(由组委会填写)

报名号#1102

目录

一、	问题重述与分析	5
1.1	问题重述	5
1.2	问题分析	5
1.2.1	问题一	5
1.2.2	问题二	5
1.2.3	问题三	5
二、	问题假设	6
三、	符号说明	6
四、	模型建立与求解	2
4.1	问题一：	2
4.1.1	宏观分析：	2
4.1.2	微观分析	4
4.1.2.1	SPSS 软件聚类分析	4
4.1.2.2	基于改进遗传思想的 K-Means 聚类算法	7
4.2	问题二	11
4.2.2.1	长期预测模型	12
4.2.2.2	短期预测模型	13
4.3	问题三	15
4.3.1	沙源地的判断	15
4.3.2	沙尘暴运动路线的判断	16
4.3.3	沙源地及运动路线判断模型验证	18
4.3.4	沙尘暴的成因及物理机制	18
4.3.5	机理模型的应用	23
4.3.6	沙尘暴的影响范围	24
五、	模型进一步讨论	24
沙尘暴防治手段：		24
六、	模型优缺点	25
优点		25
缺点		25
七、	附录	25

一、 问题重述与分析

1.1 问题重述

沙尘暴是一种灾害性天气，近年来频繁发生，对环境和人们的生产、生活都造成了影响甚至危害。对沙尘暴的发生条件需要加以研究，以便人们进行有效的预测和预防。请为沙尘暴的发生建立数学模型，具体问题如下：

问题一：对沙尘暴的发生次数进行分析，建立相应的数学模型，使人们可以根据相关气象数据的监测结果，估计出沙尘暴在某地区某段时间内发生的次数。

问题二：评估沙尘天气气象等级，预测在近期内观测站所在地发生沙尘暴的风险，并在高风险时发出警报信息。

问题三：考虑到沙尘在空气中随风传输的过程和各个观测站点的空间位置，进一步深入前面的问题，完善气候因素对沙尘暴发生频率的影响模型，完善各地区对沙尘暴的短期预报机制，并估计中国境内引起沙尘暴的重要沙源地位置及其影响范围。

1.2 问题分析

1.2.1 问题一

不同的地区气候环境不同，各个因素对沙尘暴的产生影响也不同，但沙尘暴发生的原理相同。因此，可以通过对不同站点的观测数据进行统计汇总，从中得出沙尘暴发生的趋势及其变化规律，从宏观更好地把握问题。

在获得对问题的宏观定性认识之后，通过定量分析，寻找气候因素与沙尘暴发生的关系。对于多达七个的气候因素，过多的变量必然导致拟合过程出现局部符合的现象，不适合按照常规的拟合方法来寻求它们关系。即使能够较好地拟合现有数据，但对未来沙尘暴爆发情况作出预测的可信度不高。因此，出于对数据内在关系的探索，我们决定采用数据挖掘技术中的聚类分析，同时尝试采用改进遗传算法作为聚类参数寻优的“导师”，两者有机结合来探求沙尘暴发生的规律。

1.2.2 问题二

沙尘天气的气象等级的划分与能见度相关。不同的预测时间，可分别采用不同原理建模。对于长期预测，可通过太阳辐射量、气溶胶透射比和气溶胶厚度的关系，建立模型，得到能见度与太阳辐射的关系，以此来评估沙尘天气气象等级；对于短期预测，采用遥感反演法，引入沙尘指数，通过沙尘指数与能见度的等级对应表，判断气象数据对应的沙尘天气气象等级。两个模型通过互补，能较好地沙尘暴天气进行预测，并估计沙尘天气的等级。

1.2.3 问题三

首先对某个省的所有站点沙尘暴总发生天数进行统计分析，推断出沙源地，通过各个站点月沙尘暴发生天数的相关性判断沙尘暴的运动路线，最后建立建立因风等因素起沙的动力学模型，以此求出沙粒的运动轨迹，并结合各地的地理因素沿经常性的路线漂移，直至全部降落为止。从而判断出沙尘暴的影响范围，并且可以推断出沙尘暴的高发地域。沙尘暴是风蚀荒漠化中的一种天气现象，它的形成受自然因素和人类活动因素的共同影响。

报名号#1102

二、 问题假设

1. 沙尘颗粒呈球形.
2. 颗粒群有尺寸的差别, 按固定尺寸分组, 不同尺寸分为不同的相, 其温度和物质密度相等, 即 $T_{p_1} = T_{p_2} = \cdots = T_{p_n}$, $\rho_{p_1} = \rho_{p_2} = \cdots = \rho_{p_n}$.
3. 所有颗粒相的时均速等于当地流体的相时均速, 都等于混合相的时均速, 即 $v_p = v_g = v_m$.
4. 颗粒相作为有湍流扩散的连续介质, 扩散系数相等, 即 $D_{p_1} = D_{p_2} = \cdots = D_{p_n} = D_{p_g}$.
5. 颗粒已被悬浮起来做不沉积的运动.
6. 颗粒相与流体相的阻力不计.

三、 符号说明

ETP_i	第 <i>i</i> 月的蒸发量	S'	水平面上的直接辐射
T_i	第 <i>i</i> 月平均温度	S	太阳直接辐射
P_i	第 <i>i</i> 月平均降雨量	D	散射辐射
RH_i	第 <i>i</i> 月平均相对湿度	Q	总辐射
u_i	第 <i>i</i> 月平均风速	$R_{1.6}$	可见光 1.6 μm 波段测得的反
n_i	聚类分析中第 <i>i</i> 类的特征值	射率	
d_i	第 <i>i</i> 月大风日数	T_{11}	11 μm 波段红外亮温
S_i	第 <i>i</i> 月散射量	T_{12}	12 μm 波段红外亮温
ETP_i	第 <i>i</i> 月的蒸发量	α 、 β	调节因子
T_i	第 <i>i</i> 月平均温度	V	能见度
P_i	第 <i>i</i> 月平均降雨量	ρ	颗粒物单位体积内所占的质
RH_i	第 <i>i</i> 月平均相对湿度	量	
u_i	第 <i>i</i> 月平均风速	ρ_p	颗粒密度
n_i	聚类分析中第 <i>i</i> 类的特征值	T_{p_1}	颗粒温度
d_i	第 <i>i</i> 月大风日数	D_p	扩散系数
S_i	第 <i>i</i> 月散射量	v_i	相的时均速
I_{csd}	可比沙尘强度指数	u_r	颗粒被加速的相对速度
I_{DDI}	沙尘指数	U_{g_0}	气流运动速度
Tra	气溶胶透射比	C_D	雷诺数 Re 的有关函数
τ	气溶胶光学厚度	μ	气体的动力粘性系数
		F_R	气流的曳引阻力

四、 模型建立与求解

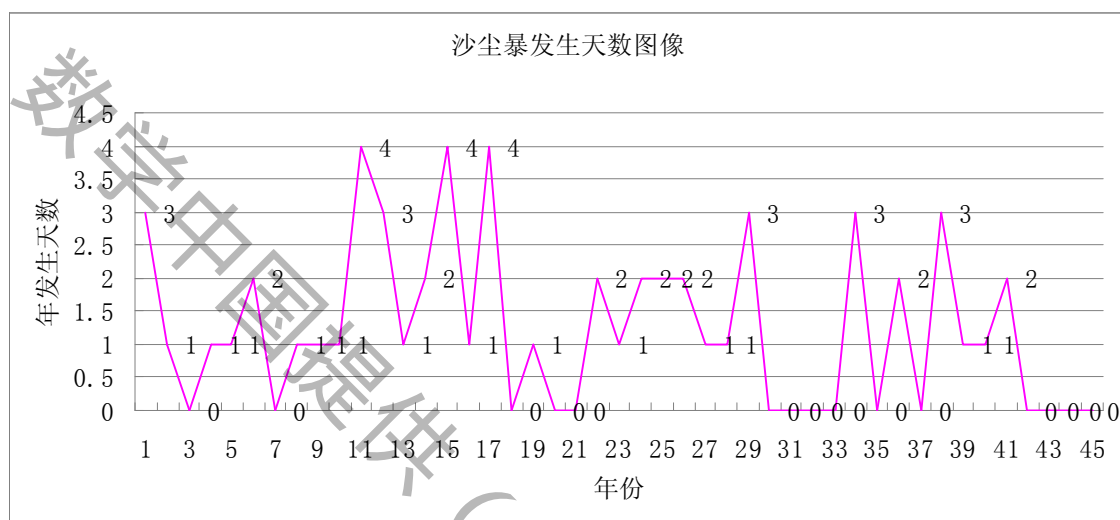
报名号#1102

4.1 问题一：

不同地区的地理气候条件不同, 导致各个气象因素对沙尘暴的产生影响程度不同, 但作用的原理一致, 根据相关气象数据的监测结果, 对历年沙尘暴发生天数的大致趋势进行聚类分析, 从而估计出沙尘暴在某地区某段时间内发生的次数. 在此, 我们任取一个观测站点, 如甘肃省的 52323 号观测站点, 对其 45 年来的观测数据进行研究.

4.1.1 宏观分析：

对 52323 号观测站点所在地区 45 年来沙尘暴发生的整体趋势进行分析, 软件统计每年沙尘暴发生的天数, 绘出图形, 如图 1 所示：

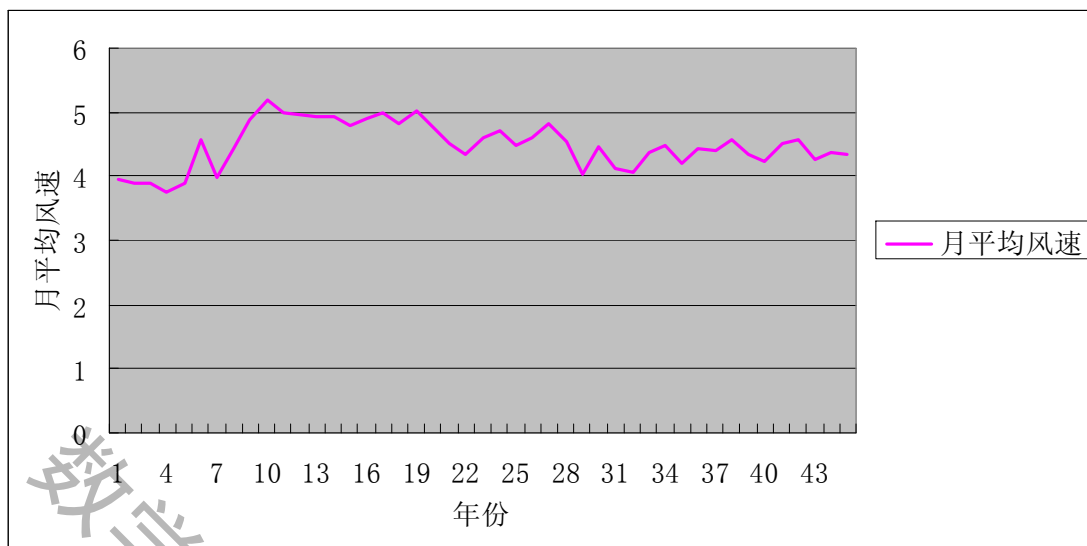


图一

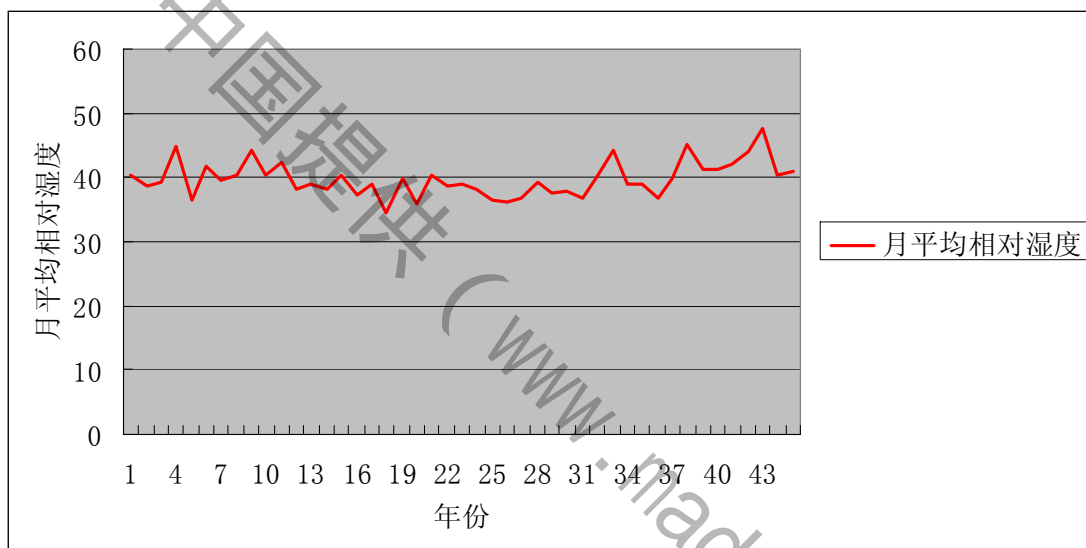
通过图一中折线的走向, 可以看出, 该地区在 60~70 年代年沙尘暴发生天数, 呈下降趋势, 20 世纪 70~80 年代沙尘暴活动频繁, 80 年代至 90 年代初沙尘暴频数基本保持稳定, 90 年代 20 世纪末沙尘暴在波动中回升, 直至 21 世纪初, 沙尘暴天数数较少. 总体上, 该地区沙尘暴天数是呈下降趋势的.

对该地区沙尘暴日数不断变化的原因, 我们可以通过该地区 45 年每月各项因素的变化来定性分析. 45 年来该地区的各个因素的大致走向如下图所示:

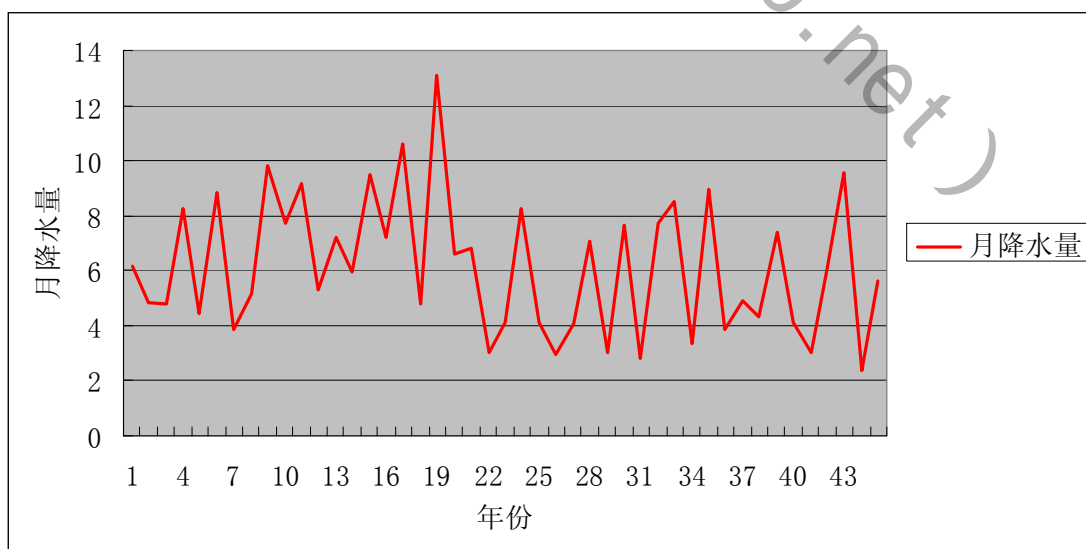
报名号#1102



图二

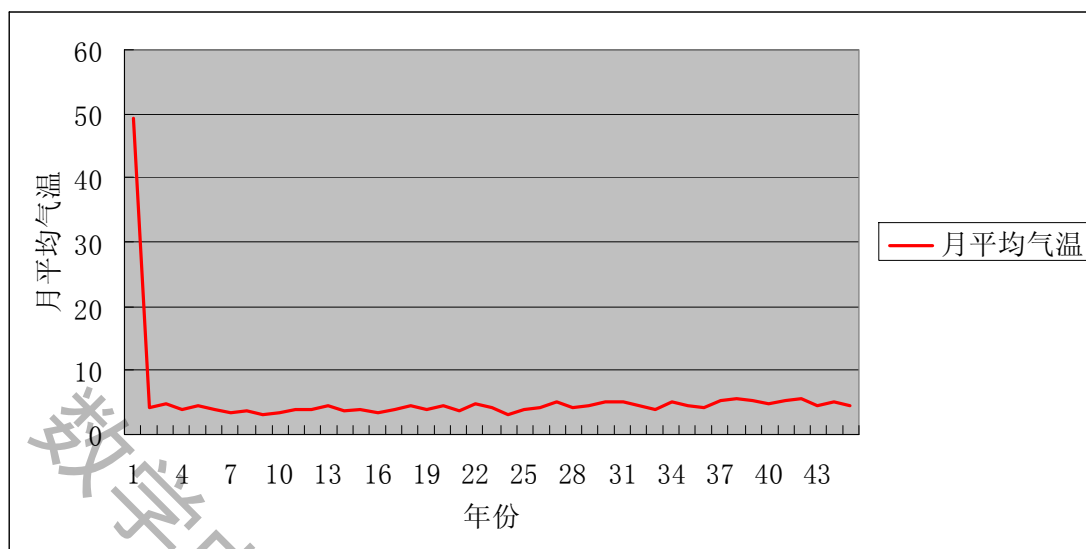


图三



图四

报名号#1102



图五

将图二、三、四、五与图一进行比较,图二、图四与图一地走向相似,即月平均风速和月降水量对沙尘暴发生的影响较为密切,但不能具体确定其他因素对其的影响大小,通过宏观分析只能定性的分析各个因素之间的关系,并不能确定各个因素对沙尘暴的产生的作用程度,故我们需要通过微观分析,量化它们之间的关系,以便较准确估计沙尘暴月发生天数.

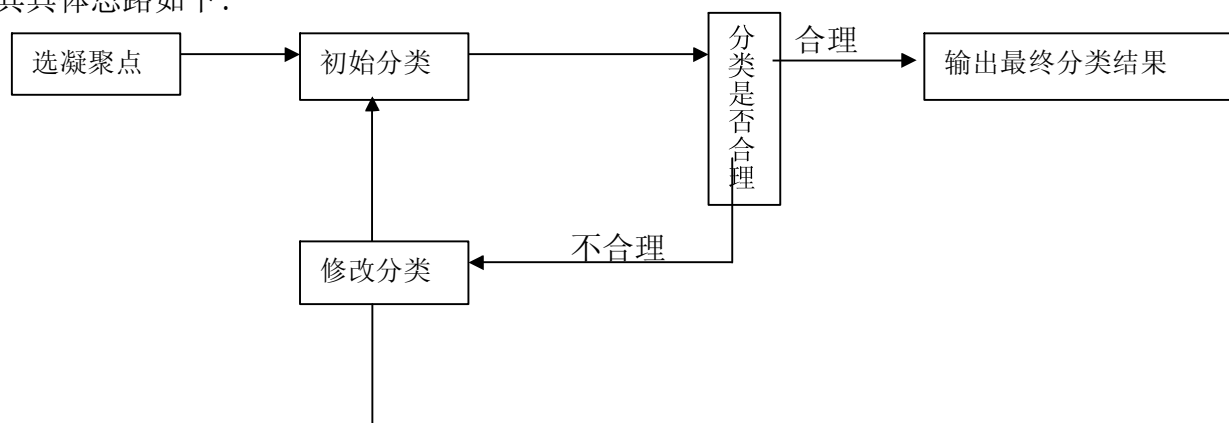
4.1.2 微观分析

对历年沙尘暴发生天数的宏观分析,只能从总体上反映每年沙尘暴发生天数的大致趋势,而不能对产生这种趋势的原因做出很好的解释,也不能根据某月相关气象数据较准确地对该月沙尘暴发生次数进行估计.在此我们通过对各个因素与沙尘暴发生之间的关系进行定量分析,以 52323 号观测站点为例,对其历年数据进行聚类分析.

4.1.2.1 SPSS 软件聚类分析

聚类分析是一种集机器学习、模式识别、统计分析和信息检索技术于一体的数据挖掘方法.其特点是在不需要训练集和预定义类别的情况下,即可从给定的空间向量集合中找到合理的聚类划分.通过数据聚类,可将空间向量划分为若干簇,并使得同一簇中的数据具有尽可能大的相似度,而簇间数据保持尽可能小的相似度,为信息的查询和检索提供了较好的优化和分析方法.

其具体思路如下:



报名号#1102

最近距离原则

将52323号站点的气象数据中七个气象因素引入聚类分析, 由于52323号观测站点所在地区月沙尘暴历年发生天数在0~3天范围内, 故我们按照沙尘暴月发生天数的种类进行分类.

通过分析得到了七个变量之间的关系, 其系数如下表所示:

表1 有关判别函数的输出

	沙尘暴天数			
	0	1	2	3
月	-.238	-.250	-.332	-.410
月平均风速	16.774	17.797	19.366	19.400
大风日数	1.489	1.556	2.083	3.101
月平均气温	.436	.311	.389	.588
月降水量	-.463	-.481	-.532	-.585
月蒸发量	.039	.048	.045	.033
月相对湿度	1.459	1.412	1.414	1.484
常数	-74.330	-79.474	-89.264	-94.570

将其转化为判别函数即为:

$$\begin{cases} n_0 = -0.238m_i + 16.774u_i + 1.489d_i + 0.436T_i - 0.463P_i + 0.039ETP_i + 1.459RH_i - 74.330 \\ n_1 = -0.250m_i + 17.797u_i + 1.556d_i + 0.311T_i - 0.481P_i + 0.048ETP_i + 1.412RH_i - 79.474 \\ n_2 = -0.332m_i + 19.366u_i + 2.083d_i + 0.389T_i - 0.532P_i + 0.045ETP_i + 1.414RH_i - 89.264 \\ n_3 = -0.410m_i + 19.400u_i + 3.101d_i + 0.588T_i - 0.585P_i + 0.033ETP_i + 1.484RH_i - 94.570 \end{cases}$$

函数式组所表征的意义:

四个函数式将该站所观测气象数据分为四类, 即分别为月沙尘暴发生 $i(i=0,1,2,3)$ 天这四类, 将该观测站的气象数据代入函数式组, 比较各个函数式值的大小, 当 n_i 最大时, 表示这组数据表征 i 类的特征值最大, 即该气象数据属于第 i 类别.

为检验其分类的正确性, 将该站点历年气象数据代入上式, 利用SPSS软件进行验证, 得到其分析结果如下表所示:

表2

报名号#1102

聚类分析验证结果^a

沙尘暴天数			Predicted Group Membership				Total
			0	1	2	3	
Original Count	0	0	203	55	6	17	281
	1	1	3	6	0	2	11
	2	2	0	2	1	3	6
	3	3	0	0	0	1	1
	Ungrouped cases		6600	0	0	0	6600
	%						
	0	0	72.2	19.6	2.1	6.0	100.0
	1	1	27.3	54.5	.0	18.2	100.0
	2	2	.0	33.3	16.7	50.0	100.0
	3	3	.0	.0	.0	100.0	100.0
	Ungrouped cases		100.0	.0	.0	.0	100.0

a. 70.6% of original grouped cases correctly classified.

由表2可知, 该模式下, 对月沙尘暴天数为0天的数据的分对率为72.2%, 月沙尘暴天数为1天的数据的分对率为54.5%, 月沙尘暴天数为2天的数据的分对率为16.7%, 月沙尘暴天数为3天的数据的分对率为100%, 总体上对气象数据的聚类分析正确率达到70.6%.

聚类误差分析

由于通过线性加权的方法对该气象数据进行聚类分析, 气象数据的分错率较大, 高达29.4%, 对于模型中的这个问题, 我们分析数据中7个变量的线性相关性.

利用SPSS软件, 通过分析得出各个因素之间的相互关系矩阵, 如下表所示:

表3

Correlation Matrix

	月	月平均风速	大风日数	月平均气温	月降水量	月蒸发量	月相对湿度
Correlation 月	1.000	-.032	-.200	.157	.056	.054	.157
月平均风速	-.032	1.000	.430	.025	-.094	.128	-.229
大风日数	-.200	.430	1.000	.141	.099	.243	-.298
月平均气温	.157	.025	.141	1.000	.493	.913	-.628
月降水量	.056	-.094	.099	.493	1.000	.390	.023
月蒸发量	.054	.128	.243	.913	.390	1.000	-.674
月相对湿度	.157	-.229	-.298	-.628	.023	-.674	1.000

当两个因素的相关性接近 ± 1 时, 表示两个因素的线性相关性越好, 但分析可知因素间相关性接近 ± 1 的较少, 通过线性估计进行聚类分析, 是导致误差较大的原因之一.

经总结:

第一, 特定地聚类分析方法, 分析结果很大程度上依赖于一系列人为主观地确定的参数, 而且没有一个很好的方法评价所得的聚类结果;

第二, 聚类谱系并不能表达各个样本间的真实关系.

第三, 采用线性聚类分析, 误差较大.

算法改进:

针对前面所存在的缺点, 利用非线性映射方法可在很大程度上克服了聚类分析的不足. 非线性映射通过某种非线性变换, 将高维空间的几何图像变换成地位空间的图像, 并且变换后仍能近似地保持原来的几何关系.

报名号#1102

为了达到上述目的, 我们引入非线性映射由高维变换到低维的约束条件:

$$K = \frac{1}{\sum_{i < j} d_{ij}^*} \sum_{i < j} \frac{(d_{ij}^* - d_{ij})^2}{d_{ij}^*} = \frac{1}{nF} \sum_{i < j} w_{ij} (d_{ij}^* - d_{ij})^2$$

其中 $nF = \sum_{i < j} d_{ij}^* = \sum_{i=1}^{N-1} \sum_{j=i+1}^N d_{ij}^*$, nF 称为标准化因子, $w_{ij} = \frac{1}{d_{ij}^*}$, w_{ij} 称为权重系数, d_{ij}^* 为

原来空间 n 个样本点间的距离, d_{ij} 为新空间中 n 个样本点的距离, K 的含义是使原空间距离与新空间距离之差平方和达到最小时来求得新空间点的几何构形.

选取甘肃省与 52323 号站点相近的 5 个站点的气象数据, 对 6 个观测站点每月的气象数据取平均值, 组合成一个新的数据样本, 从中提取了 96 组数据, 代入分析得到结果如表 4 所示:

表4

聚类分析验证结果^a

		Predicted Group Membership			Total
		.00	1.00	2.00	
Original	Count				
	.00	71	12	4	87
	1.00	1	7	0	8
	2.00	0	0	1	1
	Ungrouped cases	1	0	0	1
	%				
	.00	81.6	13.8	4.6	100.0
	1.00	12.5	87.5	.0	100.0
	2.00	.0	.0	100.0	100.0
	Ungrouped cases	100.0	.0	.0	100.0

a. 82.3% of original grouped cases correctly classified.

由表4可知, 该模式下, 对于月沙尘暴天数为0天的数据的分对率为81.6%, 月沙尘暴天数为1天的数据的分对率为87.5%, 月沙尘暴天数为2天的数据的分对率为100%, 总体上对气象数据的聚类分析正确率达到了82.3%, 即在该模型下估计的精度有了很大的提高.

4.1.2.2 基于改进遗传思想的 K-Means 聚类算法

spss软件聚类分析, 是将某些具有共同属性的对象划分成为一个群体, 按照事物间的相似性进行区分和分类的过程, 即“物以类聚”, 在这一过程中没有导师指导, 因此是一种无导师分类. 其分析中采用的算法不仅工作量巨大, 而且不能保证聚类结果的最优性, 如上面结果分析得出沙尘暴在某地区某段时间内发生次数的估计正确率仅达82.3%, 准确率较低, 实用性不高, 因此我们尝试采取一种新的聚类算法: 基于改进遗传思想的 K-Means 聚类算法

K-Means 算法的特点及其局限

k-means 聚类算法是一种硬聚类方法. 即在 n 维的欧几里得空间把 m 个样本数据分成 k 类. 首先, 由用户确定所要聚类的准确数目 k , 并随机选择 k 个对象 (样本), 每个对象称为一个种子, 代表一个类的均值或中心, 对剩余的每个对象, 根据其与各类中心的距离将它赋给最近的类. 然后重新计算每个类内对象的平均值形成新的聚类中心, 这个过程重

报名号#1102

复进行,直到满足收敛标准为止.

一般情况下,通常采用平方误差作为准则函数,其定义式为:

$$J = \sum_{i=1}^k \sum_{p \in c_i} |p - m_i|$$

式中为 J 为数据集中所有对象的平方误差的总和, P 为数据对象, m_i 为第 i 个类 c_i 的聚类中心.

它具有算法简单、局部搜索能力强且收敛速度快的特点,因而得到了广泛应用.

K-Means 算法的局限性主要如下:

1. 对象的最终聚类结果在某种程度上依赖于最初的划分
2. 产生的类的大小相差较大,对于“噪声”和孤立点数据很敏感
3. 要求用户事先给出要生成的簇的数目
4. 对非凸面形状的簇或大小差别很大的簇不适用
5. 算法经常以局部最优结束

传统遗传算法的优缺点

传统遗传算法是基于自然选择和遗传规律的一种全局搜索方法,是随机选择和适者生存理论的结合.该算法主要通过选择算子,变异算子,交叉算子的作用来产生“适者”.群体中的“适者”拥有更大的机会将其基因传给后代,因而也更能产生越加适应的“适者”.遗传算法自产生后获得广泛的应用,已被证明是一种优良的全局优化算法.但是,传统遗传算法存在以下几点不足:

1. 对问题机理不加分析,单纯进行演化运算,从而在过程中产生大量没有实际意义的染色体
2. 由于后代完全取代双亲,各代中的最优个体未能加以保留,优良染色体的结构被破坏的可能性很大,导致最终虽然得到了收敛结果,但是其适应度未必高于过程中产生的最优个体
3. 采用适应度函数作为选择种群的压力,如果压力太大,会导致没有可行解,反之压力过小,又会导致迭代过程过早收敛或者产生振荡.

基于改进遗传思想的 K-Means 算法

k-Means 聚类方法以算法简单,复杂度低,局部搜索能力强以及收敛速度快等优点成为我们的首选,但是,如何有效避免 K-Means 在聚类中陷入局部最优,我们考虑以遗传算法与之相结合,作为聚类分析的“导师”,起到判断聚类效果并对聚类参数进行优化的目的.

结合本题的实际,对传统遗传算法进行改进:

1. 判断算子 J_i

对染色体进行判断,剔除掉没有实际意义的染色体——在这里,我们认为分类数目大于 10 即为没有实际意义(因为几乎没有月沙尘暴天数超过 10 天的).

2. 双点交叉算子 T_i'

为了获取理想的搜索结果,我们对交叉算子 T_i 进行改进,由原来的单点交叉变为双点交叉 T_i' ,具体如下操作:

报名号#1102

采取轮盘赌的方式随机选取两个个体，利用 Rand (L) 产生 (0, L) 之间均匀分布的整数 $K, 1$ (L 为染色体长度)，以这两个点为交叉点依概率 p_m 进行交叉运算，重复这个过程直至生成一个新的种群

3. 梯度算子 TG_i

它以概率 1 对适应度值最高的染色体所对应的解作为初始点作梯度优化获得极值点，同时以概率 P_i 对其他串对应的解进行梯度优化，从而保证每次都能至少搜索到极值点，以此来加快收敛速度并保证结果最优。

数据的预处理

根据 Chapil 提出的气候因子概念，认为气候条件决定年均土壤的损失量，气候因子受表面土颗粒含水量和平均风速的影响，其中风速项是基于土壤的可蚀性，并随 Thornthwaite 指数变化的土壤近表层含水量的平方成反比。随后联合国粮农组织将 Chapil 公式做了修改得到气候因子：

$$C = \frac{1}{100} \sum_{i=1}^{12} u_i^3 \left[\frac{ETP_i - P_i}{ETP_i} \right] d_i$$

由气候因子的公式可知，气候因子与月平均风速非线性。针对上述模型中以多元线性模型进行分类导致的误差，我们通过对参与聚类的变量进行非线性修正——对历年月平均风速取立方，之后再对数据进行归一化处理，然后开始聚类分析。

基于改进遗传思想的K-Means聚类算法流程

Step1: 定义交叉概率 p_c ，变异概率 p_m ，种群数目 pop-size ，最大遗传代数 Max-GA ，收敛标准 EIP ，在这里我们选取

$p_c=0.7$ ， $p_m=0.3$ ， $\text{pop-size}=50$ ， $\text{Max-GA}=100$ ， $\text{EIP}=1 \times 10^{-6}$

Step2: 将 K-Means 聚类方法的各个参数进行二进制编码，同时定义遗传算法的适应度函数如下：

$$\text{Fitness}(\text{Ind}) = \frac{1}{1 + \sum_{j=1}^k \sum_{x_i \in c_j} \|x_i - z_j\|}$$

公式的含义是：计算各类中的样本到该类中心的距离，并求这些距离之和，得到各类的适应度。所有类的适应度之和加 1 并求倒数，得到染色体 Ind 的适应度。

Step3: 随机生成初始种群 pop ，以其为参数应用 K-Means 聚类方法对向量空间进行聚类，判断适应度函数是否达到收敛或者遗传代数是否达到最大值，若是，则转向 step，否则转向 step4

Step4: 采取精英保留策略，首先保留遗传过程中的精英个体。然后采用轮盘赌法，由适应度函数对应的概率分布确定把当前群体中的第 i 个体 Ind_i 按选择概率 $ps(\text{Ind}_i)$ 抽出， $ps(\text{Ind}_i)$ 由公式确定，即：

报名号#1102

$$ps(Ind_i) = \frac{Fitness(Ind_i)}{\sum_{j=1}^m Fitness(Ind_j)}$$

其中 m 为群体大小, $Fitness(Ind_i)$ 为第 i 号个体的适应度.

Step5: 采用双点交叉的方法, 利用 Rand (L) 产生均匀分布的数字 k, l , 采取轮盘赌的方式选取两个个体, 利用 Rand (L) 产生 (0, L) 之间均匀分布的整数 $K, 1$ (L 为染色体长度), 以这两个点为交叉点依概率 pm 进行交叉运算, 重复这个过程直至生成一个新的种群

Step6: 采用均匀多点变异的方法, 采用随机数生成的方法确定变异位置, 再以变异概率 pm 对其进行随机变异

Step7: 对变异之后的种群进行判断运算, 剔除没有实际意义的染色体, 计算种群适应度, 判断种群平均适应度是否达到收敛标准或者是否达到最大迭代次数 Max-GA, 若是, 则结束进化, 返回最佳种群和对应的适应度, 若不是, 则重复 step4~step6

基于改进遗传思想的 K-Means 算法计算结果

采用改进遗传算法的 K-Means 对甘肃省 52661 号站点的观测数据进行聚类分析, 得到数据中 7 个气象因素之间的关系, 其系数如表 5 所示:

表 5 有关判别函数的输出

沙尘暴天数	月	月平均风速	大风日数	月平均气温	月降水量	月蒸发量	月相对湿度	常数
0	1.0441	0.3556	-2.6313	1.3773	0.7279	0.8205	0.9619	1.2663
1	0.9436	1.3979	2.7439	0.6098	-0.9449	0.8630	0.6089	1.1437
2	1.9016	1.5086	-1.0613	-0.1845	1.1822	0.2762	0.5993	2.4966
3	1.3471	0.0674	0.0907	-2.1017	-1.4020	-2.2897	-0.4691	3.9657

其判别函数为:

$$\begin{cases} n_0 = 1.0441m_i + 0.3556u_i - 2.6313d_i + 1.3773T_i + 0.7279P_i + 0.8205ETP_i + 0.9619RH_i + 1.2663 \\ n_1 = 0.9436m_i + 1.3979u_i + 2.7439d_i + 0.6098T_i - 0.9449P_i + 0.863ETP_i + 0.6089RH_i + 1.1437 \\ n_2 = 1.9016m_i + 1.5086u_i - 1.0613d_i - 0.1845T_i + 1.1822P_i + 0.2762ETP_i + 0.5993RH_i + 2.4966 \\ n_3 = 1.3471m_i + 0.0674u_i + 0.0907d_i - 2.1017T_i - 1.402P_i - 2.2897ETP_i - 0.4691RH_i + 3.9657 \end{cases}$$

算法验证

从 52661 号站点的观测数据中取 300 组数据组成一个样本集合进行聚类验证, 验证结果如表 6 所示:

表 6

月沙尘暴天数	分对数	分错数	分对率
0	263.0000	10.0000	96.34%
1	16.0000	4.0000	80.00%
2	1	4.0000	20.00%

报名号#1102			
3	1	1	50.00%
合计	281	19	93.67%

由表6可知, 该算法下, 对于月沙尘暴天数为0天的数据的分对率高达96.34%, 月沙尘暴天数为1天的数据的分对率为80.00%, 月沙尘暴天数为2天的数据的分对率为20.00%, 月沙尘暴天数为3天的数据的分对率为50.00%, 总体上对气象数据的聚类分析正确率达到了93.67%, 在该算法下估计的精度较之SPSS软件聚类分析有了很大的提高.

4.1.3 结果比较与分析

遗传混合聚类算法首先随机产生遗传算法的第一代并开始进化, 在每代进化中都采用K-Means方法对每个个体进行进一步优化, 求得局部最优结果, 并以这些局部最优结果替换原来的个体继续进化, 直到达到最大代数或者结果符合收敛要求为止. 采用改进遗传算法对气象因子的优化选择, 从一定程度上提高了建模的快速性和预报的准确性.

通过对比结果可知, 基于改进遗传思想的K-Means聚类算法取得了良好的效果. 较SPSS软件聚类分析法, 该算法将分类精确度由82.3%提高到了93.67%, 特别在月沙尘暴天数为0天的分类上, 准确率高达96.34%, 同时月沙尘暴天数为1的分类上, 准确率也达到了80.00%, 而对于其它种类的分析, 由于抽取的样本点较少, 造成其分对率较低. 但是若以沙尘暴发生与否进行分类, 我们认为: 以目前所能达到的精度, 已经可以达到预报标准, 能够在一定程度上指导生活实践, 通过预测结果及时进行预防.

虽然通过采用基于改进遗传思想的K-Means聚类分析方法提高了估计的准确率, 但从其判别函数可知, 该算法仍然是通过线性映射的方法来进行分类, 不可避免的出现误差, 如同SPSS软件聚类算法的改进方法, 我们可以用非线性映射方法来优化算法, 提高准确率, 但由于时间关系, 在此仅提供思考方向, 不做深入研究.

4.1.4 算法时间复杂度分析

对于一个样本数为 N 的数据集, 设聚类数为 K , 数据对象的维数是 D . 如果采用K-Means算法进行聚类, 假设迭代次数为 T . 由于每轮迭代数据对象划分到某个类中, 要计算其到 K 个聚类中心的距离, 其时间复杂度为 $O(NKD)$, 而更新聚类中心的时间复杂度要小于 $O(NKD)$, 因此整个算法的时间复杂度为 $O(NKDT)$. 如果采用文中提到的基于改进遗传算法的K-Means算法进行聚类, 假设进化代数为 G . 由于每一代进化中, 均涉及到距离计算、聚类中心的调整、染色体选择、交叉和变异, 时间复杂度为 $O(NKD)$, 因此整个算法的时间复杂度为 $O(NKDG)$. 由于HPGA算法进化代数和K-Means算法迭代次数处于同一数量级, 可以得出这样的结论: 基于改进遗传算法的K-Means算法在实现全局寻优的同时, 并没有增加算法的时间复杂度, 它和普通K-Means算法一样具有较高的效率, 能够快速收敛并且效果最优.

4.2 问题二

4.2.1 沙尘暴天气等级

根据国家气象局关于沙尘天气气象等级定义, 可以将沙尘天气分为4个强度等级即:

浮尘: 尘土、细沙均匀地浮游在空中, 使水平能见度小于10公里的天气现象;

报名号#1102

扬沙：风将地面尘沙吹起，使空气相当混浊，水平能见度在1公里至10公里以内的天气现象；

沙尘暴：强风将地面大量尘沙吹起，使空气很混浊，水平能见度小于1公里的天气现象；

强沙尘暴：大风将地面尘沙吹起，使空气很混浊，水平能见度小于500米的天气现象。

4.2.2 沙尘暴天气气象等级评估及预测模型

本文针对不同的预测周期，将预测分为长期和短期预测两个模型，对于长期预测模型，本文通过气溶胶厚度与太阳辐射的关系来预测沙尘暴的等级；而针对短期模型，通过卫星遥感反演法，建立模型，预测沙尘天气气象等级。

4.2.2.1 长期预测模型

沙尘暴是由不同大小的沙尘粒子构成的，其在空中浓度很强时会形成一个密度很厚的沙尘层，这个地表和太阳辐射，同时产生反射、散射，并向外发射辐射。因此我们可以通过分析沙尘暴形成的沙尘层向外发射的辐射量，将国家气象局关于沙尘天气气象等级转化为能见度等级。

气溶胶是指悬浮在空气中的由固体和液体微粒与气体载体共同组成的多相体系，是大气中的一种粒子成分。沙尘层吸收地表和太阳的辐射量与气溶胶的光线透射比及气溶胶光学厚度有关，而通过气溶胶的光线透射比可确定气溶胶光学厚度，即通过空气中沙尘的光线透射比来确定沙尘的浓度，确定每个沙尘天气的等级。

根据资料[1]气溶胶光学厚度大时，气溶胶透射比小；气溶胶光学厚度小时，气溶胶透射比大。其中气溶胶透射比：

$$Tra = \exp[-\tau \cdot m(\theta)]$$

$m(\theta)$ 代表太阳天顶角为 θ 时的大气光学质量，气溶胶光学厚度大时，气溶胶透射比就小。而对于水平太阳光束面上的直接辐射 S' 与气溶胶透射比互为近似线性关系：

$$S'/Q = a \cdot Tra + b$$

$$\text{其中 } a = 0.92, b = 0.18$$

由于气溶胶透射比 Tra 与 S'/Q 、 D/Q 之间有良好的线性关系，随着气溶胶透射比的增大，水平面上的直接辐射 S' 对总辐射的贡献增大，而散射辐射 D 对总辐射的贡献减小。这是由于气溶胶透射比大时，其对太阳直接辐射的削弱就小，因而地面接收到的直接辐射大，散射辐射小，结果使水平面上的直接辐射对总辐射的贡献增大，而散射辐射对总辐射的贡献减小。

根据上述式子进行变形可得：

$$S'/Q = a \cdot \exp[-\tau \cdot m(\theta)] + b$$

$$\tau = \frac{S'/Q - b}{a} \cdot \frac{1}{m(\theta)}$$

$$\text{故 } \tau = -\frac{S'/Q - b}{a} \cdot \cos \theta, \text{ 其中 } \frac{1}{m(\theta)} = -\cos \theta$$

报名号#1102

故可以得到能见度与辐射的关系：

$$c(vv)^d = \frac{S'/Q - b}{a} \cdot \cos \theta \cdot \lambda^\alpha$$

在此，我们引入大气柱光学厚度[2]，其是采用太阳光度计测量太阳直射光强，并由Langley方法反演而得，并假设光学厚度由下所决定：

$$\tau = c(vv)^d \cdot \lambda^{-\alpha}$$

β 为大气浑浊度参数； α 为波长指数。

引用AlmeidaE的经验公式：

$$\beta = c(vv)^d$$

其中 vv 表示水平能见度，单位km， $c = 2.26, d = -0.73$ 。

故可以导出能见度与辐射的关系式：

$$vv = 1/d \cdot \ln \frac{S'/Q - b}{ac} \cdot \cos \theta \cdot \lambda^\alpha$$

其中透射光线波长 $\lambda = 550nm, \alpha = 1.3$ 。

根据国家气象局沙尘天气气象等级的划分，我们可以得出：

$$vv = \begin{cases} \text{浮尘, } (0 \leq vv \leq 10\text{km}) \\ \text{扬沙, } (1 \leq vv \leq 10\text{km}) \\ \text{沙尘暴, } (0 \leq vv \leq 1\text{km}) \\ \text{强沙尘暴, } (0 \leq vv \leq 0.5\text{km}) \end{cases}$$

根据观测站的气象数据值，我们可以得到其对应的水平能见度，将其划分到所属的能见度等级，以此来评估沙尘天气气象等级。

4.2.2.2 短期预测模型

由于沙尘暴是由不同大小的沙尘粒子构成的，其在空中浓度很强时会形成一个密度很厚的沙尘层，这个沙尘层会强烈地吸收地表和太阳辐射，同时产生反射、散射，并向外发射辐射。

因此我们通过分析沙尘暴形成的沙尘层向外发射的辐射量，建立起与沙尘暴强度的函数模型。而在现实生活中，测定沙尘暴强度，大致可以分为直接观测和遥感反演两种方法，在此我们应用遥感反演的方法进行测定。

专家通过研究沙尘的光谱分布特征和粒子构成的物理特征表明，沙尘信息在遥感数据的各波段上均有强弱不同的反映，因此在沙尘的识别过程中，必须综合考虑各有关波段的特性。

由于沙尘粒子的分布跨度较大，观测到的粒子半径 r 可以从 $0.1 \mu m - 100 \mu m$ ，较强沙尘天气粒子半径分布最大值常在 $5 - 10 \mu m$ 之间。大气分子与微粒气溶胶对可见光有较

报名号#1102

强的散射，成为大气沙尘遥感的干扰因素. 在我们使用的所有遥感数据中，位于 $3.7\mu\text{m}$ 和 $1.6\mu\text{m}$ 波段附近的中红外和近红外数据对沙尘信息有最明显反应.

而实际应用中 $1.6\mu\text{m}$ 波段仍然受到空间、时间、卫星等方面因素的干扰，监测结果仍不具可比性，不能满足多源（多时次，多平台）遥感数据在沙尘暴动态监测中定量分析的要求. 为达此目的，经专家反复试验，利用 $1.6\mu\text{m}$ 波段特性，构建了可比沙尘强度指数 I_{csd}

即：

$$I_{csd} = \alpha \times (e^{\beta \times R_{1.6}} - 1)$$

其中 α 和 β 为调节因子，通过查阅相关资料知，其一般取值分别为 10 和 0.8. 且 $R_{1.6}$ 与 I_{csd} 在各个地区的经验值如下表所示：

表7

	东北地区	华北地区	南疆地区
$R_{1.6}$	0.45	0.72	0.83
I_{csd}	0.42	0.41	0.45

由于 $R_{1.6}$ 的取值变化非常小，所以 I_{csd} 的值无明显变化，故可取

$$I_{csd} = 0.42$$

沙尘指数计算公式

利用可比沙尘强度指数的稳定性，同时又利用 $11\mu\text{m}$ 、 $12\mu\text{m}$ 波段探测值的差值对沙尘信息的敏感性，结合二者的优点，最终得到一个可定量描述沙尘暴识别信息和强度信息的沙尘指数，其计算公式如下：

$$I_{DDI} = I_{csd} \cdot [(T_{12} - 1) / T_{11} - 1] \cdot 100 - 100.$$

沙尘指数的计算结果是一个 0-100 之间的无量纲数值，值越大表示沙尘强度越强. 查阅相关资料得到能见度与沙尘指数对照表.

表8

	正常	浮尘	扬沙	沙尘暴	强沙尘暴	特强沙尘暴
沙尘指数	1-10	11-30	31-50	51-70	71-85	> 90
能见度(km)	≈2.0	2.0-1.5	1.5-1.0	1.0-0.5	0.5-0.1	< 0.1

鉴于此，解决沙尘暴的分类的核心问题是求解 T_{11} 、 T_{12}

根据联合国粮农组织修改的 Chepil 公式：

气候因子：

$$C = \frac{1}{100} \sum_{i=1}^{12} u_i^3 \left[\frac{ETP_i - P_i}{ETP_i} \right] d_i$$

潜在蒸发：

$$ETP_i = \begin{cases} 0.19(20 + T_i)^2(1 - RH_i) & , T_i > -20 \\ 0 & , T_i < -20 \end{cases}$$

当潜在蒸发小于降水时，由于地面不会产生实际的沙尘暴，虽然风蚀的潜在能力依

报名号#1102

然存在，但可认为风蚀指数为 0
因此可以设

$$T_{11} = k_1 \ln \left\{ u_i^3 \left[\frac{ETP_i - P_i}{ETP_i} \right] d_i \cdot S_i \right\}$$

$$T_{12} = k_2 \ln \left\{ u_i^3 \left[\frac{ETP_i - P_i}{ETP_i} \right] d_i \cdot S_i \right\}$$

利用 Lingo 软件，按 $I_{DDI} = I_{csd} \cdot [(T_{12} - 1) / (T_{11} - 1)] \cdot 100 - 100$ ，对一个省的月平均散射辐射量和其他各个气象因子月平均值进行拟合得到 $k_1 = 0.208$, $k_2 = 0.4438$ 。

因此，气象数据代入求解出的 I_{DDI} 公式中，将所得的函数值与表 8 对比，即可判别天气的气象等级。

4.2.3 模型分析

长期预测模型基于气溶胶厚度与太阳辐射的关系，建立能见度与太阳辐射的关系，对要进行评估的地区，选取该地区的气象数据，得出其对应的能见度，根据模型中所定义的沙尘天气气象等级标准，就能直观地对沙尘气象天气进行判断。但该模型只考虑了太阳辐射，对于其他因素的影响，并未考虑，因此下一步可以通过灵敏度分析，得出对其影响较大的因素，对该模型进行修正。

短期预测模型基于遥感反演法，通过沙尘暴形成的沙尘层向外发射的辐射量，建立起与沙尘暴强度的函数模型，并引入沙尘指数，将沙尘暴强度转化为沙尘指数，并通过沙尘指数建立与能见度的联系，最终将沙尘暴强度通过能见度等级划分出来。该模型因受沙尘的光谱分布特征影响，且采用的波段易受到空间、时间、卫星等方面因素的干扰，在实际应用方面受到了一定限制，但其较好的实时性，仍然值得我们去推广和应用。

两个模型只能对沙尘天气气象等级进行估计，对沙尘暴发生的风险值未能做出研究，因此，在下一步工作中，我们应该在此模型的基础上，着重解决沙尘暴发生概率的研究，但由于时间有限，不做进一步探讨。

4.3 问题三

为了判断中国境内引起沙尘暴的重要沙源地位置及其影响范围进行分析，我们必须分析沙尘暴的成因及物理机制及其运动路线。首先判断沙源地的位置，再通过确定沙尘暴的运动路线，根据路线中各种气象因素，及沙粒的漂移和沉淀机制的共同作用，判断沙尘暴的影响范围。

4.3.1 沙源地的判断

沙源地的位置与沙尘暴爆发的频率密切相关，通过沙尘暴爆发的次数，我们可以较准确地判断沙源地的位置，本文以内蒙古各个站点的沙尘天气气象观测数据为例，通过统计分析内蒙古省内的沙源地。

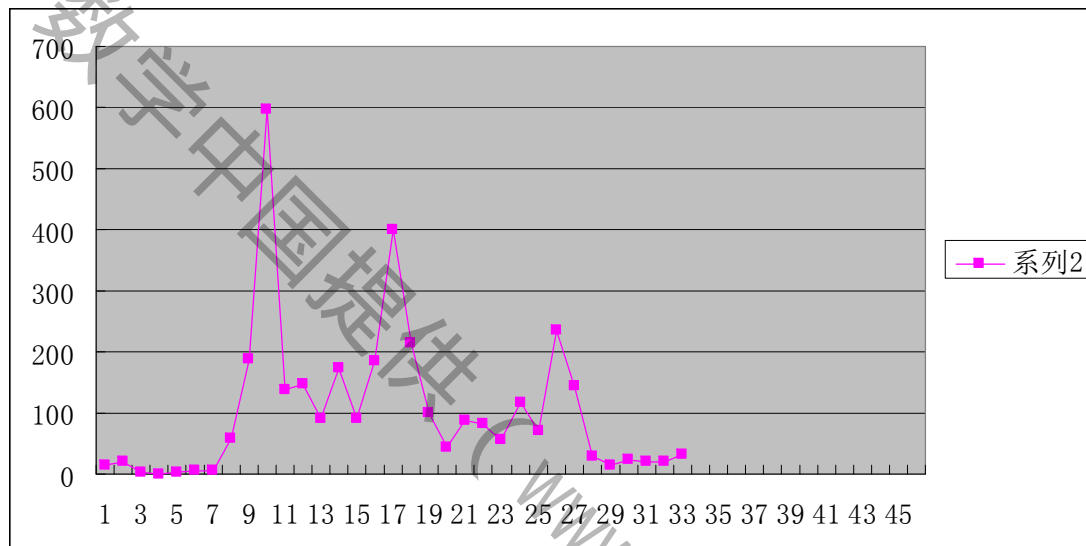
对内蒙古1981至2005年每个观测站点所在地区25年共发生沙尘暴天数进行统计，结果如表9所示：

表9 各个站点25年总发生沙尘暴天数

站点	50425	50514	50548	50632	50727	50834	50838	50915
----	-------	-------	-------	-------	-------	-------	-------	-------

	报名号#1102							
发生次数	15	20	4	0	2	7	6	59
站点	52267	52378	52576	53068	53083	53149	53192	53195
发生次数	189	597	137	148	92	174	90	185
站点	53231	53276	53352	53362	53391	53446	53480	53502
发生次数	401	214	100	43	87	82	57	117
站点	53513	53529	53602	54012	54115	54135	54213	54218
发生次数	71	236	143	29	16	24	22	21

由于上述站点按照序号数序相邻，将其转化为图表，如下图所示：



图六

根据图六可知，该图出现了3个峰值点，而这些站点两侧相邻站点所在地区爆发沙尘暴次数逐渐减少，即该站点产生的沙尘暴对其他站点的影响随着距离的变大而逐渐减小，故本文认为3个峰值点即内蒙古省沙尘暴的沙源地，它们从左至右分别是：拐子湖、海力素、鄂托克旗。

根据沙尘暴发生频率、强度、沙尘物质组成与分布、生态现状、土壤水分含量、水土利用方式和强度，结合区域环境背景将我国西北方划分出4个主要沙尘暴中心和源区：1、甘肃河西走廊及内蒙古阿拉善盟；2、南疆塔克拉玛干沙漠周边地区；3、内蒙古阴山北坡及浑善达克沙地毗邻地区；4、蒙陕宁长城沿线。上述沙尘暴多发地区的沙尘也常随西风和西北气流输送到华北及长江中下游，形成沙尘天气。

4.3.2 沙尘暴运动路线的判断

对内蒙古1981至2005年每个观测站点所在地区25年月沙尘暴天数进行统计，利用SPSS软件分析每个站点月发生沙尘暴的相关性，分析结果如下表所示：

表 10

	索伦	乌兰浩特	东乌珠穆沁旗	额济纳旗	拐子湖	阿拉善右旗	二连浩特	那仁宝力格	满都拉	阿巴嘎旗	苏尼特左旗
索伦	1	0.1276	0.1492	-0.045	-0.05	0.142	0.1406	0.1254	0.3039	0.259	0.2266
乌兰浩特	0.1276	1	0.336	-0.020	-0.015	0.1679	0.2458	0.5232	0.2808	0.3572	0.2591
东乌珠穆沁旗	0.1492	0.336	1	-0.064	-0.090	0.3681	0.4804	0.5223	0.3715	0.4049	0.4348

报名号#1102

额济纳旗	-0.045	-0.020	-0.0635	1	0.460	0.0411	-0.0888	-0.1197	-0.118	-0.15	-0.1483
拐子湖	-0.05	-0.015	-0.0902	0.4595	1	-0.067	-0.1229	-0.1017	-0.117	-0.1102	-0.1527
阿拉善右旗	0.142	0.1679	0.3681	0.0411	-0.07	1	0.2973	0.3588	0.2078	0.2844	0.164
二连浩特	0.141	0.2458	0.4804	-0.089	-0.12	0.2973	1	0.5965	0.5825	0.5652	0.6398
那仁宝力格	0.125	0.5232	0.5223	-0.120	-0.10	0.3588	0.5965	1	0.5295	0.6972	0.622
满都拉	0.304	0.2808	0.3715	-0.1177	-0.12	0.2078	0.5825	0.5295	1	0.5753	0.5772
阿巴嘎旗	0.259	0.3572	0.4049	-0.15	-0.11	0.2844	0.5652	0.6972	0.5753	1	0.6064
苏尼特左旗	0.227	0.2591	0.4348	-0.148	-0.15	0.164	0.6398	0.622	0.5772	0.6064	1

当两个站点的相关系数很大时，即两个站点的沙尘暴发生具有很大的相关性，随着季风的影响，可以根据风向，及各个站点的相互关系，定性的分析出沙尘暴的运动轨迹. 取二连浩特为沙尘暴初始运动站点，沙尘暴具体运动轨迹为：

二连浩特 → 那仁宝力格 → 满都拉 → 阿巴嘎旗 → 苏尼特左旗

继续对上述几个站点及其相邻站点进行相关性分析，分析结果如表 11 所示

表 11

	满都拉	阿巴嘎旗	苏尼特左旗	海力素	朱日和	乌拉特后旗
满都拉	1.0000	0.5756	0.5732	0.5099	0.5428	0.3722
阿巴嘎旗	0.5756	1.0000	0.6052	0.3981	0.4417	0.2347
苏尼特左旗	0.5732	0.6052	1.0000	0.3065	0.3252	0.2229
海力素	0.5099	0.3981	0.3065	1.0000	0.3283	0.2184
朱日和	0.5428	0.4417	0.3252	0.3283	1.0000	0.4625
乌拉特后旗	0.3722	0.2347	0.2229	0.2184	0.4625	1.0000

同样通过此表结合季风走向，可以推断出沙尘暴在这几个站点间的运动轨迹，具体路线为：

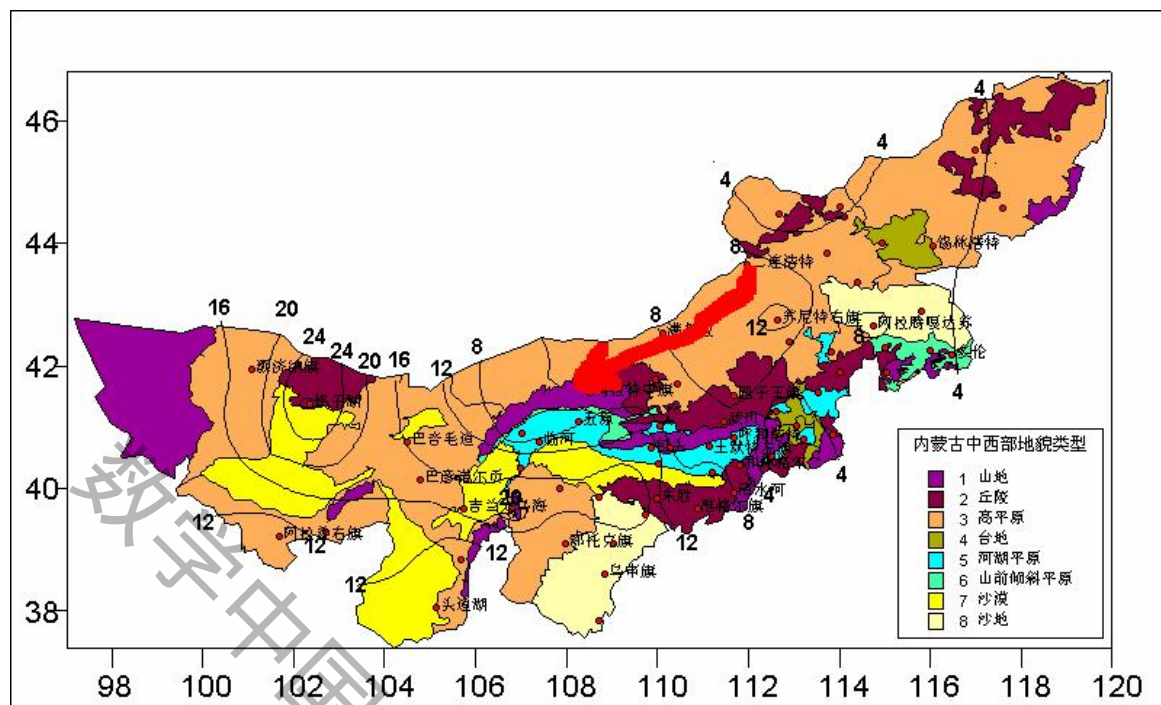
苏尼特左旗 → 海力素 → 朱日和 → 乌拉特后旗

综合两次的分析结果，可以得出内蒙古北部沙尘暴的运动轨迹，由于二连浩特并非沙源地，故可能二连浩特是其它地区沙源地沙尘暴运动过程中的一个站点. 其具体运动轨迹为：

二连浩特 → 那仁宝力格 → 满都拉 → 阿巴嘎旗 → 苏尼特左旗 → 海力素 → 朱日和 → 乌拉特后旗

即沙尘暴在内蒙古北部的运动轨迹如图七（运用相关性可判断出内蒙古的沙尘暴的沙源地如图中所示）中红色箭头所示：

报名号#1102



图七

4.3.3 沙源地及运动路线判断模型验证

经查阅相关气象数据及历史气候资料验证,该沙尘暴运动路线的沙源地位于蒙古国中南部,沙尘暴的南下运动首先影响中国的二连浩特,然后经该区域影响张家口,继续影响北京,从而验证了模型的准确性,对于其他省份及其他站点,我们可以采用同样的方式,推断沙尘暴的运动轨迹.

4.3.4 沙尘暴的成因及物理机制

沙尘暴天气成因

沙尘暴的发生一般需要强劲的风力、丰富的沙尘源和不稳定的空气层结三个必备条件,其中强风是产生和发生沙尘暴的动力条件,局地大气强热力不稳定是发生对流、沙尘垂直输送的热力条件,以沙尘为主的疏松的地表物质是沙尘暴发生的物质基础.

沙尘暴形成的物理机制

在极有利的大尺度环境、高空干冷急流和强垂直风速、风向切变及热力不稳定层结条件下,引起锋区附近中小尺度系统生成、发展,加剧了锋区前后的气压、温度梯度,形成了锋区前后的巨大压温梯度.在动量下传和梯度偏差风的共同作用下,使近地层风速陡升,掀起地表沙尘,形成沙尘暴或强沙尘暴天气.

起沙尘机制及模式

在风沙运动过程中,一些沙粒在风力作用下直接起动,这种起动叫气动起动或流体起动.一旦沙粒脱离地表进入气流,就在气流中很容易被风加速,从而从气流中获得动量.一部分沙粒会获得足够的动量,以致当它们返回地表与地表撞击时会发生回弹或使其它地表颗粒因颗粒冲击而起动进入气称之为回弹起动和溅射起动或冲击起动.

1、起沙的条件

报名号#1102

► 运动形式定义

不同粒径的沙粒有不同的运动方式，拜格诺[3]将沙粒的运动形式分为三种，即悬浮，跃移和蠕移。悬浮的粒子直径一般小于 $70\mu\text{m}$ ，这种粒子因为重量轻，沉降速度小，可以在大气中悬浮。跃移的粒子直径大约在 70 到 $1000\mu\text{m}$ 之间。跃移的粒子在风蚀表面作弹跳运动。蠕移的粒子直径大于 $1000\mu\text{m}$ ，这个尺度的粒子因为太重而在地表作滚动运动。

► 起沙判断条件

从微观分析：

当 $f_a > f_g + f_i$ 时粒子脱离地面，起沙

当 $f_a < f_g + f_i$ 时粒子无法脱离地面，不起沙

从宏观分析：

是否起沙决定于实际的摩擦速度 u_* （即空气动力）和沙粒脱离地表的临界摩擦速度 u_{*t} （即临界空气动力）。当 $u_* > u_{*t}$ 时，空气动力大于沙粒的重力和内部粘性力的合力，跳跃运动发生，地面会因此而起尘；当 $u_* < u_{*t}$ 时，空气动力小于或等于微粒的重力和内部粘性力的合力，跳跃运动不能发生，地面也不会因此而起尘。

2、临界摩擦速度 u_{*t} 的作用

临界摩擦速度的取决因素：

- (1) 气象条件：风速、气温、降水
- (2) 土壤类型
- (3) 土壤状态：水分、紧固度、聚集度
- (4) 地面微观条件：粗糙度、植被覆盖等
- (5) 地面宏观条件：地貌、防风林等

3、垂直沙尘通量 F 的计算

当跃移粒子在风的作用下在土壤表面跳跃时，粒子从空气里吸收了动量并在垂直方向产生一定的动量输送。当跃移粒子冲击地面时，使更多的粒子飞溅到空气中。相比之下，较小的尘粒，一旦悬浮起来，便可以悬浮在空中一段时间，且由于湍流作用可扩散到界外层以外，长时间传输到下游较远的地方。

研究表明， F 主要包含三个部分：

(1) 在微弱的风蚀条件下，没有跳跃运动，地表土壤中的尘粒直接在空气动力的作用下被抬升到大气中的部分 F_a ，理论上，沙尘粒子可以被空气动力直接抬离地面，对于较小粒子重力和空气动力的重要性相对减小，而粒子间的作用力变得相对重要，沙尘粒子不易被风吹离地面，因此直接由空气动力造成的起沙尘率比较小；

(2) 地表土壤中的尘粒由于沙粒的跳跃碰撞而悬浮于空中的部分 F_b ；

(3) 覆盖在集合体表面的尘粒由于沙粒的跳跃碰撞而悬浮于空中的部分 F_c 。

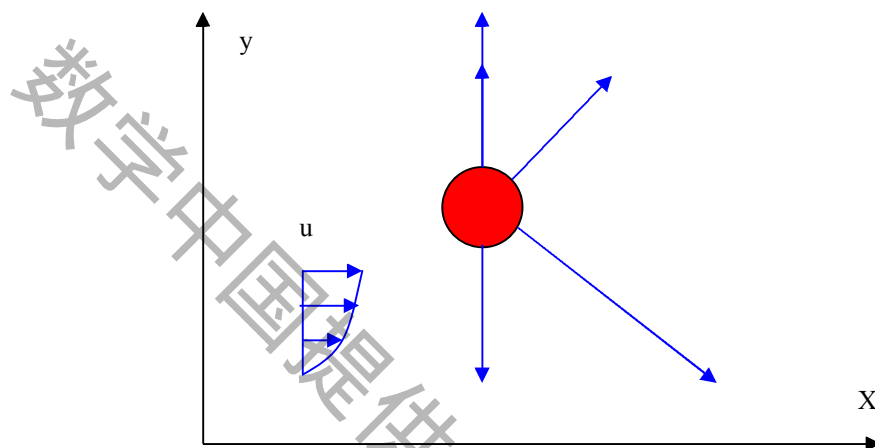
在较强的风蚀起沙尘过程中， F_a 相对很小，可以忽略不计， F_b 和 F_c 则量级相当，是地表尘粒排放的主要部分。

4、起沙模式及其结构

► 单颗沙粒受力分析

(1) 把粒子抬离地表的力 f_a (空气中粒子的浮力和拖曳力), 与近地层的风切变有关, 因此其为摩擦系数 u_* 的函数

(2) 阻碍粒子运动的力, 重力 f_g 和粒子间作用力 f_i (范德瓦尔斯力、静电力、毛细管力等)



图八 单颗沙粒的受力分析图

$$\text{空气阻力 } F_{di} = \frac{1}{2} C_d \rho \Delta v \left(\frac{di}{dt} - u \right) \pi \left(\frac{d}{2} \right)^2$$

$$\text{Magnus 升力 } F_{mi} = \frac{1}{8} \pi \rho d^3 \frac{di}{dt} \omega$$

当 Magnus 力大于等于重力时, 不再受临界粒径的约束, 而仅受临界相对垂直速度的影响. 当 Magnus 力与重力相等时, 临界相对垂速与粒径无关. 这说明, 此时任何粒径的沙粒均可以稳定的上升, 不存在风对沙粒的筛选问题. 当 Magnus 力大于重力时, 临界相对垂速与粒径成正比. 这意味着, 粒径越大越不容易维持上升, 粒径越小越容易维持上升.

则沙粒的运动的基本方程为:

$$m \frac{d^2 x}{dt^2} = -F_{dx} + F_{my} \quad (1)$$

$$m \frac{d^2 y}{dt^2} = -mg - F_{dy} - F_{mx} \quad (2)$$

其中, $d = 0.25 \text{ mm}$, $\rho_p = 2.65 \times 10^3 \text{ kg/m}^3$, $\rho = 1.29 \text{ kg/m}^3$, $\gamma = 0.133 \text{ cm}^2/\text{s}$.

Δv 为沙粒与风之间的相对速度可表示为: $\Delta v = \left[\left(\frac{dx}{dt} - u \right)^2 + \left(\frac{dy}{dt} \right)^2 \right]^{1/2}$

C_d 为阻力系数, 且可表示为: $C_d = \frac{24}{Re} + \frac{6}{1 + (Re)^{1/2}} + 0.4$, Re 是雷诺数, 可由式求得:

报名号#1102

$Re = \frac{d}{\gamma} \Delta v$ ，来流风速为： $u = u_* \ln(y/y_0)/k$ ， u_* 是摩阻风速， k 是冯·卡门常数 0.4，粗糙度 y_0 一般取为 $d/30$ 。

以此计算水平位移与竖直方向的位移的关系，从而给定参数即可求出沙尘的水平漂移的位置

➤ 沙粒通量

风沙物理学中用顺风向沙粒通量 Q 表征进行跳跃运动的沙粒的量。是沙粒漂移强度 $q(z)$ 沿高度的积分，其物理意义是指单位时间内单位宽度从地表到积分高度处所形成的平面（与水平风向垂直）内的沙粒质量。具体计算时，对于只包含单一粒子尺度的土壤，最常用的是 *Owen* 方程：

$$Q = \begin{cases} \frac{c_s \rho u_*^3}{g} \left[1 - \left(\frac{u_{*t}}{u_*} \right)^2 \right] & u_* > u_{*t} \\ 0 & u_* \leq u_{*t} \end{cases}$$

其中 ρ 为空气的密度， c_s 是经验常数，为 0.8。

Lu 等人在考虑土壤的可侵蚀性和塑性压力的基础上，从理论上计算了沙粒碰撞土壤表面时所产生的凹坑的体积，并建立了一个沙尘排放模式：

$$F(d) = \frac{5gf\rho_b}{2p} \left(0.24 + 1.37u \sqrt{\frac{\rho_p}{p}} \right) Q$$

f 是地表土壤中尘粒的质量分数，由粒子尺度分布情况决定； ρ_b 和 ρ_p 分别是土壤的体积密度和微粒密度， $\rho_p = 2650 \text{ kg} \cdot \text{m}^{-3}$ ； p 是土壤的塑性压力，是表征地面阻力大小的一个物理量，由土壤的致密程度决定，松软的土壤阻力较小， p 也较小，致密的土壤阻力较大， p 也较大； ρ_b 和 p 对于不同的土壤类型有不同的取值。

Shao 等人在充分考虑地表土壤风蚀起沙的物理机制和风蚀起沙尘过程中地表土壤粒子尺度分布的基础上，提出了一个更完善的计算方案：

$$F(d) = c_y \frac{p(d)}{p_f(d)} \cdot \frac{Qg}{u_*^2 m} (\rho_b \eta_f \Omega + \eta_c m)$$

$$p(d) = \gamma p_m(d) + (1 - \gamma) p_f(d)$$

其中：

- c_y 是常数，为 0.114；
- d 是尘粒的直径；
- $p(d)$ 是地表土壤粒子尺度分布的概率密度函数；
- $p_m(d)$ 是地表土壤遭受最小程度破坏时粒子尺度分布的概率密度函数；

报名号#1102

- $p_f(d)$ 是地表土壤遭受最大程度破坏时粒子尺度分布的概率密度函数；
 - η_f 指地表土壤中能够排放到大气中的尘粒的质量分数；
 - η_c 指覆盖于集合体表面的尘粒的质量分数；
 - ρ_b 为土壤的体积密度；
 - m 是直径为 d 的沙粒的质量, 具体计算时将沙粒近似作为球形；
 - Ω 是沙粒碰撞土壤表面时所产生的凹坑的体积, 可根据 Lu 等的计算方法确定。
- $p_m(d)$ 和 $p_f(d)$ 表征的是地表土壤两种理想的粒子尺度分布状态, 对不同类型的地表土壤其分布特征也不同; γ 和 $(1-\gamma)$ 分别是 $p_m(d)$ 和 $p_f(d)$ 在风蚀起沙尘时所占的权重, $\gamma = e^{-t(u_* - u_{*t})^n}$, t 和 n 是两个经验系数, 近似取 27.3 和 3, η_f 与 η_c 二者均决定于地表土壤的粒子尺度分布;

对包含多种尺度粒子的自然界土壤, 可将上式根据地表土壤中沙粒的尺度分布 $p(d)$ 进行积分, 得到:

$$F = \int_{d_1}^{d_2} F(d)p(d)\delta d$$

$$p(d) = \frac{1}{d} \sum \frac{w}{\sqrt{2\pi}} \exp\left(-\frac{(\ln d - \ln D)^2}{2\sigma^2}\right)$$

其中 d_1 、 d_2 是沙粒直径的下限和上限, 根据风沙物理学中对地表沙粒的划分, 分别近似取 $60 \mu m$ 和 $1000 \mu m$ 。

5、起沙参数的确定

垂直尘粒通量的确定需要以下 4 个参数: 土壤类型、风蚀起沙尘过程中地表土壤的粒子尺度分布 $p(d)$ 、摩擦速度 u_* 和临界摩擦速度 u_{*t} 。

➤ 摩擦速度

摩擦速度 u_* 可以用中性的近地层风廓线公式计算:

$$u = \frac{u_*}{\kappa} \ln \frac{z}{z_0}$$

式中:

κ 为冯·卡门常数;

u 是高度为 z 处的风速;

z_0 是地面粗糙度。

➤ 临界摩擦速度

临界摩擦速度表征的是地表对风蚀起沙的阻碍能力, u_{*t} 的大小随粒子尺度的变化而变化, 对于一定尺度的沙粒, 其脱离地表的临界摩擦速度还决定于土壤水分含量 w 以及地表植被覆盖度 λ 等, 对此 Shao 等曾给出计算方法:

$$u_{*t} = H(w)R(\lambda)\sqrt{0.0123\left(\sigma_p g d + \frac{3 \times 10^{-4}}{\rho d}\right)}$$

其中 $H(w)$ 和 $R(\lambda)$ 分别表征土壤水分和植被覆盖对风蚀起沙的阻碍作用, σ_p 是土壤

报名号#1102

微粒密度和空气密度的比值， ρ 是空气密度，为 $1.23 \text{ kg}\cdot\text{m}^{-3}$ 。

➤ 土壤类型

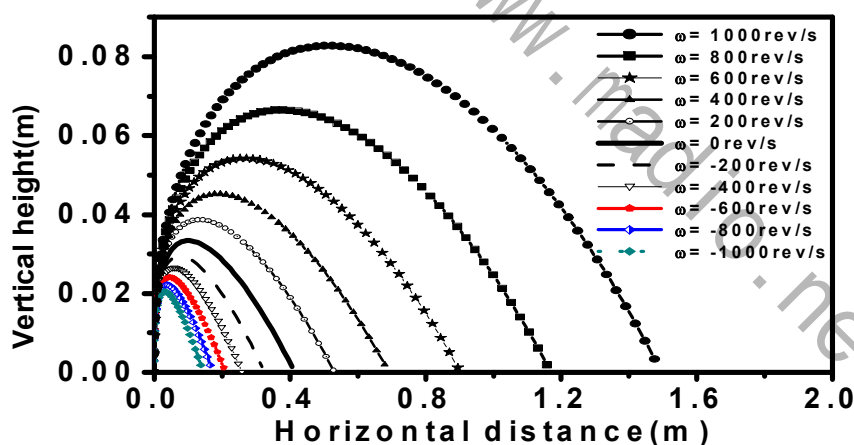
风蚀起沙尘过程中地表土壤的粒子尺度分布依据当地的土壤类型和状态决定

模型计算

在给定初始条件为 $t=0$ 、 $x=0$ 、 $y=0$ 、 $\frac{dx}{dt}=0$ 、 $\frac{dy}{dt}=1 \text{ m/s}$ 、 $u_*=0.3 \text{ m/s}$ 时，用 Rungar-kutta 方法计算得到沙粒运动轨迹如下：

计算步骤如下：

1. 给出输入参数沙粒起跳速度 $\frac{dy}{dt}$ ，空气动力粘性系数 γ ， k ， d ， m ， ρ_p ， ρ 的值，计算初始风速分布。
2. 在初始条件下，利用 Rungar-kutta 方法求解方程得到 $x_i(t)$ 、 $y_i(t)$ 。
3. 将 $y_i(t)$ 带入来流风速的式子，以得到变化的来流风速 $u_i(t)$ 。
4. 将 $u_i(t)$ 带入方程，在初始条件下，得到沙粒的跃移轨迹，并将结果记为
5. 用 $x_i(t+1)$ 、 $y_i(t+1)$ 分别代替 $x_i(t)$ 、 $y_i(t)$ ，重复进行步骤 3—5，直到 $y_i(t+1) < 1.0\text{e-}30$ 。



图九 不同沙粒自旋角速度下沙粒运动的轨迹

图九给出了不同沙粒自旋角速度下沙粒的运动轨迹。随着沙粒自旋角速度的增加，沙粒的最大水平位移、最大跃移高度都在增大，而且逆时针的自旋角速度对沙粒运动轨迹的影响比顺时针的影响更大。

4.3.5 机理模型的应用

运用沙尘通量对气象等级的划分

报名号#1102

Carson 使用极轨卫星资料计算的光学厚度值, 分析了撒哈拉沙尘的爆发. 用光学厚度 τ 和沙尘浓度 M 的关系 $M=3.75\tau$, 与沙尘通量成线性关系, 即 $F=d\cdot\tau$, 从而根据沙尘通量计算出光学厚度, 得到可见度, 判断出气象等级

运用沙尘通量估计沙尘暴的发生概率

利用沙尘暴的发生阈值建立其发生的灾害指数, 假设其为 F_1 时发生刚好必然发生, 为 F_2 时发生刚好必然不发生, 则其在期间服从正态分布, 则可以计算出发生的概率, 并同时得到其等级

4.3.6 沙尘暴的影响范围

针对沙尘暴运动路线中的不同气候条件, 可以根据建立的因风等因素起沙的动力学模型, 求出沙粒的运动轨迹, 并结合各地的地理因素沿经常性的路线漂移, 推算出沙粒全部降落的地点, 从而判断出沙尘暴的影响范围, 并且可以推断出沙尘暴的高发地域.

五、模型进一步讨论

沙尘暴防治手段:

1) 切实控制沙区人口的持续增长

人口持续增长是沙区自然资源破坏, 生态环境恶化的第一压力, 人口的增加, 加重了社会负担, 加剧了对自然资源掠夺式开发经营, 造成了环境的恶化, 形成了越垦越穷, 越穷越垦的恶性循环

2) 重视农田沙化防治, 加强农田水利建设, 保护基本农田

推广留茬免耕、覆盖种植、节水灌溉等技术, 加强基本农田建设, 调整种植结构, 培肥地力, 提高农田单产, 解决人地矛盾, 是防止农田沙化的有效途径.

3) 发展舍饲圈养, 减轻草场压力, 防止草场沙化、退化

北方荒漠化地区是我国主要的牧业区, 畜牧业在国民产值中占举足轻重的地位, 然而长期以来, 不合理的放牧方式和超载过牧, 使草场退化沙化严重, 必须加大高效集约化人工草场的建设, 大力发展舍饲圈养, 减轻草场压力, 防止草场沙化.

4) 做好退耕还林、植树种草工作, 建立高效防风固沙林体系

治沙止漠刻不容缓, 绿色屏障势在必建. 退耕还林, 植树种草, 保护和恢复林草植被, 建立有效的防风固沙林体系, 是当前防沙治沙, 生态建设的重要举措, 同时也是解决农村问题、增加农民收入的一项最直接、最有效的办法.

5) 大力发展沙产业, 加快沙区经济发展

沙漠治理既要讲“被子” (给土地盖被子), 又要讲“票子” (给农民增加票子), 只有大力发展沙产业, 提高农民经济收入, 加快农民脱贫致富, 才能从根本上解决“五滥”现象 (滥垦、滥伐、滥牧、滥采、滥灌), 实现生态经济持续发展.

6) 加快荒漠化防治法的实施, 加强执法力度

2002年1月1日起, 我国正式实施了《中华人民共和国防沙治沙法》, 从而使我国的荒漠化防治进一步走向法制化轨道. 必须加大宣传和执法力度, 坚持“谁破坏, 谁治理”、“谁治理, 谁受益”, 加速生态环境建设.

7) 加强和深入荒漠化防治技术体系的研究

控制荒漠化和沙尘暴是今后相当长时期内生态环境建设的首要任务. 而荒漠化与沙

报名号#1102

尘暴防治必须依靠先进的科学技术,因而加强荒漠与沙尘暴防治技术体系研究是生态环境建设的客观要求.今后荒漠化防治研究应继续注重于以下几方面研究:①农田与草原牧场沙化综合配套防治技术体系研究;②水资源合理调配和高效利用技术研究;③荒漠化地区植被快速重建与可持续经营技术研究;④抗逆性植物种引种选育及产业化开发技术研究;⑤荒漠化防治与区域社会经济可持续发展战略研究;⑥荒漠化及沙尘暴监测系统及评价技术研究.

六、 模型优缺点

优点

1. 问题一中采用聚类分析法对数据进行分类,较直观,易于操作.
2. 通过非线性映射分析,克服了原来的线性聚类分析的不足,通过某种非线性变换降低了样本的报错率.
3. 采用基于改进遗传的K平均算法,极大提高了聚类分析的准确率.
4. 问题二中采用长期和短期预测模型,较全面地对沙尘暴进行预测.
5. 问题三中,采用相关性分析,比较直观地找出沙源地,并推断出沙尘暴运动路线.
6. 对问题三文章运用机理模型对沙尘暴的等级和概率进行了评估,很贴近实际情况

缺点

1. 在聚类分析法中,用多元线性模型对数据进行分类,忽略了各个因素之间的非线性关系,增大误差.
2. 问题一中认为因子的重要性是固定不变的,而在估计过程中,有时会因因子集合中的一个或某几个“最佳”因子的表现不佳而导致该次预报的失败,多元线性模型难以对这种现象进行合理的判断和动态处理.
3. 虽然采用基于改进遗传的K平均算法,但仍不可避免线性聚类分析导致的误差.

七、 附录

- [1]李韧 季国良 敦煌地区大气气溶胶的辐射效应 太阳能学报 第25卷第3期 2004年6月
- [2]李霞 胡秀清 崔彩霞 李娟 南疆盆地沙尘气溶胶光学特性及我国沙尘天气强度划分标准的研究 中国沙漠 第25卷 第4期 2005年7月
- [3]王 炜 方宗义 沙尘暴天气及其研究进展综述 应用气象学报 第15卷3期 2004年6月