

DOI:10.13657/j.cnki.gxkxyxb.20170726.003

王超华,李国东,徐文霞,等.基于距离判别法的雹云识别模型[J].广西科学院学报,2018,34(2):151-155.

WANG C H, LI G D, XU W X, et al. Hail cloud recognition model based on distance discriminant analysis[J]. Journal of Guangxi Academy of Sciences, 2018, 34(2): 151-155.

基于距离判别法的雹云识别模型*

Hail Cloud Recognition Model based on Distance Discriminant Analysis

王超华^{1**}, 李国东^{1,2}, 徐文霞³, 马莉¹

WANG Chaohua¹, LI Guodong^{1,2}, XU Wenxia³, MA Li¹

(1. 新疆财经大学应用数学学院, 新疆乌鲁木齐 830012; 2. 新疆财经大学新疆社会经济统计研究中心, 新疆乌鲁木齐 830012; 3. 新疆维吾尔自治区气象局人工影响天气办公室, 新疆乌鲁木齐 830002)

(1. Applied Mathematics, Xinjiang University of Finance and Economics, Urumqi, Xinjiang, 830012, China; 2. Research Center of Xinjiang Social and Economic Statistics of Xinjiang University of Finance and Economics, Urumqi, Xinjiang, 830012, China; 3. Xinjiang Weather Modification Office, Urumqi, Xinjiang, 830002, China)

摘要:【目的】提高雹云识别准确率,降低因冰雹造成的经济损失。【方法】依据气象雷达反射率图像,利用 K-means 聚类提取云层的内外轮廓,计算其距离方差;利用雷达软件提取云高数据并计算云高的一阶统计测度;将云层内外层轮廓的距离方差与云高一阶统计测度相结合,构造雹云判别模型。【结果】利用此模型对已有样本检测,可知该模型识别率判别准确率为 88.75%,准确率较高。【结论】通过内外轮廓方差与云高一阶统计测度构造的距离判别模型有较好的判别效果。

关键词:一阶统计测度 距离判别法 内外轮廓方差

中图分类号:O212.5 文献标识码:A 文章编号:1002-7378(2018)02-0151-05

Abstract:【Objective】Improving the accuracy of hail cloud identification can reduce the economic losses caused by hail. 【Methods】According to the meteorological radar reflectivity image, the inner and outer contours of the clouds were extracted using K-means clustering, and the distance variance was calculated. The radar software was used to extract the cloud height data and calculate the first-order statistical measure of the cloud height. The distance variance of the inner and outer layers of the cloud was combined with the first-order statistical measure of the cloud height to construct a hail cloud recognition model. 【Results】Using this model to detect existing samples, we could see that the recognition accuracy of the model was 88.75%, and the accuracy was higher. 【Conclusion】The results showed that the distance recognition model constructed by the internal and external contour variance and the first-order statistical measure of cloud height had a better recognition effect.

Key words: first-order statistical measure, distance estimation, internal and external contour variance

收稿日期:2017-05-11

作者简介:王超华(1993-),女,硕士研究生,主要从事数据分析与图像处理研究,E-mail:1730230882@qq.com。

*国家自然科学基金(11461063),国家社科基金(14BTJ021)和新疆维吾尔自治区普通高等学校人文社会科学重点研究基地基金(050315B03)资助。

**通信作者。

0 引言

【研究意义】冰雹是由强对流天气系统造成的,

它的产生会造成严重的经济损失,给农业、旅游业等产业带来不可估量的损失。新疆虽地处内陆干旱、沙漠地区,但仍是冰雹多发区,深受冰雹灾害的困扰。冰雹虽然来时快、持续时间短,但会造成持续性、毁灭性的灾害。因此,研究如何有效地防雹减灾来降低经济损失具有重要意义。【前人研究进展】国外自20世纪中期已开始冰雹的预测研究。Johnson等^[1]针对强对流天气的回波识别阈值研究已取得很大进展,并且该方法在美国的实地测验过程中也取得较好的结果;Cotton等^[2]利用冰晶的浓度进行预报;Jr Auer等^[3],Smyth等^[4],Witt等^[5]结合回波反射率和云顶温度进行冰雹识别。国内虽然在这方面起步较晚,但近些年也有很大进展:夏文梅等^[6]研究V型缺口在C波段多普勒雷达中的关系,路志英等^[7]将暴雨冰雹的强度特征和纹理特征与探空数据(0℃和-20℃温度层高度)结合,利用粗糙集理论进行数据挖掘并建立识别模型。冰雹暴雨等强对流天气的回波图像具有云层图像的客观规律,利用这些规律也可得到判别雹云的数据^[8-9]。【本研究切入点】目前,利用K-means聚类和CNN边缘检测对独立云层单体图像进行处理,可得到云层的轮廓方差特征,利用该特征可区别雹云与非雹云^[10-11]。但将云层图像的内外轮廓距离差与云高数据相结合,有可能得到更准确的预测,所以本研究将两者相结合来分析雹云与非雹云。【拟解决的关键问题】利用云层图像内外轮廓距离差的方差及云层云高的一阶统计测度,构造距离判别函数对雹云与非雹云进行判别。距离判别的雹云识别原理如图1所示。

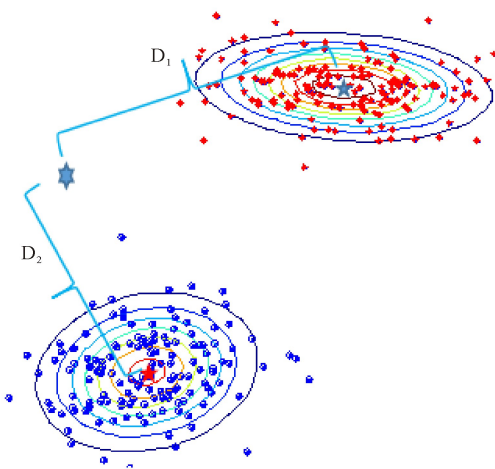


图1 距离判别的雹云识别解释图

Fig. 1 Interpretation map of hail cloud recognition based on distance identification

1 雹云内外轮廓距离差算法

在计算雹云内外轮廓距离差前,要进行K-means聚类,以各像素间的相似度距离作为不同颜色间的差距。为让聚类结果更加准确,一般需要将图像像素值从RGB模型转到Lab模型。由于a、b层的像素矩阵 N_a 、 N_b 已经包含所有像素信息,所以不需要将L层的信息表达出来。RGB模型转到Lab模型公式如下:

$$N_a = 1.4749(0.2213R - 0.3390G + 0.1177B) + 128, \quad (1)$$

$$N_b = 0.6245(0.1914R - 0.6057G - 0.8006B) + 128. \quad (2)$$

将K-means聚类后的图像采用canny算法进行边缘提取,将得到的边缘图像叠加得到一个二层轮廓的单体,计算内外轮廓的距离差,利用距离差计算内外轮廓的方差。内外层轮廓距离差算法步骤如下。

Step 1:将聚类分割单体得到的图像按公式(1)(2)的逆运算得到RGB值,按照下式将图像二值化处理。

$$\text{Gray} = 0.01 * R + 0.59 * G + 0.3 * B, \\ g(x, y) = \begin{cases} 0, & \text{Gray} > 127 \\ 255, & \text{Gray} \leq 127 \end{cases} \quad (3)$$

Step 2:将得到的二值图像运用canny算法进行边缘检测,得到黄色区域轮廓、蓝色区域轮廓。将两个轮廓进行叠加,得到两层轮廓单体图像TZE。

Step 3:在TZE图像上以黄色轮廓的几何中心O为原点,建立直角坐标系,其中 x_i, y_i 表示像素值为1的纵横坐标。

$$O = \left(\frac{\max \{x_i\} - \min \{x_j\}}{2}, \frac{\max \{y_i\} - \min \{y_j\}}{2} \right). \quad (4)$$

Step 4:从原点O射出一条射线,得到射线与两个轮廓的交点坐标 (x_{Yi}, y_{Yi}) (x_{Bi}, y_{Bi}) ,绕原点每隔 6° 旋转画一射线,共得到60个交点,通过下式计算内外层轮廓距离差序列。

$$d_i = \sqrt{(x_{Bi} - x_{Yi})^2 + (y_{Bi} - y_{Yi})^2}, i = 1, 2, \dots, 60. \quad (5)$$

Step 5:依据距离差序列计算方差,方差公式如下:

$$\sigma_i = \frac{1}{59} \sum_{j=1}^{60} (M_i(j) - \bar{M})^2, \quad (6)$$

其中 i 表示图像序号, j 表示射线的旋转次数。

根据以上步骤即可算出石河子地区雹云与非雹云图像内外轮廓差的距离方差。

2 雹云单体特征提取

2.1 云层图像云高数据的提取

在强冰雹云中,云顶的温度可达 -40°C 以下,云低温度在 0°C 以上,云中有一条 0°C 等温线。所以各云层云滴结构也不同,顶部是冰晶和雪花,中部是过冷水滴和冰晶,底部是水滴。根据探空资料统计发现^[12],一般 0°C 高度在4 km左右, -20°C 层高度在7 km左右,所以根据云层的高度可反映雹云的部分特征。

以计算轮廓方差时提取的云层图像作为云高数据提取的原图。利用雷达软件,在一幅云层图像上不重复的选取 k 个值作为云层图像的云高数据(本研究设定 $k=20$,表1)。

表1 云高数据(节取,km)

Table 1 Data of cloud height(section, km)

雹1云高 Hail 1 cloud height	雹2云高 Hail 2 cloud height	非雹1云高 Non hail 1 cloud height	非雹2云高 Non hail 2 cloud height
4	4	4	4
4	6	5	5
7	4	5	5
:	:	:	:
:	:	:	:
7	4	6	4
6	5	5	4
6	5	5	6
5	5	4	4

2.2 云高的一阶统计测度

为掌握数据的规律,有必要进一步讨论和分析得到的数据。

利用概率统计中的分布列,选取集中趋势中的算数平均值、离散趋势中的方差以及数据分布形状中的偏度和峰度这4个指标作为判断雹云与非雹云的统计测度。

均值:云高的均值。

$$E = \sum_{i=1}^k i \times p_i. \quad (7)$$

方差:用来度量云高与其均值之间的偏离程度。

$$D = \sum_{i=1}^k (i - \text{mean})^2 \times p_i. \quad (8)$$

偏度:用于度量分布偏斜方向和程度的测度,是次数分配的非对称程度。

$$\text{skewness} = \frac{1}{(\text{variance})^{\frac{3}{2}}} \cdot \sum_{i=1}^k (i - \text{mean})^3 \cdot p_i. \quad (9)$$

峰度:放映随机变量的分布形状量。

$$\text{kurtosis} = \frac{1}{(\text{variance})^2} \cdot \sum_{i=1}^k (i - \text{mean})^4 \cdot p_i. \quad (10)$$

利用表1中的数据可计算得到云高的均值、方差、偏度及峰度的一阶统计测度。

3 基于距离分析的雹云判别模型

判别分析是多元统计分析中用于判别样本所属类型的一种方法,是将已知研究对象分成若干类型、且已得到各种类型样本的观测数据情况下,同时指定一种判别规则,建立判别式,对未知类型的样本进行判别分类。

3.1 距离判别法

其基本思想是由训练样本得出每个分类的重心坐标,然后对待测样本求出它们离各个类别重心的距离远近,从而归入离得最近的类。最常用的距离是马氏距离,因为马氏距离可消除离散程度不一致所造成的影响。

本研究有两个已知总体 G_1 (雹云)与 G_2 (非雹云),一个待测样本 X 。判断测试样本 X 属于总体 G_1 (雹云)还是总体 G_2 (非雹云),主要看 X 到总体 G_1 的重心 $\bar{x}^{(1)}$ 近还是到总体 G_2 的重心 $\bar{x}^{(2)}$ 近。比较两者之间的距离 $D(X, G_1)$ 与 $D(X, G_2)$, 如果 $D(X, G_1) < D(X, G_2)$, 则 X 属于类 G_1 ; 如果 $D(X, G_1) > D(X, G_2)$, 则 X 属于类 G_2 ; 如果 $D(X, G_1) = D(X, G_2)$, 则样本 X 待判。

定义 X 到第 i 类重心 $\bar{x}^{(i)}$ 的距离为马氏距离,即

$$D^2(X, G_i) = (X - \bar{x}^{(i)})' (\sum_i)^{-1} (X - \bar{x}^{(i)}) \quad (i=1, 2), \quad (11)$$

则判别准则可以写成如下形式:

$$W(X) = D^2(X, G_2) - D^2(X, G_1) \begin{cases} X \in G_1, \text{当 } W(X) > 0 \\ X \in G_2, \text{当 } W(X) < 0. \\ \text{待判, 当 } W(X) = 0 \end{cases} \quad (12)$$

设有2个五维总体: G_1 (雹云), G_2 (非雹云), 它们的样本估计均值分别为 μ_1 和 μ_2 , 样本估计协方

差矩阵分别为 \sum_1 和 \sum_2 ($\sum_1, \sum_2 > 0$)。对任意给定的五维样本 $X = (x_1, x_2, x_3, x_4, x_5)$, 判定它来自哪个总体。按照距离最近的原则对 X 进行判别归类时, 首先计算样本 X 到 2 个总体的马氏距离 $D_i^2(X), i=1, 2$, 然后再把 X 判别到距离最小的那个样本。如果

$$D_l^2 = \min_{i=1,2} \{D_i^2(X)\}, \quad (13)$$

则 $X \in G_l$ 。

距离判别的特点是直观、简单, 它对变量的分布类型无严格要求, 尤其是并不要求总体协方差阵

表 2 石河子地区雹云与非雹云图像的特征值

Table 2 Eigenvalues of hail cloud and non hail cloud in Shihezi area

样本类型 Sample type	样本类型 Sample type	均值 Mean	方差 Variance	峰度 Kurtosis	偏度 Skewness	距离方差 Distance variance
降雹 Hail	1	5.633	1.275	-1.282	-0.279	0.048
	1	5.133	1.223	-1.021	0.539	0.067
	1	4.967	0.930	-0.755	0.564	0.048
	1	5.033	0.447	-0.589	-0.037	0.064
	1	5.750	1.039	-0.755	-0.435	0.038
	1	5.250	1.566	-1.554	0.369	0.103
	1	6.263	1.205	-0.541	-0.304	0.044
	1	5.632	0.912	-0.241	-0.848	0.072
	1	5.579	0.813	-0.336	-0.517	0.074
	1	5.895	1.544	-0.757	0.026	0.079
非雹 Non hail	2	5.467	1.568	1.383	1.267	0.021
	2	4.033	0.033	30.000	5.477	0.038
	2	4.567	0.461	-0.402	0.805	0.024
	2	4.053	0.053	19.000	4.359	0.038
	2	4.526	0.374	-0.312	0.703	0.032
	2	5.000	0.778	-1.777	0.000	0.051
	2	4.158	0.140	2.410	2.041	0.058
	2	4.150	0.134	2.776	2.123	0.085

表 3 线性判别函数

Table 3 Linear discriminant function

变量 Variable	标签 Label	降雹 Hail	非雹 Non hail
常数 Constant		-348.202	-326.753
x_1	均值 Mean	135.89	132.654
x_2	方差 Variance	-83.774	-85.444
x_3	峰度 Kurtosis	-7.359	-8.006
x_4	偏度 Skewness	56.591	60.605
x_5	距离方差 Distance variance	603.445	489.303

相等。

3.2 模型构造

本研究利用 2009 年至 2011 年 3 年的统计数据, 新疆阿克苏市、阿瓦提县、拜城的降雹图像以及非降雹图像提取云高数据共 18 副图像作为训练样本, 由第 1 节及 2.2 节计算步骤及方法可得到如表 2 所示的雹云与非雹云图像特征值。利用此特征值对雹云与非雹云进行判别分析并构造判别模型, 结果如表 3 所示(所用软件为 sas 9.2)。

别函数分别是

$$f_1 = -348.202 + 135.890x_1 - 83.774x_2 - 7.359x_3 + 56.591x_4 + 603.445x_5,$$

$$f_2 = -326.753 + 132.654x_1 - 85.444x_2 - 8.006x_3 + 60.605x_4 + 489.303x_5.$$

其判别原则是: 如果样本的 f_1 (雹云) $>$ f_2 (非雹云), 则属于 f_1 (雹云); 如果 f_1 (雹云) $<$ f_2 (非雹云), 则属于 f_2 (非雹云)。

3.3 模型检验

利用已知的分组数据对上述判别函数 f_1, f_2 进行检验, 数据选取阿克苏地区 2009 年至 2011 年 3 年的另外 4 组统计数据作为待测样本(表 4), 得到如表 5 所示的分组结果。

由表 3 中的判别函数的系数可以得到两组的判

表 4 待测样本一阶统计测度及距离方差

Table 4 First order statistical measure and distance variance of the sample to be measured

样本类型 Sample type	均值 Mean	方差 Variance	峰度 Kurtosis	偏度 Skewness	距离方差 Distance variance
待测样本 Sample to be measured	5.100	1.042	-0.905	0.442	0.139
	4.105	0.211	19.000	4.359	0.061
	4.900	1.779	3.620	1.828	0.064
	6.000	2.222	-1.436	0.000	0.079

表 5 分组结果

Table 5 Results of grouping

分组前 Before grouping	分组后 After grouping	
	非雹 Non hail	降雹 Hail
降雹 1 Hail 1	0	1
非雹 2 Non hail 2	0.997 8	0.002 2
非雹 3 Non hail 3	0.568 3	0.431 7
降雹 4 Hail 4	0.000 1	0.999 9

从模型检验结果可知,测试样本集中共有 4 个样本,第 1 个和第 4 个被判成降雹,第 2 个及第 3 个被判为无雹,测试样本均正确归类。另从分组的出错估计可知该模型将降雹错判为非雹的概率是 10%,将非雹错判为降雹的概率是 12.5%,该模型的判别准确率为 88.75%,准确率较高,表明此模型具有一定的判别效果。但模型存在一定的误判,将雹云判为非雹云;且因为样本容量小,所以判别函数在一定程度上受到限制。

4 结论

基于距离判别法的雹云分类模型能有效地对雹云与非雹云进行分类;雹云与非雹云内外轮廓距离差的反差大小可反映降雹的可能性,云层高度信息也可作为是否降雹的一个依据,将两者相结合可增大判别的精准率,减少因冰雹灾害产生的损失。由于本研究样本容量有限,所以判别样本库的建立是提高判别准确率的有效途径。

参考文献:

- [1] JOHNSON J T, MACKKEEN P L, WITT ARTHUR, et al. The storm cell identification and tracking algorithm: An enhanced WSR-88D algorithm[J]. *Weather and Forecasting*, 1998, 13(2): 263-276.
- [2] COTTON W R, TRIPOLI G J, RAUBER R M, et al. Numerical simulation of the effects of varying ice crystal nucleation rates and aggregation processes on orographic snowfall[J]. *Journal of Applied Meteorology*, 1986, 25(11): 1658-1680.
- [3] AUER A H JR. Hail recognition through the combined use of radar reflectivity and cloud-top temperatures[J]. *Monthly Weather Review*, 1994, 122(9): 2218-2221.
- [4] SMYTH T J, BLACKMAN T M, ILLINGWORTH A

- J. Observations of oblate hail using dual polarization radar and implications for hail-detection schemes[J]. *Quarterly Journal of the Royal Meteorological Society*, 1999, 125(555): 993-1016.
- [5] WITT A, EILTS M D, STUMPF G J, et al. An enhanced hail detection algorithm for the WSR-88D[J]. *Weather and Forecasting*, 1998, 13(2): 286-303.
- [6] 夏文梅, 王晓君, 孙康远, 等. V 型缺口在 C 波段多普勒雷达中的应用研究[J]. *气象*, 2016, 42(1): 67-73.
- XIA W M, WANG X J, SUN K Y, et al. Application study of 'V' notch used in C band doppler radar[J]. *Meteorological Monthly*, 2016, 42(1): 67-73.
- [7] 路志英, 刘海, 贾惠珍, 等. 基于雷达反射率图像特征的冰雹暴雨识别[J]. *物理学报*, 2014, 63(18): 189201.
- LU Z Y, LIU H, JIA H Z, et al. Recognition of hail and rainstorm based on the radar reflectivity image features[J]. *Journal of Physics*, 2014, 63(18): 189201.
- [8] 李少云, 王德良, 樊志超, 等. 郴州市冰雹天气预测预警及人工防雷方法研究[J]. *南方农业*, 2014, 8(27): 156-158.
- LI S Y, WANG D L, FAN Z C, et al. Research on hail weather predict and forecast and artificial hail suppression method in Chenzhou city[J]. *South China Agriculture*, 2014, 8(27): 156-158.
- [9] 徐艳华. 冰雹的预测及其防治[J]. *气象水文海洋仪器*, 2009, 26(3): 159-161.
- XU Y H. Hail prediction and prevention[J]. *Meteorological, Hydrological and Marine Instruments*, 2009, 26(3): 159-161.
- [10] 王雪, 李国东, 廖飞佳. 雹云图像的识别指标设计[J]. *哈尔滨理工大学学报*, 2016, 21(1): 46-50.
- WANG X, LI G D, LIAO F J. Design the recognition index of hail cloud image[J]. *Journal of Harbin University of Science and Technology*, 2016, 21(1): 46-50.
- [11] 王雪, 李国东. 快速聚类 and 3 阶 CNN 算法在雹云判别中的应用[J]. *湖南科技大学学报: 自然科学版*, 2016, 31(3): 123-128.
- WANG X, LI G D. Application of K-means clustering and CNN algorithm in hail cloud determination[J]. *Journal of Hunan University of Science & Technology: Natural Science Edition*, 2016, 31(3): 123-128.
- [12] 刘鹏, 蔡衡, 黄天福. 冰雹云识别和作业时机的选择[J]. *农业服务*, 2011, 28(2): 255-256.
- LIU P, CAI H, HUANG T F. Identification of hail clouds and timing of operations[J]. *Agricultural Services*, 2011, 28(2): 255-256.

(责任编辑:米慧芝)