

文章编号:1001-9014(2011)01-0068-06

# 联合星载毫米波雷达和激光雷达资料的 云相态识别技术

严卫, 任建奇, 陆文, 吴限  
(解放军理工大学气象学院, 江苏南京 211101)

**摘要:**依据星载偏振激光雷达云相态识别原理,借鉴星载毫米波雷达温度阈值云相态识别方法,利用支持向量机(SVM)构建了联合CloudSat和CALIPSO卫星资料的云相态识别模型并进行了实例反演验证。SVM方法训练和测试样本集采用了CloudSat的2B-GEOPROF-LIDAR云廓线数据、CALIPSO的2级1km云层数据以及欧洲中期天气预报中心的辅助温度数据,识别结果与温度阈值法得到的CloudSat云相态产品、CALIPSO云相态产品以及相关资料进行了对比验证,结果表明联合两种雷达数据的支持向量机云相态识别技术具有较高的识别精度,能够更为准确地反演云相态的垂直分布信息。

**关键词:**毫米波雷达;激光雷达;支持向量机;云相态识别  
**中图分类号:**P426.5 **文献标识码:**A

## Cloud phase discrimination technology based on spaceborne millimeter wave radar and lidar data

YAN Wei, REN Jian-Qi, LU Wen, WU Xian  
(Institute of Meteorology, PLA University of Science & Technology, Nanjing 211101, China)

**Abstract:** According to cloud phase discrimination theory of spaceborne polarization lidar, and use of the method of temperature threshold for spaceborne millimeter wave radar for reference, a cloud phase discrimination algorithm using CloudSat and CALIPSO Satellite data based on support vector machines (SVM) method was established. The training and testing data of samples used for establishing SVM model were mainly derived from CloudSat 2B-GEOPROF-LIDAR, CALIPSO level 2 1km cloud layer, and ECMWF auxiliary temperature data products. The discrimination result was compared with CloudSat cloud phase product retrieved by temperature threshold method, CALIPSO cloud phase product and other relevant data. The research results show that this cloud phase discrimination technique, e. g. SVM method with combined data of radar and lidar detection, has a superior accuracy and can provide better vertical distribution information of cloud phase.

**Key words:** millimeter wave radar; lidar; support vector machine (SVM); cloud phase discrimination

**PACS:** 92.60.Jq

### 引言

云相态是指云所处的热力学状态,即液态或固态,不同的相态类型具有不同的吸收和散射特性,其变化伴随着复杂的热力学过程,将影响各种尺度天气系统的形成与演变。识别云的相态是反演其它云微物理参数的前提,各种反演模型都是根据不同的相态类型而建立的,准确识别云的相态对于提高有效粒子半径、液态水含量等云微物理参数的反演精

度尤为重要。此外,获取云的相态信息对于飞机潜在结冰区的监测与预报也有着十分现实的意义。

传统的以被动遥感为代表的云相态反演技术已有30年的历史<sup>[1,2]</sup>,主要依据冰云和水云在热红外波段的不同散射特性而建立,虽然不断得到改善,但仍存在分辨率不高、对复杂云场景反演效果差、无法获取云顶以下相态信息等问题。2006年4月,美国NASA的CloudSat和CALIPSO卫星发射升空,CloudSat搭载了一部94GHz的毫米波云雷达,用于

收稿日期:2010-04-14,修回日期:2010-12-18

基金项目:国家自然科学基金资助(41076118)

作者简介:严卫(1961-),男,浙江宁波人,解放军理工大学气象学院教授,博士,主要研究领域为卫星大气、海洋遥感, E-mail: weiyuan2002net@yahoo.com.

Received date: 2010-04-14, revised date: 2010-12-18

获取云垂直结构的观测数据,弥补了被动遥感的不足. CALIPSO 的主要载荷为正交偏振双波段(532nm,1064nm)激光雷达,能提供云和气溶胶 30m 垂直分辨率、333m 水平分辨率的观测信息,其 532nm 波段的退偏振比可用于识别云相态.

Hu 等<sup>[3]</sup>考虑了多次散射效应的影响,对 CALIPSO 云相态识别算法进行了研究,总结了不同相态类型云粒子的雷达回波与偏振信号之间的关系.严卫等<sup>[4]</sup>根据温度与云相态的经验关系,利用 CloudSat 云廓线数据以及辅助温度数据得到了基于云廓线和温度阈值的云相态识别结果,并指出该方法对于混合云的识别不够细致,可以融合星载激光雷达资料对混合云的相态作进一步分析.利用上述两种星载雷达资料识别云相态分别取得了较好的效果.由于波长不同,两个雷达分别对特定大小的云粒子比较敏感,在单独观测时经常会出现各自的“盲区”,无法再现云场景的真实情况.文章就联合星载毫米波雷达和激光雷达资料反演云相态的方法进行了研究,并得到了相应的结论.

## 1 原理与方法

### 1.1 基本原理

地基偏振激光雷达的观测实验证明了利用雷达探测到的退偏振比能够识别云的相态<sup>[5]</sup>,其原理是非球形粒子使激光雷达接收的后向散射光产生偏振而球形粒子则不会产生这种现象.由于水云通常由球形粒子组成,而冰云由非球形粒子组成,因此利用雷达的偏振信息可以有效的区分出水云和冰云.然而该理论只有在后向散射信号以单次散射为主时才有效,对于星载偏振激光雷达而言,由于观测视场较大,回波信号的多次散射效应明显,当多次散射发生时,即便是球形粒子也会产生退偏振信号,多次散射给云相态的识别带来较大的误差.在这种情况下很难根据偏振信息完成相态的识别,因此必须考虑多次散射带来的影响.

从理论上计算多次散射对于激光雷达回波信号的贡献十分费时,业务上通常引入一个反映多次散射效应的后向散射因子  $A_s$ ,它表示单次散射占雷达接收到总后向散射的比率. Hu 等<sup>[3]</sup>利用蒙特卡洛模拟和地基多视场激光雷达实测验证的方法,得到了球形粒子发生多次散射时退偏振比  $\delta$  与后向散射因子  $A_s$  之间的关系式:

$$A_s = \frac{\gamma_{ss}}{\gamma'} 0.999 - 3.906\delta + 6.263\delta^2 - 3.554\delta^3 \quad (1)$$

式中, $\gamma'$ 为激光雷达接收到的总后向散射,它是多次散射  $\gamma_{ms}$ 与单次散射  $\gamma_{ss}$ 的总和,后来该式被进一步简化为:

$$\frac{1}{A_s} = \frac{\gamma_{ss} + \gamma_{ms}}{\gamma_{ss}} = \frac{\gamma'}{\gamma_{ss}} = \left(\frac{1+\delta}{1-\delta}\right)^2 \quad (2)$$

这个关系和地基雷达观测的结果以及 CALIPSO 的观测结果较为一致<sup>[6]</sup>,同时利用最小二乘法拟合得到了不透明冰云的退偏振比和衰减后向散射的关系式:

$$\gamma' = \frac{1}{1+88\delta} \quad (3)$$

以上两个关系式对应图 1 中的两个特征曲线.该图中横、纵坐标分别表示层积分退偏振比和层积分衰减后向散射,图中曲线归纳了水云和冰云在层积分退偏振比和衰减后向散射平面上的分布特征.为了提高云相态识别的准确性,Hu 等<sup>[7]</sup>还在云相态识别工作中引入了 1064nm 的衰减后向散射以及退偏振比等信息,并取得了较好的效果,这些研究为星载偏振激光雷达云相态识别工作提供了依据.

此外,温度也是云相态判识的重要依据,Pruppacher 等<sup>[8]</sup>利用 MODIS 云产品统计得到云相态类别的出现频率与温度之间的关系,如图 2 所示,图中横坐标表示云的温度,纵坐标表示冰、水云出现的概率,研究表明:如果云顶温度高于 0℃,那么云层对应的是水云,如果云底温度低于 -40℃,则对应的云层为冰云,如果云顶和云底的温度落于两个阈值之间则可能为水云、冰云或者是混合云.

### 1.2 支持向量机

支持向量机(Support Vector Machines, SVM)方法<sup>[9]</sup>是 Vapnik 等根据统计学习理论提出的一种机器学习方法,其基本思想是:定义最优线性超平面,并把寻找最优线性超平面的算法归结为求解一个凸规划问题.进而基于 Mercer 核展开定理,通过非线性映射  $\phi$ ,把样本空间映射到一个高维乃至无穷维的特征空间(Hilbert 空间),使在特征空间中可以应用线性学习机的方法解决样本空间中的高度非线性分类和回归等问题. SVM 的构建最终等价于求解一个线性约束的二次规划问题,进而构造一个超平面作为决策平面,使得特征空间中两类模式之间的距离最大,而且保证得到的解为全局最优解,其具体步骤如下:

首先通过非线性映射将输入的低维欧式空间映射到高维内积线性特征空间,定义内积函数  $K(x_i, x_j)$  为满足 Mercer 条件的核函数,使得低维空间的非

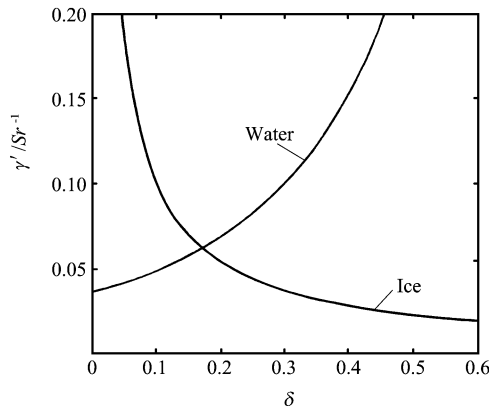


图1 不同相态下激光雷达回波信号与偏振信号的关系<sup>[6]</sup>  
Fig.1 Relationship between lidar backscatter and depolarization signal for cloud phase discrimination

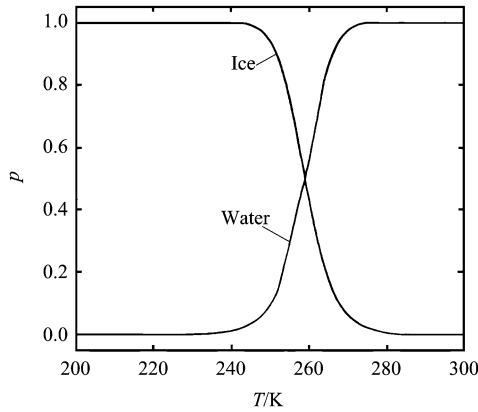


图2 利用 MODIS 数据统计得到的云相态概率分布<sup>[8]</sup>  
Fig.2 Frequency distribution of different cloud phase using MODIS data

线性分类问题转化为高维空间的线性分类问题。

$$K(x_i, x_j) = \langle \varphi(x_i), \varphi(x_j) \rangle \quad (4)$$

从而最优分类面求解转化为求解目标函数  $Q(\alpha)$  的最大值:

$$Q(\alpha) = \sum_{i=1}^l \alpha_i - \frac{1}{2} \sum_{i=1}^l \sum_{j=1}^l \alpha_i \alpha_j y_i y_j K(x_i, x_j) \quad (5)$$

由 Mercer 条件可知, 矩阵  $G(y_i y_j K(x_i, x_j))_{i,j=1}^l$  是一个对称的半正定矩阵. 因此是一个凸优化问题, 具有唯一的全局最优解, 避免了可能出现的陷于局部最优解的问题.

所求的最优分类函数为:

$$g(x) = \text{sgn} \left\{ \sum_{x_i \in SV} \alpha_i^* y_i k(x_i, x) + b^* \right\} \quad (6)$$

其中,  $x_i$  为支持向量,  $b^*$  为偏移量, 利用任意支持向量  $f$  在约束条件  $y_j f(x_j)$  下求得, 即:

$$b^* = y_j - \sum_{x_i \in SV} \alpha_i^* y_i k(x_i, x_j) \quad (7)$$

在最优分类面中采用适当的核函数  $K(x_i, x_j)$  就可以实现某一非线性变换后的线性分类, 避免求解映射函数  $\phi$  的具体形式, 将高维特征空间中的点积运算转换为低维输入空间中的核函数运算, 且并不增加计算复杂度, 本文即采用 SVM 完成云相态的识别.

## 2 云相态识别模型的建立

### 2.1 数据来源

本文所用的卫星资料包括: CloudSat 的 2B-GEOPROF-LIDAR 数据, 该数据为利用 CloudSat 毫米波雷达和 CALIPSO 激光雷达联合观测资料得到的云廓线产品; 与 CloudSat 相匹配的 ECMWF 辅助数据, 该数据是欧洲中期天气预报中心提供的温度产品, 通过它可以得到观测廓线所对应的温度信息; CALIPSO 的 2 级 1km 云层数据, 该数据提供了激光雷达对云整层积分的 532nm 和 1064nm 两个波段的衰减后向散射、退偏振比以及颜色比 (1064nm 和 532nm 两个波段衰减后向散射的比值) 信息.

### 2.2 方案设计

根据云相态的识别原理, 结合 SVM 的技术特点, 设计了云相态识别的具体步骤 (如图 3 所示):

(1) 利用 2B-GEOPROF-LIDAR 云廓线数据提供的云轮廓信息对 CALIPSO 激光雷达 2 级 1km 云层数据进行提取, 由于两种雷达观测分辨率的不同, 所以在提取的过程中考虑了空间分辨率的匹配问题, 从而确保空间分辨率的一致性.

(2) 利用 2B-GEOPROF-LIDAR 云廓线数据相对

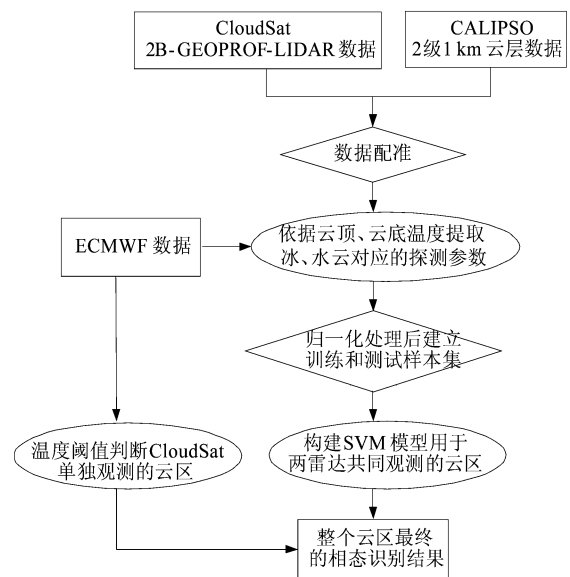


图3 云相态识别的方案设计  
Fig.3 The flow chart of cloud phase discrimination

应的 ECMWF 辅助温度数据提取得到每个云层云顶和云底所对应的温度值。以温度为判据,将不同相态类型云的探测值进行提取,最终得到水云和冰云整层积分的特征参量,分别是:532nm 衰减后向散射、1064nm 衰减后向散射、退偏振比、颜色比以及温度。

(3) 对数据进行归一化处理。将每一维数据归一化到 $[0, 1]$ 之间,将特征参数中无效的值进行整组剔除后进行归一化处理,同时建立相应数量的标识数组,完成训练样本集和测试样本集的构建。

(4) 选择核函数。核函数一般情况下选择径向基函数作为核函数<sup>[10]</sup>。

(5) 按照输入特征参数的内容分组进行训练,将特征参数和标识参数分别输入到 SVM 进行训练获得分类器参数,利用训练好的 SVM 对测试样本进行测试,反复调整相关参数直到取得最优效果,最终得到每组最优结果所对应的模型参数。

(6) 将所得识别模型用于两个雷达同时观测的云区,对于毫米波雷达单独观测的部分采用温度阈值法加以补充,最终完成整个云区的相态识别。

### 2.3 实验结果

在 SVM 模型构建过程中,输入特征向量的选取非常重要,它反映了输入参量与输出参量之间的映射关系,将直接关系到云相态识别模型的构建效果,其确定是根据与云相态类别的密切程度而定。依照前人的研究成果以及本文所得的统计结论,将 CALIPSO 激光雷达 532nm 衰减后向散射、退偏振比作为基本的输入参量,在此基础上分别增加温度、1064nm 衰减后向散射,以及颜色比 3 个参数。此外,核函数类型、形状参数以及惩罚因子的选择对于识别结果的影响也不容忽视,对于它们的选择目前还没有很好的理论方法,只能采取试验的办法,通过比较不同参数设置情况下的测试效果来加以确定。按照上述方案,选取不同的特征参数,分别设定形状参数和惩罚因子后进行分组训练和测试,得到不同输入特征参数下的最好结果,如表 1 所示:

从表中可以看到,选取简单的二维特征向量即 532nm 衰减后向散射和退偏振比,经过多次参数设置反复训练可达到 73.2% 的正确率;输入特征向量增加到三维后达到的效果要优于二维特征向量的情况,识别正确率分别为 85.9%、83.3% 以及 84.4%,从增加的参数信息和得到的识别效果可以看出,三种组合分别增加了不同的特征参数,其中增加温度参数的一组效果最为明显为 85.9%,说明温度与云相态的关联程度相对较高,也反映出增加与相

态相关的输入信息对于提高相态识别正确率的作用;当输入的特征向量增加到四维后,识别效果进一步提高,分别达到 87.6% 和 88.1%,进一步证明了增加特征向量的维数对于改善识别效果的作用,这也是 SVM 用于模式识别时将低维线性不可分问题转化为高维线性可分问题思想的体现;当五个特征参数全部输入时,经过训练优化得到了最好的效果为 89.3%,并最终确定为云相态识别模型。

表 1 各类特征向量对应的参数设置及识别效果  
Table 1 Parameters of different kinds of characteristic vectors and discrimination results

特征向量维数	特征向量	形状参数 $\gamma$	惩罚因子 C	正确率
2	532nm 后向散射,退偏比	0.1	100	73.2%
	532nm 后向散射,退偏比,温度	1	1	85.9%
3	532nm 后向散射,退偏比,1064nm 后向散射	0.5	10	83.3%
	532nm 后向散射,退偏比,颜色比	2	100	84.4%
4	532nm 后向散射,退偏比,温度,1064nm 后向散射	0.5	0.5	87.6%
	532nm 后向散射,退偏比,温度,颜色比	0.1	1	88.1%
5	532nm 后向散射,退偏比,温度,1064nm 后向散射,颜色比	0.5	10	89.3%

### 3 个例分析

所用个例为 CloudSat 和 CALIPSO 卫星 2008 年 3 月 29 日 22:44:50 至 22:52:19 的一轨数据,图 4 给出空间范围为 $(39.2 \sim 54.6^\circ\text{S}, 37.6 \sim 31.6^\circ\text{W})$ 的云场景,图 a 为本文得到的云相态识别结果。由于没有实测或地面观测资料作为“真值”检验的依据<sup>[4]</sup>,通常采用 A-train 编队卫星资料之间对比验证的办法,为了说明问题,从两个方面予以讨论:

(1) 与云分类产品比较。图 b 是基于模糊逻辑技术和 2B-GEOPROF-LIDAR 云廓线数据的云分类结果,该产品将云分为 8 种类别如图所示,通过对比发现,对于分类图中 6km 高度以上的卷云(Ci)部分,图 a 对应的是冰云或是不确定的混合云,这些云区对应激光雷达较强的后向散射和退偏振比以及毫米波雷达较弱的回波信号,如图 d、f、g 所示;图 a 中两个深对流云(Deep Convection)的顶部为冰云,对应较强的激光雷达衰减后向散射和退偏振比以及毫米波雷达信号,而中下部为水云或混合云,此处的激光雷达信号被衰减;对于水平范围较大的雨层云(Ns),其中部为水云,顶部和下部为混合云,该云区下部的混合云很有可能是以水云为主,而顶部的混合态很可能对应云顶边缘微小的冰晶颗粒;对于图

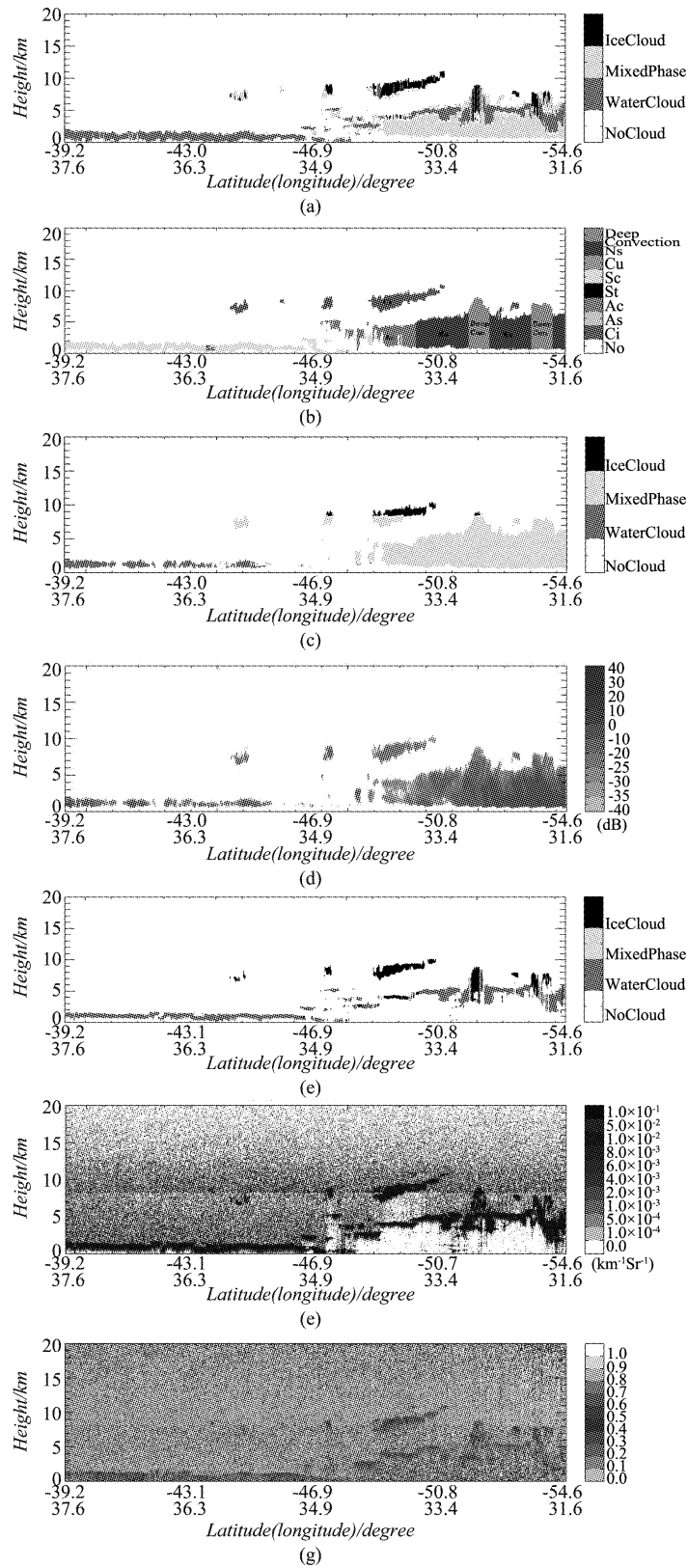


图4 云相态识别结果对比分析(a)基于SVM和2B-GEOPROF-LIDAR云廓线数据的云相态识别结果(b)基于模糊逻辑技术和2B-GEOPROF-LIDAR云廓线数据的云分类结果(c)基于温度阈值的毫米波雷达云相态识别结果(d)CloudSat雷达反射率因子(e)CALIPSO激光雷达2级1km云相态产品(f)CALIPSO532nm衰减后向散射(g)CALIPSO的退偏振比

Fig.4 Comparison of cloud phase discrimination results (a)Discrimination results using SVM method and 2B-GEOPROF-LIDAR data (b)Classification results of cloud type based on fuzzy logical method and 2B-GEOPROF-LIDAR data (c)Discrimination results using temperature threshold method and 2B-GEOPROF-LIDAR data (d)CloudSat radar reflectivity factor (e)CALIPSO lidar level 2 1km cloud phase product (f)CALIPSO 532nm attenuated backscatter (g)CALIPSO depolarization ratio

a 中的高积云 (Ac) 和层积云 (Sc) 则以水云为主, 这些识别结果与该类别云的属性比较一致, 对应较弱的毫米波雷达信号以及较强的激光雷达信号, 该部分云区中激光雷达退偏振比在云顶处具有较大值, 主要由水云的多次散射效应引起。

(2) 与毫米波雷达、激光雷达云相态产品比较。图 c 显示了利用温度阈值法得到的毫米波雷达云相态识别结果, 图中暗灰色部分为温度大于  $0^{\circ}\text{C}$  的水云区, 黑色部分为温度低于  $-40^{\circ}\text{C}$  的冰云区, 灰色部分为温度介于  $0$  至  $-40^{\circ}\text{C}$  之间的混合云区。与温度阈值法得到的毫米波雷达云相态产品相比, 本文首先采用 SVM 对毫米波雷达和激光雷达同时观测到的云区进行相态识别, 而对于毫米波雷达单独观测的部分则采用温度阈值法进行处理, 使两种方法有效的结合起来得到更为详细的相态分布信息, 与 CloudSat 毫米波雷达和 CALIPSO 激光雷达单独获得的相态产品相比具有较好的宏观一致性, (如图 c 和 e 所示), 该方法最大的特点是联合毫米波雷达和激光雷达云廓线数据得到的云场景比两种雷达单独探测的云场景范围要大, 充分发挥了两个传感器的探测能力, 毫米波雷达增大了对浓厚云层的穿透能力, 得到了浓厚云层的内部结构, 而激光雷达对纤细云顶和薄卷云十分敏感, 两个雷达相结合不仅扩大了云的轮廓边界, 还观测到了新的云团, 例如: 图 4b 中  $(45.8, 32.4)$   $8\text{km}$  处以及  $(52.8, 32.4)$  处  $6\text{km}$  处的卷云 (Ci), 这正是由于两种雷达的波长不同而导致的观测结果。

#### 4 结语

从星载毫米波雷达和激光雷达云相态识别的基本原理出发, 采用 SVM 构建了联合两种雷达资料的云相态识别模型并对识别个例进行了讨论, 实验结果表明: 在 SVM 的训练过程中以输入特征参量的类别为单位进行分组, 分别进行核函数参数设置试验, 最后得到本组实验的最优结果以达到最优识别效果, 通过测试样本验证及个例分析说明 SVM 模型具有较好的识别能力, 可用于星载雷达资料的云相态识别; 在 SVM 识别模型的构建过程中增加了与云相态相关的特征信息如:  $1064\text{nm}$  衰减后向散射、颜色

比, 有效的提高了云相态识别的正确率, 这与模式识别中将低维线性不可分问题转化到高维特征空间中完成分类的思想是一致的; 利用两个星载雷达同步观测的技术特点, 得到联合毫米波云雷达、激光雷达两个资料的云相态垂直分布, 对云边界轮廓定位更加准确, 得到了更为真实的相态识别结果, 这也是联合两种资料的优势所在。

**致谢** 本文所用卫星资料由 NASA Cloudsat 数据处理中心及兰利研究中心提供, 特此感谢。

#### REFERENCES

- [1] LIOU K N. On the radiative properties of cirrus in the window region and their influence on remote sensing of the atmosphere [J]. *Journal of Atmosphere Science*, 1974, 31(2): 522—532.
- [2] ACKERMAN S A, STRABALA K I, MENZEL W P, et al. Discriminating clear-sky from clouds with MODIS [J]. *Journal of Geophysical Research*, 1998, 103(D24): 32141—32157.
- [3] HU Y. Depolarization ratio effective lidar ratio relation: Theoretical basis for space lidar cloud phase discrimination [J]. *Geophysical Research Letters*, 2007, 34(11): 11812-1—11812-4.
- [4] YAN Wei, YANG Han-le, YE Jing. Analysis of cloud properties based on satellite-borne millimeter cloud radar measurements [J]. *Journal of Remote Sensing* (严卫, 杨汉乐, 叶晶. 星载毫米波测云雷达资料的云特征分析. *遥感学报*), 2009, 13(4): 575—579.
- [5] SCHOTLAND R M, SASSEN K, STONE R. Observations by lidar of linear depolarization ratios for hydrometers [J]. *Journal of Applied Meteorology*, 1971, 10(5): 1011—1017.
- [6] HU Y, VAUGHAN M, LIU Z, et al. The depolarization-attenuated backscatter relation: CALIPSO lidar measurements vs. theory [J]. *Optics Express*, 2007, 15(9): 5327—5332.
- [7] HU Y, WINKER D, VAUGHAN M, et al. CALIPSO/CALIOP cloud phase discrimination algorithm [J]. *Journal of Atmospheric and Oceanic Technology*, 2009, 26(11): 2293—2309.
- [8] PRUPPACHER H. A new look at homogeneous ice nucleation in supercooled water drops [J]. *Journal of Atmosphere Science*, 1995, 52(11): 1924—1933.
- [9] BURGE C J C. A tutorial on support vector machines for pattern recognition [J]. *Data Mining and Knowledge Discovery*, 1998, 2(2): 121—167.
- [10] VAPNIK V N. *The Nature of Statistical Learning Theory* [M]. New York: Springer-Verlag, 1995.