

# 利用激光云高仪和自动气象站数据识别南极吹雪

叶进<sup>1</sup>, 刘磊<sup>1</sup>, 杨琬莹<sup>1</sup>, 任洪<sup>1</sup>

(1.国防科技大学 气象海洋学院, 湖南 长沙 410073)

**摘要:** 吹雪是南极洲常见的天气现象, 在南极水汽循环和冰盖质量平衡中发挥着重要作用。尽管吹雪过程对南极洲的气候有重大影响, 但人们对这一过程知之甚少。目前使用激光云高仪来检测吹雪, 很难将雾事件与吹雪事件区分开来。在本研究中, 基于激光云高仪, 进一步结合自动气象站观测到的气象参数, 使用AdaBoost算法来实现吹雪和雾的检测, 以目测的结果作为“真值”。本文使用了ARM AWARE (Atmospheric Radiation Measurement West Antarctic Radiation Experiment) 站点2016年1月1日至2016年12月31日的数据, 其中, 四分之三的数据用作训练集, 其余数据用作测试集。所提出的算法对测试集的分类准确率约为94%。与Loeb方法相比, 所提出的算法可以检测89.12%的吹雪事件和76.10%的雾事件, 而Loeb方法只能识别64.29%的吹雪和31.87%的雾事件。

**关键词:** 吹雪; 雾; 激光云高仪; 自动气象站

## Blowing Snow Detection in Antarctica Based on Ceilometer and Surface Meteorology Systems

Ye Jin<sup>1</sup>, Liu Lei<sup>1</sup>, Yang Wanying<sup>1</sup>, Ren Hong<sup>1</sup>

(1. College of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China)

**Abstract:** Blowing snow is a common weather phenomenon in Antarctica and plays an important role in the water vapor cycle and ice sheet mass balance. Although it has a significant impact on the climate of Antarctica, people do not know much about this process. Fog events are difficult to distinguish from blowing snow events using existing detection algorithms by a ceilometer. In this study, based on ceilometer, the meteorological parameters observed by surface meteorology systems are further combined to detect blowing snow and fog using the AdaBoost algorithm. The weather phenomena recorded by human observers are ‘true’. The dataset is collected from 1 January 2016 to 31 December 2016 at the AWARE site. Among them, three-quarters of the data are used as the training set and the rest of the data as the testing set. The classification accuracy of the proposed algorithm for the testing set is about 94%. Compared with the Loeb method, the

proposed algorithm can detect 89.12% of blowing snow events and 76.10% of fog events, while the Loeb method can only identify 64.29% of blowing snow events and 31.87% of fog events.

**Keywords:** blowing snow; fog; ceilometer; surface meteorology systems

## 1 研究背景

吹雪是一种极端的天气现象<sup>[1]</sup>, 在高纬度地区, 吹雪会显著影响大气的热力学性质和边界层中粒子的运动状态。吹雪的发生会增大颗粒间的摩擦阻力, 从而影响垂直风速<sup>[2]</sup>。在南极洲, 强烈的下降风将大量的雪从内陆带到海岸<sup>[3],[4]</sup>。当风速超过阈值时, 雪颗粒会跃变并悬浮在大气中<sup>[5]</sup>。

吹雪的检测主要是通过地基和卫星遥感的方式开展进行的。目前, 许多研究人员采用最小风速阈值作为吹雪发生的条件, 这与雪表面性质有关, 不同季节的风速阈值可能不同。Palm 等人利用 CALIPSO 卫星的后向散射信号并设置风速阈值来对南极洲的吹雪事件进行检测。虽然卫星数据可以提供较大的空间覆盖范围, 但吹雪的检测仅限于晴空且吹雪层达到 30 m 以上的吹雪事件<sup>[6]</sup>。此外, 约 90% 的吹雪事件发生在多云场景<sup>[7]</sup>。因此, 开展包括激光云高仪在内的地基研究是至关重要的。Gossart 等人利用激光云高仪的后向散射廓线来对极地吹雪事件进行探测, 但无法区分吹雪和雾事件<sup>[8]</sup>。Loeb 和 Kennedy 指出, 3 m/s 的风速可以作为雾消散的条件。但夏季风速在 3 m/s 以上的平流雾在麦克默多站较为常见。由于这个原因, Loeb 和 Kennedy 认为可以使用附加的 90% 相对湿度阈值来区分雾和吹雪。因此, 他们在 Gossart 的基础上, 引入地面气象站的数据, 通过设置风速、湿度和能见度阈值, 对吹雪和雾进行粗略区分<sup>[7]</sup>。本节以地面气象站的观测资料和激光云高仪的后向散射廓线为基础, 采用 Adaboost 机器学习的方法构建了吹雪和雾的检测算法并与 Loeb 的方法进行了比较。

本文致力于利用机器学习算法, 基于激光云高仪和自动气象站的观测要素, 建立一种快速、稳健、准确的吹雪和雾的检测算法。我们将人工观测的结果作为“真值”, 并将天气现象分为三类: 吹雪、雾和其他情况(晴空或降雪)。基于 AdaBoost 算法, 进一步利用激光云高仪和自动气象站的数据构建了与天气现象之间的映射关系。本文的结构如下: 第 2 节简要介绍了本文中使用的仪器和数据。第 3 节构建了吹雪和雾的检测的算法。第 4 节评估了算法在不同风速和不同相对湿度下的准确性。最后, 第 5 节为总结和结论。

## 2 仪器和数据

美国能源部 (U.S. Department of Energy, DOE) ARM AWARE 计划于 2016 年全年, 整

个 ARM 移动设备 (AMF2) 部署在罗斯岛的麦克默多站 (77°51'S, 166°40'E)<sup>[9]</sup>。麦克默多站部署了多种仪器, 如维萨拉 CL31 激光云高仪、自动气象站以及 OTT 雨滴谱仪等。此外, 威斯康星大学麦迪逊分校的南极气象研究中心 (AMRC) 提供了自 1999 年以来麦克默多站每三小时的人工观测结果<sup>[7]</sup>。

维萨拉 CL31 激光云高仪 (CEIL) 是一种全天候的自动测量装置, 装备了 10 kHz 脉冲的砷掺杂镓 (InGaAs) 二极管激光器。它能发射 910±10 nm 的近红外脉冲, 脉冲宽度为 100 ns<sup>[10]</sup>。CEIL 能够接收云、雨、雾和气溶胶的后向散射信号, 可以每 16 s 生成一个 0~7700 m 的垂直后向散射信号剖面 (单位: 1/sr/km/10000), 垂直分辨率为 10 m<sup>[11]</sup>。在整个测量范围内, 信号强且稳定, 使得接收到的信号质量良好<sup>[12]</sup>。

自动气象站主要由传统的原位传感器组成, 能够测量 1 分钟内基本气象要素的平均值, 包括 10 米高度的矢量平均风速和矢量平均风向 (维萨拉 WS425)、2 米高度的温度和相对湿度 (维萨拉 HMP155) 和水汽压。同时, 天气现象探测器 (维萨拉 PWD22) 可以提供能见度和降水强度等相关信息<sup>[13]</sup>。

Parsivel2 是一种光学的雨滴谱测量仪器, 采用了激光光源和光电二极管探测器<sup>[14]</sup>。650 nm 的激光分布在光源和探测器之间的 27 mm×180 mm 的水平薄片上。当没有粒子时, 光电二极管的电压信号通常较大。如果粒子穿过视场, 光电二极管的输出会减少, 从而测量出粒子的阴影。Parsivel2 可以直接测量粒子的尺寸和粒子下落速度, 此外还可以得到其他衍生的参数, 例如粒径谱、降水类型和强度、雷达反射率等<sup>[7]</sup>。

### 3 算法构建

#### 3.1 特征筛选及训练集和测试集的建立

人工观测记录的是麦克默多站每 3 个小时内吹雪、雾等现象的持续时间。但是如果吹雪或雾只在 3 小时内的某段时间发生, 则无法确定什么时刻经历了吹雪或者雾事件。如图 1 所示, 图中展示了 2016 年 3 月 14 日 00:00~12:00 UTC 期间人工观测的结果。在 06:00~09:00 UTC 期间仅发生 1.5 小时的吹雪事件, 09:00~12:00 UTC 期间仅发生 1.8 小时的雾事件。因此, 研究中选择在 3 小时内都经历某一个特定的事件进行研究, 例如 09:00~12:00 UTC 期间的吹雪事件以及 00:00~06:00 UTC 期间的其他现象事件。此外, 当吹雪出现的时候也可能出现雾, 在文中将这类数据归为吹雪事件, 仅当出现雾而没有出现吹雪的事件被归类为雾事件。

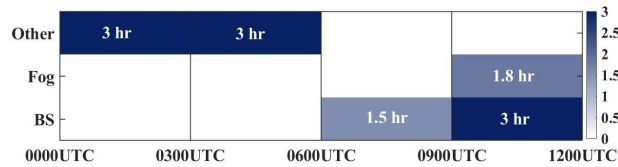


图1 3月14日00:00~12:00UTC, 吹雪(BS)、雾(Fog)和其他现象(Other)的持续时间

为提高分类算法的准确度、降低过拟合的风险、减少模型的计算时间,需要对输入量进行特征筛选<sup>[15]</sup>。图2给出了激光云高仪在5米(C1)、15米(C2)、25米(C3)、35米(C4)、45米(C5)、55米(C6)、65米(C7)和75米(C8)的后向散射信号以及自动气象站观测得到的风速(W\_speed)、相对湿度(Rh)、能见度(Vb)、温度(Temp)和降水量(Pr)的相关性。可以看出,这三种情况下激光云高仪的后向散射信号在每个高度上都表现出显著的正相关。这些相关的特征增加了模型的冗余度,因此,本节选择一个较低高度(C2)和一个较高高度(C6)的后向散射信号、风速、相对湿度、能见度、温度和降水量这7个参量作为训练特征。

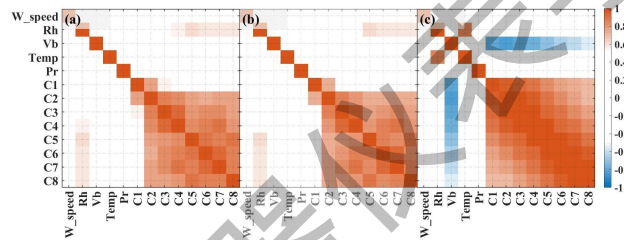


图2 各要素在(a)其他现象与吹雪, (b)其他现象与雾, (c)吹雪与雾条件下的相关性

本节随机选择每个月75%的吹雪、雾以及其他现象作为训练集,其余的为测试集。如表1所示,训练集中包含228980组其他现象、12559组吹雪事件以及5161组雾事件。为了保持训练集的数据平衡,从训练集中随机选择13000组其他现象加入到训练集中。此外,对训练集中的雾事件进行smote过采样处理<sup>[16]</sup>,使其数据量翻倍。训练集中处理后的数据在表1括号中给出。

表2.2 训练集和测试集数据在对应天气现象下的组数

天气现象	训练集	测试集
其他现象	228980 (13000)	76327
吹雪	12559	4187
雾	5161 (10322)	1721

### 3.2 吹雪和雾检测算法的构建

采用 Adaboost 算法构建的吹雪和雾检测算法的主要流程如下所示:

- 1、选取训练集中25%的其他现象、雾和吹雪事件作为验证集;

2、将训练集的数据代入到 AdaBoost 算法（选择 50 个弱分类器）中，构建所选输入参数与天气现象之间的映射关系。然后，基于得到的吹雪和雾检测算法对验证集中的数据进行分类验证；

3、为了验证算法的普适性，将步骤 1 和步骤 2 重复 100 次；

4、计算步骤 3 中得到的 100 个算法模型对应的验证集的准确性。结果表明，每个算法模型对应验证集的准确率都在 89%~92%之间，揭示了利用机器学习算法检测吹雪和雾的潜力；

5、选择准确率最高的算法模型作为最终的吹雪和雾检测算法模型，其对应的验证集准确率为 91.53%；

6、使用吹雪和雾检测算法对测试集进行分类。

### 3.3 Loeb 和 Kennedy 提出的吹雪检测算法（Loeb 算法）

首先比较激光云高仪在 10~20 m 之间的后向散射信号的平均值是否大于晴空阈值，在 AWARE 站点的晴空阈值通常取 21/sr/km/10000。进一步比较激光云高仪在 15 m 高度处的后向散射信号和 30~80 m 内的后向散射信号的平均值。如果前者较大，则判断为吹雪事件。为了进一步区分吹雪和雾事件，算法进一步引入了气象阈值。在能见度小于 10 km 的前提下，当风速小于 3 m/s 或者当相对湿度大于 90%时，算法将吹雪事件校正为雾事件。当 30~80 m 内的后向散射信号的平均值较大时，激光云高仪将其判断为其他现象（云或降雪），但是无法具体区分是云还是降雪事件<sup>[7]</sup>。

## 4 结果分析

利用本文提出的算法对测试集中的数据进行分类（图 3），可以看出，在测试集中两种算法的准确率都在 94.5%左右，但是在对吹雪和雾的识别上存在着比较大的差异。与 Loeb 算法相比，本文提出的方法大大提高了对吹雪和雾事件的识别能力，尤其对于雾的检测，准确率甚至提高了近 50%。但是对其他情况的识别中，本文算法的准确率比 Loeb 算法略低 2%左右。为了深入了解算法在不同气象条件下的性能，本节使用随机森林算法计算训练集中每个输入的“基尼重要性（Gini importance）”，得到不同场景下不同输入的权重<sup>[17]</sup>，如图 4 所示。激光云高仪的后向散射系数在区分其他情况和吹雪事件、其他情况和雾事件中起到了至关重要的作用，但是仅靠后向散射信号难以区分吹雪和雾现象。很明显，风速的引入可以很好的弥补激光云高仪的这个不足。此外，Loeb 算法将相对湿度大于 90%作为判断雾出现的条件，可能导致相对湿度小于 90%的雾事件无法识别。为此，下文分别从不同风速和不同相对湿度两方面来评估算法的准确性。

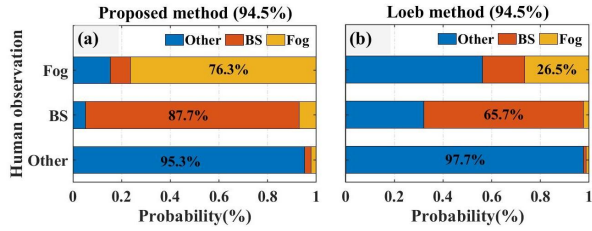


图3 测试集中本文提出的算法与 Loeb 算法的分类对比结果

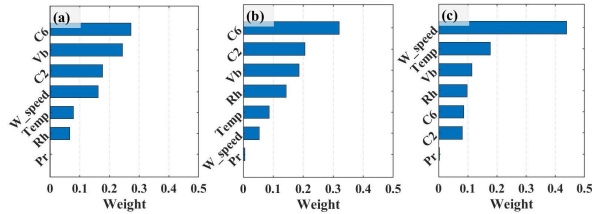


图4 在(a)其他情况 vs 吹雪, (b)其他情况 vs 雾, (c)吹雪 vs 雾条件下, 训练集各参数的权重及其重要性排序

#### 4.1 算法在不同风速条件下的准确性评估

吹雪的发生与风密切相关, 如图 5a 所示, 吹雪基本上发生在风速大于 6 m/s 的情况下, 并且在 11 m/s 处吹雪发生的频率最高。其他情况和雾事件的分布类似, 几乎都分布在小于 10 m/s 的风速范围内。

从图 5b 中可以看出, 本文提出的算法在识别高风速条件下的其他情况准确性显著下降。这可能由于风速在 12 m/s 以上的其他情况发生的概率仅不到 2%, 远远小于吹雪发生的频率。因此, 在高风速下吹雪样本数占绝对的优势, 从而导致其他情况下的误判。同时, 当风速小于 10 m/s 时, 在对其他情况的误判中雾事件占据 73%, 而吹雪事件占据 27%。这是由于在低风速下雾和其他情况分布较为一致, 从而降低了其他情况和雾的区分性。

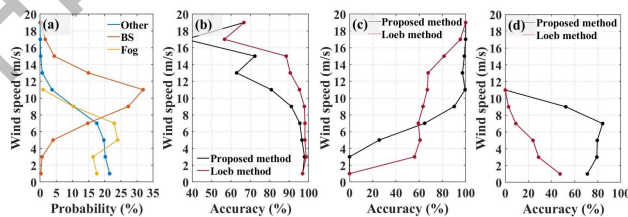


图5 (a)不同风速下其他情况、吹雪和雾的分布比例。两种算法在不同风速下对其他(b)、吹雪(c)和雾(d)分类准确性

从图 5c 中可以看出, 本文提出的算法几乎可以识别所有 10 m/s 风速以上的吹雪事件, 但对于此类吹雪事件, Loeb 算法的识别能力明显降低。图 5a 中显示 10 m/s 风速以上的吹雪事件发生的概率超过 60%。这意味着本文提出的算法对伴有强风的吹雪事件有较好的识别效果, 表明了风速在基于激光云高仪检测吹雪的过程中起到了重要的辅助作用。但是, 对于

风速较小的吹雪事件，算法的精度较低。其中一个原因是图 6 中显示的激光云高仪本身的后向散射剖面的限制。在 2016 年 12 月 6 日 00:00~00:18 UTC 之间，AWARE 站点经历了较强的吹雪事件（风速约为 4 m/s），OTT 雨滴谱仪显示此时的粒子数浓度超过了  $28 \text{ m}^{-3}\text{mm}^{-1}$ 。虽然激光云高仪在较低高度处出现较强的后向散射信号，但是在 15 m 处的后向散射信号小于 30~80 m 范围内的后向散射信号的平均值（如图 6(c)所示），这导致本文提出的算法和 Loeb 算法将吹雪事件误判为其他情况。另一个重要原因是低风速会降低吹雪和雾的区分度，在小于 10 m/s 的吹雪误判样本中，有超过 60% 的样本被误判为雾事件。因此，对于低风速的吹雪事件，本文提出算法的识别准确度小于 Loeb 算法。然而，低风速的吹雪事件在南极洲并不常见。

本文算法虽然对高风速条件下雾事件的识别能力有所降低，但是其对雾的识别准确性在任何风速条件下都远大于 Loeb 算法。

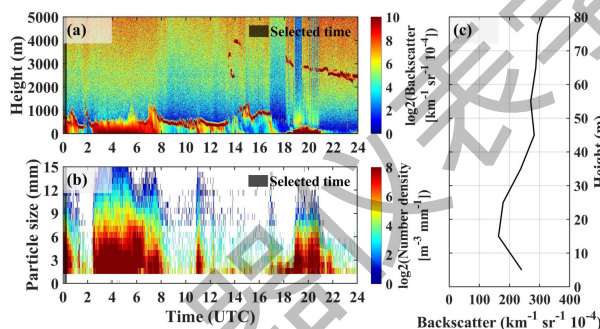


图 6 2016 年 12 月 6 日 00:12UTC 激光云高仪和 OTT 雨滴谱仪的测量数据

## 4.2 算法在不同相对湿度条件下的准确性评估

如图 7a 所示，相对湿度在 90% 以上时，只发生不到 30% 的雾事件。因此，Loeb 方法设置 90% 的相对湿度阈值会导致大量的雾事件难以检测。在其他情况下，相对湿度对识别性能影响较小，两种算法的精度相近。但在对吹雪或雾的检测上，两种算法存在较大差异。从图 7c 和 7d 可以看出，在任意相对湿度范围内，本文算法对吹雪和雾的检测能力都明显高于 Loeb 方法。与简单的阈值设置相比，使用机器学习来区分吹雪和雾具有更高的准确性。当相对湿度在 90% 以上时，Loeb 方法几乎无法识别吹雪事件。然而，本文提出的算法几乎可以识别绝大多数相对湿度在 90% 以上的吹雪事件，不受到高湿度的限制。这是由于在高湿条件下，吹雪和雾发生时的风速存在显著差异(图 8)。AWARE 站点观测到的高湿度条件下的吹雪事件的风速几乎都大于 7 m/s，而雾对应的风速大多在 7 m/s 以下。



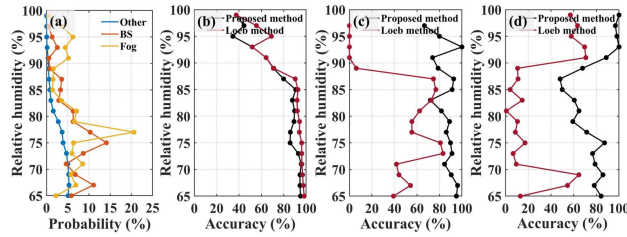


图 7 (a)不同相对湿度下其他、吹雪和雾的分布比例。两种算法在不同湿度下对其他 (b)、吹雪(c)和雾(d)

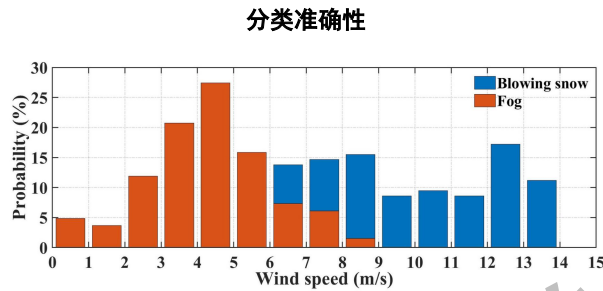


图 8 相对湿度 90%以上，不同风速下吹雪和雾事件发生概率

## 5 结论

本文利用 AdaBoost 算法，根据激光云高仪和自动气象站的数据，构造了吹雪和雾的检测算法。气象参数的引入可以解决仅用激光云高仪难以区分吹雪和雾的问题。与 Loeb 方法相比，本文提出的算法大大提高了对吹雪和雾的检测精度。所提出的算法可以检测 89.12% 的吹雪事件和 76.10% 的雾事件，而 Loeb 方法只能识别 64.29% 的吹雪和 31.87% 的雾事件。研究结果显示，所提出的算法可以识别高风速下几乎所有的吹雪事件。同时，与设置湿度阈值来区分吹雪和雾的 Loeb 方法相比，所提出的算法不受高湿度的影响。这是由于发生吹雪和雾事件时风速的显著差异造成的。此外，在任何风速或相对湿度下，本文算法对雾的检测能力都优于 Loeb 方法。这些都是本文算法的优点，也揭示了使用机器学习进行吹雪和雾检测的潜力和优势。

值得注意的是，本文存在两个挑战。首先是在高风速下检测晴空或降雪和在低风速下检测吹雪事件的能力下降。这是因为这些情况在南极洲并不常见，导致训练集中的相应数据很少。此外，大雪也会干扰所提出算法的准确性。第二个挑战是，所提出的算法仅适用于雾和吹雪的检测，但不能计算吹雪层的高度或雾的厚度。因此，未来的工作将进一步结合激光云高仪和自动气象站，准确反演吹雪层高度和雾的厚度。

## 参考文献

[1] Nishimura K, Nemoto M. Blowing snow at Mizuho station, Antarctica[J]. Philos Trans A Math



Phys Eng Sci, 2005, 363(1832): 1647-1662.

- [2] Bintanja R. Modification of the wind speed profile caused by snowdrift: Results from observations[J]. Quarterly Journal of the Royal Meteorological Society, 2001, 127(577): 2417-2434.
- [3] Mann G W, Anderson P S, Mobbs S D. Profile measurements of blowing snow at Halley, Antarctica[J]. Journal of Geophysical Research: Atmospheres, 2000, 105(D19): 24491-24508.
- [4] Pokharel B, Geerts B, Kristovich D a R. Blowing Snow as a Natural Glaciogenic Cloud Seeding Mechanism[J]. Monthly Weather Review, 2015, 143(12): 5017-5033.
- [5] Jing Y, Yau M K. A New Triple-Moment Blowing Snow Model[J]. Boundary-Layer Meteorology, 2008, 126(1): 137-155.
- [6] Palm S P, Kayetha V, Yang Y. Toward a Satellite - Derived Climatology of Blowing Snow Over Antarctica[J]. Journal of Geophysical Research: Atmospheres, 2018, 123(18): 10301-10313.
- [7] Loeb N A, Kennedy A. Blowing Snow at McMurdo Station, Antarctica During the AWARE Field Campaign: Surface and Ceilometer Observations[J]. Journal of Geophysical Research: Atmospheres, 2021, 126(7): e2020JD033935.
- [8] Gossart A, Souverijns N, Gorodetskaya I V, et al. Blowing snow detection from ground-based ceilometers: application to East Antarctica[J]. The Cryosphere, 2017, 11(6): 2755-2772.
- [9] Lubin D, Zhang D, Silber I, et al. AWARE: The Atmospheric Radiation Measurement (ARM) West Antarctic Radiation Experiment[J]. Bulletin of the American Meteorological Society, 2020, 101(7): E1069~E1091.
- [10] Young J S, Whiteman C D. Laser Ceilometer Investigation of Persistent Wintertime Cold-Air Pools in Utah's Salt Lake Valley[J]. Journal of Applied Meteorology and Climatology, 2015, 54(4): 752-765.
- [11] Morris V R. Ceilometer Instrument Handbook[R]. Washington, D.C: U.S. Department of Energy, 2016: 1-19.
- [12] O'dowd C D, Milroy C, Martucci G. Detection of Cloud-Base Height Using Jenoptik CHM15K and Vaisala CL31 Ceilometers[J]. Journal of Atmospheric and Oceanic Technology, 2010, 27(2): 305-318.

- [13] Ritsche M T, Prell J. ARM Surface Meteorology Systems (MET) Handbook[R]. Washington, D.C: U.S. Department of Energy, 2011: 1-13.
- [14] Bartholomew M J. Laser Disdrometer Instrument Handbook[R]. Washington, D.C: U.S. Department of Energy, 2020: 1-15.
- [15] Peng H, Long F, Ding C. Feature selection based on mutual information criteria of max-dependency, max-relevance, and min-redundancy[J]. IEEE Transactions on Pattern Analysis & Machine Intelligence, 2005, 27(8): 1226-1238.
- [16] Chawla N V, Bowyer K W, Hall L O, et al. SMOTE: synthetic minority over-sampling technique[J]. Journal of artificial intelligence research, 2002, (16): 321-357.
- [17] Menze B H, Kelm B M, Masuch R, et al. A comparison of random forest and its Gini importance with standard chemometric methods for the feature selection and classification of spectral data[J]. BMC Bioinformatics, 2009, (10): 213.

中国仪器仪表表学