

王珊珊,雷彦森,方鸿斌,等,2022. 基于 CatBoost 算法的长江中游冬季降水相态预报方法研究[J]. 气象,48(9):1153-1161.
Wang S S,Lei Y S,Fang H B,et al,2022. Method of winter precipitation type prediction in the middle reaches of the Yangtze River based on CatBoost algorithm[J]. Meteor Mon,48(9):1153-1161(in Chinese).

基于 CatBoost 算法的长江中游冬季降水 相态预报方法研究^{*}

王珊珊¹ 雷彦森¹ 方鸿斌¹ 孟英杰¹ 章翠红¹ 刘文婷¹ 李康丽²

1 武汉中心气象台,武汉 430074

2 湖北省荆门市气象台,荆门 448124

提 要: 利用 2000—2014 年地面观测资料和欧洲中期天气预报中心(ECMWF)ERA5 再分析资料,选取表征四类降水相态(雨、雪、雨夹雪、冻雨)的温度、湿度、微物理特征的 43 个特征量,使用精细地形高度订正,利用 CatBoost 算法开展长江中游降水相态预报方法研究。结果显示:此方法对雨、雪、冻雨有较好的分类和预报效果。使用精细地形高度预处理后的特征量,能够提高降水相态判别的准确率和空间精细度。雨、雪、冻雨的预报准确率与 ECMWF 预报产品相比分别提高了 9.9%、39.1%、11.1%,但对雨夹雪的改进不明显。

关键词: 降水相态,Catboost 算法,不平衡数据处理

中图分类号: P456

文献标志码: A

DOI: 10.7519/j.issn.1000-0526.2022.041301

Method of Winter Precipitation Type Prediction in the Middle Reaches of the Yangtze River Based on CatBoost Algorithm

WANG Shanshan¹ LEI Yansen¹ FANG Hongbin¹ MENG Yingjie¹
ZHANG Cuihong¹ LIU Wenting¹ LI Kangli²

1 Wuhan Central Meteorological Observatory, Wuhan 430074

2 Jingmen Meteorological Observatory of Hubei Province, Jingmen 448124

Abstract: Based on the surface observation and ERA5 hourly reanalysis data from 2000 to 2014, 43 factors were selected to characterize the temperature, humidity and microphysical characteristics of four types of precipitation phases (rain, snow, sleet, freezing rain). The objective forecast method of precipitation type in the middle reaches of Yangtze River is studied by using CatBoost algorithm and the fine terrain height pre-processing. The results show that this method has good classification and forecast effect for rain, snow and freezing rain. The accuracy and spatial fineness of precipitation type discrimination can be improved by using the fine terrain height pre-processing. Compared with the ECMWF prediction, the prediction accuracy of rain, snow and freezing rain has increased by 9.9%, 39.1% and 11.1% respectively, but the improvement of sleet was not obvious.

Key words: precipitation phase, CatBoost algorithm, unbalanced data processing

^{*} 国家重点研发计划(2018YFE0196000)、中国气象局气象预报业务关键技术发展专项[YBGJXM(2020)3A-08]、中国气象局预报员专项(CMAYBY2019-072)和湖北省气象局重点课题(2019Z01)共同资助

2021 年 10 月 13 日收稿; 2022 年 4 月 13 日收修定稿

第一作者:王珊珊,主要从事中短期天气和短时临近预报业务研究. E-mail:52862776@qq.com

引 言

长江流域地处中纬度地带,冬季降水相态复杂多变,同样量级的降水,由于雨、雪、雨夹雪、冻雨等不同的降水相态,其对人类的生产和活动的影响也不同,如 2008 年初发生在中国南方的低温雨雪冰冻灾害,使得城乡交通、电力、通信等遭受重创,百姓生活受到严重影响,经济损失巨大。降水相态预报一直是长江流域冬季降水预报的难点,很多学者进行了大量研究。关于雨雪的判别,国内很多研究针对温度层结和特定气压层的厚度进行了大量统计和分析,徐辉和宗志平(2014)在 2012 年 11 月 3—4 日华北地区降水相态转换过程中温度垂直结构特征分析中指出,当 850~1000 hPa 的厚度差 <1300 gpm,且 700~850 hPa 的厚度差 <1530 gpm 时,地面的降水相态类型以雪或雨夹雪为主,反之,则以雨为主。张琳娜等(2013)通过对近 10 年北京地区雨雪转换过程分析,得到了 850 hPa 温度、925 hPa 温度、1000 hPa 温度、1000~700 hPa 厚度、1000~850 hPa 厚度和地面湿温条件(2 m 温度、2 m 相对湿度的结合量)与雨、雨夹雪、雪三种降水相态的关系。漆梁波和张瑛(2012)考虑温度和厚度因子提出了中国东部降水相态的识别判据。陈双和符娇兰(2021)、陈雷等(2012)、郑丽娜等(2016)、杨成芳等(2015a;2015b)、许爱华等(2006)基于个例分析,给出了温度和厚度的识别指标。

另外还有一些学者通过开展客观算法研究实现降水相态预报,Dai(2008)通过 30 年的地面观测建立了降雪频率与地面气温和气压的关系,Bourgouin(2000)认为降水相态的变化与垂直方向上高于和低于 0℃ 的面积相关,因此利用环境融化参数研发了降水相态识别算法。陈双等(2019)通过对我国临界气温条件下降雪的时空分布特征分析,引入湿球温度,利用决策树方法对临界条件下雪和雨进行了判别分析。

近年来机器学习算法在气象领域广泛应用,朱文刚等(2020)发现利用 DNN 法对山东雨、雪、雨夹雪的预报较 ECMWF 预报有明显的提高。董全等(2013)对相同条件下线性回归法和人工神经网络的预报效果进行对比,指出人工神经网络法要优于线性回归,同时也指出南方的雨雪分界线没有北方的预报效果好。杨璐等(2021)分别基于 XGBoost、SVM、DNN 三种机器学习方法建立了降水相态的

高分辨率客观分类模型,通过对比指出 XGBoost 和 DNN 都很好地实现降水相态的分类。黄骄文等(2021)应用深度学习网络技术构建雨、雪判识模型,雨雪分界线比 ECMWF 预报更接近实况。

综上所述,国内外很多判断降水相态的判据,不管是人工判断还是客观算法判断,主要是对温度和厚度的垂直分布的分析,因此降水相态的转换更加依赖对温度的分析,对于数值预报而言,看似很小的温度误差,也会导致错误的预测(Frick and Wernli, 2012)。

除温度外,相对湿度由于能影响降水粒子融化蒸发等过程,也会影响到达地面的降水相态(Kain et al, 2000; Stewart et al, 2015)。也有研究表明造成降水相态不同的关键在于云中的成雪机制以及雪花下落过程中发生的变化(廖晓农等, 2013),也就是云微物理机制。

因此,降水相态的预报需综合考虑温度、湿度和云物理等要素。目前在实际的预报业务中发现,长江中下游降水相态精细化预报主要有两个难点:一是地形复杂,降水相态精细化格点预报模型不好建立;二是温度的预报准确率对降水相态预报影响较大。本文采用 5 km 分辨率的地形对训练数据进行预处理,将降水相态分为四类(雨、雪、雨夹雪、冻雨)进行建模,使用订正后的地面温度逐小时预报产品,选用 CatBoost(党存禄等, 2020)算法进行预报,提供更加准确和精细的格点降水相态客观预报产品。

1 资料和方法

1.1 资 料

所用资料包括 2000—2014 年冬季(11 月至次年 2 月)范围为 27°~36°N、108°~118°E 的地面观测站逐小时的 2 m 气温(T_{2m})、降水相态、高空观测数据以及地面观测站所在高度;ECMWF ERA5 再分析资料(以下简称 ERA5),包括 1000、975、950、925、900、875、850、825、800、775、750、700、650、600、550、500、400、300、200 hPa 温度(T_{1000} 、 T_{975} 、 T_{950} 、 T_{925} 、 T_{900} 、 T_{875} 、 T_{850} 、 T_{825} 、 T_{800} 、 T_{775} 、 T_{750} 、 T_{700} 、 T_{650} 、 T_{600} 、 T_{550} 、 T_{500} 、 T_{400} 、 T_{300} 、 T_{200})、比湿、云冰和云水混合比,并计算得到 700 hPa 以下大于 0℃ 的层数(以下简称暖层层数)和小于 0℃ 的层数(以下简称冷层层数)。ERA5 的空间分辨率为 0.25°×0.25°,时间分辨率为 1 h。分析时利用反距离权重

插值法将 ERA5 逐小时数据插值到上述范围内的自动气象站数据,与观测的降水相态进行时间匹配。

1.2 CatBoost 算法

降水相态的判断是个复杂的过程,要综合考虑各层的温度、湿度的配置,单一的决策树方法往往难以判断不同温湿状态下的降水相态。因此本文采用集成学习算法,它使用一系列的学习器进行学习,并使用某种规则将各个学习器的学习结果进行整合,从而获得比单个学习器更好的学习效果。集成学习模型主要分为 Bagging 和 Boosting。基于 Bagging 模型的方差较小,但偏差较大。Boosting 可以降低模型偏差,它通过迭代训练一系列的分类器,每个分类器采用的样本分布都和上一轮的学习结果有关,对基分类器的准确性要求较低。基于 Boosting 模型的算法主要有 AdaBoost, GBDT, XGBoost, LightGBM, CatBoost 等算法。CatBoost 是基于梯度提升决策树(GBDT)(刘顺祥,2018)的机器学习方法,相对其他 Boosting 算法的优势是无需调参即可获得较高的模型质量。降水相态预报中需要寻找各个要素的联系,如温度和湿度,该算法使用了组合类别特征,可以利用特征之间的联系,极大丰富特征的维度。CatBoost 通过添加先验分布项的方式减少了噪声和低频数据对于数据分布的影响,对于类别数较少的特征,先验项的添加有利于噪音数据的减少。传统的 Boosting 算法计算的是平均数,而 CatBoost 在这方面做了优化,采用 oblivious 树(党存禄等,2020)作为基学习器,oblivious 树中每个叶子节点的索引被编码为长度与树深度相等的二进制矢量,这种计算节点值的方式避免了直接计算过拟合的问题。

1.3 检验方法算法

本文将降水相态划分为四类:雨、雪、雨夹雪、冻雨,对某一类降水相态检验其准确率、漏报率和空报率。以雪为例,当观测为雪、预报也是雪时,判别为正确,记为 NA;当观测为雪、预报为其他相态时,判别错误,记为 NC;当预报为雪,观测为其他相态时,判别错误,记为 NB。任意一类降水相态的检验公式如下:

$$TS = \frac{NA}{NA + NB + NC} \times 100\%$$

$$PO = \frac{NC}{NA + NC} \times 100\%$$

$$FAR = \frac{NB}{NA + NB} \times 100\%$$

式中:TS 为某一类降水相态的准确率,PO 为某一类降水相态的漏报率,FAR 为某一类降水相态的空报率。

2 主要气象要素统计特征

选取地面观测中的 T_{2m} 和天气现象(将天气现象分为四类降水相态),ERA5 的 1000~700 hPa 温度、850 hPa 以下的云冰混合比和云水混合比,及利用 ERA5 各层温度统计的冷暖层的层数作为降水相态判别的特征量,提取各个站点海拔高度以上的数值作统计,以确保数据真实性和可用性。

为了检验 ERA5 要素的可用性,首先将实况探空 500、700、850、925 hPa 的温度与 ERA5 相同层次对比,图 1a~1d 分别为 2000—2014 年长江中游探空站雨、雪、雨夹雪、冻雨相态对应的 500、700、850、925 hPa 的实况探空温度和 ERA5 相同层次温度。可以看出雨、雪、雨夹雪实况观测的数据与 ERA5 观测的数据分布基本一致,雨的 850 hPa 和 925 hPa 温度的第 50%分位均高于 0℃,雪的各层温度均低于 0℃,但是冻雨实况观测的各层温度第 50%分位比 ERA5 低 1~2℃。对比发现,除了冻雨观测与 ERA5 略有偏差外,其他几种相态偏差很小,由于 ERA5 的垂直层次、水平分辨率以及时间分辨率更精细,因此可以使用 ERA5 的数据与国家站逐 3 h 的地面观测数据一一对应,建立降水相态数据集。

2.1 各层气温与降水相态的统计特征

图 2a 为 2000—2014 年不同降水相态 T_{2m} 的箱线图,粗柱分别代表第 25%分位和第 75%分位,可以看出,75%以上的降雨样本 $T_{2m} > 4℃$,降雪样本 $T_{2m} < 0.6℃$,冻雨样本 $T_{2m} < -0.2℃$,雨夹雪样本 $T_{2m} < 2℃$,可见,四类降水相态的 T_{2m} 有一定的差异,但是也有交叉,特别是雪和冻雨在 25%~75%样本的 T_{2m} 分布范围基本一致。图 2b 为各类相态的 925 hPa 温度箱线图,75%以上的降雨样本 $T_{925} > 0℃$,而 75%以上的雨夹雪、雪、冻雨样本的 T_{925} 在 -7~0℃,交叉范围大,无法确定阈值分类,但从分布看,冻雨的低层温度要比雨夹雪更低一些。图 2c 为 775 hPa 温度箱线图,75%以上的降雨和冻雨样本 $T_{775} > 0℃$,而 75%以上的降雪和雨夹雪样本 $T_{775} < 0℃$ 。表 1 为 1000~700 hPa 各类降水相态的第 50%分位温度值,可以看出,低层温度很难区分雪和冻雨,但是从 875 hPa 开始,雪的温度依然维

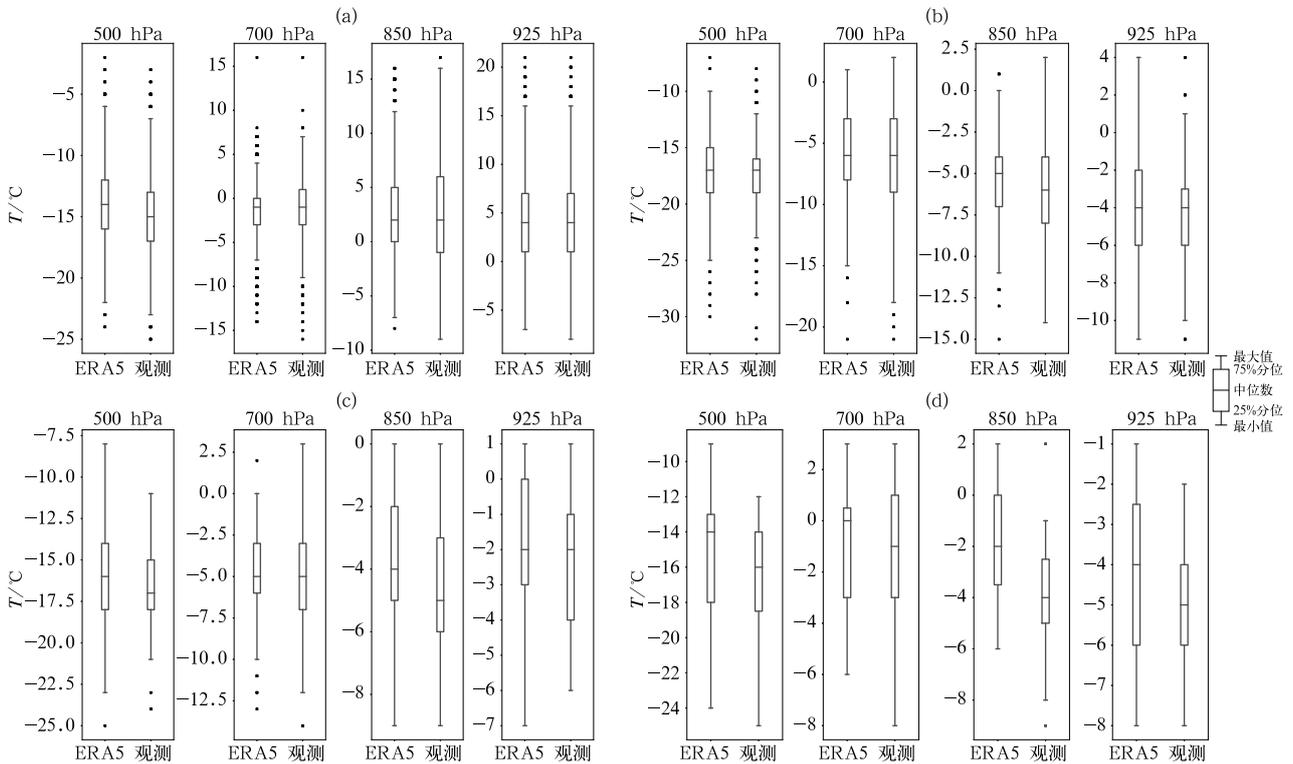


图 1 2000—2014 年不同降水相态(a)雨,(b)雪,(c)雨夹雪,(d)冻雨高空观测与 ERA5 数据的 500、700、850、925 hPa 温度箱线图

Fig. 1 Boxplots of temperature at 500, 700, 850, 925 hPa between observation and ERA5 data for (a) rain, (b) snow, (c) sleet, (d) freezing rain during 2000—2014

持在 -5°C 以下,而冻雨的温度快速升高,在 775 hPa 达到最高,为 1.7°C 。雨夹雪不管是底层温度还是中层温度一直处于雪和冻雨之间。从以上的分析可以看出,为某一层温度设定一个阈值来判断降水相态是不合理的,在判断降水相态时,需要综合各层温度的配置,因此可选用各层温度作为机器学习输入的一个特征量。

2.2 冷、暖层厚度和强度与相态的统计特征和差异分析

为了描述温度在垂直方向的分布对降水相态的影响,统计 700 hPa 以下暖层 ($\geq 0^{\circ}\text{C}$) 和冷层 ($< 0^{\circ}\text{C}$) 的层数,以反映垂直方向上大气的冷暖结构。图 3a、3b 分别为 2000—2014 年 ERA5 资料冬季(11 月至次年 2 月)不同降水相态对应的冷层、暖层层数箱线图,可以看出在雨和雪的判断上,冷层层数和暖层层数上基本无交叉,降雪冷层大于 10 层,降雨有 25% 样本冷层大于 4 层。雨夹雪和冻雨有交叉,冻雨的暖层层数要比雨夹雪多,第 50% 分位为 5 层,而雨夹雪为 2 层,同时暖层最高温度也高于雨夹雪(图 3c)。总体上看,暖层层数和冷层层数对

降水相态有较好的指示意义,由于机器学习算法寻找的是特征量与结果的数学关系,而气象预报关注的是物理意义,因此该带有物理意义的特征量可以作为机器学习算法的输入量。

2.3 云中冰水混合比与降水相态的统计特征

由于每次过程云冰和云水含量差别比较大,单独分析某层云冰或云水的含量,数值变化范围很大,并不容易总结出阈值。因此分析四种相态各层水或冰所占云中水物质的比值,以 925 hPa(图 4a)和 800 hPa(图 4b)为例,可以看到雨和雪仍然是比较容易区分的,但是雪和雨夹雪、冻雨和雨则存在很多交叉。越到低层,雪和雨夹雪的含冰量占比分布范围比较宽,为 25%~100%,而雨的含冰量基本为 0%,冻雨的含冰量占比主要在 0%~25%,而随着高度的增加,雨夹雪的含冰量占比降到 50% 以下,而冻雨的含冰量占比降到 0%。根据逐层分析(图略),可以看出,850 hPa 以下的云冰和云水占比对降水相态有一定的指示意义,特别是利用从低到高云冰和云水占比的变化,可以对四类相态做区分。

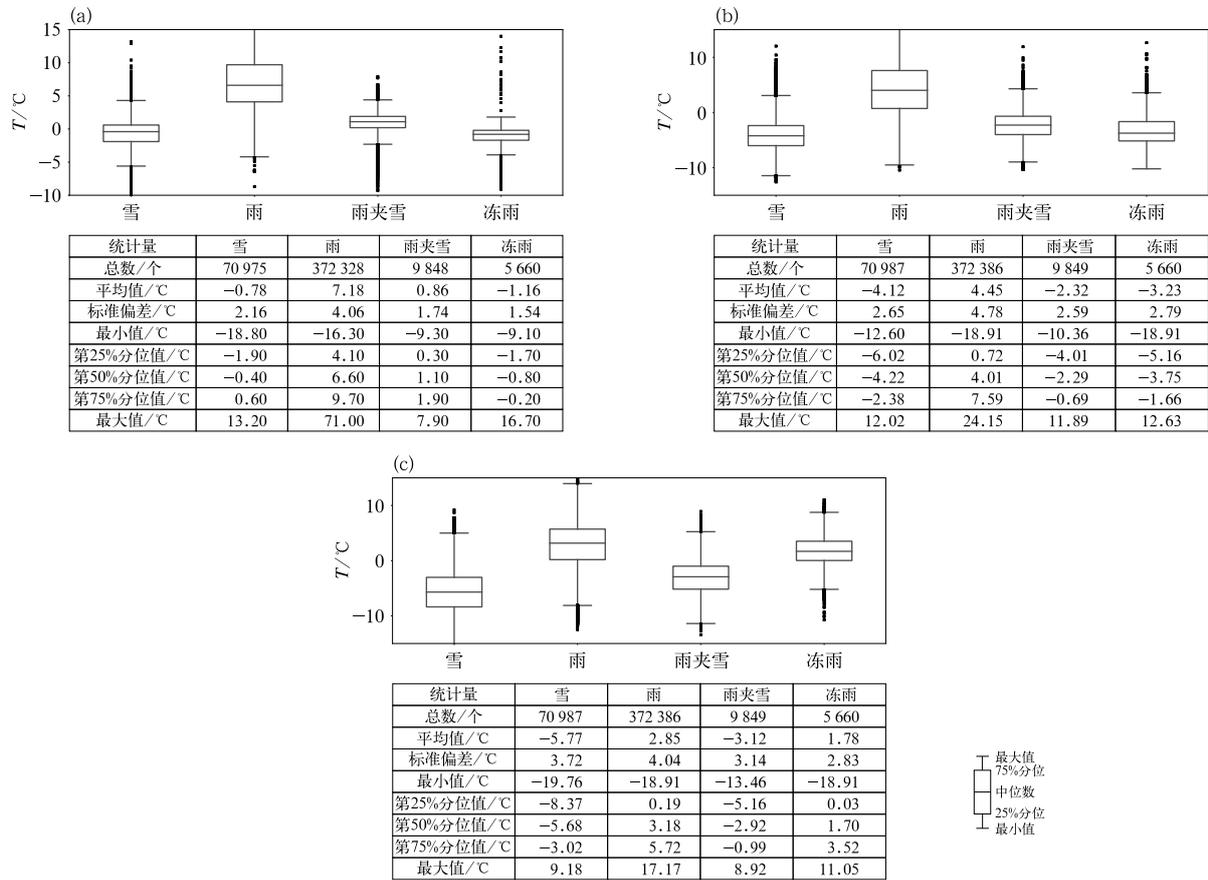


图 2 2000—2014 年不同降水相态(a) T_{2m} , (b) ERA5 T_{925} , (c) ERA5 T_{775} 的箱线图

Fig. 2 Boxplots of (a) T_{2m} , (b) ERA5 T_{925} , (c) ERA5 T_{775}

for different precipitation types during 2000—2014

表 1 1000~700 hPa 温度中位数 (单位: °C)

Table 1 The mean values of temperature from 1000 hPa to 700 hPa (unit: °C)

| 相态 | p/hPa | | | | | | | | | | | |
|-----|---------|------|------|------|------|------|------|------|------|------|------|------|
| | 700 | 750 | 775 | 800 | 825 | 850 | 875 | 900 | 925 | 950 | 975 | 1000 |
| 雪 | -5.7 | -5.5 | -5.7 | -5.8 | -5.9 | -5.7 | -5.4 | -4.9 | -4.2 | -3.3 | -2.0 | -0.5 |
| 雨夹雪 | -3.4 | -2.8 | -2.9 | -3.2 | -3.4 | -3.6 | -3.5 | -3.0 | -2.3 | -1.4 | -0.3 | 1.0 |
| 冻雨 | 0.2 | 1.6 | 1.7 | -1.3 | 0.4 | -0.9 | -2.5 | -3.6 | -3.8 | -3.1 | -1.8 | -0.4 |
| 雨 | 0.1 | 2.4 | 3.2 | 3.6 | 3.7 | 3.7 | 3.6 | 3.7 | 4.0 | 4.7 | 5.8 | 7.1 |

3 算法设计

3.1 基于地形高度的特征量预处理

首先需要提取每个站点相对地面以上的特征量,主要原因在于:(1)对于海拔高度高的站点,边界层(如 925 hPa)的特征量在地面以下,该层数据对降水相态没有指示意义,不能作为训练数据;(2)由于预报的分辨率是 $5\text{ km} \times 5\text{ km}$,地形起伏比较大,

长江中下游部分地区海拔高度可达 3 km 以上,而用于训练的国家站均在 3 km 高度以下,建立的模型用于更高海拔高度台站的降水相态预报误差会比较大。基于以上两点原因,首先把特征量在垂直方向上进行插值,然后选取每个站点地面以上固定层数的特征量,形成训练数据集。

3.2 训练模型

通过上文降水相态与温度、湿度及微物理特征量的分析,利用不同特征量组合测试,找到最优的输

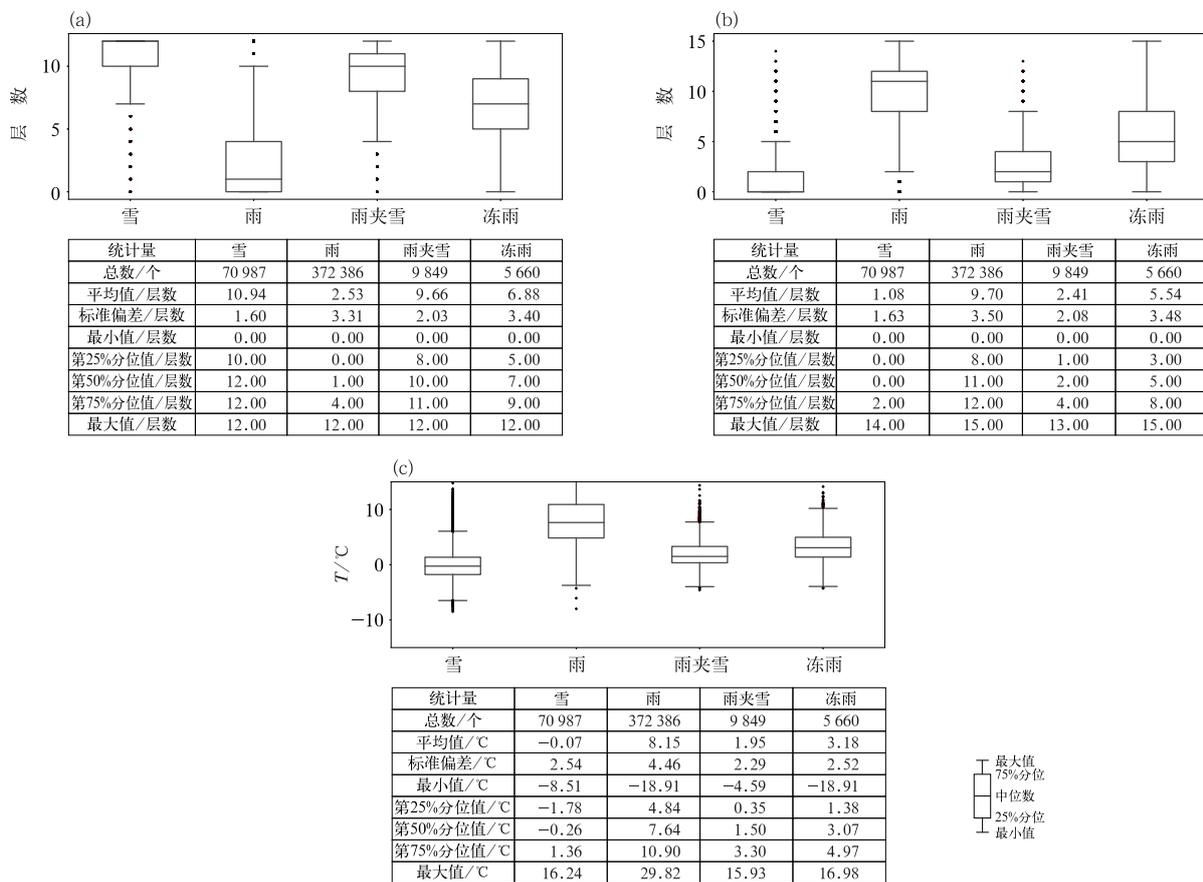


图 3 2000—2014 年不同降水相态 700 hPa 以下 (a) 冷层层数, (b) 暖层层数, (c) 暖层最高温度箱线图

Fig. 3 Boxplots of (a) cold layer thickness, (b) warm layer thickness, (c) maximum temperature of warm layer for different precipitation types during 2000—2014

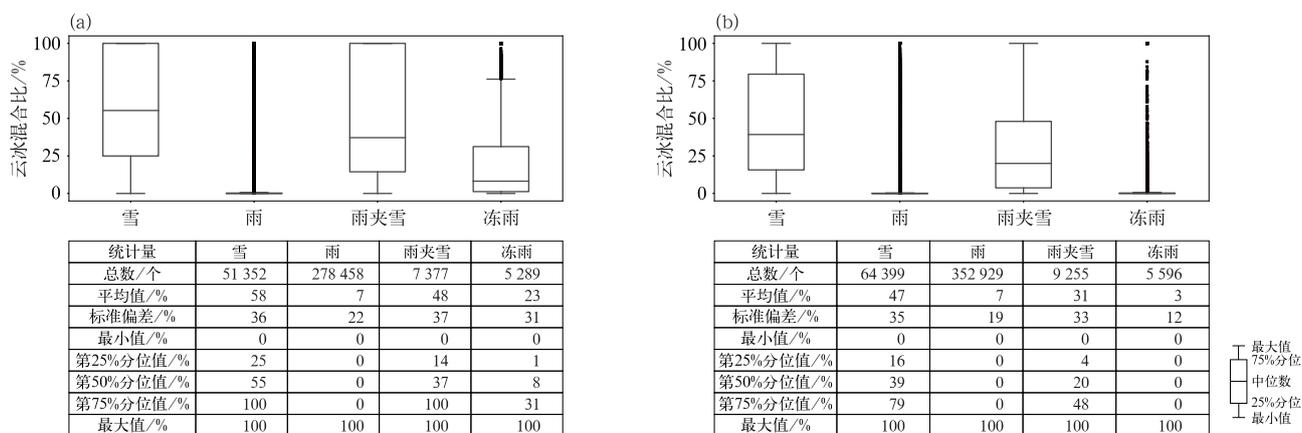


图 4 2000—2014 年不同降水相态 (a) 925 hPa, (b) 800 hPa 云冰混合比箱线图

Fig. 4 Boxplots of specific cloud ice mixing ratio of (a) 925 hPa and (b) 800 hPa for different precipitation types during 2000—2014

入特征量。选取表征四类降水相态(雨、雪、雨夹雪、冻雨)的温度、湿度、微物理特征的 43 个特征量,包括地面上 0~3.6 km 每 300 m 间隔高度上的温度、云冰混合比、云水混合比,600~400 hPa 的比湿,0~3 km 冷层和暖层的层数。其中之所以采用 600~400 hPa 的比湿是由于在预报个例总结中,高层的比湿反映了水汽伸展的高度,体现了高层冰晶的含量,对降水相态的判断也有一定的辅助作用。利用 CatBoost 算法,建立降水相态预报模型。

以 2000—2014 年的数据训练模型,经统计,2000—2014 年区域内雨、雪、雨夹雪、冻雨分别有 372 386、70 987、9 849、5 660 个站次。通过分析也发现,雨、雪、雨夹雪、冻雨样本比例为 66 : 13 : 2 : 1,存在明显的数据不平衡现象,因此在训练前首先进行数据不平衡处理。常见的处理不平衡数据多采用 SMOTE (synthetic minority oversampling technique) (郝晓红, 2019),该算法的缺点是生成的少数类样本容易与周围的多数类样本产生重叠导致难以分类,本文采用 SMOTE+ENN (郝晓红, 2019) 方法,该算法是 SMOTE 和 k 近邻算法(刘顺祥, 2018)的结合。先用 SMOTE 方法生成新的少数类样本,获得新的数据集对其中每一个样本使用 k 近邻预测,若预测结果和实际类别标签不同则剔除该样本,对于属于多数类的一个样本,如果其 k 个近邻点有超过一半都不属于多数类,则这个样本会被剔除,以达到数据类别均匀的目的。

为了能够得到比较理想的结果,需要不断尝试不同的组合参数值,Python 提供了网格搜索法(刘顺祥, 2018),可以快速进行各参数组合试错,最终得到最佳参数组合值,本文通过该方法选取最优参数组合为: iterations=1000, depth=10, learning rate=0.3, loss function='multiclass'。

4 应用效果评估

4.1 独立样本检验

为了更加真实地反映模型的性能,不使用随机抽取样本方式检验,而是采用独立样本方式检验,本文使用 2015—2019 年的样本,并将检验数据范围扩大为 25°~36°N、105°~118°E,包括湖南和贵州的大部站点,以获取更多的降水相态数据进行检验。共有降雨样本 204 151 个站次,降雪样本 28 627 个站次,雨夹雪样本 5 045 个站次,冻雨样本 2 238 个站

次。表 2 为 2015—2019 年利用上述方法训练出的模型对降水相态判别的结果,可以看出雨、雪的准确率比较高,分别为 95.1% 和 75.8%,冻雨准确率为 34.1%,最低的雨夹雪仅为 12.3%。总体来看雨一般以空报为主,雪、雨夹雪、冻雨均是以漏报为主。

表 2 2015—2019 年 CatBoost 算法识别降水相态检验结果

Table 2 The verification results of precipitation type identified by CatBoost algorithm during 2015—2019

| 相态 | 准确率/% | 空报率/% | 漏报率/% |
|-----|-------|-------|-------|
| 雨 | 95.1 | 4.4 | 0.5 |
| 雪 | 75.8 | 6.7 | 19.8 |
| 雨夹雪 | 12.3 | 75.3 | 80.3 |
| 冻雨 | 34.1 | 25.9 | 61.3 |

4.2 个例检验

将以上模型投入实时业务运行,需要使用数值模式要素作为特征量输入模型。由于不同模式预报偏差不同,会影响降水相态的判别,因此在输入模型前,首先对各个模式的要素预报效果进行检验。通过对几种业务运行的数值模式检验,ECMWF 细网格模式要素预报要优于其他几个业务模式,因此,使用 ECMWF 细网格模式要素预报来驱动降水相态预报模型。另外,地面 2 m 气温使用武汉中心气象台自主研发融合影响因子及模式偏差订正的机器学习格点气温预报产品,该产品在湖北省特别是西部山区的绝对误差低于 ECMWF 细网格的预报,因此用该产品代替模型中 ECMWF 细网格模式的地面 2 m 气温。

2020 年 1 月 6—10 日湖北出现一次大范围雨、雪、雨夹雪和冻雨天气过程。表 3 为 2020 年 1 月 5 日 20 时至 1 月 9 日 20 时起报的未来 12~36 h 逐 3 h 的 ECMWF 细网格模式与本方法的降水相态预报准确率对比,可见,与 ECMWF 细网格模式预报的降水相态对比,雨准确率提高了 9.9%,雪提高了 39.1%,冻雨提高了 11.1%,雨夹雪的准确率相当。图 5 为 CatBoost 算法和 ECMWF 2020 年 1 月 8 日 08 时起报 18 h 和 24 h 降水相态和实况观测对比图,可见,1 月 9 日 02 时(18 h)湖北中部一线已经转为纯雪,CatBoost 算法对降雪与降雨的分界线预报的比较准确,特别是孝感和襄阳,而 ECMWF 对襄

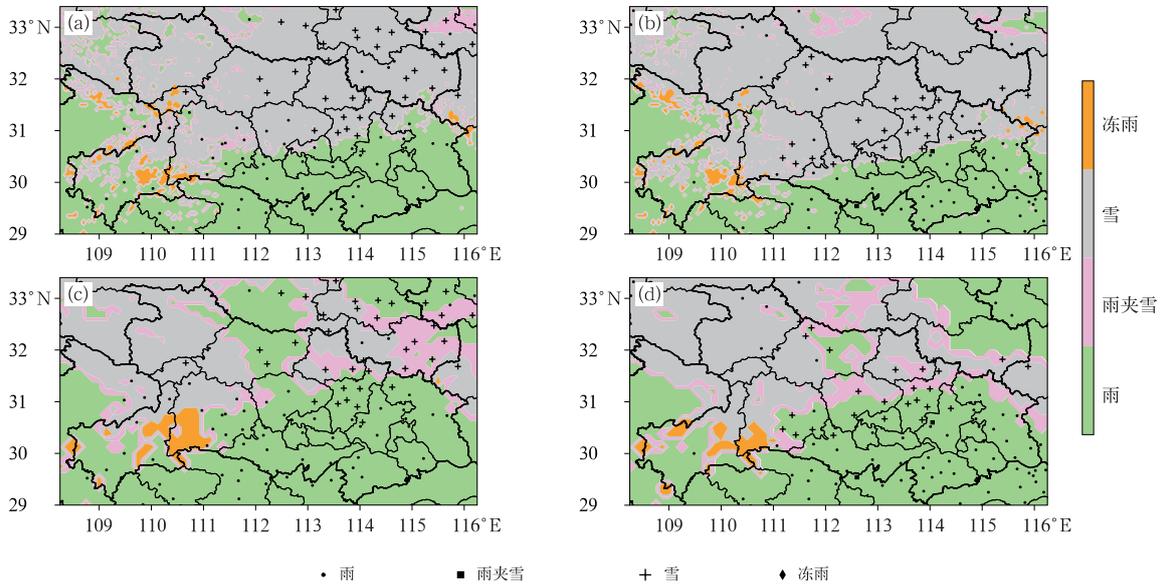


图 5 (a,b)CatBoost 算法和(c,d)ECMWF 在 2020 年 1 月 8 日 08 时起报
(a,c)18 h 和(b,d)24 h 降水相态预报(填色)与实况观测(符号)

Fig. 5 The forecast precipitation types with (a, c) 18 h and (b, d) 24 h
leadtimes starting from 08:00 BT 8 January 2020 by
(a, b) CatBoost algorithm, (c, d) ECMWF and the observation
(colored: forecast, symbol: observation)

表 3 2020 年 1 月 6—10 日 CatBoost 算法与 ECMWF
细网格模式降水相态准确率

Table 3 Precipitation types identified by
CatBoost algorithm and ECMWF
during 6—10 January 2020

| 相态 | CatBoost 方法/% | ECMWF 细网格模式/% |
|-----|---------------|---------------|
| 雨 | 87.3 | 77.4 |
| 雪 | 54.1 | 15.0 |
| 雨夹雪 | 3.1 | 3.4 |
| 冻雨 | 11.1 | 0.0 |

阳和孝感的降雪出现了漏报,1月9日08时(24 h),孝感至宜昌一线均转为纯雪,ECMWF的纯雪有明显的漏报,而 CatBoost 算法预报襄阳、孝感、宜昌和荆门的转雪时间基本与实况一致。通过对比,可以看出 CatBoost 算法对相态的预报能力要明显优于 ECMWF 的预报。

5 结 论

(1)通过对 2000—2014 年冬季不同降水相态的温度、湿度和微物理特征的统计分析,得到降水相态机器学习算法输入的 43 个特征量,利用 CatBoost 算法,建立湖北省降水相态格点预报模型,可以提供

时间和空间更精细的降水相态预报。

(2)利用 5 km 分辨率的地形高度数据,将要素订正到相对地形高度以上等高度的层次上,与使用固定层次的特征量训练的模型相比,对降水相态的判别和预报准确率均有提高,同时能反映湖北省西部高山和高原相态的区别。

(3)由于地面 2 m 温度受地形影响偏差较大,利用本地 2 m 温度客观产品代替 ECMWF 的地面 2 m 温度预报产品,能够更好地反映地形的 2 m 温度的变化,从而提高降水相态的预报准确率,特别是雨雪分界线也较 ECMWF 有更好的订正。

(4)利用建立的模型对 2020 年 1 月 6—10 日大范围雨雪过程做订正预报,并与 ECMWF 的降水相态预报产品对比,对雨、雪、冻雨的准确率有明显的提高,雨夹雪准确率仍比较低。

参考文献

- 陈雷,戴建华,韩雅萍,2012.上海地区近 10 年冷流降雪天气诊断分析[J].气象,38(2):182-187. Chen L, Dai J H, Han Y P, 2012. Diagnosis analysis of cold airflow snowfall events over Shanghai Area in recent 10 years[J]. Meteor Mon, 38(2):182-187(in Chinese).
- 陈双,谌芸,何立富,等,2019.我国中东部平原地区临界气温条件下降水相态判别分析[J].气象,45(8):1037-1051. Chen S, Chen

- Y, He L F, et al, 2019. Discrimination analysis of snow and rain occurring under critical temperature conditions in central and eastern China[J]. Meteor Mon, 45(8):1037-1051(in Chinese).
- 陈双, 符娇兰, 2021. 华北地区雪密度不同的两次降雪过程对比分析[J]. 气象, 47(1):36-48. Chen S, Fu J L, 2021. Comparative analysis of two snow events with different snow density in North China[J]. Meteor Mon, 47(1):36-48(in Chinese).
- 党存禄, 武文成, 李超锋, 等, 2020. 基于 CatBoost 算法的电力短期负荷预测研究[J]. 电气工程学报, 15(1):76-82. Dang C L, Wu W C, Li C F, et al, 2020. Short-term load forecasting based on CatBoost algorithm[J]. J Electr Eng, 15(1):76-82(in Chinese).
- 董全, 黄小玉, 宗志平, 2013. 人工神经网络法和线性回归法对降水相态的预报效果对比[J]. 气象, 39(3):324-332. Dong Q, Huang X Y, Zong Z P, 2013. Comparison of artificial neural network and linear regression methods in forecasting precipitation types[J]. Meteor Mon, 39(3):324-332(in Chinese).
- 郝晓红, 2019. 不平衡数据的研究及应用[D]. 武汉: 华中科技大学: 11-14. Hao X H, 2019. Research of imbalanced data and its application[D]. Wuhan: Huazhong University of Science and Technology: 11-14(in Chinese).
- 黄骄文, 蔡荣辉, 姚蓉, 等, 2021. 深度神经网络在降水相态判别和预报中的应用[J]. 气象, 47(3):317-326. Huang J W, Cai R H, Yao R, et al, 2021. Application of deep learning method to discrimination and forecasting of precipitation type[J]. Meteor Mon, 47(3):317-326(in Chinese).
- 廖晓农, 张琳娜, 何娜, 等, 2013. 2012年3月17日北京降水相态转变的机制讨论[J]. 气象, 39(1):28-38. Liao X N, Zhang L N, He N, et al, 2013. Analysis on the mechanism of the 17 March 2012 precipitation type variety in Beijing[J]. Meteor Mon, 39(1):28-38(in Chinese).
- 刘顺祥, 2018. 从零开始学 Python 数据分析与挖掘[M]. 北京: 清华大学出版社. Liu S X, 2018. Start Learning Python Data Analysis and Mining[M]. Beijing: Tsinghua University Press(in Chinese).
- 梁漆波, 张瑛, 2012. 中国东部地区冬季降水相态的识别判别研究[J]. 气象, 38(1):96-102. Qi L B, Zhang Y, 2012. Research on winter precipitation types' discrimination criterion in eastern China[J]. Meteor Mon, 38(1):96-102(in Chinese).
- 许爱华, 乔林, 詹丰兴, 等, 2006. 2005年3月一次寒潮天气过程的诊断分析[J]. 气象, 32(3):49-55. Xu A H, Qiao L, Zhan F X, et al, 2006. Diagnosis of a cold wave weather event in March 2005[J]. Meteor Mon, 32(3):49-55(in Chinese).
- 徐辉, 宗志平, 2014. 一次降水相态转换过程中温度垂直结构特征分析[J]. 高原气象, 33(5):1272-1280. Xu H, Zong Z P, 2014. Analysis on characteristics of thermal vertical structure evolution during the transition of precipitation type in winter[J]. Plateau Meteor, 33(5):1272-1280(in Chinese).
- 杨成芳, 周淑玲, 刘畅, 等, 2015a. 一次入海气旋局地暴雪的结构演变及成因观测分析[J]. 气象学报, 73(6):1039-1051. Yang C F, Zhou S L, Liu C, et al, 2015a. Case study of the cause and the dynamic structure for a small-scale snowstorm event associated with a cyclone[J]. Acta Meteor Sin, 73(6):1039-1051(in Chinese).
- 杨成芳, 周雪松, 李静, 等, 2015b. 基于构成要素的一次切变线暴雪天气分析[J]. 高原气象, 34(5):1402-1413. Yang C F, Zhou X S, Li J, et al, 2015b. Study on shear snowstorm by ingredients-based methodology[J]. Plateau Meteor, 34(5):1402-1413(in Chinese).
- 杨璐, 南刚强, 陈明轩, 等, 2021. 基于三种机器学习方法的降水相态高分辨率格点预报模型的构建及对比分析[J]. 气象学报, 79(6):1022-1034. Yang L, Nan G Q, Chen M X, et al, 2021. The construction and comparison of high resolution precipitation type prediction models based on three machine learning methods[J]. Acta Meteor Sin, 79(6):1022-1034(in Chinese).
- 张琳娜, 郭锐, 曾剑, 等, 2013. 北京地区冬季降水相态的识别判别研究[J]. 高原气象, 32(6):1780-1786. Zhang L N, Guo R, Zeng J, et al, 2013. Research on discrimination criterion of precipitation types in Beijing in winter[J]. Plateau Meteor, 32(6):1780-1786(in Chinese).
- 郑丽娜, 杨成芳, 刘畅, 2016. 山东冬半年回流降雪形势特征及相关降水相态[J]. 高原气象, 35(2):520-527. Zheng L N, Yang C F, Liu C, 2016. Study on circulation features of return-flow and related precipitation types during cold seasons in Shandong[J]. Plateau Meteor, 35(2):520-527(in Chinese).
- 朱文刚, 李昌义, 曲美慧, 等, 2020. 深度神经网络方法在山东降水相态判别中的应用[J]. 干旱气象, 38(4):655-664. Zhu W G, Li C Y, Qu M H, et al, 2020. Application of deep neural networks method in precipitation phase identification in Shandong Province[J]. J Arid Meteor, 38(4):655-664(in Chinese).
- Bourgouin P, 2000. A method to determine precipitation types[J]. Wea Forecasting, 15(5):583-592.
- Dai A G, 2008. Temperature and pressure dependence of the rain-snow phase transition over land and ocean[J]. Geophys Res Lett, 35(12):L12802.
- Frick C, Wernli H, 2012. A case study of high-impact wet snowfall in northwest Germany (25-27 November 2005): observations, dynamics, and forecast performance[J]. Wea Forecasting, 27(5):1217-1234.
- Kain J S, Goss S M, Baldwin M E, 2000. The melting effect as a factor in precipitation-type forecasting[J]. Wea Forecasting, 15(6):700-714.
- Stewart R E, Theriault J M, Henson W, 2015. On the characteristics of and processes producing winter precipitation types near 0°C[J]. Bull Amer Meteor Soc, 96(4):623-639.

(本文责编:王婷波)