

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*

杨璐 南刚强 陈明轩 宋林焯 刘瑞婷 程丛兰 曹伟华

YANG Lu NAN Gangqiang CHEN Mingxuan SONG Linye LIU Ruiting CHENG Conglan CAO Weihua

北京城市气象研究院, 北京, 100089

Institute of Urban Meteorology, China Meteorological Administration, Beijing 100089, China

2020-11-25 收稿, 2021-07-15 改回.

杨璐, 南刚强, 陈明轩, 宋林焯, 刘瑞婷, 程丛兰, 曹伟华. 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. 气象学报, 79(6): 1022-1034

Yang Lu, Nan Gangqiang, Chen Mingxuan, Song Linye, Liu Ruiting, Cheng Conglan, Cao Weihua. 2021. The construction and comparison of high resolution precipitation type prediction models based on three machine learning methods. *Acta Meteorologica Sinica*, 79(6):1022-1034

Abstract Refined and objective prediction of precipitation type and its transition time in winter is of great practical significance for improving the quality of forecast service. This paper establishes a high-resolution precipitation type prediction model based on temperature and weather phenomena data collected at 174 national automatic weather stations for the period 1955–2019 in Beijing-Tianjin-Hebei and the high-resolution forecast products of rapid update multi-scale analysis and forecast system-integrated subsystem (RMAPS-IN) using three machine learning methods, i.e., the XGBboost, the support vector machine (SVM) and the depth neural network (DNN) prediction models. Firstly, differences in spatial distribution between various precipitation types and corresponding climatologically mean probabilities of air temperature and wet bulb temperature at 174 national stations in Beijing-Tianjin-Hebei region are statistically analyzed. The fine integrated products provided by RMAPS-IN, i.e., 2 m air temperature, dew point temperature, relative humidity, snowline height, the ratio of frozen precipitation to total precipitation in the near surface atmosphere for different precipitation types, and the analysis fields of three-dimensional meteorological elements such as temperature and wet bulb temperature are analyzed. The observational weather phenomena, climatological characteristics of precipitation type over complex terrain and high-resolution model output products are taken as feature vectors. The classification model of precipitation type is then established based on the XGBboost, SVM and DNN, and the prediction effects of three machine learning algorithms on rain, sleet and snow are compared and evaluated. The results show that: (1) the accuracy of the three machine learning methods for rain, sleet and snow prediction can be significantly improved by adding climatological features of precipitation type over complex terrain to the feature parameters; (2) the prediction ability of the three machine learning methods for rain and snow is better than that for sleet; (3) the XGBboost and DNN have the same prediction ability, which are obviously better than SVM.

Key words Precipitation phase, Objective forecast, Numerical model, Climate statistics, Machine learning method

摘要 冬季降水相态及其转变时间的精细化客观预报对提高气象预报和服务质量具有重要的现实意义。利用京津冀地区国家级自动气象站观测资料及网格化快速更新精细集成产品, 统计分析了京津冀地区复杂地形下各类降水相态温度和湿球温度平均气候概率的分布差异及不同降水相态时网格化快速更新精细集成产品中可能影响降水相态判断的特征信息。然后将地面观测天气现象资料、复杂地形下降水相态气候特征及高分辨率模式输出产品作为特征向量, 分别基于梯度提升(XGBoost)、支

* 资助课题: 国家重点研发计划项目(2018YFF0300102)、国家自然科学基金青年基金项目(41805034、41705091)。

作者简介: 杨璐, 主要从事多源资料融合分析与应用研究。E-mail: lyang@ium.cn

持向量机(SVM)、深度神经网络(DNN)3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,并对同样条件下3种机器学习方法对雨、雨夹雪和雪3种京津冀主要降水相态的预报效果进行了对比检验,进一步提升了雨夹雪复杂降水相态的客观分类预报技巧。

关键词 降水相态, 客观预报, 数值模式, 气候统计, 机器学习方法

中图法分类号 P456 P458.1

1 引言

降水是地球水循环和能量循环的关键环节之一(Zhang, 2005),冬季降水相态主要包括雨、雪、雨夹雪、冰粒、冻雨等(蔡雪薇等, 2013),不同相态的降水对陆地表面物质和能量循环的影响巨大(Wu, et al, 2009),相同的降水量不同的相态所产生的影响存在显著差异(孙燕等, 2014; 王春乙等, 2005)。

随着近年来北京城市化进程的不断加快,大城市面对极端气象灾害的脆弱性也日益凸显,冬季降雪、冻雨甚至雨夹雪天气会给城市交通、电力通信及人民生活、生活带来巨大危害。即使是很弱的降雪过程,如果降水相态预报失败或漏报,也会给城市运行带来严重的不利影响(孙继松等, 2003),另外,2022年第24届冬季奥林匹克运动会将在北京和张家口举办,冬奥会对气象条件要求严苛,尤其是精细化的温度、湿度及降水相态预报服务等,将是冬奥会气象服务保障的重点(Chen, et al, 2018)。因此,做好冬季降水相态及其转变时间的精细化预报,研究京津冀地区冬季降水相态预报方法,并开发适用于京津冀地区的降水相态高精度格点分类客观预报算法,对提升现代气象预报服务质量、保障国家重大活动、防灾减灾等具有重要现实意义。

降水相态的形成机制包含大气垂直热力学和微物理过程两个方面。一方面,到达地表的降水类型很大程度上取决于大气的垂直温度,而大气的垂直温度通常由高空的暖层和下面的再冻结层组成。大气垂直温度一个微小的变化都会导致地表附近降水类型的变化(Sankaré, et al, 2016)。而区分降水类型转化区的一个重要的热力信息即融化层($T_w > 0^\circ\text{C}$)相对于地面高度的位置(Stewart, et al, 2015)。另一方面,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,粒子在降落的过程中涉及到扩散生长、增生、聚集、融化、再冻结等许多复杂的物理过程(Thériault, et al, 2010)。

国际上利用数值模式开展了许多针对降水相态精细化预报的后处理算法,为降水相态的分类预报提供了丰富的科学依据,如一些学者(Bourgouin, 2000; Shafer, 2010, 2015)通过提取模式输出产品温度和湿度廓线中的特征量,利用算法、决策树或统计学的方法建立特征量与降水相态的关系,实现雨、雪、冻雨、冰粒的区分。还有一些学者(Ramer, 1993; Schuur, et al, 2012; Elmore, et al, 2015)通过计算或推导下落水成物中的液态水含量,实现雨、雪、冻雨、冰粒的区分。另外,一些学者(Thompson, et al, 2004, 2008; Ikeda, et al, 2013)通过改进模式中的微物理方案,利用混合相云微物理方案预测的水成物信息,如云水、雨水、冰晶、雪、霰和水汽含量等计算出的混合比预报物理量作为诊断量区分降水相态。

中国学者从不同方面开展了大量研究工作,主要包括降雪气候分布及其变化趋势的相关研究(段长春等, 2011; 刘玉莲等, 2012),不同地区和台站雨雪转换的经验温度阈值的研究(李江波等, 2009; 尤凤春等, 2013; 董全等, 2013; 杨成芳等, 2015)。如漆梁波等(2012)通过对中国东部冬季降水相态的研究,认为温度平均廓线对雨和雪的区分较好,雪和雨夹雪在低层的大气冷暖状态较相似。张琳娜等(2013)在建立北京地区冬季降水相态识别指标时,除了将不同高度层上的温度和位势厚度作为判据,还加入了地面2 m气温和相对湿度,为模式制作客观预报产品提供了参考依据。崔锦等(2014)利用数值模式的云微物理量输出产品,对东北地区的冬季降水相态进行了预报试验。董全等(2013)选取降水发生时和发生前6 h的地面2 m气温、露点温度作为预报因子,对相同条件下线性回归法和人工神经网络法对降水相态的预报效果进行了对比检验。陈双等(2019)基于2001—2003年地面观测和探空资料,对地面气温位于0—2℃中国降雪的时、空分布及其与降雨的垂直热力特征进行了研究,引入了决策树判别方法对上述条件下雪和雨进行了判别分析。这些研究结果表明:温度和湿球温

度的垂直廓线、地面 2 m 气温、露点温度、相对湿度, 数值模式的云微物理量输出产品等对于地面降水相态的客观诊断具有很好的指示意义。

对于京津冀地区降水相态的高分辨率格点客观分类预报模型, Yang 等 (2021) (以下简称 Y20) 基于中国气象局北京快速更新循环数值预报系统 (CMA-BJ) (原华北区域快速更新循环数值预报系统, RMAPS-ST) 预报的雪、雨、冰、霰降水混合比及粒子降落末速度结合快速更新多尺度分析和预报系统集成子系统 (RMAPS-IN) 分析和预报的高分辨率湿球温度廓线、雪线高度等联合地面观测订正, 开展了降水相态诊断变量最优阈值的研究分析, 构建了降水相态综合诊断算法, 可提供覆盖京津冀全域、空间分辨率 1 km、时间分辨率 10 min 的雪线高度及降水相态的高分辨率诊断分析及 0—12 h 预报。客观检验结果表明该算法能够较好地诊断冬季降水类型, 尤其是雨和雪, 准确率均超过 90%, 雨夹雪的准确率相对较低 (41%)。

实际的预报和研究中, 预测降水类型的难度主要在于雨雪过渡区。而雨雪相态转换阶段, 850 和 925 hPa 温度对于雨、雪、雨夹雪的识别没有明显指示性 (杨成芳等, 2015), 地面气温往往在 0°C 附近, 有时甚至还会出现地面气温在 0°C 以上却发生降雪的情况 (董全等, 2013)。另外, 京津冀地区地形复杂, 地形的特征差异对地面的辐射加热、地形抬升等的影响会进一步造成不同海拔高度下雨雪相态转变规律及阈值的细微差异 (Rajagopal, et al, 2016)。模式热力学和微物理输出信息的不确定以及由网格插值引起的诊断类型和观测类型的空间偏移是 Y20 诊断算法中降水类型误诊的重要原因之一, 另外, 诊断算法中京津冀范围内选取和使用同一阈值也会给降水类型的诊断带来一定的误差。

为进一步提高冬季降水相态客观预报的准确率, 将在 Y20 基础上, 充分利用 RMAPS-IN 系统生成的高分辨率温度、相对湿度、湿球温度及雨、雪混合比等网格化快速更新精细集成产品 (Haiden, et al, 2011), 结合自动气象站观测资料, 进一步通过机器学习方法构建、测试和优化京津冀地区降水相态客观预报算法。

文中首先利用京津冀地区国家级气象站观测资料及网格化快速更新精细集成产品, 统计分析了京津冀地区复杂地形下各类降水相态温度和湿球

温度平均气候概率的分布差异、不同降水相态时 RMAPS-IN 提供的网格化快速更新精细集成产品中 7 个可能影响降水相态判断的特征信息, 包括地面 2 m 气温、露点温度、相对湿度、雪线高度、近地面大气层中冻结部分降水混合比在可凝结成降水的水汽混合比中的比例以及气温和湿球温度三维气象要素客观分析场等。然后将地面观测天气现象资料、复杂地形下降水相态气候特征及高分辨率模式输出产品作为特征向量, 分别基于梯度提升 (XGBoost)、支持向量机 (SVM)、深度神经网络 (DNN) 3 种机器学习方法建立降水相态的客观分类模型, 并对同样条件下 3 种机器学习方法对雨、雨夹雪和雪 3 种京津冀主要降水相态的预报效果进行对比检验。

2 数据及其统计特征

2.1 数据

观测数据包括 2 部分: (1) 2016 年 1 月—2019 年 4 月冬季 15 个降水日京津冀地区 174 个国家自动气象站天气现象资料, 此数据分别用于对 RMAPS-IN 网格化快速更新精细集成产品降水相态分类模型样本的筛选及客观检验。这部分的天气现象观测资料采样处理为 10 min 时间间隔, 即原始观测的天气现象数据集是根据人工判断的每一个降水日的雪、雨、雨夹雪具体起止时间进行记录, 如 A 站记录的降雪时间为 08 时 (北京时, 下同)—09 时 10 分, B 站记录的降雨时间是 08 时—09 时 10 分, 那么在进行采样处理时, 从 08 时到 09 时 10 分的每 10 min, 如 08 时—08 时 10 分、08 时 10 分—08 时 20 分, …… , 09 时—09 时 10 分, A 站的降水类型记录为雪, B 站记录为雨。(2) 1955 年 11 月至 2019 年 4 月京津冀地区 174 个国家观测站的日天气现象及其对应的地面气温、气压、相对湿度观测数据 (其中 1955—2003 年为 02、08、14、20 时 4 个时次, 2004—2019 年为逐时)。虽然选取的站点数据已经经过质量控制, 但由于天气现象为人工判断和记录, 且天气现象为日数据, 所以文中根据以往关于降水相态的研究 (许美玲等, 2015), 对筛选出的数据做了进一步的质量控制。京津冀地区国家级站点分布如图 1 所示。

文中用到的模式数据为 2016 年 1 月—2019 年 4 月冬季 15 个降水日 RMAPS-IN 提供的网格化快

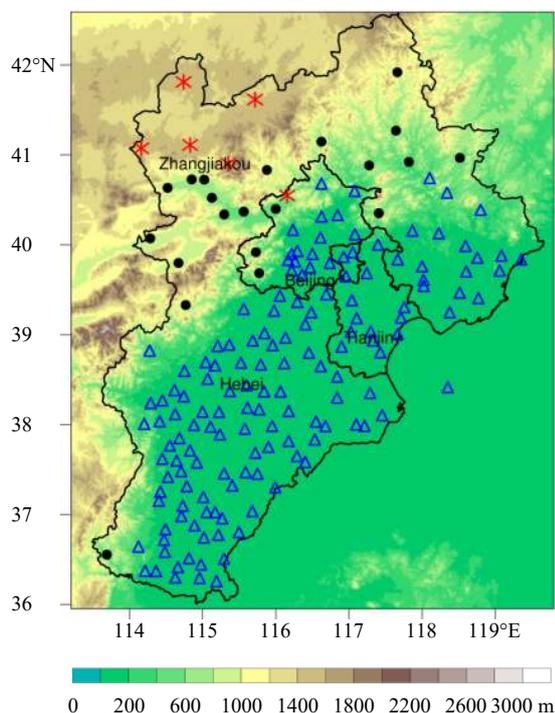


图1 京津冀地区174个国家级气象站分布(红色雪花表示海拔高度在1000—1500 m的站点;黑色圆点表示海拔高度在400—1000 m的站点,蓝色三角表示海拔高度在400 m以下的站点,色阶为海拔高度)

Fig. 1 Locations of 174 stations in Beijing-Tianjin-Hebei region (red asterisk indicates the altitude is 1000—1500 m, black circles indicate the altitude is 400—1000 m, and blue triangles indicate the altitude is 0—400 m; color shadings show terrain height, shaded: altitude)

速更新精细集成产品, RMAPS-IN 系统(杨璐等, 2019; 程丛兰等, 2019; 宋林烨等, 2019)利用最新的自动气象站和雷达、下垫面地形地表信息等其他观测资料进行精细化融合分析后, 结合线性外推技术和数值预报结果提供一个更接近真实大气的0—12 h 短时天气预报, 水平方向上分辨率为1 km, 垂直方向上分辨率为200 m, 共21层, 时间分辨率为10 min, 建模中使用的分析场数据包括地面2 m气温(T_{2m})、露点温度(T_d)、湿球温度(T_w)、相对湿度(RH)、雪线高度(Z_s)、雪混合比占雨和雪混合比的比例(Snf), 以及气温和湿球温度三维气象要素; 预报场数据包括 T_{2m} 、 T_d 、 T_w 、RH、 Z_s 及 Snf。

其中, 雪线高度 Z_s 定义为雪开始融化($T_w = 0^\circ\text{C}$) 时对应的高度。基于 RMAPS-IN 系统不同气压层上的高分辨率温度和比湿数据, 用二分算法(李金霞等, 2012)反复迭代, 自上向下计算各层的湿球温

度。当某层湿球温度首次出现大于临界值 0°C 时, 用该层和上一层线性插值得到的高度来计算雪线高度。若计算得到的雪线高度低于最低地形高度时则设为0, 计算出的雪线高度为海拔高度, 所以若想得到融化层相对于地面高度的位置, 需要用计算出的雪线高度 Z_s 减去地面高度。

基于 Thompson 混合相云微物理方案的高分辨率数值模式对近地面大气层中混合比的预报结果, 将模式预报的雪、雨、冰、霰降水混合比通过三维插值到1 km 分辨率网格点上, 计算得到 Snf, Snf 的取值范围为0—1。

以上采集到的观测数据和模式数据, 都将作为特征向量, 用于3种机器学习方法对京津冀地区降水相态客观预报算法的构建和测试。

2.2 降水相态的气候分布特征

2.2.1 不同降水相态的空间分布特征

基于1955—2019年京津冀地区174个国家级气象站逐日天气现象数据, 统计了雨、雨夹雪及雪的空间分布, 1955—2003年天气现象观测资料采样处理为对应的02、08、14、20时4个时次, 即对应时次有天气现象时, 记录1次; 2004—2019年天气现象观测资料采样处理为逐时数据, 即每个时次对应有天气现象时, 记录1次。从京津冀地区国家级气象站雨、雨夹雪及雪的空间分布(图2)可以看出, 京津冀地区雪的出现次数远多于雨和雨夹雪, 雨夹雪的出现次数最少。其中, 北京地区雨和雨夹雪的出现次数较河北地区少, 最多降雨次数主要出现在河北的南部及东北部平原地区; 海拔高度大于400 m 的山区站, 雨夹雪和雪的出现次数高于海拔高度400 m 以下的平原站, 张北(位于河北北部, 海拔1393 m)雨夹雪和雪的次数位居京津冀首位, 说明不同海拔高度下各类降水相态平均气候概率分布不同, 地形对于京津冀地区降水的多少也有一定的影响。

2.2.2 不同降水相态湿球温度的空间分布特征

降水相态的判别紧紧依赖于相对湿度、海拔高度及气温(Ding, et al, 2014), 而湿球温度包含了气温、相对湿度、气压信息。在 Y20 的基础上, 进一步利用各气象站的本站气压、相对湿度、气温数据分别计算了各站对应时次的湿球温度(Ding, et al, 2014), 重点关注地面湿球温度为0—1 $^\circ\text{C}$ 各类型降水相态的空间分布特征, 统计临界湿球温度条件下

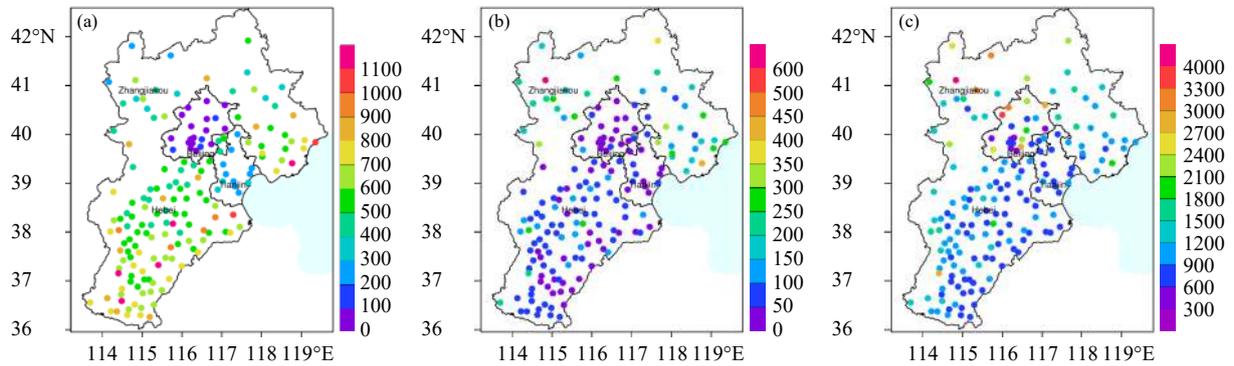


图2 1955—2019年京津冀地区雨(a)、雨夹雪(b)、雪(c)3种降水相态次数空间分布

Fig. 2 Spatial distributions of precipitation type frequencies of rain (a), sleet (b) and snow (c) in Beijing-Tianjin-Hebei region from 1955 to 2019

与降水相态的关系,并将复杂地形下降水相态的气候特征作为特征向量用于3种机器学习方法对京津冀地区降水相态客观预报算法的构建和测试。

图3、4、5分别给出了雨、雨夹雪和雪各降水相态 $T_w < 0^\circ\text{C}$ 、 $0 \leq T_w < 1.0^\circ\text{C}$ 和 $T_w \geq 1^\circ\text{C}$ 三个区间的概率分布情况。从图3可以看出, $T_w < 0^\circ\text{C}$ 的条件下,京津冀地区除河北东北部个别站外,其他站雪出现的概率基本都在0.9以上,雨夹雪和雨的概率都在0.1以下,可以区分出90%以上的雨和雪。 $0 \leq T_w < 1.0^\circ\text{C}$ 的条件下,雪出现的概率开始降低(0.5—0.7),但还是高于雨和雨夹雪的出现概率(0.1—0.5),海拔400 m以下的站,雨出现的概率相对较高(0.2—0.4),海拔400 m以上的山区站,雨夹雪出现的概率高于雨出现的概率。 $T_w \geq 1^\circ\text{C}$ 条件下,除河北北部海拔高于1000 m的几个站及河北南部个别站外,雨出现的概率基本都在0.9以上。

2.3 RMAPS-IN 高分辨率格点数据的统计特征

文中利用2016年1月—2019年4月冬季15个降水日京津冀地区174个国家级自动气象站采样处理好的逐10 min天气现象资料,提取各站周围对应的RMAPS-IN系统8个网格点上相应的分析场和1—12 h预报场资料。

温度垂直分布是降水相态的主要决定因素,本研究从京津冀地区RMAPS-IN高分辨率气温和湿球温度三维客观分析场样本中随机抽取雨、雨夹雪和雪各30组样本,分别绘制并分析了3种降水相态对应的气温和湿球温度随高度的变化。图6给出了不同降水相态对应的湿球温度随高度的变化。

从图6可以看出,不同降水相态湿球温度的三维廓线分布特征不同。雨的30组样本,基于地面高度的0 m湿球温度都在 0°C 以上,湿球温度在低层高于 0°C 的面积明显大于雨夹雪,开始融化的高

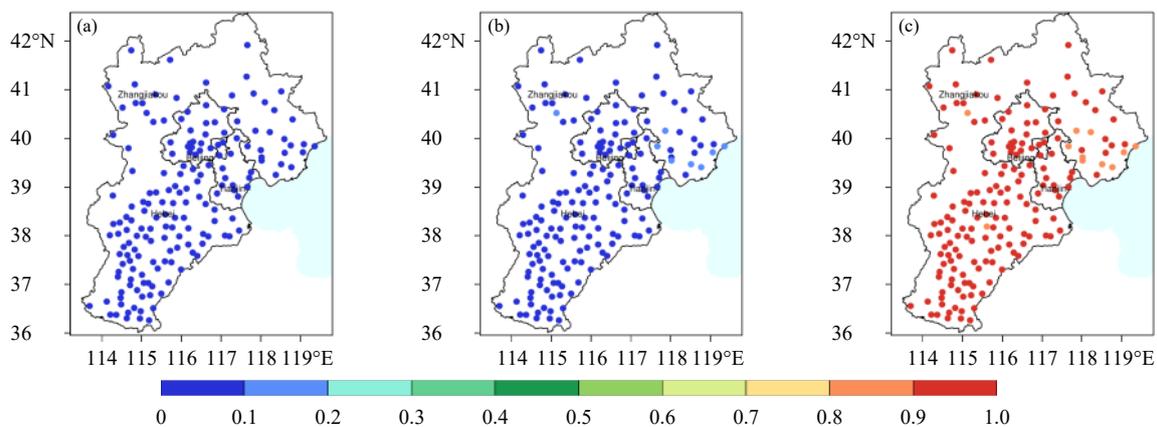


图3 雨(a)、雨夹雪(b)、雪(c)3种降水相态在 $T_w < 0^\circ\text{C}$ 时的概率空间分布

Fig. 3 Probability spatial distributions of rain (a), sleet (b) and snow (c) at $T_w < 0^\circ\text{C}$

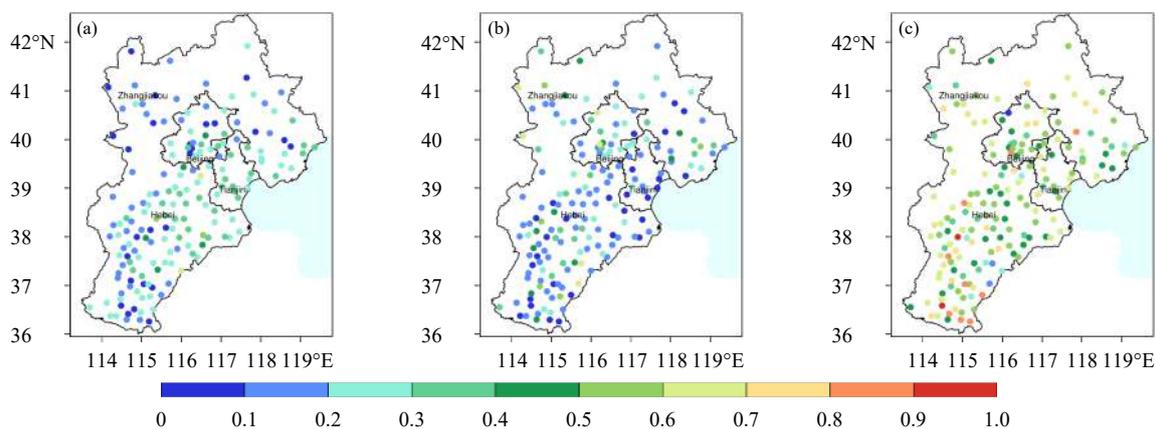


图 4 雨 (a)、雨夹雪 (b)、雪 (c) 3 种降水相态在 $0 \leq T_w < 1.0^\circ\text{C}$ 的概率空间分布

Fig. 4 Probability spatial distributions of rain (a), sleet (b) and snow (c) at $0 \leq T_w < 1.0^\circ\text{C}$

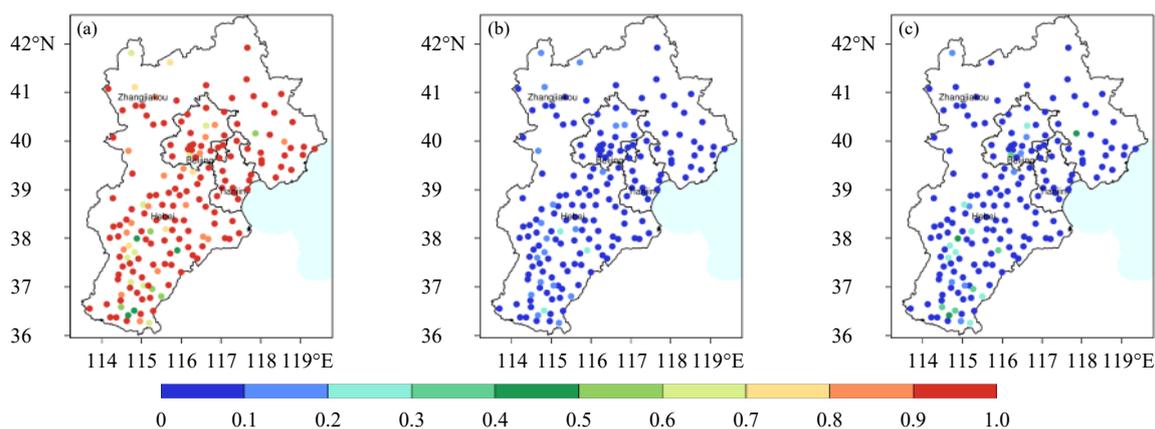


图 5 雨 (a)、雨夹雪 (b)、雪 (c) 3 种降水相态在 $T_w \geq 1^\circ\text{C}$ 的概率空间分布

Fig. 5 Probability spatial distributions of rain (a), sleet (b) and snow (c) at $T_w \geq 1^\circ\text{C}$

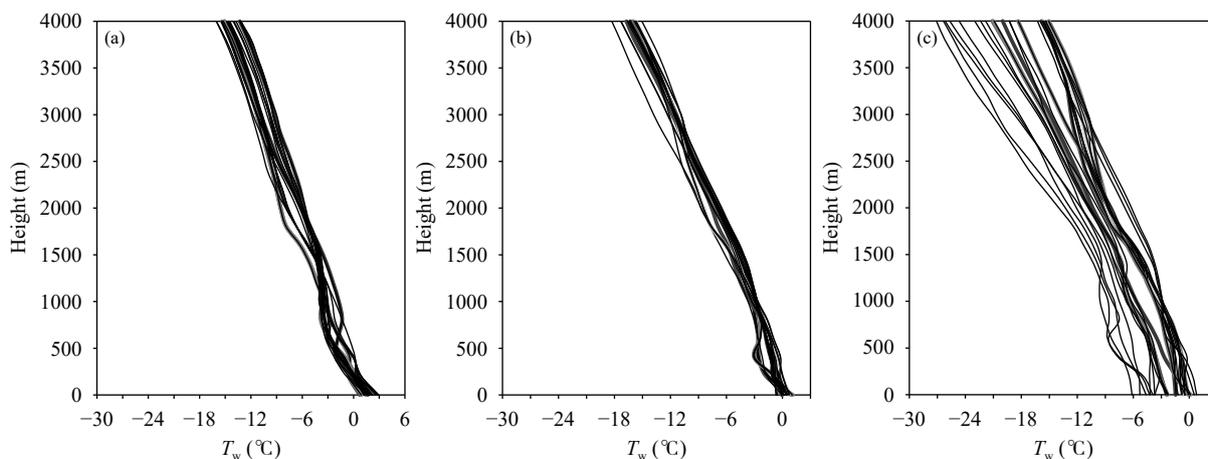


图 6 京津冀地区雨 (a)、雨夹雪 (b) 和雪 (c) 对应的湿球温度随高度的变化

Fig. 6 Vertical profiles of wet bulb temperature corresponding to rain (a), sleet (b) and snow (c) in Beijing-Tianjin-Hebei region

度较雨夹雪和雪高很多,暖层厚度相比雨夹雪更深,基本在离地 250 m 以上,低层温度的垂直递减率较大。雨夹雪的 30 组样本,0—500 m 高度,大部分廓线有弱的逆温层,0 m 湿球温度分布在-1—2℃,近地层都存在一个暖层,暖层相比雨较浅薄,分布在 0—100 m。雪的 30 组样本中大多样本整层湿球温度都在 0℃ 以下,有极少数的廓线样本近地层湿球温度高于 0℃,有非常浅薄的暖层。从三者湿球温度的垂直分布来看,湿球温度差异比较明显的区域主要位于 500 m 以下,尤其是近地面层,500 m 以上 3 种不同降水相态湿球温度基本都是冷冻层。

图 7 是京津冀地区 3 种降水相态对应的 T_{2m} 、 T_d 、RH、雪线高度和地面高度(Z_s-Z)以及 Snf 总样本的箱线图。方框中间的横线为中值(样本覆盖率达到 50% 的值),方框的上边界和下边界分别为 25% 和 75% 分位,上须和下须分别为样本中的最大值和最小值。

从图 7 可以看出,雨、雨夹雪和雪 2 m 气温中位数分别为 2.04℃、1.59℃、-0.09℃,25%—75% 分位的范围只有雨夹雪与雨有小范围的交叉,雪的 75% 分位在 0.61℃ 以下,雨的 25% 分位在 1.66℃ 以上,雨夹雪 25%—75% 分位在 1.2—1.86℃。2 m 露点温度雪的 75% 分位在 -0.27℃ 以下,雨的 25% 分位在 0.37℃ 以上,雨夹雪 25%—75% 分位在 -0.11—0.63℃。从相对湿度来看,雪、雨夹雪和雨的中位数比较接近,3 种相态的分布交叉范围较大。 Z_s-Z 表示从开始融化到地面的高度,不同相态 Z_s-Z 的分布特征比较明显,雨、雨夹雪和雪的中位数分别为 279 m、146 m、-45 m,25%—75% 分位的范围只有雨夹雪和雨有小范围交叉,雪的 25%—75% 分位在 -461—12 m,雨夹雪的 25%—75% 分位在 5—276 m,雨的 25%—75% 分位在 213—424 m。Snf 雨和雪的特征比较明显,雪和雨的中位数分别为 1 和 0,降水相态为雪时,Snf 的值主要集

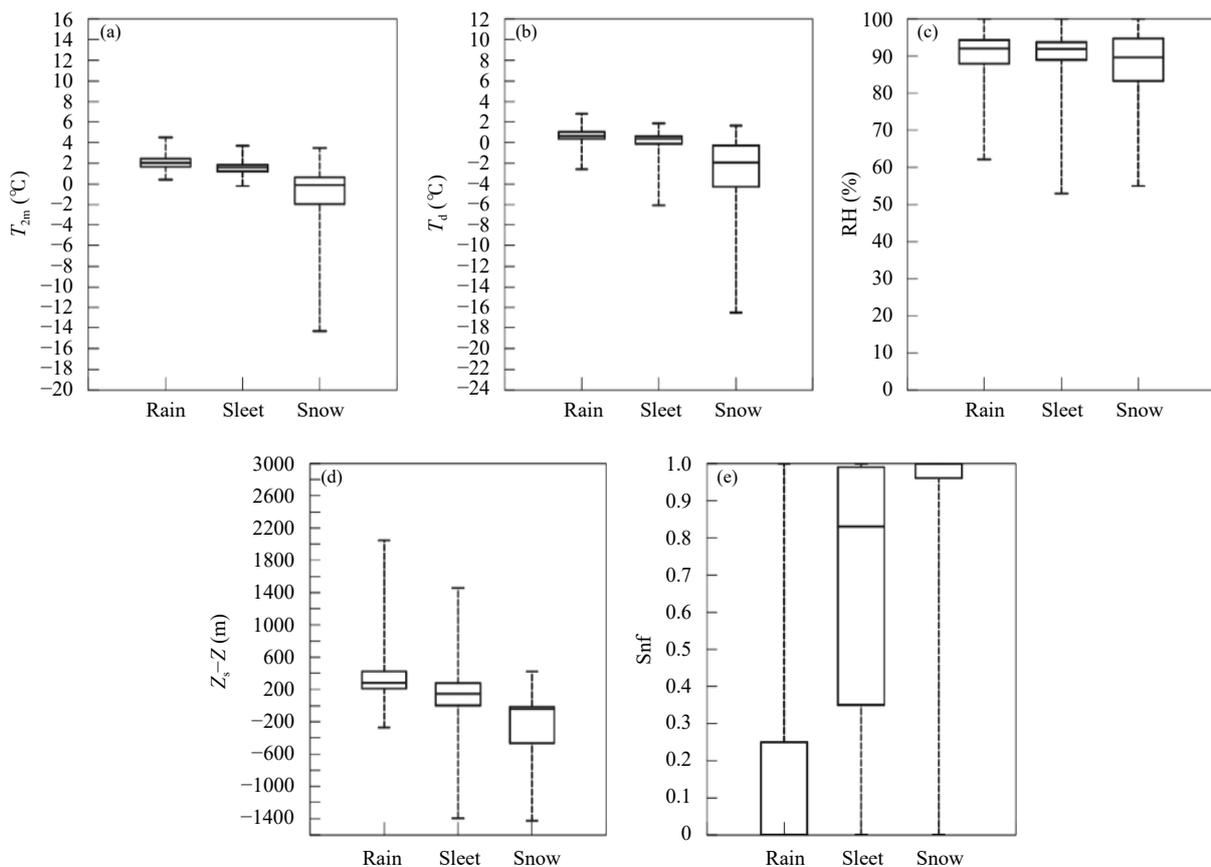


图 7 京津冀地区 3 种降水相态对应的 T_{2m} (a)、 T_d (b)、RH (c)、 Z_s-Z (d) 和 Snf (e) 箱线图

Fig. 7 Boxplots of T_{2m} (a), T_d (b), RH (c), Z_s-Z (d) and Snf (e) corresponding to three precipitation types in Beijing-Tianjin-Hebei region

中在 0.9—1, 降水相态为雨时, S_{nf} 的值主要集中在 0—0.25, 雨夹雪分布在 0—1。

3 机器学习理论基础

3.1 XGBoost 算法原理

XGBoost(Chen, et al, 2016) 是基于梯度提升框架的一种高度可扩展的树结构增强模型, 对稀疏数据的处理能力卓越, 算法原理是将原始数据集分割成多个子数据集, 将每个子数据集随机分配给基分类器进行预测, 然后将弱分类的结果按照一定的权重进行计算, 它由模型、参数和目标函数组成。其中, 模型是根据给定的输入样本因子去预测输出的结果, 参数是指最终建立的最优模型对应的系数, 而目标函数的优化情况则决定了模型的准确性, 目标函数优化的越好, 预测结果就越接近真实值。目标函数 $Obj(\theta)$ 由两部分组成, 如式(1)所示

$$Obj(\theta) = \sum_{i=1}^n L(y_i, \hat{y}_i) + \sum_{k=1}^m \Omega(f_m) \quad (1)$$

式中, $\sum_{i=1}^n L(y_i, \hat{y}_i)$ 代表损失函数, n 表示所用的样本数量, y_i 表示真实值, \hat{y}_i 表示预测值; $\sum_{k=1}^m \Omega(f_m)$ 是正则化项, m 表示当前迭代的次数, f_m 表示当前迭代的误差。

文中选择 CART 回归树作为模型的基函数, 单个 CART 第 m 次预测的结果可以表示为式(2)

$$f_m(x) = T(x, \theta_m) \quad (2)$$

式中, T 为决策树, m 代表基分类器的数量, θ 代表决策树的划分路径, 每棵决策树一个一个往里面加, 最后预测结果为前一次的预测结果加上当下的。误差项可以表示为

$$L(y_i, \hat{y}_i) = L(y_i, f_{m-1}(x)) + T(x, \theta_m) \quad (3)$$

式中, $L(y_i, \hat{y}_i)$ 为真实值 (y_i) 和预测值 (\hat{y}_i) 的差值之和。

3.2 支持向量机

支持向量机 (Support Vector Machine, SVM) 由 Cortes 等 (1995) 提出, 是一种建立在统计学习理论和结构风险最小化原理基础上的小样本学习方法。它最初是为二值分类问题设计的, 当处理多类问题时, 就需要构造合适的多类分类器。文中采用 libsvm 中的一对一法, 其做法是在任意两类样本之间设计一个 SVM, 因此 k 个类别的样本就需要设计 $k(k-1)/2$ 个 SVM。当对一个未知样本进行分类

时, 最后得票最多的类别即为该未知样本的类别。

3.3 DNN 深度神经网络

DNN (Deep Neural Networks) 是深度学习较为常见的也是最为基本的网络结构, 由输入层、隐藏层、输出层组成。输入层是经过预处理的输入数据, 紧接着是隐藏层, 隐藏层可以是一层, 也可以多层级联。网络的最后一层是输出层, 输出层则需要针对不同的预测目标来设计其节点数, 例如是手写体数字识别则属于十分类任务, 可以输出 10 个结点代表十分类的概率分布。DNN 的训练可以分为两个过程, 输入信息的前向传播和基于误差的反向传播。试验中所用深度神经网络结构是一个 4 层的全连接神经网络, 包含两个隐藏层, 输入层节点数为训练样本的特征数, 训练样本从输入层传入到整个神经网络, 输出层含有 3 个节点, 对应 3 分类。设 W_{ij} 为连接节点 i 与 j 的权值, b_j 则是节点 j 对应的偏置, a_j 为节点最终的激活值, σ 代表激活函数, 通常选用 Relu 函数或者是 sigmoid 函数。正向传播的公式如下

$$a_j = \sigma \left(\sum_{i=1}^m w_{ij} x_i + b_j \right) \quad (4)$$

误差的反向传播通常采用经典的 BP 算法 (Rumelhart, et al, 1986), 通过计算损失函数, 使用梯度下降法来不断的调整网络中的权值以减少输出结果的误差, 例如 d 是期望输出, y 是网络输出, 损失函数的定义可以是平方差损失

$$loss = \frac{1}{2} \sum_{k=1}^n (d - y)^2 \quad (5)$$

整个训练以梯度下降法 (Gradient Descent Optimizer) 作为优化器、以交叉熵作为损失函数进行网络的优化, 使用带指数衰减的学习率设置、L2 正则化来避免过度拟合, 并使用滑动平均模型来使得最终得到的模型具有更强的鲁棒性。

4 降水相态分类模型构建及结果检验对比

4.1 不同特征参数组的选取方法

使用 2016—2019 年 1 月—次年 3 月的高分辨率模式产品和同期的国家级自动气象站观测资料, 将分析场 ($t=0$ h) 按近似 7:3 的比例随机划分为训练集和测试集; 同样将预报场 ($t=1-12$ h) 样本按近

似 7 : 3 的比例随机划分为训练集和测试集, 分别基于分析场和预报场 70% 的样本建立降水相态模型, 并利用剩余的 30% 的样本开展独立检验。分析场和预报场总样本数如表 1 所示。

表 1 分析场和预报场样本总数
Table 1 Total number of samples of analysis field and prediction field

	雨	雨夹雪	雪
分析场	97758	51219	201924
预报场	504495	209031	1215927

通过对与预报对象有明确意义的各种特征参数的选取, 最终选取 41 个分析场特征, 11 个预报场特征, 各特征的物理意义如下:

(1) RMAPS-IN 模式输出产品: 分析场样本包括 T_{2m} 、 T_d 、 T_w 、RH、 Z_s 、Snf 及气温和湿球温度三维气象要素; 预报场样本包括 T_{2m} 、 T_d 、 T_w 、RH、 Z_s 及 Snf。

(2) 地面实况观测数据: 实况天气现象。

(3) 复杂地形下降水相态气候特征: 基于 1955—2019 年自动气象站观测数据, 得到京津冀地区各国家级站点 T_{2m} 和 T_w 在 $[-10^{\circ}\text{C}, 10^{\circ}\text{C}]$ 每 0.5°C 间隔内雨、雨夹雪和雪的出现概率。

为比较不同特征参数构建对不同机器学习方法降水相态分类预报模型的影响, 文中设计了 2 组 (test1 和 test2) 特征参数组的构建方法 (表 2), 其中 test2 中增加了复杂地形下降水相态气候特征, 即根据 RMAPS-IN 模式输出的 2 m 气温和湿球温度, 匹配 $[-10^{\circ}\text{C}, 10^{\circ}\text{C}]$ 每 0.5°C 间隔内雨、雨夹雪和雪的出现概率, 并利用 XGBoost、SVM 和 DNN 三种机器学习方法, 针对 test1 和 test2 对应的两种不同特征参数组, 分别进行建模和检验。

表 2 不同模型不同特征参数组的构建方法

Table 2 Construction methods for different characteristic parameter groups in different models

		RMAPS-IN 模式 输出产品	地面实况 观测数据	复杂地形下降水 相态气候特征
XGBoost	test1	√	√	
	test2	√	√	√
SVM	test1	√	√	
	test2	√	√	√
DNN	test1	√	√	
	test2	√	√	√

4.2 数据归一化

由于奇异样本数据会导致训练时间增大和模型无法收敛的情况, 为消除其带来的不良影响, 使用 min-max 标准化将全部样本的每个特征映射到 $[0, 1]$ 。

$$X_{\text{norm}} = \frac{X - X_{\text{min}}}{X_{\text{max}} - X_{\text{min}}} \quad (6)$$

式中, X_{min} 是原始特征 X 的最小值, X_{max} 是原始特征的最大值, X_{norm} 是标准化结果。

4.3 预测结果

由于雨、雨夹雪和雪的数据集不平衡, 为了客观评测不同模型的性能差异, 文中利用混淆矩阵进行评价。混淆矩阵是衡量分类模型准确度中最基本、最直观、计算最简单的方法, 即分别统计分类模型归错类、归对类的观测值个数, 然后把结果放在一个表里。混淆矩阵的具体定义如表 3 所示, 其中 TP 表示正类样本中分类正确的数量, FN 表示对正类误分为负类的样本数, FP 表示将负类错分为正类的样本数, TN 表示负类正确分类的样本数。因为样本数量大, 为方便对比, 数字下面加“()”里标上相对于总体的百分数, 表示预测准确率, 如表 4 和表 5 所示。

表 4 中 test1 特征参数组为 RMAPS-IN 分析场样本和地面实况观测的天气现象, test2 特征参数组中增加了复杂地形下降水相态气候特征。针对 2 种不同特征参数组构建得到的 3 种不同降水相态混淆矩阵可以看出, XGBoost 和 DNN 两种机器学习方法对于雨、雨夹雪和雪的预测准确率相当, 都明显高于 SVM 方法。特征参数中增加复杂地形下降水相态气候特征, 可以明显提升 3 种机器学习方法对于雨、雨夹雪和雪的预测准确率。基于此, 针对预报场样本, 只对 test2 特征参数组方案 (RMAPS-IN 预报场样本、地面实况观测的天气现象及复杂地形下降水相态气候特征) 进行了建模和检测, 混淆矩阵如表 5 所示。从表中可以看出, XGBoost 和 DNN 两种机器学习方法对于雨、雨夹雪和雪的预测准确率相当, 都明显高于 SVM 方法, 整体预报准确率略低于基于分析场样本建立的降水相态模型。

表 3 混淆矩阵

Table 3 Confusion matrix

	预测为真	预测为假
实际为真	TP (True Positive)	FN (False Negative)
实际为假	FP (True Positive)	TN (False Negative)

表 4 基于分析场样本建立的不同降水相态模型的混淆矩阵
Table 4 Confusion matrix of different precipitation type models based on analysis field samples

		XGBoost预测值			SVM预测值			DNN预测值		
		雨	雨夹雪	雪	雨	雨夹雪	雪	雨	雨夹雪	雪
test1	真实值 雨	27418 (93.5%)	1773 (6%)	136 (0.5%)	25471 (86.9%)	3412 (11.6%)	444 (1.5%)	27931 (95.2%)	1164 (4.0%)	232 (0.8%)
	真实值 雨夹雪	1365 (8.9%)	12050 (78.4%)	1951 (12.7%)	3409 (22.2%)	7229 (47.0%)	4728 (30.8%)	1038 (6.8%)	13011 (84.7%)	1317 (8.6%)
	真实值 雪	140 (0.2%)	1365 (2.3%)	59072 (97.5%)	377 (0.6%)	2592 (4.3%)	57608 (95.1%)	305 (0.5%)	999 (1.6%)	59273 (97.8%)
test2	真实值 雨	28162 (96.0%)	1036 (3.5%)	129 (0.4%)	25987 (88.6%)	2790 (9.5%)	547 (1.9%)	28573 (97.4%)	588 (2.0%)	163 (0.6%)
	真实值 雨夹雪	789 (5.1%)	13251 (86.2%)	1326 (8.6%)	3003 (19.5%)	9288 (60.4%)	3064 (19.9%)	627 (4.1%)	13803 (89.8%)	925 (6.0%)
	真实值 雪	84 (0.1%)	584 (1.0%)	59909 (98.9%)	564 (0.9%)	2236 (3.7%)	57770 (95.4%)	147 (0.2%)	550 (0.9%)	59873 (98.8%)

表 5 基于预报场样本建立的不同降水相态模型的混淆矩阵
Table 5 Confusion matrix of different precipitation type models based on prediction field samples

		XGBoost预测值			SVM预测值			DNN预测值		
		雨	雨夹雪	雪	雨	雨夹雪	雪	雨	雨夹雪	雪
test2	真实值 雨	141481 (93.5%)	5368 (3.5%)	4499 (3.0%)	134258 (88.7%)	8367 (5.5%)	8723 (5.8%)	141079 (93.2%)	6371 (4.2%)	3559 (2.4%)
	真实值 雨夹雪	3183 (5.1%)	44797 (71.5%)	14711 (23.5%)	7990 (12.7%)	30073 (48.0%)	24628 (39.3%)	4752 (7.6%)	42926 (68.5%)	14367 (22.9%)
	真实值 雪	1601 (0.4%)	5829 (1.6%)	357348 (98.0%)	4206 (1.1%)	8978 (2.5%)	351594 (96.4%)	2164 (0.6%)	6811 (1.9%)	355186 (97.4%)

另外,根据分析场和预报场 test2 方案混淆矩阵计算了不同降水相态模型的命中率(probability of detection, POD)、虚警率(false alarm ratio, FAR)和临界成功指数(critical success index, CSI) (Chen, et al, 2017),如表 6 和表 7 所示。基于分析场样本, SVM 模型 3 种降水相态的整体命中率为 88.4%, XGBoost 和 DNN 模型整体命中率分别为 96.3% 和 97.1%, 明显优于 SVM 模型。基于预报场样本, SVM 模型 3 种降水相态的整体命中率为 89.1%, XGBoost 和 DNN 模型整体命中率分别为 93.9% 和 93.4%。针对不同的降水类型, 3 种模型对于雨和雪的命中率都明显优于雨夹雪。另外, 基于分析场和预报场样本建立的模型, 针对 3 种不同的降水类型, XGBoost 和 DNN 模型虚警率都低于 SVM 模型。基于预报场数据得到的模型命中率略低于分析场, 虚警率略高于分析场。一方面是由于 RMAPS-IN 系统格点分辨率高, 分析场数据本身应用 5 min 自动气象站观测资料进行了数据融合和快速订正, 大多数站点观测值可作为“真值”处理, 数据的精度和准确性要优于预报场数据; 另外, 基于分析场样本建模时, 共选取了 41 个分析场特征, 其

中包括了气温和湿球温度的三维气象要素, 而由于 RMAPS-IN 系统没有三维气象要素的预报场, 所以预报场样本只选取了 11 个预报场特征, 特征向量相对较少, 对模型整体预测的正确性也有一定影响。

整体来看, 基于 XGBoost 和 DNN 预报场模型, 对于雨诊断的命中率为 93%, 略低于 Y20 方案 (94%), 基于 SVM 预报场模型, 雨诊断的命中率较 Y20 方案偏低 5%; 对于雨夹雪, 基于 XGBoost 和 DNN 预报场模型, 命中率可以达 70% 左右, 高于 Y20 方案 (41%); 对于雪, 3 种机器学习模型命中率都高于 Y20 方案, 其中 XGBoost 和 DNN 命中率较 Y20 提升 7%、SVM 提升 5%。这也进一步说明, 面对不同气候背景和海拔高度的站点, 将每个诊断变量设置为同一阈值本身也会给降水相态的客观诊断带来误差, 而选取合适的特征参数, 基于机器学习方法对样本进行训练和学习, 能够更好地解决客观模型在不同区域的适用性和预报能力。

5 个例检验

利用 Y20 及 3 种机器学习方法建立的降水相态的高分辨率客观分类模型, 对 2021 年 2 月 13—

表 6 基于分析场样本建立的模型预测评分
Table 6 Model prediction scores based on analysis field samples

降水类型	POD			FAR			CSI		
	XGBoost	SVM	DNN	XGBoost	SVM	DNN	XGBoost	SVM	DNN
雨	0.96	0.88	0.97	0.03	0.12	0.03	0.93	0.79	0.95
雨夹雪	0.86	0.60	0.90	0.11	0.35	0.08	0.78	0.46	0.84
雪	0.99	0.95	0.98	0.02	0.06	0.02	0.97	0.90	0.97

表 7 基于预报场样本建立的模型预测评分
Table 7 Model prediction scores based on prediction field samples

降水类型	POD			FAR			CSI		
	XGBoost	SVM	DNN	XGBoost	SVM	DNN	XGBoost	SVM	DNN
雨	0.93	0.89	0.93	0.03	0.08	0.05	0.91	0.82	0.89
雨夹雪	0.71	0.48	0.69	0.20	0.37	0.23	0.61	0.38	0.57
雪	0.98	0.96	0.98	0.05	0.09	0.05	0.93	0.88	0.93

15日降水相态进行了预测和检验。表 8 给出了 Y20、XGBoost、SVM、DNN 这 4 种降水相态客观分类模型对这次过程的预测准确率评分。从表 8 可以看出,这次过程 Y20 预测雪的准确率(0.89)高于 3 种机器学习方法,其中 DNN 方法预测的雪准确率(0.77)略优于 XGBoost(0.75)和 SVM(0.73);Y20 方案对雨预测的准确率(0.79)低于 3 种机器学习方法,SVM 对于雨预测的准确率(0.98)最高,其次为 DNN(0.93)。

表 8 2021 年 2 月 13—15 日 Y20、XGBoost、SVM、DNN 四种降水相态客观分类模型预测准确率评分
Table 8 Prediction accuracy score of four objective classification models for precipitation type based on Y20, XGBoost, SVM and DNN during 13—15 February 2021

降水类型	POD			
	Y20	XGBoost	SVM	DNN
雨	0.79	0.89	0.98	0.93
雨夹雪	/	/	/	/
雪	0.89	0.75	0.73	0.77

6 结论与讨论

利用京津冀地区国家级自动气象站观测资料及网格化快速更新精细集成产品,统计分析了京津冀地区复杂地形下各类降水相态温度和湿球温度平均气候概率的分布差异、不同降水相态时网格化快速更新精细集成产品中可能影响降水相态判断的特征信息。将地面观测天气现象资料、复杂地形下降水相态气候特征及高分辨率模式输出产品作

为特征向量,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,并对同样条件下 3 种机器学习方法对雨、雨夹雪和雪 3 种京津冀主要降水相态的预报效果进行了对比检验。

(1)从 1955—2019 年京津冀地区雨、雨夹雪及雪的空间分布来看,海拔高度高于 400 m 的山区站点,雨夹雪和雪的出现次数都高于海拔高度 400 m 以下的平原站点,张北(位于河北北部,海拔 1393 m)雨夹雪和雪的次数均居京津冀之首,说明不同海拔高度下各类降水相态平均气候概率分布不同,地形对于京津冀地区降水相态有一定的影响。

(2)通过统计长时间序列京津冀地区国家级站点观测资料降水相态与湿球温度(T_w)的关系,发现在 $T_w < 0^\circ\text{C}$ 时,京津冀地区除河北东北部个别站点,其他站点雪出现的概率基本在 0.9 以上,雨夹雪和雨的概率在 0.1 以下。 $0 \leq T_w < 1.0^\circ\text{C}$ 时,雪出现的概率开始降低,但还是高于雨和雨夹雪的出现概率,雨和雨夹雪出现的概率开始升高,雨出现的概率在大部分海拔 400 m 以下的站点相对较高,海拔 400 m 以上的山区站点,雨夹雪出现的概率高于雨出现的概率。 $T_w > 1^\circ\text{C}$ 时,除河北北部海拔高于 1000 m 的几个站点及河北南部个别站点,雨出现的概率基本在 0.9 以上。整体来看,湿球温度低于 0°C 时,可以区分出 90% 以上的雨和雪。

(3)3 种机器学习方法对 3 种降水相态都具有较好的预报能力,对降雨和降雪预报最好,其次是雨夹雪,并且 XGBoost 和 DNN 的预报能力相当,都明显优于 SVM。SVM 本身是一个二分分类器,要

实现多分类必须构造合适的多类分类器, 一般有一对一法和一对多法, 本试验中使用一对一法, 其缺陷就是使得训练时间和测试时间较长, 再者 SVM 缺失数据敏感, 对参数和核函数的选择敏感, 对于大规模训练样本难以实施。

(4) 构建的特征参数中增加复杂地形下降水相态气候特征, 可以明显提升 3 种机器学习方法对于雨、雨夹雪和雪的命中率。基于分析场样本, SVM 模型 3 种降水相态的整体命中率为 88.4%, XGBoost 和 DNN 模型整体命中率分别为 96.3% 和 97.1%。基于预报场样本, SVM 模型 3 种降水相态的整体命中率为 89.1%, XGBoost 和 DNN 模型整体命中率分别为 93.9% 和 93.4%。

(5) 在本试验中, XGBoost 和 DNN 都很好地实现了降水相态的分类。XGBoost 是 GBDT 的一种实现, 既可以解决分类问题, 也可以解决回归问题。XGBoost 算法在目标函数中加入了正则化防止过拟合, 使得算法的健壮性更好, 在处理每个特征列时可以做到并行, 并且考虑了训练数据稀疏值的情况, 可以为缺失值或指定的值指定分支的默认方向, 这大幅度提升了算法的效率。而 DNN 是深度学习方法, 网络结构搭建灵活, 可调控参数较多, 并且有丰富的优化器, 可以深度挖掘数据特征, 并且可以利用 GPU 加速计算, 适用于大数据特征挖掘。所以在其他区域构建降水相态模型时, DNN 不失为一种更便捷的方法。另外, 在以后研究中还可以另辟新路, 构建各物理特征的多维样本, 利用卷积神经网络(CNN)实现降水相态的高准确度分类。

参考文献

- 蔡雪薇, 白佳蕴. 2013. 2012 年 11 月吉林省东南山区两次雨转暴雪过程对比分析. *气象与减灾研究*, 36(3): 43-49. Cai X W, Bai J Y. 2013. The comparative analysis of two raining to snowing processes occurred in the southeastern mountain area of Jilin province during november 2012. *Meteor Disaster Reduction Res*, 36(3): 43-49 (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. *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. *Meteor Mon*, 45(8): 1037-1051 (in Chinese)
- 程丛兰, 陈敏, 陈明轩等. 2019. 临近预报的两种高时空分辨率定量降水预报融合算法的对比试验. *气象学报*, 77(4): 701-714. Cheng C L, Chen M, Chen M X, et al. 2019. Comparative experiments on two high spatiotemporal resolution blending algorithms for quantitative precipitation nowcasting. *Acta Meteor Sinica*, 77(4): 701-714 (in Chinese)
- 崔锦, 周晓珊, 阎琦等. 2014. WRF 模式不同微物理过程对东北降水相态预报的影响. *气象与环境学报*, 30(5): 1-6. Cui J, Zhou X S, Yan Q, et al. 2014. Impact of different microphysical processes of WRF model on precipitation phase forecasting in Northeast China. *J Meteor Environ*, 30(5): 1-6 (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. *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. *Meteor Mon*, 39(3): 324-332 (in Chinese)
- 段长春, 段旭, 段苏琴等. 2011. 近 50 年云南省降雪的气候变化特征. *气象*, 37(5): 599-606. Duan C C, Duan X, Duan S Q, et al. 2011. Climate variational characteristics of snowfall in Yunnan province for the last 50 years. *Meteor Mon*, 37(5): 599-606 (in Chinese)
- 李江波, 李根娥, 裴雨杰等. 2009. 一次春季强寒潮的降水相态变化分析. *气象*, 35(7): 87-94. Li J B, Li G E, Pei Y J, et al. 2009. Analysis on the phase transformation of precipitation during a strong cold wave happened in spring. *Meteor Mon*, 35(7): 87-94 (in Chinese)
- 李金霞, 王海丰. 2012. 二分查找判定树的构造方法研究及应用. *软件*, 33(4): 20-22. Li J X, Wang H F. 2012. Study on construction methods of binary search decision tree and its applications. *Software*, 33(4): 20-22 (in Chinese)
- 刘玉莲, 任国玉, 于宏敏. 2012. 中国降雪气候学特征. *地理科学*, 32(10): 1176-1185. Liu Y L, Ren G Y, Yu H M. 2012. Climatology of snow in China. *Scientia Geogr Sinica*, 32(10): 1176-1185 (in Chinese)
- 漆梁波, 张瑛. 2012. 中国东部地区冬季降水相态的识别判据研究. *气象*, 38(1): 96-102. Qi L B, Zhang Y. 2012. Research on winter precipitation types' discrimination criterion in eastern China. *Meteor Mon*, 38(1): 96-102 (in Chinese)
- 宋林焯, 陈明轩, 程丛兰等. 2019. 京津冀夏季雷达成量降水估测的误差统计及定量气候校准. *气象学报*, 77(3): 497-515. Song L Y, Chen M X, Cheng C L, et al. 2019. Characteristics of summer QPE error and a climatological correction method over Beijing-Tianjin-Hebei region. *Acta Meteor Sinica*, 77(3): 497-515 (in Chinese)
- 孙继松, 梁丰, 陈敏等. 2003. 北京地区一次小雪天气过程造成路面交通严重受阻的成因分析. *大气科学*, 27(6): 1057-1066. Sun J S, Liang F, Chen M, et al. 2003. An analysis on serious city traffic trouble caused by light snow. *Chinese J Atmos Sci*, 27(6): 1057-1066 (in Chinese)
- 孙燕, 尹东屏, 顾沛澍等. 2014. 华东地区冬季不同降水相态的时空变化特征. *地理科学*, 34(3): 370-376. Sun Y, Yin D P, Gu P S, et al. 2014. The spatial and temporal variations of different precipitation phases in eastern China during winter. *Scientia Geogr Sinica*, 34(3): 370-376 (in Chinese)
- 王春乙, 王石立, 霍治国等. 2005. 近 10 年来中国主要农业气象灾害监测预警与评估技术研究进展. *气象学报*, 63(5): 659-671. Wang C Y, Wang S L, Huo Z G, et al. 2005. Progress in research of agro-meteorological disasters in China in recent decade. *Acta Meteor Sinica*, 63(5): 659-671 (in Chinese)

- 许美玲, 梁红丽, 金少华等. 2015. 昆明冬季降水相态识别判据研究. *气象*, 41(4): 474-479. Xu M L, Liang H L, Jin S H, et al. 2015. Study on identification criterion for precipitation phase in winter Kunming. *Meteor Mon*, 41(4): 474-479 (in Chinese)
- 杨成芳, 周淑玲, 刘畅等. 2015. 一次入海气旋局地暴雪的结构演变及成因观测分析. *气象学报*, 73(6): 1039-1051. Yang C F, Zhou S L, Liu C, et al. 2015. Case study of the cause and the dynamic structure for a small-scale snow storm event associated with a cyclone. *Acta Meteor Sinica*, 73(6): 1039-1051 (in Chinese)
- 杨璐, 陈敏, 陈明轩等. 2019. 高空分辨率三维风场在强对流天气临近预报中的融合应用研究. *气象学报*, 77(2): 243-255. Yang L, Chen M, Chen M X, et al. 2019. Fusion of 3D high temporal and spatial resolution wind field and its application in nowcasting of severe convective weather. *Acta Meteor Sinica*, 77(2): 243-255 (in Chinese)
- 尤凤春, 郭丽霞, 史印山等. 2013. 北京降水相态判别指标及检验. *气象与环境学报*, 29(5): 49-54. You F C, Guo L X, Shi Y S, et al. 2013. Discrimination index of precipitation phase state and its verification in Beijing. *J Meteor Environ*, 29(5): 49-54 (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据研究. *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. *Plateau Meteor*, 32(6): 1780-1786 (in Chinese)
- Bourgouin P. 2000. A method to determine precipitation types. *Wea Forecast*, 15(5): 583-592
- Chen M, Bica B, Tüchler L, et al. 2017. Statistically extrapolated nowcasting of summertime precipitation over the Eastern Alps. *Adv Atmos Sci*, 34(7): 925-938
- Chen M X, Quan J N, Miao S G, et al. 2018. Enhanced weather research and forecasting in support of the Beijing 2022 winter Olympic and Paralympic games. *WMO Bull*, 67(2): 58-61
- Chen T Q, Guestrin C. 2016. XGBoost: A scalable tree boosting system // Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. San Francisco: ACM, 785-794
- Cortes C, Vapnik V. 1995. Support-vector networks. *Mach Learn*, 20(3): 273-297
- Ding B H, Yang K, Qin J, et al. 2014. The dependence of precipitation types on surface elevation and meteorological conditions and its parameterization. *J Hydrol*, 513: 154-163
- Elmore K L, Grams H M. 2015. Using mPING data to drive a forecast precipitation type algorithm // Proceedings of the 13th Conference on Artificial Intelligence. Phoenix: American Meteorological Society
- Haiden T, Kann A, Wittmann C, et al. 2011. The integrated nowcasting through comprehensive analysis (INCA) system and its validation over the Eastern Alpine region. *Wea Forecast*, 26(2): 166-183
- Ikeda K, Steiner M, Pinto J, et al. 2013. Evaluation of cold-season precipitation forecasts generated by the hourly updating high-resolution rapid refresh model. *Wea Forecast*, 28(4): 921-939
- Rajagopal S, Harpold A A. 2016. Testing and improving temperature thresholds for snow and rain prediction in the western United States. *J Ame Water Resour Assoc*, 52(5): 1142-1154
- Ramer J. 1993. An empirical technique for diagnosing precipitation type from model output // Proceedings of the 5th International Conference on Aviation Weather Systems. Vienna: American Meteorological Society, 227-230
- Rumelhart D E, Hinton G E, Williams R J. 1986. Learning representations by back-propagating errors. *Nature*, 323(6088): 533-536
- Sankaré H, Thériault J M. 2016. On the relationship between the snowflake type aloft and the surface precipitation types at temperatures near 0°C. *Atmos Res*, 180: 287-296
- Schuur T J, Park H S, Ryzhkov A V, et al. 2012. Classification of precipitation types during transitional winter weather using the RUC model and polarimetric radar retrievals. *J Appl Meteor Climatol*, 51(4): 763-779
- Shafer P E. 2010. Logit transforms in forecasting precipitation type // Proceedings of the 20th Conference on Probability and Statistics in the Atmospheric Sciences. Atlanta: American Meteorological Society, 222
- Shafer P E. 2015. A sample size sensitivity test for MOS precipitation type // Proceedings of the Special Symposium on Model Postprocessing and Downscaling. Phoenix: American Meteorological Society
- Stewart R E, Thériault J M, Henson W. 2015. On the characteristics of and processes producing winter precipitation types near 0°C. *Bull Amer Meteor Soc*, 96(4): 623-639
- Thériault J M, Stewart R E, Henson W. 2010. On the dependence of winter precipitation types on temperature, precipitation rate, and associated features. *J Appl Meteor Climatol*, 49(7): 1429-1442
- Thompson G, Rasmussen R M, Manning K. 2004. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part I: Description and sensitivity analysis. *Mon Wea Rev*, 132(2): 519-542
- Thompson G, Field P R, Rasmussen R M, et al. 2008. Explicit forecasts of winter precipitation using an improved bulk microphysics scheme. Part II: Implementation of a new snow parameterization. *Mon Wea Rev*, 136(12): 5095-5115
- Wu B Y, Yang K, Zhang R H. 2009. Eurasian snow cover variability and its association with summer rainfall in China. *Adv Atmos Sci*, 26(1): 31-44
- Yang L, Chen M X, Wang X L, et al. 2021. Classification of precipitation type in North China using model-based explicit fields of hydrometeors with modified thermodynamic conditions. *Wea Forecast*, 36(1): 91-107
- Zhang T J. 2005. Influence of the seasonal snow cover on the ground thermal regime: An overview. *Rev Geophys*, 43(4): RG4002