

基于机器学习的青岛市区近岸海雾集成预报方法

高松^{1,2}, 徐江玲^{1,2}, 刘桂艳^{1,2}, 毕凡^{1,2}, 张薇^{1,2}, 白涛^{1,2}

(1. 国家海洋局北海预报中心, 山东 青岛 266061; 2. 山东省海洋生态环境与防灾减灾重点实验室, 山东 青岛 266061)

摘要: 利用 2014—2017 年青岛小麦岛海洋站观测资料, 采用机器学习方法建立了青岛市区近岸海雾集成预报模型, 通过主成分分析方法对预报因子进行了优选。结果表明: 采用能见度、风向、风速、气压、露点、气温、海温、气温露点差、气海温差、相对湿度、云量、气温 24h 变温 12 个预报因子建立的海雾集成预报模型, 对 2018 年海雾预报的 TS 评分约为 0.64, 海雾预报正确率约为 0.783, 具有较好的预报能力, 为海雾预报提供了新的方法。

关键词: 海雾; 机器学习; 集成预报; 青岛沿海

中图分类号: P732 **文献标识码:** A **文章编号:** 1000-3096(2021)03-0033-10

DOI: 10.11759/hyxx20200310004

海雾是指在海洋的影响下, 在海上、岛屿或沿海地区形成的雾, 我国沿海自南向北均有海雾发生。其中, 黄海是我国海雾发生范围最广、频率最高的海域^[1-3]。青岛位于黄海西部, 是受黄海海雾影响最为严重的地区之一^[4]。海雾对青岛市的公众社会活动、海上交通运输、海上活动等均有重要影响。海雾生消机制非常复杂, 目前业务预报水平还远不能满足社会服务的需求。

现今在国内外气象部门业务运行中, 对海雾的预报方法主要有天气学方法、数值预报方法和统计方法^[5]。天气学方法主要运用天气学基本原理和天气学分析方法, 考虑天气系统与海雾发生之间的联系, 对预报员的主观经验依赖性高, 预报准确率和可靠性难以保障。近年来随着计算机技术的进步以及对海雾宏观、微观结构和发生、发展规律的深入认识, 数值预报研究成果丰硕, 部分成果实现了业务应用, 逐渐成为海雾预报的重要方法^[6-7]。统计预报方法通过历史资料的统计分析, 建立海雾与预报因子之间的统计模型对海雾进行预报, 该方法在我国沿海地区海雾的业务预报中也取得了较好的成效^[8-10]。常见的建模方法有逐步回归、模糊和神经网络^[11]、支持向量机^[12]、策树方法^[9]等, 可对未来是否有雾进行判断。目前沿海海雾不同客观预报方法预报准确率约在 66%, 均存在空报、漏报现象, 预报员仍需在综合分析各种客观预报结果基础上进行主观订正, 提高海雾和能见度的精细化预报精度^[13]。

本文旨在借鉴气象集成预报^[14-15]的思路, 采用多种机器学习方法建立青岛市区近岸海雾集成预报模型, 为海雾预报提供新方法。利用多年青岛小麦岛海洋观测站气象海洋观测资料, 采用决策树、判别分析、回归树、支持向量机、K 近邻、分类集成等 6 类 23 个机器学习模型进行遴选, 获得预报评分前 8 名的模型作为基础模型集。再通过主成分分析和比较实验方法进行预报因子优选, 获得最终集成预报模型。

1 资料与方法

1.1 资料

本文所用资料为 2014—2018 年每天 02、05、08、11、14、17、20、23 时(北京时, 下同)青岛小麦岛海洋站观测资料。青岛小麦岛海洋站为自然资源部标准海洋站, 位于青岛市区沿岸中段(36°03'N, 120°25.5'E), 离岸约 150 m(见图 1), 具有较好的代表性。观测要素包括: 能见度、风速、风向、气压、露点、气温、相对湿度、云量、降雨、海温、盐度、有效波高、有效波周期和海雾。经数据质量控制处理后, 共获得有效记录 14 600 余组。其中, 海雾观测为目测, 记录了每次海雾发生和消亡的时间段。

收稿日期: 2020-03-10; 修回日期: 2020-07-14

基金项目: 国家重点研发计划资助(2016YFC1401409)

[Foundation: The National Key Research and Development Plan, No. 2016YFC1401409]

作者简介: 高松(1980—), 男, 辽宁省盖州市人, 硕士, 高级工程师, 研究方向为海洋灾害预测、突发事件应急, E-mail: 41267946@qq.com

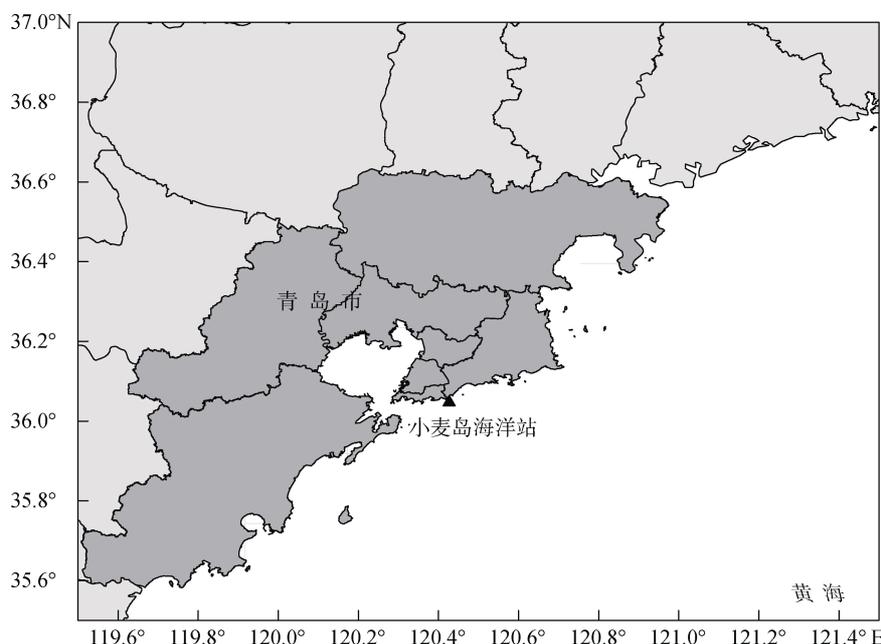


图 1 青岛小麦岛海洋站地理位置
Fig. 1 Location of the Xiaomaidao marine station in Qingdao

1.2 机器学习方法

机器学习方法是研究在计算机上从数据中产生“模型”的算法,即“学习算法”^[16]。简单地说,机器学习就是把无序的数据转换成有用的信息^[17]。从20世纪50年代发展至今,机器学习已经成为一个相当大的科学领域,与普通人的生活密切相关。例如在天气预报、能源勘探、环境监测、商业决策等方面都有广泛的应用^[16]。机器学习在气象领域的应用从20世纪80年代开始,目前已在短期雷暴、强对流天气、降水、海雾等方面的预报中取得了较好的应用成果^[14]。本文研究的海雾预报问题,属于机器学习中的监督学习分类问题。本文采用了较为经典的决策树、判别分析、回归树、支持向量机、K近邻、分类集成等6类算法。具体算法原理参见文献^[15]和^[16]。

1.3 集成预报方法

集合预报方法最早在二十世纪六七十年代提出,最初是为了克服数值模式初始场的不确定性对预报的影响问题,后来逐渐扩展到解决数值模式中物理过程参数化方案等不确定性和随机性对预报的影响问题^[17]。这类集合预报是从“单一”模式出发,通过考虑模式初始场、物理过程参数化方案等不确定性和随机性,得到“一群”预报值的方法。集合预报的

另一类可以称为“多模式集成”方法,也称为集成预报,是一种提高模式预测准确率的非常有效的后处理统计方法^[15]。该方法是通过同时采用“一群”模式,得到“一群”预报值。集合预报在本世纪初得到欧美许多发达国家气象界的高度重视,目前已在世界各国的预报机构中得到广泛的应用^[18]。本文采用的集合预报方法为“多模式集成”方法,即采用不同的机器学习方法,建立多个海雾预报模型,通过输入相同的预报因子数据,获得多个海雾预报结果,再通过对多个预报结果的统计分析确定最终的预报结果。

1.4 评分方法

本文采用TS评分^[19]、海雾预报正确率、非海雾预报正确率和总预报正确率四个指标,评估分析预报结果的优劣。其中,TS评分在对海雾预报评定中的表达式为:

$$T_s = \frac{N_c}{N_o + N_f - N_c}, \quad (1)$$

式中的 N_c 为报对海雾时次数, N_o 为观测海雾时次数, N_f 为预报海雾时次数。

海雾预报正确率为报对海雾时次数与预报海雾时次数之比。非海雾预报正确率为报对非海雾时次数与预报非海雾时次数之比。总预报正确率为报对海雾时次数、非海雾时次数之和与预报时次数之比。

2 青岛市区近岸海雾特征及水文气象条件

2.1 雾日数

青岛市区沿岸海雾具有高发性特点, 根据小麦岛海洋站 2014—2018 年海雾观测数据, 年均海雾发生天数为 55 d, 其中 2017(2016)年海雾发生天数最少(多)为 43(77) d。海雾发生存在着明显的季节变化, 雾

季一般从 4 月开始至 7 月结束, 占全年发生海雾发生天数的 64%, 其中以 6、7 月发生天数最多; 1、2、3、12 月占全年海雾发生天数的 35%; 8、9、10、11 月偶有海雾发生。从海雾出现的时段来看, 海雾最易发生在凌晨 02—08 时段, 11—17 时段相对较少, 与气温日变化规律相反(见图 2)。上述结论与马艳采用 1978—2007 年 30 年的气候资料分析的青岛市区的海雾发生规律基本一致^[4]。

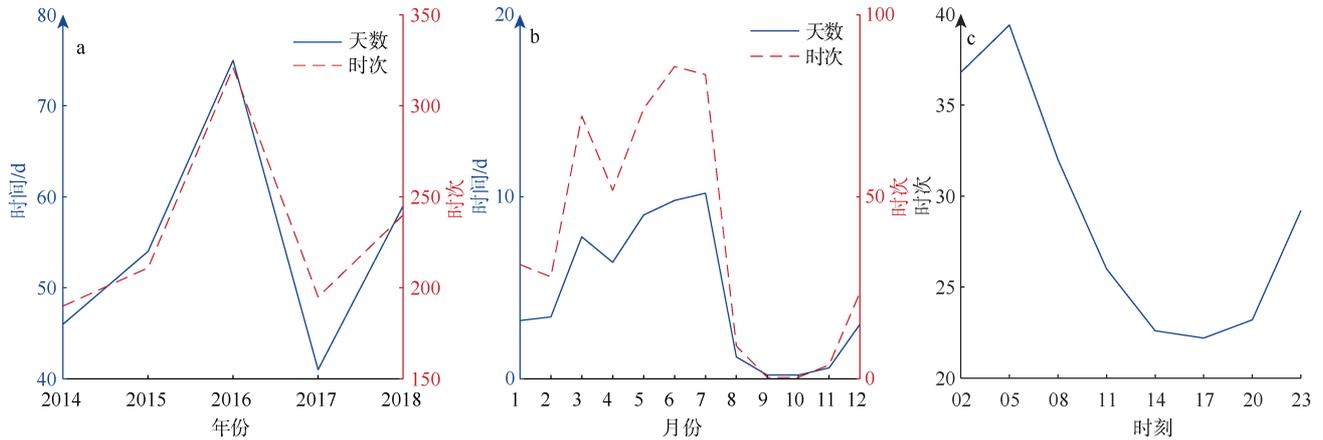


图 2 2014—2018 年海雾天数和次数变化曲线

Fig. 2 Cumulative lasting days and times of sea fog during 2014–2018

注: a: 逐年累积天数和时次数; b: 累年平均逐月天数和时次数; c: 累年平均逐时段时次数

2.2 气海温差

严格意义下的平流冷却雾是指在暖洋面上形成的暖湿气流输送到冷水面上, 低层空气温度下降至露点而成雾^[20]。青岛市区近岸发生海雾时, 气温为

2~28 °C, 海温为 4~27 °C。同时, 气海温差对海雾形成和演变有着重要作用^[21]。图 3 给出了海雾时次各月气温与海温差的分布, 2—8 月 69%(80%)的海雾发生在气海温差-0.5~2.5 °C(-0.5~4.0 °C), 符合

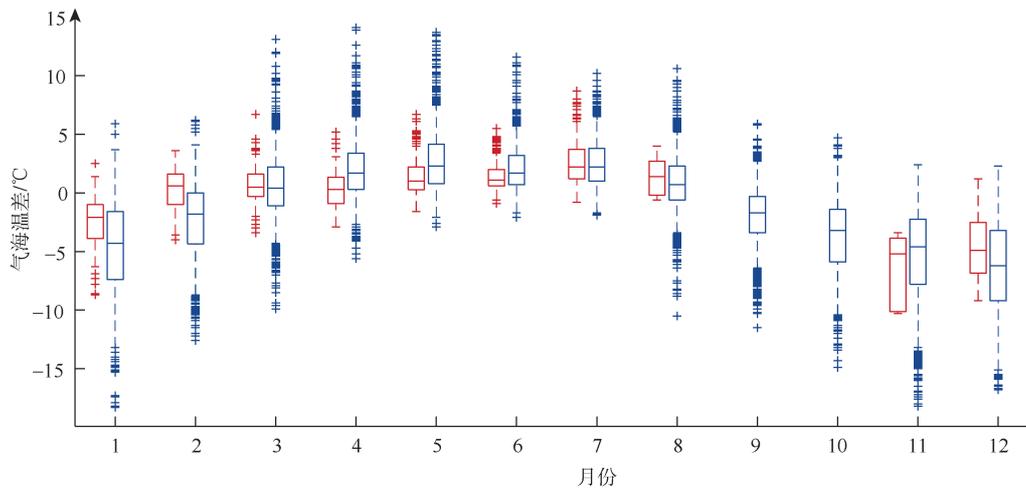


图 3 逐月气海温差分布盒须图

Fig. 3 Box plot of monthly differences between the air and sea temperatures

注: 红色: 海雾时次, 蓝色: 全时次。图中盒子中心标记表示中间值, 盒子的底部和顶部边缘分别表示第 25%和 75%的值。盒须延伸到数据极值, “+”符号单独绘制异常值

我国近海平流冷却雾成雾的条件^[1]。而 11—1 月的海雾发生时的气海温差为-1~10 °C，对于此类海雾的形成条件和机制仍需进一步研究。此外，虽然青岛市气温季节变化明显，各月海雾发生时的气温范围也不相同，但 95% 以上的海雾发生在月平均气温±5 °C 的范围内。这一规律也可作为数据样本剔除时的一个条件。

2.3 湿度和云量

水汽是海雾生成的物质基础，相对湿度大小及其分布是海雾能否生成的物理基础^[22]。统计表明，2—8 月的海雾发生时，相对湿度均在 80% 以上，气温露点差在-1~3 °C；11—1 月的海雾发生时，相对湿度在 75%~90%，气温露点差在 1~5 °C；海雾发生时的云量均在 7 层以上(图 4)。

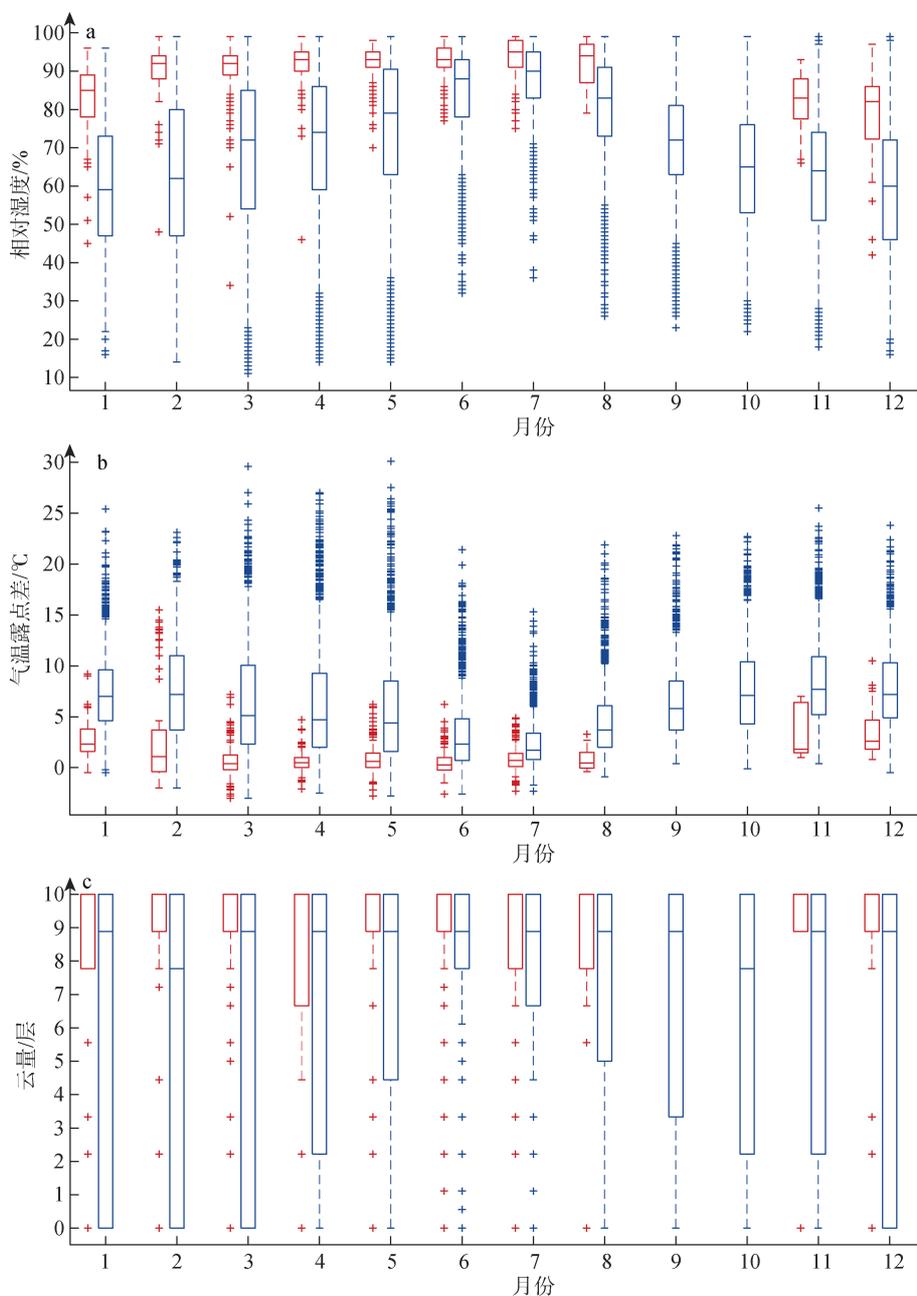


图 4 逐月相对湿度(a)、气温露点差(b)和云量(c)分布盒须图

Fig. 4 Box plot of monthly relative humidity (a), depression of the dew point (b), and cloud cover (c)

注：红色：海雾时次，蓝色：全时次

2.4 风

适宜的风向和风速将暖湿气流向冷水面输送是沿海海雾产生的重要条件^[23]。海雾发生时段小麦岛海域以东风为主, 约占 39%。其次是东南风, 约占 20%。南风 and 东北风, 约占 9%和 10%。北风和西北

风发生频率 6%和 9%, 主要发生在 1、2、12 月。从风速方面来看, 海雾主要发生在 3 级及以下的风力情况下, 其中 1 级风(0.3~1.6 m/s)约占 34%, 2 级风(1.6~3.4 m/s)约占 48%, 3 级风(3.4~5.5 m/s)约占 13%。4 级—6 级风约占 5%, 主要发生在 1、2、12 月(图 5)。

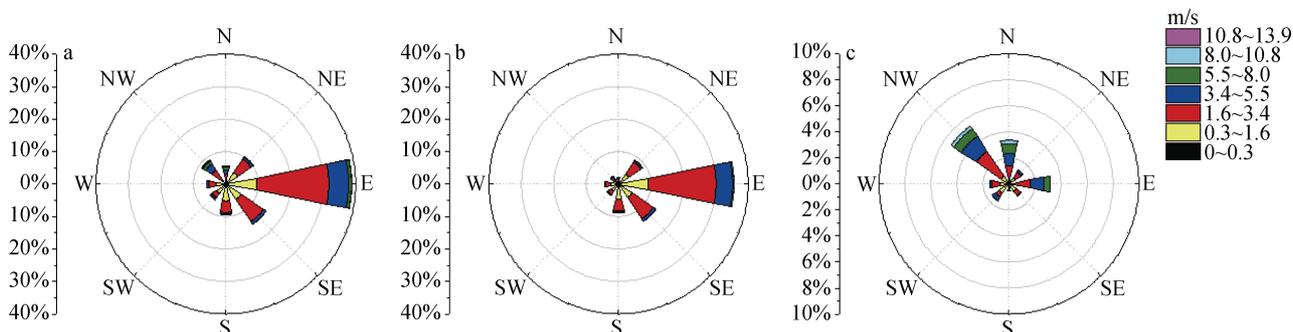


图 5 海雾发生时次海面风玫瑰图
Fig. 5 Wind rose charts for sea fog cases
注: a: 全年; b: 3—9 月; c: 1、2、12 月

3 预报模型与实验

3.1 数据样本构造

3.1.1 数据样本构建

海雾的发生属于低概率事件, 在总的 14 608 组观测数据中, 海雾发生时次数据为 1 155 组, 约占 8%。正如上文分析, 海雾发生时各变量的数值范围, 小于当月所有时次数据的数值范围。因此选择各变量每月海雾发生期间 99%概率阈值作为截断值, 剔除阈值之外的数据, 以减少原始数据中的无效数据。将处理后的剩余观测数据 4 117 组作为样本数据, 其中海雾发生时次数据 1 069 组, 约占 26%。然后, 将样本数据进行规范化处理后, 分为训练数据和检验数据两组。训练数据为 2014 年 1 月—2017 年 12 月的 3 208 组数据, 其中海雾发生时次数据 852 组, 约占 27%。检验数据为 2018 年 1 月—12 月的 909 组数据, 其中海雾发生时次数据 217 组, 约占 24%。

3.1.2 影响因子选取

考虑海雾发生所需的热力、动力、水汽和稳定性等 4 方面条件, 以及所能获得的有限的观测要素, 我们共选取了 18 个影响因子, 分为直接因子和间接因子两类。直接因子是指有观测仪器直接观测的要素, 分别为能见度(V_i)、风速(W_s)、风向(W_D)、气压(P_T)、露点(D_p)、气温(A_T)、相对湿度(R_c)、云量(C_i)、降雨(R_a)、海温(W_T)、盐度(S_a)、有效波高(H_s)、有效

波周期(H_p)等共 13 个。间接因子是指由直接观测要素之间或直接观测要素自身不同时刻之间的关系引申而来的间接要素, 分别为 3 小时变压(P_{r3})、气温 24 小时变温(A_{T24})、海温 24 小时变温(W_{T24})、气温温差(A_T)、气温露点差(A_D)5 个。实验目标因子为海雾是否发生。

3.2 模型选择

作为一款为科学和工程计算而专门设计的大型交互式计算工具软件, MATLAB 集数据处理、数值计算、算法开发、可视化及计算机仿真为一体。本文采用 MATLAB 2018b 提供的分类学习器(classification learner app)进行模型的搭建和训练^[25]。具体分类模型包括: 决策树、判别分析、回归树、支持向量机、K 近邻、分类集成等 6 类 23 个模型, 模型设置见表 1。通过对训练样本分类的综合评分分析, 确定了评分前 8 位的模型作为预报基础模型, 分别为 Model-02、Model-06、Model-07、Model-08、Model-11、Model-19、Model-20、Model-23。其中决策树模型 1 个、回归树模型 1 个、支持向量机 3 个、分类集成模型 3 个。

3.3 调训实验

3.3.1 预报因子分析

为了进一步简化基础模型, 优选影响因子, 去掉多余的影响因子。文本采用主成分分析^[24]对 18 个影响因子进行主成分分析, 通过对评估各影响因子

表 1 机器学习模型参数设置

Tab. 1 Functions and parameters of machine learning models

| 模型名 | 算法 | 函数名 | 主要参数设置 |
|----------|--|--------------|--|
| Model-01 | | | SplitCriterion: 'gdi'; MaxNumSplits: 100 |
| Model-02 | 决策树 | fitctree | SplitCriterion: 'gdi'; MaxNumSplits: 20 |
| Model-03 | | | SplitCriterion: 'gdi'; MaxNumSplits: 4 |
| Model-04 | | | |
| Model-05 | 判别分析 | fitcdiscr | DiscrimType: 'quadratic' |
| Model-06 | 回归树 | fitglm | Distribution: 'binomial' |
| Model-07 | | | Kernel function: 'linear' |
| Model-08 | | | Kernel function: 'quadratic' |
| Model-09 | 支持向量机 | fitcsvm | Kernel function: 'cubic' |
| Model-10 | | | Kernel function: 'Gaussian'; KernelScale: 1.1 |
| Model-11 | | | Kernel function: 'Gaussian'; KernelScale: 4.2 |
| Model-12 | | | Kernel function: 'Gaussian'; KernelScale: 17 |
| Model-13 | | | Distance: 'Euclidean'; NumNeighbors: 1; DistanceWeight: 'Equal' |
| Model-14 | Distance: 'Euclidean'; NumNeighbors: 17; DistanceWeight: 'Equal' | | |
| Model-15 | K 近邻 | fitcknn | Distance: 'Euclidean'; NumNeighbors: 100; DistanceWeight: 'Equal' |
| Model-16 | | | Distance: 'Cosine'; NumNeighbors: 10; DistanceWeight: 'Equal' |
| Model-17 | | | Distance: 'Minkowski' NumNeighbors: 10; DistanceWeight: 'Equal' |
| Model-18 | | | Distance: 'Euclidean' NumNeighbors: 10; DistanceWeight: 'SquaredInverse' |
| Model-19 | | | Method: 'AdaBoostM1' |
| Model-20 | | | Method: 'Bag' |
| Model-21 | 分类集成 | fitcensemble | Method: 'Subspace' |
| Model-22 | | | Method: 'Subspace' |
| Model-23 | | | Method: 'RUSBoost' |

的贡献, 确定最终的输入模型的预报因子, 同时也达到降维的目的。利用 MATLAB 2018b 的主成分分析函数对影响因子进行主成分分析, 结果显示前 5 个主成分贡献率分别为 39%、17%、11%、9%和 5%, 累计贡献率已经超过 80%。如表 2 所示, 通过对 5 个主成分中各影响因子的权重分析, 能见度、风向、气压、露点、气温、海温、气温露点差、气海温差、相对湿度、云量、气温 24 小时变温等 11 个影响因子的权重较高。其中除了上节分析中与海雾生成密切相关的气温、海温、气海温差、相对湿度、露点、气温露点差、风向因子外, 还包括了能见度、气压和气温 24 小时变温 3 个影响因子。其中, 海雾的产生将导致能见度的降低, 两种之间有密切的关系。比较稳定的大气环流背景场是海雾产生的前提条件^[1], 而气压变化和气温 24 小时变温是大气环流稳定度的一种表现, 气压和气温变化越小越有利于海雾的生成。此外, 虽然风速(W_S)权重较小, 但从上节的分析可以看出海雾的发生与风速有较强的相关性, 同时风速和风向一般都是同时使用。因此, 最后确定的强影响因子为 12 个。

表 2 前 5 个主成分的权重矩阵

Tab. 2 Weight matrix of the first five principal components

| 主成分 | 1 | 2 | 3 | 4 | 5 |
|-----------|-------|-------|-------|-------|-------|
| V_i | 0.01 | -0.21 | -0.18 | 0.85 | -0.25 |
| W_D | -0.07 | -0.03 | 0.97 | 0.14 | -0.10 |
| W_S | -0.09 | 0.01 | -0.06 | 0.08 | -0.18 |
| P_r | -0.35 | 0.02 | -0.02 | 0.06 | 0.09 |
| P_{r3} | 0.00 | -0.02 | 0.00 | -0.03 | 0.05 |
| D_p | 0.48 | -0.06 | 0.03 | -0.04 | -0.02 |
| A_T | 0.50 | -0.14 | 0.04 | 0.08 | 0.11 |
| A_{T24} | -0.07 | -0.10 | 0.03 | 0.11 | 0.71 |
| W_T | 0.52 | -0.11 | 0.07 | 0.11 | 0.24 |
| W_{T24} | 0.04 | -0.02 | -0.02 | 0.02 | -0.01 |
| A_D | -0.14 | -0.13 | 0.00 | 0.21 | 0.23 |
| A_w | 0.17 | -0.08 | -0.03 | -0.02 | -0.12 |
| S_a | -0.01 | -0.02 | -0.05 | -0.12 | 0.25 |
| H_S | 0.05 | 0.02 | -0.02 | 0.04 | -0.28 |
| H_p | 0.04 | 0.00 | 0.02 | 0.00 | -0.06 |
| R_e | 0.20 | 0.13 | 0.00 | -0.24 | -0.29 |
| C_1 | 0.14 | 0.93 | 0.00 | 0.29 | 0.14 |
| R_a | 0.01 | 0.01 | 0.00 | -0.01 | -0.01 |
| 贡献率/% | 38.55 | 16.94 | 11.10 | 9.27 | 4.97 |
| 累计贡献率/% | 38.55 | 55.49 | 66.60 | 75.87 | 80.84 |

3.3.2 调训实验设计

根据影响因子贡献率分析, 将预报因子分为 3 种情况, 进行预报模型的调训试验设计(见表 3)。实验 1(Test1): 采用全部 18 个影响因子作为预报因子。实验 2(Test2): 采

用 12 个强影响因子作为预报因子。实验 3(Test3): 采用主成分累积贡献率达 95% 的前 10 个主成分作为预报因子。三组试验均采用上文确定 8 个预报基础模型进行检验数据期间(2018 年 1—12 月)的海雾发生预报。

表 3 调训模型实验设计
Tab. 3 Setup of training models

| 实验名 | 模型名 | 预报因子 | 目标因子 | 训练数据和检验数据 |
|-------|---|---|-----------------|------------------------------|
| Test1 | Model-02, Model-06, Model-07, Model-08, | $P_r, D_p, A_T, W_T, A_W, R_e, V_i, W_D, W_S, C_1, A_{T24}, A_D, P_{r3}, S_a, H_s, H_p, R_a, W_{T24}$ | | 训练数据: 2014 年 1 月—2017 年 12 月 |
| Test2 | Model-11, Model-19, Model-20, Model-23 | $P_r, D_p, A_T, W_T, A_W, R_e, V_i, W_D, W_S, C_1, A_{T24}, A_D$ | Fog(无: 0; 有: 1) | 检验数据: 2018 年 1—12 月 |
| Test3 | | 前 10 个主成分 | | |

3.3.3 结果分析

从 3 组调训实验的预报结果来看(表 4), Test1 实验预报结果最优; Test2 实验预报结果略差, 平均 TS 得分达到 Test1 实验的约 97%, 且个别模型略优于 Test1 实验模型结果。可以看出强影响因子在海雾的预报中起到重要的作用, 其他影响因子则在海雾的预报中可以忽略。Test3 实验结果也有很好的预报

精度, 但其在预报准确率主要表现在对非雾情况预报较好, 在海雾预报正确率方面明显低于 Test1 实验和 Test2 实验。综合考虑预报结果评估情况和预报因子的数量, 选择 Test2 实验作为集成预报模型集。

表 4 调训实验预报结果比较

Tab. 4 Comparison of the forecasting results of training models

| 模型名 | TS 得分 | | | 海雾预报正确率 | | |
|----------|----------|-------|-------|---------|-------|-------|
| | Test1 | Test2 | Test3 | Test1 | Test2 | Test3 |
| Model-02 | 0.613 | 0.617 | 0.553 | 0.728 | 0.754 | 0.606 |
| Model-06 | 0.677 | 0.658 | 0.661 | 0.784 | 0.772 | 0.745 |
| Model-07 | 0.667 | 0.647 | 0.657 | 0.771 | 0.753 | 0.744 |
| Model-08 | 0.671 | 0.647 | 0.682 | 0.755 | 0.785 | 0.732 |
| Model-11 | 0.652 | 0.665 | 0.660 | 0.767 | 0.758 | 0.694 |
| Model-19 | 0.658 | 0.613 | 0.626 | 0.772 | 0.733 | 0.705 |
| Model-20 | 0.636 | 0.609 | 0.599 | 0.789 | 0.775 | 0.701 |
| Model-23 | 0.657 | 0.631 | 0.562 | 0.714 | 0.676 | 0.579 |
| 平均值 | 0.654 | 0.636 | 0.625 | 0.760 | 0.751 | 0.688 |
| 模型名 | 非海雾预报正确率 | | | 总预报正确率 | | |
| | Test1 | Test2 | Test3 | Test1 | Test2 | Test3 |
| Model-02 | 0.927 | 0.921 | 0.945 | 0.871 | 0.877 | 0.821 |
| Model-06 | 0.941 | 0.936 | 0.947 | 0.898 | 0.891 | 0.888 |
| Model-07 | 0.940 | 0.937 | 0.945 | 0.893 | 0.885 | 0.886 |
| Model-08 | 0.949 | 0.926 | 0.966 | 0.892 | 0.890 | 0.891 |
| Model-11 | 0.934 | 0.944 | 0.973 | 0.889 | 0.891 | 0.877 |
| Model-19 | 0.936 | 0.926 | 0.944 | 0.891 | 0.872 | 0.870 |
| Model-20 | 0.920 | 0.912 | 0.929 | 0.888 | 0.878 | 0.862 |
| Model-23 | 0.959 | 0.963 | 0.978 | 0.881 | 0.864 | 0.810 |
| 平均值 | 0.938 | 0.933 | 0.953 | 0.888 | 0.881 | 0.863 |

3.4 集成预报方法

多模式集成方法是一种提高模式预测准确率非常有效的后处理统计方法^[18]。从表 5 中可以看出, 当集成预报模型集中预报出海雾的模型数 ≥ 5 个时, TS 评分最优, 且海雾预报正确率相对较高。因此, 最终确定出雾模型数 ≥ 5 个, 作为预报海雾发生的判断条件。最终 2018 年海雾集成预报的 TS 评分为 0.643, 略优于 Test2 实验的 TS 评分平均值 0.636。集成预报的海雾预报正确率为 0.783, 同样优于 Test2 实验的平均值 0.751。与高荣珍等^[9]采用决策树方法的青岛

表 5 不同集成样本集预报结果比较

Tab. 5 Comparison of the forecasting results of different model samples

| 出雾模型数 | 正确数 | 误报数 | 漏报数 | TS 评分 | 海雾预报正确率 |
|----------|-----|-----|-----|-------|---------|
| 8 | 144 | 23 | 77 | 0.590 | 0.862 |
| ≥ 7 | 160 | 34 | 61 | 0.627 | 0.825 |
| ≥ 6 | 166 | 40 | 55 | 0.636 | 0.806 |
| ≥ 5 | 173 | 48 | 48 | 0.643 | 0.783 |
| ≥ 4 | 180 | 61 | 41 | 0.638 | 0.747 |
| ≥ 3 | 191 | 76 | 30 | 0.643 | 0.715 |
| ≥ 2 | 196 | 90 | 25 | 0.630 | 0.685 |
| ≥ 1 | 208 | 106 | 13 | 0.636 | 0.662 |

海雾预报模型的预报准确率基本相当,表明本文方法对于海雾预报业务工作具有很好的参考意义。

4 结论与展望

本文利用2014—2018年青岛市区近岸的小麦岛海洋站观测资料,对青岛市区近岸的海雾特征及海雾生成的水文气象条件进行了分析,探讨了采用机器学习方法建立海雾集成预报模型的可行性,形成了如下的结论:

(1) 青岛市区近岸海雾以平流冷却雾为主。主要发生在4—7月,占全年的发生天数的64%;1、2、3、12月占全年的发生天数的35%,8、9、10、11月偶有海雾发生。海雾最易发生在凌晨02—08时段,11—17时段相对较少。

(2) 青岛市区近岸海雾生成的水文气象条件特征,95%以上的海雾发生在月平均气温 $\pm 5^{\circ}\text{C}$ 的范围内,80%以上的海雾发生在气海温差在 $-0.5\sim 4^{\circ}\text{C}$ 时。80%以上的海雾发生在气温露点差在 $-1\sim 3^{\circ}\text{C}$ 时。95%以上的海雾发生在1~3级风。海雾发生时次风向以东风和东南风为主,分别约占39%和20%。

(3) 通过主成分分析发现了与海雾发生密切相关的12个强影响因子,具体为能见度、风向、风速、气压、露点、气温、海温、气温露点差、气海温差、相对湿度、云量、气温24小时变温。这些强影响因子基本体现了海雾发生所需的热力、动力、水汽和稳定性条件,具有比较明确的物理意义。

(4) 基于机器学习建立的海雾集成预报方法,对2018年海雾预报的TS评分约为0.64,海雾预报正确率约为0.783,具有较好的预报能力。集合预报方法较单一模型预报,在预报结果的稳定性和准确性上均有一定的提升。

通过本文的研究可以看出,基于机器学习的海雾集成预报方法确实可以充分地挖掘预报因子和目标因子之间的重要关系,实现对海雾的有效预报,对于海雾预报业务工作具有很好的参考意义,且具有广泛的拓展空间和应用前景。但是,不同海域海雾的生成条件不同,因此,本文预报方法推广使用到其他海域仍需要重新建立模型。另外,对于青岛冬季海雾的生成机制研究和预报精度的进一步提高,仍将是今后研究的难点和重点。

参考文献:

[1] 王彬华. 海雾[M]. 北京: 海洋出版社: 1983, 352.

Wang Binhua. Sea Fog[M]. Beijing: China Ocean Press: 1983, 352.

[2] Fu G, Zhang S P, Gao S H, et al. Understanding of Sea Fog over the China Seas[M]. Beijing: China Meteorological Press, 2012.

[3] Yi L, Zhang S P, Thies B, et al. Spatio-temporal detection of fog and low stratus top heights over the Yellow Sea with geostationary satellite data as a precondition for ground fog detection—A feasibility study[J]. Atmospheric Research, 2015, 151: 212-223.

[4] 马艳, 郝燕, 王颖, 等. 青岛地区海雾分布特征及风险评估[J]. 中国海洋大学学报(自然科学版), 2014, 44(11): 11-15.

Ma Yan, Hao Yan, Wang Ying, et al. Characteristics of sea fog and risk assessment for fog disaster in Qingdao[J]. Periodical of Ocean University of China, 2014, 44(11): 11-15.

[5] 史得道, 吴振玲, 高山红, 等. 海雾预报研究综述[J]. 气象科技进展, 2016, 6(2): 51-57.

Shi Dedao, Wu Zhenlin, Gao Shanhong, et al. A summary of research on sea fog forecasting[J]. Advances in Meteorological Science and Technology, 2016, 6(2): 51-57.

[6] 杨悦, 高山红. 黄前海雾 WRF 数值模拟中垂直分辨率的敏感性研究[J]. 气象学报, 2016, 74(6): 974-988.

Yang Yue, Gao Shanhong. Sensitivity study of vertical resolution in WRF numerical simulation for sea fog over the Yellow Sea[J]. Acta Meteorologica Sinica, 2016, 74(6): 974-988.

[7] 饶莉娟. YSU 与 MYNN 边界层方案的黄前海雾模拟效果比较[D]. 青岛: 中国海洋大学, 2014.

Yao Lijuan. Comparison of the effect of YSU and MYNN PBL schemes on the modeling of the Yellow Sea fogs[D]. Qingdao: Ocean University of China, 2014.

[8] 胡波, 杜惠良, 郝世峰, 等. 一种统计技术结合动力释用的沿海海雾预报方法[J]. 海洋预报, 2014(5): 82-86.

Hu Bo, Du Huiliang, Hao Shifeng, et al. A forecast method of coastal sea fog based on the combination of statistical technique and dynamical interpretation[J]. Marine Forecasts, 2014(5): 82-86.

[9] 高荣珍, 李欣, 任兆鹏, 等. 青岛沿海海雾决策树预报模型研究[J]. 海洋预报, 2018, 33(4): 80-87.

Gao Rongzhen, Li Xin, Ren Zhaopeng, et al. Study of the sea fog prediction by classification and regression tree(CART) analyses in Qingdao Coastal area[J]. Marine Forecasts, 2018, 33(4): 80-87.

[10] 江海英, 张韧, 刘科峰, 等. 青岛海雾的自适应神经模糊推理系统建模[J]. 海洋科学, 2005, 29(2): 26-31.

Jiang Haiying, Zhang Ren, Liu Kefeng, et al. Anfis modeling on Qingdao sea fog[J]. Marine Sciences, 2005,

- 29(2): 26-31.
- [11] 徐志鹏, 张苏平, 衣立, 等. 青岛近岸能见度逐小时分级预报模型初探[J]. 海洋湖沼通报, 2018, (1): 9-17.
Xu Zhipeng, Zhang Suping, Yi Li, et al. A tentative establishment of a hourly classification prediction model for Qingdao coastal water visibility[J]. Transactions of Oceanology and Limnology, 2018, (1): 9-17.
- [12] 江敦双, 张苏平, 陆惟松. 青岛海雾的气候特征和预测研究[J]. 海洋湖沼通报, 2008, (3): 7-12.
Jiang Dunshuang, Zhang Suping, Lu Weisong. The climatic characteristics and forecast study of the sea fog in Qingdao[J]. Transactions of Oceanology and Limnology, 2008, (3): 7-12.
- [13] 任兆鹏, 高荣珍, 时晓曦, 等. 2018年6月青岛近岸一次海雾导致能见度变化的成因分析[J]. 海洋湖沼通报, 2019, (4): 81-90.
Ren Zhaopeng, Gao Rongzhen, Shi Xiaomeng, et al. Analysis of visibility variation caused by a sea fog event near Qingdao coast in June 2018[J]. Transactions of Oceanology and Limnology, 2019, (4): 81-90.
- [14] 朱玲, 吴心玥. 人工智能在气象领域的应用述评[J]. 广东气象, 2019, 41(1): 35-39.
Zhu Ling, Wu Xinyue. Review on the application of artificial intelligence in meteorological field[J]. Guangdong Meteorology, 2019, 41(1): 35-39.
- [15] 周志华. 机器学习[M]. 北京: 清华大学出版社, 2016.
Zhou Zhihua. Machine Learning[M]. Beijing: Tsinghua University Press, 2016.
- [16] Harrington P, 著. 李锐, 李鹏, 曲亚东, 等译. 机器学习实战[M]. 北京: 人民邮电出版社, 2013.
Harrington P. trans. Li Rui, Li Peng, Qu Yadong, et al. Machine learning in action[M]. Beijing: Posts & Telecom Press, 2013.
- [17] 杜钧. 集合预报的现状和前景[J]. 应用气象学报, 2002, 13(1): 16-28.
Du Jun. Present situation and prospects of ensemble numerical prediction[J]. Journal of Applied Meteorological Science, 2002, 13(1): 16-28.
- [18] 智协飞, 彭婷, 李刚, 等. 多模式集成的概率天气预报和气候预测研究进展[J]. 大气科学学报, 2014, 37(2): 248-256.
Zhi Xiefei, Peng Ting, Li Gang, et al. Advances in multimodel ensemble probabilistic prediction[J]. Transactions of Atmospheric Sciences, 2014, 37(2): 248-256.
- [19] 牟惟丰. 预报评分方法述评和方案建议[J]. 气象, 1986, 12(2): 45-49.
Mu Weifeng. Review of forecast scoring method and proposal[J]. Meteorological, 1986, 12(2): 45-49.
- [20] 张苏平, 鲍献文. 近十年中国海雾研究进展[J]. 中国海洋大学学报(自然科学版), 2008, 38(3): 359-366.
Zhang Suping, Bao Xianwen. The main advances in sea fog research in China[J]. Periodical of Ocean University of China, 2008, 38(3): 359-366.
- [21] 王帅, 傅聘, 陈德林, 等. 2009年春季一次黄海海雾的观测分析及数值模拟. 大气科学学报, 2012, 35(3): 282-294.
Wang Shuai, Fu Dan, Chen Delin, et al. An observation and numerical simulation of a sea fog event over the Yellow Sea in the spring of 2009. Transactions of Atmospheric Sciences, 2012, 35(3): 282-294.
- [22] 杨悦, 高山红. 黄海海雾天气特征与逆温层成因分析[J]. 中国海洋大学学报(自然科学版), 2015, 45(6): 19-30.
Yang Yue, Gao Shanhong. Analysis on the synoptic characteristics and inversion layer formation of the Yellow Sea fog[J]. Periodical of Ocean University of China, 2015, 45(6): 19-30.
- [23] 徐燕峰, 陈淑琴, 戴群英, 等. 舟山海域春季海雾发生规律和成因分析[J]. 海洋预报, 2002, 19(3): 59-64.
Xu Yanfeng, Chen Shuqin, Dai Qunying, et al. Regularity and formation cause analyses of fog in Zhoushan sea area in spring[J]. Marine Forecasts, 2002, 19(3): 59-64.
- [24] 施能. 气象科研与预报中的多源分析方法[M]. 北京: 气象出版社, 2002.
Shi Neng. Multivariate Analysis Method in Meteorological Research and Forecast[M]. Beijing: China Meteorological Press, 2002.
- [25] MathWorks 软件公司. 关于分类学习器工具包原理的介绍[EB/OL]. [1994-2021]. <https://www.mathworks.com/help/stats/classification-learner-app.html>.

Ensemble forecast of sea fog in Qingdao coastal area based on machine learning

GAO Song^{1, 2}, XU Jiang-ling^{1, 2}, LIU Gui-yan^{1, 2}, BI Fan^{1, 2}, ZHANG Wei^{1, 2},
BAI Tao^{1, 2}

(1. North Sea Marine Forecast Center of State Oceanic Administration, Qingdao 266061, China; 2. Shandong Provincial Key Laboratory of Marine Ecological Environment and Disaster Prevention and Mitigation, Qingdao 266061, China)

Received: Mar. 10, 2020

Key words: sea fog; machine learning; ensemble forecast; Qingdao shelf

Abstract: Based on the 2014–2017 observation data of the Xiaomaidao marine station in Qingdao Shelf, ensemble forecast models of sea fog in the Qingdao coastal area were established using machine learning. The principal component analysis method was adopted to optimize the prediction factors. Results show that the prediction model consisting of 12 forecasting factors, namely, visibility, wind, air pressure, dew point, air temperature, sea temperature, depression of the dew point, air–sea temperature difference, relative humidity, cloud volume, and air–sea 24-h delayed difference—performed best. The threat score (TS) of this model for sea fog prediction was approximately 0.64 in 2018, and the accuracy rate was approximately 0.783. This method shows good performance for sea fog prediction. Moreover, it provides a new approach for the operational forecasting of sea fog.

(本文编辑: 丛培秀)