

浙江中南部海雾预报决策树模型研究

俞涵婷¹, 廖晨昕², 王可欣¹, 陈华忠³

(1. 椒江区气象局, 浙江 台州 318000; 2. 余杭区气象局, 浙江 杭州 311100; 3. 玉环市气象局, 浙江 玉环 317600)

摘要: 利用椒江大陈沿海航线上重要站点—江山岛、大陈站和头门岛 2015—2018 年 2—6 月的海雾历史观测资料和 NCEP/NCAR FNL 再分析资料, 从海雾的成因中找出大气与海雾的关系。分析的影响因子包括: 地面与高空温差 ($T_1\ 000\ \text{hPa}$ — $T_2\ \text{m}$)、($T_{925}\ \text{hPa}$ — $T_2\ \text{m}$)、($T_{850}\ \text{hPa}$ — $T_2\ \text{m}$)、($T_{975}\ \text{hPa}$ — $T_2\ \text{m}$) 和 $1\ 000\ \text{hPa}$ 相对湿度, 低层上升速度分析等。结论如下: (1) 暖湿气流本身强弱对大雾无影响, 温差才是形成大雾的重要因素, 近地层的逆温有利于大雾形成, 越低层逆温越强越有利于大雾形成; (2) 大雾形成时所需相对湿度基本集中在 90 以上, $950\ \text{hPa}$ 上较弱的上升速度利于大雾的形成, 散度条件对海雾的影响差别不大; (3) 通过训练集数据参与模型的建立, 模型整体的学习准确率为 0.85。将此测试集数据运用于 2019 年 2—6 月的大雾数据检验中, 成功率为 0.8。决策树模型建立的海雾判别流程可在业务中用于浙中南有无海雾的判别。

关键词: 浙江; 海雾; 分类与回归树; 决策树模型

中图分类号: P732.4 **文献标识码:** A **文章编号:** 1003-0239(2020)06-0096-06

1 引言

海雾是海上灾害性天气现象之一。它给近海航运、港区作业和渔业生产带来很大影响^[1]。它的复杂性给预报带来一定的困难。在中低纬度海域, 大范围海雾以平流冷却雾出现的机会最多, 雾也最浓, 成为海雾预报的重点^[2-3]。海雾一般多出现在低于 $20\ ^\circ\text{C}$ 的海区里, 高于 $20\ ^\circ\text{C}$ 时雾逐渐减少, 超过 $25\ ^\circ\text{C}$ 不再有雾^[4]。张苏平等^[5]通过天气研究预报模式 (Weather Research and Forecasting Model, WRF) 模拟得出: 黄海海温升高 ($0.5\sim 2\ ^\circ\text{C}$), 海雾面积相应减小; 升温程度越大, 面积减小的越多。在湿度较小的薄海雾区, 海温变化对稳定度影响稍大; 而在湿度较大的浓海雾区, 海温变化对稳定度的影响不大。这说明了下垫面温度、湿度、稳定度相互之间以及它们分别对海雾的影响。郑怡等^[6]对渤海沿岸的海雾分析认为: 沿岸相对湿度在 90% 以上, 近海相对湿度在 80% 以上, 低层大气存在逆温层, 有

利于渤海海雾的生成和发展。黄健等^[7]利用 2000—2008 年 1—5 月的资料, 采用分类与回归树方法 (Classification And Regression Trees, CART) 对海雾生成前 24 h 的海洋气象条件进行分析, 利用近地面预报因子, 如: $10\ \text{m}$ 风向和风速、 $1\ 000\ \text{hPa}$ 和 $850\ \text{hPa}$ 的温度、湿度和风场等建立海雾预报决策树模型, 对不同地域海雾预报的准确率在 73% 以上。史达伟等^[8]基于 5 种机器学习算法 (如线性支持向量机及多层神经网络等) 对特强浓雾开展诊断, 测试认为 CART 算法易于使用, 且效果较佳。高荣珍等^[9]肯定了用 CART 方法建立青岛沿海海雾预报模型的可行性, 并指出 $850\ \text{hPa}$ 风向在青岛沿海海雾决策树预报模型中也很关键。胡波等^[10]利用液态水含量和云滴粒子密度两个参数优化了大雾估算方法, 从而提高了大雾预报的准确率。

本研究通过平流雾的形成原因, 结合天气学分析方法, 利用浙中南沿海海域 2015—2018 年 2—6 月的海雾历史观测资料和对应的美国国家环境预

收稿日期: 2019-11-01; 修回日期: 2020-04-24。

基金项目: 台州市椒江区科技计划项目 (192032); 台州市气象科技计划项目 (TZ2018QN02)。

作者简介: 俞涵婷 (1989-), 女, 工程师, 本科, 主要从事海洋预报和天气预报工作。E-mail: 495234225@qq.com

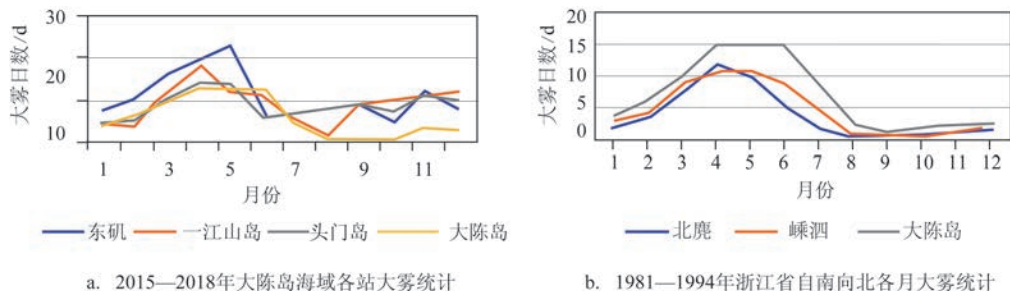


图1 浙中南近海雾季的月际变化

报中心/美国国家大气研究中心(National Centers for Environmental Prediction/National Center for Atmospheric Research, NCEP/NCAR) FNL再分析资料寻找海雾预报因子,计算因子具体阈值,并形成海雾预报决策树,以期对实际海雾预报业务有所助益。

2 海雾特征及环流背景

据统计,浙江省每年雾季雾日的多少,在南北地域上具有很强的相关性,且具有明显的系统性,即先南后北^[11]。这表明整个海雾产生的气候背景是一致的。浙中南近海雾季存在明显的月际变化,每年的2—6月是一年中雾日最多的月份(见图1)。张苏平等^[12]指出,浙江沿海海雾天气形势主要出现在入海变性冷高压西部、气旋和低压槽东部、副热带高压西部、静止锋或冷锋前部。平流雾没有日变化,可以维持多日。统计显示浙江省沿海以平流雾为主^[13-14]。

理论上来说,海雾形成于地面偏南风并略带风速的环境下,但从观测实况上看,2—6月任何风向下都有可能生成海雾,特别是在入海变性高压后部,弱形势场地面风一般为风速不大的偏北风情况下,生成雾的几率反而更多(见图2)。黄克慧等^[15]提到,当背景场(一般如925 hPa、850 hPa)是西南暖湿气流,且暖湿气流输送条件已具备时,不论地面是何风向都不影响雾的生成和维持。

3 研究方法 with 过程

基于上述事实,由于海雾形成时天气为一静稳状态,观测数据中温度、湿度和风场变化不大,风向、风速也难求规律;在预报上地面气象要素的得

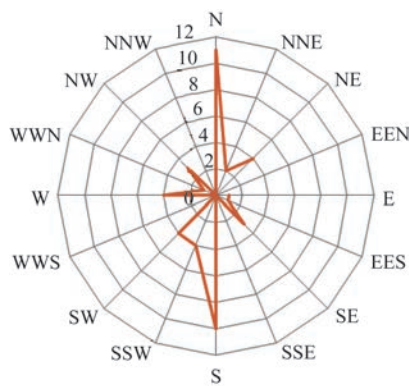


图2 一江山岛2017年大雾与风向的关系

出是基于模式产品,准确率不高。本文通过高空近地层资料探究,从模式已有输出因子(高低空要素场)入手,通过对平流冷却雾的本身定义推导,筛选海雾预报因子,运用CART方法得出大陈海域海雾预报决策树。

3.1 研究资料与数据基础

本文将能见度小于1 km记为有雾。选取浙江中南部大陈岛海域附近二岛2015—2018年2—6月的观测站地面能见度和天气现象资料进行研究。大陈岛虽有近60 a历史资料,但多年实际研究中发现大陈岛能见度观测仪设在海拔87.2 m的山坡顶,因海拔高度过高,大雾记录多为山顶层云遮盖所致,海面上实时能见度较好,观测的大陈岛大雾记录远大于真实海面大雾记录。鉴于大陈岛大雾记录有误差,本文以该海域附近站点一江山岛和头门岛雾记录为准。此二岛距离较近,海拔高度都约在50 m左右。因头门岛能见度记录从2015年2月12日16时开始,经查大陈岛2015年2月1—12日都无大雾天气(大陈岛雾记录远多于其他二岛),因此

2015 年以头门岛资料为准。一江山岛记录从 2016 年开始,之后头门岛、一江山岛资料相互替换,并相应选取 2015—2018 年 2—6 月 NCEP 的 $1^{\circ} \times 1^{\circ}$ 的再分析资料,时间间隔为 6 h,通过插值方法将 NCEP 格点资料插值到各个站点。

3.2 候选预报因子的设计

黄辉军等^[16-17]用近地层温差因子改进的广东沿海海雾区域预报取得了不错的效果。Huang 等^[18]通过再分析资料和观测资料,分析了多个预报变量,例如低层风矢量与暖湿平流、低层基本气象要素^[19]等因子作为表征海雾形成时所需的天气形势和外界环境。在垂直方向的温差因子中,选取了 925 hPa 与 1 000 hPa 的温度差值,以及 1 000 hPa 与地面 2 m 的温度差值,用来反映有雾时近地层的温度特性。研究证实了近地层 1 000 hPa 与地面 2 m 的温差是有效的预报变量因子之一。本研究从平流雾定义出发:当暖湿空气平流到较冷的下垫面上,下部冷却形成的雾。我们用低层温度与地面 2 m 温度的温差值来寻找本地海雾形成时的温差阈值,并用 2015—2018 年 2—6 月各层温差值统计大雾的结果。同时,由于大雾形成需要稳定的天气形势,任何冷空气的侵入^[15]及近地层降水对流等新系统的入侵都会破坏大雾的天气形势。所以我们还需考虑大雾形成时的天气背景:近地层的温度结构、地面湿度、上升速度和散度等,并通过这些数值预报结果确定大雾的关键预报因子,并在实际预报中做出判断。

3.3 高空气象要素与大雾发生特征的统计分析

由于暖湿气流经过冷地表面,本文分别选取了 850 hPa、925 hPa、950 hPa、975 hPa、1 000 hPa 与地面 2 m 的温度差值;湿度是大雾形成的必要条件,选取了地面 1 000 hPa 相对湿度(1 000 U)、925 hPa 相对湿度(925 U)和 850 hPa 相对湿度(850 U);根据天气的稳定程度,适当的逆温层也给大雾提供了稳定条件:选取了 850 hPa 与 925 hPa 的温差(T850 hPa—T925 hPa)、925 hPa 与 975 hPa 的温差(T925 hPa—T975 hPa)和 950 hPa 与 1 000 hPa 的温差(T950 hPa—T1 000 hPa);同时我们还选取了前后时次的 925 hPa 温差(T925 hPa—T925 hPa)代表暖湿气流本身强度

对大雾的影响;从垂直速度场观察,大雾形成后有微弱的上升运动,因此选取了 950 hPa 的上升速度 950 vvel;大雾形成时的静稳物理量:950 hPa 散度为 D950 hPa。

图 3 和图 4 是根据 2015—2018 年资料统计的核密度图,代表大陈航线上不同变量在有无大雾时的分布情况,纵坐标是密度,即位于此点的数据占全体数据的比例。图 3a—e 定量反映了暖湿气流对大雾的影响,其中各个高度与 2 m 地面的温差与大雾都约为中等强度相关,各个高度与 2 m 地面的温差与大雾相关性相近。图中也显示有雾时比无雾时气温温差向右位移了 2~5 $^{\circ}\text{C}$,有的甚至为正值,说明暖湿气流与地面温度形成的温差是大雾生成的重要因素。图 3f 是 925 hPa 前后时次的温差,它与大雾的相关性弱说明暖湿气流本身的强弱对大雾无影响,这也更说明温差才是形成大雾的重要因素。图 4c 的相关性大于 4a,说明近地层的逆温有利于大雾形成,越低层逆温越强越有利于大雾形成;大雾形成时需要 1 000 hPa 的相对湿度基本集中在 90 以上,相关性为强相关(见图 4d),而大雾与 925 hPa 和 850 hPa 的相对湿度的相关性远弱于 1 000 hPa 的相对湿度;950 hPa 上较弱的上升速度利于大雾的形成(见图 4e),950 hPa 上升速度与大雾的相关性高于 700 hPa 上升速度。对于平流雾来说散度等物理量影响较小,不论是 925 hPa 还是 950 hPa,散度条件对海雾的影响差别不大。

4 CART 决策树算法运用及检验

CART 算法是机器学习中用来处理分类和回归问题的方法^[20],适用于离散型变量和连续型变量的分类。CART 决策树的生成就是递归地构建二叉决策树的过程。对分类树来说,CART 用 Gini 系数最小化准则进行特征选择来生成二叉树。

基于 CART 决策树算法对大雾天气建立模型,模型的输入变量为 2015—2018 年大陈航线以上气象要素。将之前处理好的训练集输入 CART 算法,经过多次交叉检验得到决策树。模型见图 5。

决策树中基本反映了前面分析的分布情况,并且更加明确了湿度是大雾形成的必要条件,沿海海面湿度需大于 90 才能执行该模型;其次是近地层的

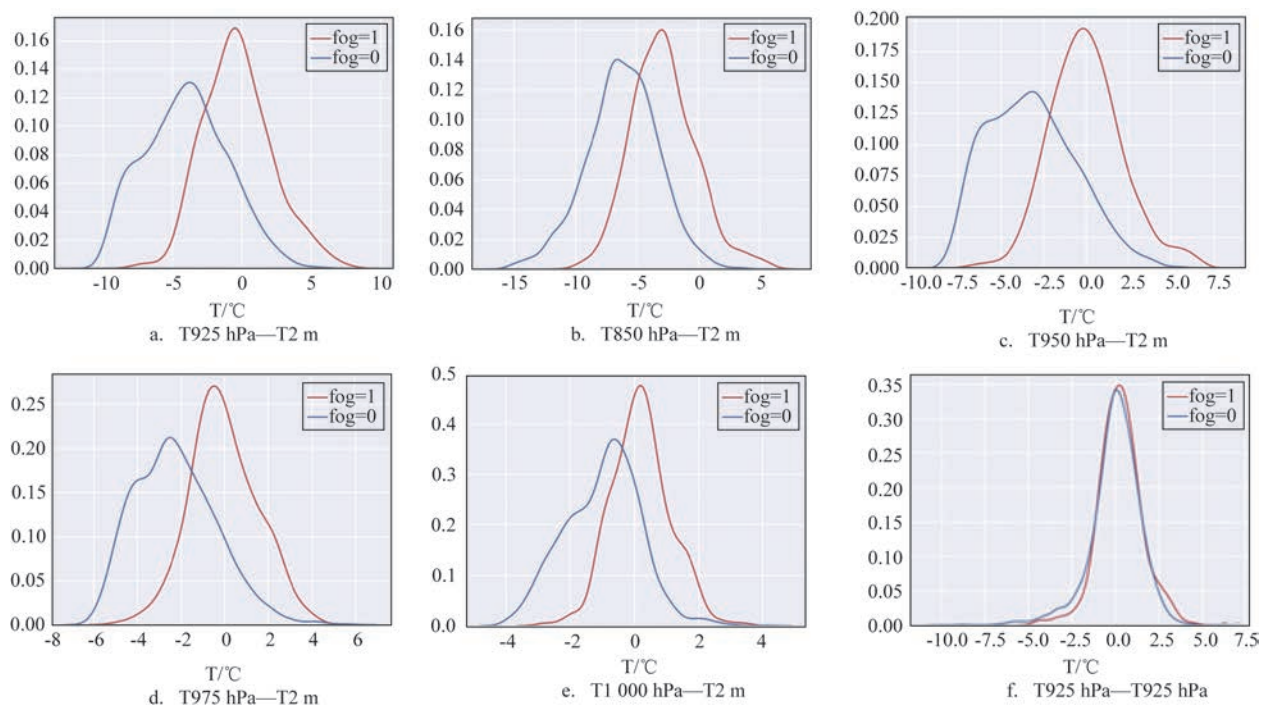


图3 2015—2018年大陈航线不同高度与2 m地面温差和925 hPa前后
 时段温差与有无大雾的分布

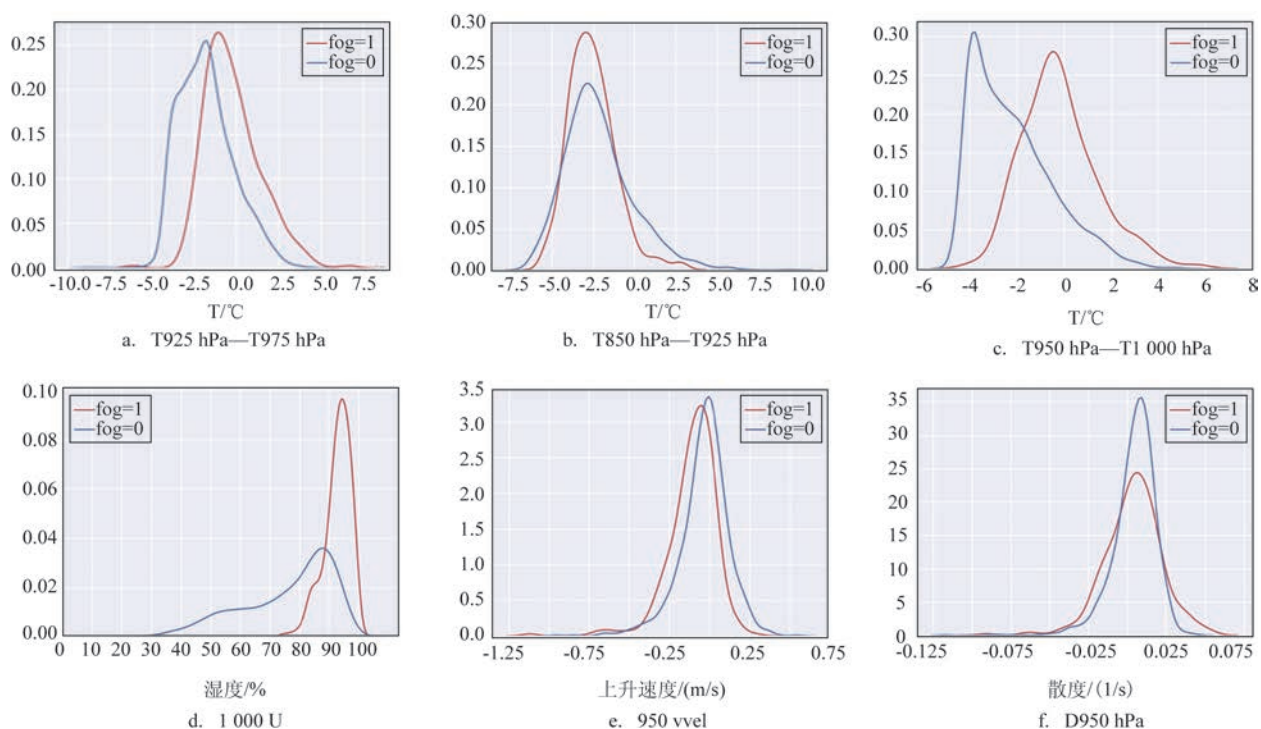


图4 2015—2018年大陈航线不同高度温度差、1 000 hPa湿度、950 hPa上升速度和950 hPa
 散度与有无大雾的分布

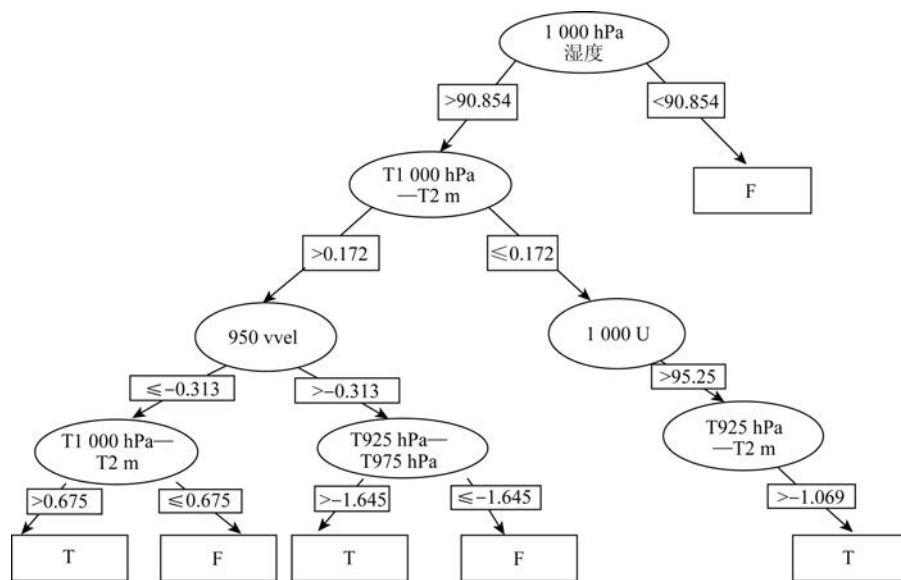


图5 2015—2018年椒江沿海航线大雾气象观测要素特征诊断决策树模型

温差情况;另外,低层上升速度、低层逆温等都对大雾形成有影响。训练集数据模型整体学习准确率为0.85。将此测试集数据运用于2019年2—6月的大雾数据检验中,成功率为0.8,模型达到不错的泛化能力。此模型右支也相似于张苏平等^[12]的模拟结果,湿度越大时对温度要求越低。左支则更考虑上升运动和温差的相互影响。该模型说明了湿度是海雾出现的必要条件,而海雾形成和维持是低层相关因子相互作用、相辅相成的结果。

5 结论与缺陷

利用大陈岛海域各站点2015—2018年2—6月的海雾历史观测资料和NCEP/NCAR FNL再分析资料,通过海雾基本定义和形成所需的气象环境条件,选取高空气象要素对海雾预报进行分析研究。结果表明:暖湿气流本身强弱对大雾无影响,温差才是大雾形成的重要因素,近地层的逆温有利于大雾的形成,越低层逆温越强越有利于大雾形成;大雾形成时所需相对湿度基本集中在90以上;950 hPa上较弱的上升速度利于大雾的形成;散度条件对海雾的影响差别不大。利用训练集数据参与模型建立,模型整体的学习准确率为0.85。将此测试集数据运用于2019年2—6月的大雾数据检验中,成功率为0.8。但此模型仍有一些缺陷:

(1)数据是个连续的过程:当冷空气已下,风向突转,雾即已消散,但此时温度和湿度却还没及时降下来,数值仍高居不下;虽然经过个别人为剔除,但仍有少数出现与结论相悖的情况,会影响模型判别。

(2)由于天气影响要素众多,选取要素有限,并没有选取足够多的天气要素来对大雾进行影响分析和判别。

(3)本文运用的数据为再分析资料,与实况有一定误差,模型结果还需在长期实际运用中进行测试。

(4)大雾数据不够:2 404个数据样本中只有283个大雾样本,有、无雾数据相差太大,影响模型判断。

(5)形成大雾有众多的充分条件,但大雾消散却有很多充要条件。在考虑大雾预报时也需要着重考虑这些不利影响,如冷空气的渗透即破坏大雾的稳定,强的上升速度则表示有强降水不会出现大雾天气等。

参考文献:

- [1] 施素芬,陈鸣星. 浙江中部沿海海雾短期预报方法[J]. 海洋预报, 1992, 9(1): 80-83.
- [2] 周发琇. 第四讲 海雾预报[J]. 海洋预报, 1989, 6(4): 71-78.
- [3] 周发琇. 第一讲 海雾及其分类[J]. 海洋预报, 1988, 5(1): 78-84.
- [4] 周发琇. 第三讲 海雾的水文气象特征[J]. 海洋预报, 1988, 5(4):

- 84-94.
- [5] 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, 2008, 38(3): 359-366.
- [6] 郑怡, 李冉, 史得道, 等. 渤海中西部近海与沿岸海雾的特征分析[J]. 海洋预报, 2016, 33(6): 74-80.
- [7] 黄健, 黄辉军, 黄敏辉, 等. 广东沿岸海雾决策树预报模型[J]. 应用气象学报, 2011, 22(1): 107-114.
- [8] 史达伟, 颜佳任, 王伟健, 等. 连云港大雾天气背景下的特强浓雾的人工智能本地化诊断研究[C]//第十五届长三角气象科技论文集. 2018: 468-471.
- [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, 2016, 33(4): 80-87.
- [10] 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, 31(5): 82-86.
- [11] 周发琇, 刘龙太. 第七节 海雾[J]. 山东海洋学院学报, 1986, 16(1): 114-127, 131.
- [12] 张苏平, 任兆鹏. 下垫面热力作用对黄海春季海雾的影响——观测与数值试验[J]. 气象学报, 2010, 68(4): 439-449.
- [13] 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(1): 59-64.
- [14] 李昀英, 王汉杰. 台湾海峡地区雾形成的天气类型分析[J]. 热带海洋, 2000, 19(4): 65-70.
- [15] 黄克慧, 张意权, 周功铤, 等. 浙南沿海海雾特征分析[J]. 浙江气象, 2017, 28(1): 18-22.
- [16] 黄辉军, 黄健, 刘春霞, 等. 用近地层温差因子改进广东沿海海雾区域预报[J]. 热带气象学报, 2013, 29(6): 907-914.
- [17] 黄辉军, 黄健, 刘春霞, 等. 用GRAPES模式输出变量因子作广东沿海海雾预报[J]. 热带气象学报, 2010, 26(1): 31-39.
- [18] Huang H J, Liu H N, Jiang W M, et al. characteristics of the boundary layer structure of sea fog on the coast of southern China [J]. *Advances in Atmospheric Sciences*, 2011, 28(6): 1377-1389.
- [19] 胡基福, 郭可彩, 鄢利农. 应用模式输出统计海雾出现判别预报[J]. 青岛海洋大学学报, 1996, 26(4): 439-445.
- [20] Breiman L, Friedman J H, Olshen R, et al. *Classification and regression trees*[M]. Belmont, Calif: Wadsworth International Group, 1984.

A decision-tree forecasting model of sea fog for the central and southern coastal area of Zhejiang Province

YU Han-ting¹, LIAO Chen-xin², WANG Ke-xin¹, CHEN Hua-zhong³

(1. Jiaojiang Meteorological Bureau, Taizhou 318000 China; 2. Yuhang Meteorological Bureau, Hangzhou 311100 China; 3. Yuhuan Meteorological Bureau, Yuhuan 317600 China)

Abstract: Based on the conventional observational data in Yijiangshan Island, Dachen Island and Toumen Island between February and June from 2015 to 2018 and the NCEP $1^{\circ} \times 1^{\circ}$ reanalysis data, the relationship between marine atmospheric conditions and sea fog is revealed in this paper. The influence factors analyzed include the temperature difference between 1 000 hPa, 975 hPa, 925 hPa, T850 hPa and the surface layer, relative humidity at 1 000 hPa and the ascending speed of the low level air. The results show that the strength of warm and moist advection itself has no effect on the fog and the temperature difference is an important factor in the formation of sea fog. The inversion of the near-surface temperature is conducive to sea fog. The stronger the inversion temperature is, the more favorable it is for the formation of sea fog. The relative humidity required for the formation of sea fog is basically above 90. The weaker ascending air speed at 950 hPa is more conducive to the formation of sea fog, while the divergence conditions have no significant influence to sea fog. A decision-tree model is established by dataset training with an overall learning accuracy rate of 0.85. The model is validated using the sea fog observation from February to June 2019, which reveals a success rate of 0.8. Therefore, the decision-tree model can be operationally applied to identify the sea fog events in the central and southern coastal area of Zhejiang province.

Key words: Zhejiang; sea fog; Classification and Regression Trees; decision-tree mode