

黄健, 黄辉军, 黄敏辉, 等. 广东沿岸海雾决策树预报模型. 应用气象学报, 2011, 22(1): 107-114.

## 广东沿岸海雾决策树预报模型

黄 健<sup>1)\*</sup> 黄辉军<sup>1)</sup> 黄敏辉<sup>2)</sup> 薛登智<sup>2)</sup> 毛伟康<sup>1)</sup> 白玉洁<sup>3)</sup>

<sup>1)</sup> (中国气象局广州热带海洋气象研究所, 广州 510080)

<sup>2)</sup> (广州中心气象台, 广州 510080) <sup>3)</sup> (广东省气象局培训中心, 广州 510080)

### 摘 要

利用汕头、珠海和湛江地面观测站 2000—2008 年 1—5 月的海雾历史观测资料和 NCEP/NCAR FNL 再分析资料, 采用分类与回归树 (CART) 方法对海雾及其生成前 24 h 的海洋气象条件进行分类分析, 建立了海雾决策树预报模型; 并根据现有的海雾理论认识, 对海雾预报规则包含的物理意义进行讨论。10 次交叉检验的结果表明: 采用 CART 方法建立的海雾决策树预报模型有较好的预报性能, 对广东沿岸海雾的预报准确率总体上可达到 73% 以上。根据决策树预报模型建立的海雾预报判别流程, 可在业务工作中直接用于有雾/无雾判别。海雾预报判别流程同时也具有明确的物理意义, 能够较好地反映水汽与海表冷却条件对平流冷却雾形成的重要性, CART 方法可作为海雾业务预报的有效建模工具。

**关键词:** 广东沿岸海雾; 分类与回归树; 决策树预报模型; 判别流程

### 引 言

海雾是海洋大气边界层中的一种水汽凝结或凝华现象, 是春季广东沿岸地区的主要灾害性天气之一<sup>[1-2]</sup>。海雾导致水平和垂直能见度降低, 对海上渔业、平台作业、航运、军事行动以及沿岸航空和公路交通造成很大影响, 也是引发重大事故的主要原因之一。但是, 与其他灾害性天气相比, 海雾的业务预报仍然处于相当低的水平<sup>[3-4]</sup>。

海雾的生成、发展和消散涉及到湍流输送、辐射、雾滴沉降、风切变、雾顶夹卷等物理过程及其复杂的相互作用<sup>[5-6]</sup>。中尺度数值模式还难以准确刻画海雾中复杂的边界层过程, 海雾业务数值预报还存在很多困难<sup>[7-9]</sup>。目前, 最有效的海雾预报方法仍然以传统的统计预报方法为主<sup>[5, 10]</sup>。统计预报方法的建模手段多采用逐步回归等统计分析方法, 如胡基福等<sup>[11]</sup>、王厚广等<sup>[12]</sup>建立的青岛地区海雾预报方法, 于润玲等<sup>[13]</sup>开发的黄海和东海海雾短期客观预报方法等。近年来, 其他一些统计建模方法, 如模糊

逻辑仿真、FUZZY 和神经网络等方法也用于开发、建立雾的预报模型<sup>[14-17]</sup>。上述统计预报方法在我国沿海地区海雾的业务预报中取得了一定成效, 但同时也存在建模过程复杂、预报变量的权重值意义模糊和预报流程逻辑结构不清等缺点。分类与回归树 (CART) 方法是近年来普遍采用的一种数据分类方法, 分类准确性较高、计算过程简单, 尤其是其分类结果以二叉决策树表示, 具有易于解释或理解的优点<sup>[18]</sup>。目前, 采用 CART 技术建立的海雾决策树预报模型, 已在外国一些地区投入业务应用<sup>[19-20]</sup>。

本研究的目的是探讨采用分类技术建立广东沿岸海雾预报模型的可行性, 为海雾预报技术的研究提供新思路。利用 2000—2008 年 1—5 月广东沿岸 3 个观测站的海雾观测资料, 以及同期的 NCEP/NCAR FNL 再分析资料, 在采用 CART 方法对海雾与其生成前 24 h 的海洋气象条件进行分类分析的基础上, 建立了海雾决策树预报模型; 并采用 10 次交叉检验法对预报模型的预测性能进行检验。最后, 根据现有的海雾理论认识, 分析了海雾预报流程所包含的物理意义。

2010-04-29 收到, 2010-12-06 收到再改稿。

资助项目: 中国气象局新技术推广项目 (CMAFG 2007M23), 广东省科技三项“珠江口海雾监测预警系统研究”, 广东省科技厅计划项目 (2006B36702005)

\* E-mail: hj9387@21cn.com

## 1 资料与方法

### 1.1 资料

#### 1.1.1 海雾观测资料

本文采用的海雾观测资料为2000—2008年1—5月汕头、珠海和湛江3个地面观测站的地面能见度和天气现象资料。以能见度和天气现象作为挑选海雾个例的依据,并以二元变量表示。当能见度低于1 km,且天气现象记录为雾时,记为1;当能见度高于1 km,或能见度小于1 km,但天气现象为降水时,记为0。另外,由于沿岸地区的观测中有可能混杂了锋面雾、辐射雾和海雾等不同性质的雾<sup>[1, 21]</sup>,而锋面雾、辐射雾往往与来自北方的高压气团有关,风向多为偏北风或离岸风。因此,在海雾个例的筛选过程中,本文将汕头、珠海和湛江3个站的地面风向限制在一定范围内,其中汕头站的风向范围为 $45^{\circ}\sim 240^{\circ}$ ,珠海站的风向范围为 $60^{\circ}\sim 240^{\circ}$ ,湛江站的风向范围为 $60^{\circ}\sim 270^{\circ}$ 。经筛选后,汕头、珠海和湛江的分类样本数分别2510,4613组和3833组,有雾样本分别为110,133组和313组。

#### 1.1.2 海洋气象资料

目前,中尺度业务预报模式的同化分析产品已经非常丰富,但由于同化产品的积累时间不长,本文采用了与海雾观测同期的NCEP/NCAR FNL再分析资料,作为分析海雾与海洋气象条件关系的基本变量。资料的空间分辨率为 $1^{\circ}\times 1^{\circ}$ ,时间分辨率为6 h/次。

预报参考点的位置分别位于汕头、珠海和湛江观测站附近的海洋上。其中,汕头、珠海和湛江的预报参考点分别为A点( $23^{\circ}\text{N}, 118^{\circ}\text{E}$ )、B点( $22^{\circ}\text{N}, 114^{\circ}\text{E}$ )和C点( $21^{\circ}\text{N}, 112^{\circ}\text{E}$ ) (图1)。在预报变量的选择上,考虑到海雾发生在海洋大气边界层中,海雾的生成和维持一方面取决于天气系统控制下的边界层条件;另一方面也与海洋与大气之间的热量和水汽交换有密切关系<sup>[19]</sup>。本文挑选了海上低层大气(海面至850 hPa)的风、温、湿要素以及海表温度等作为预报变量。另外,还增加了近海面层气温与海表温度差和露点与海表温度差,以反映海气交换过程对海雾形成的作用。挑选的预报变量共15个,具体内容可参见表1。

本文以实测海雾事件作为分类分析的目标变量,以海洋气象要素作为预报变量。在分类样本的

构成上,海洋气象要素作为海雾生成的前提条件,在时间上比海雾事件提前了24 h。例如2008年1月1日08:00(北京时,下同)的NCEP/NCAR再分析资料与2008年1月2日08:00的地面观测构成1组分类样本。因此,分类结果反映了海雾事件出现前24 h的海洋气象要素状况,对海雾具有预报意义。

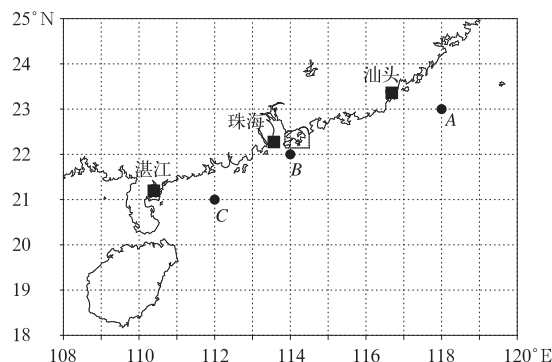


图1 地面观测站与预报参考点地理位置图

Fig. 1 The location of the surface observation stations and forecasting reference points

表1 用于CART方法的预报变量

Table 1 The predictor variables used in CART analysis

序号	预报变量名	海洋气象要素
1	$T_{\text{sea}}$	海表温度,单位: $^{\circ}\text{C}$
2	$T_{2\text{ m}}$	2 m 高度气温,单位: $^{\circ}\text{C}$
3	$D_{2\text{ m}}$	2 m 高度露点温度,单位: $^{\circ}\text{C}$
4	$\alpha_{10\text{ m}}$	10 m 高度风向,单位: $(^{\circ})$
5	$V_{10\text{ m}}$	10 m 高度风速,单位: m/s
6	$T_{1000}$	1000 hPa 高度气温,单位: $^{\circ}\text{C}$
7	$D_{1000}$	1000 hPa 高度露点温度,单位: $^{\circ}\text{C}$
8	$\alpha_{1000}$	1000 hPa 高度风向,单位: $(^{\circ})$
9	$V_{1000}$	1000 hPa 高度风速,单位: m/s
10	$T_{850}$	850 hPa 高度气温,单位: $^{\circ}\text{C}$
11	$D_{850}$	850 hPa 高度露点温度,单位: $^{\circ}\text{C}$
12	$\alpha_{850}$	850 hPa 高度风向,单位: $(^{\circ})$
13	$V_{850}$	850 hPa 高度风速,单位: m/s
14	$T_{2\text{ m}} - T_{\text{sea}}$	气温—海表温度差,单位: $^{\circ}\text{C}$
15	$D_{2\text{ m}} - T_{\text{sea}}$	露点—海表温度差,单位: $^{\circ}\text{C}$

### 1.2 CART方法

CART方法是一种非参数化的回归及分类方法,不需要预先假设目标变量与预报变量之间的关系,而是根据目标变量,利用递归划分法,将由预报变量定义的空间划分为尽可能同质的类别。每一次划分都由预报变量的一个最佳划分阈值来完成,并将数据分成两个部分,重复此过程,直到数据不可再分为止。CART方法建立决策树的过程,由树的生

长和树的剪枝两个步骤组成<sup>[18]</sup>。

### 1.2.1 决策树的生长

对于二元分类问题, CART 方法通常采用 Gini, Twoing 和熵系数等方法进行样本分类, 其中, Gini 系数方法尤其适用于样本数量较少的分类问题<sup>[22]</sup>。本文采用 Gini 系数方法的节点不纯度指数作为样本分类和建立最佳分类规则的准则。Gini 系数的不纯度指数表达式为

$$i(t) = \sum_{i \neq j} p(i|t)p(j|t) = 1 - \sum_j p^2(j|t). \quad (1)$$

其中,  $i(t)$  为节点  $t$  的不纯度指数;  $p(i|t)$  为节点  $t$  中属于  $i$  类的样本所占的比例;  $p(j|t)$  是节点  $t$  中属于  $j$  类的样本所占的比例。假定节点  $t$  的下一代子节点分为  $t_L$  和  $t_R$ , 则不纯度指标的降低量表示为

$$\Delta i(t) = i(t) - i(t_L)p(i|L) - i(t_R)[1 - p(i|L)]. \quad (2)$$

其中,  $i(t)_L$  和  $i(t)_R$  分别是左、右子节点的不纯度指数;  $p(i|L)$  为节点  $t$  上划分到左子节点的样本所占比例。

### 1.2.2 决策树的剪枝

理想的决策树不仅有较强的分类性能, 同时也应有较少的分枝或中间节点。因此, 需要对决策树进行剪枝, 在分类准确性与节点数两者之间取得平衡<sup>[18]</sup>。相对成本反映了决策树的节点数与整体误分率的关系<sup>[22]</sup>, 一般可作为评估为决策树分类性能以及剪枝的标准。

相对成本可表示为

$$C_{\text{rel}} = \sum (\partial_L \cdot (1 - p(n_0|L)) + \partial_R \cdot (1 - p(n_0|R))). \quad (3)$$

其中,  $p(n_0|L)$  和  $p(n_0|R)$  为 Bayes 条件性概率:

$$p(n_0|L) = \frac{p(L|n_0)p(n_0)}{p(L|n_0)p(n_0) + p(L|n_1)p(n_1)},$$

$$p(n_0|R) = \frac{p(R|n_0)p(n_0)}{p(R|n_0)p(n_0) + p(R|n_1)p(n_1)},$$

故  $(1 - p(n_0|L))$ ,  $(1 - p(n_0|R))$  分别表示当类别 1 错分为类别 0 时的误分率, 类别 0 错分为类别 1 时的误分率。  $\partial_L$ ,  $\partial_R$  分别为错分惩罚系数, 改变  $\partial_L$ ,  $\partial_R$  的大小, 可提高不同类别的事件在分类过程中的重要性。在本文中, 有雾或无雾的分类错分惩罚系

数  $\partial_j$  均取值为 1, 故  $C_{\text{rel}}$  与误分率有相同意义。

### 1.3 决策树的分类准确性检验

决策树建立后, 需要对决策树的分类准确性, 尤其是有雾类别的分类准确性进行评估。在样本数量较少, 特别是分类/预测对象为小概率事件的条件下, 10 次交叉检验法的检验结果, 可避免分类/预测对象时间分布不均匀而导致的不确定性<sup>[22]</sup>。采用 10 次交叉检验法, 对决策树分类性能进行检验的步骤包括: 将样本集分为数量大致相等的 10 个子集, 每个子集中的目标变量分布基本一致。首先, 以 9 个子集作为训练子集建立决策树, 并计算出决策树自身的误分率; 用剩下的 1 个子集作为验证子集, 检验决策树的分类/预测误分率。然后, 重复上述过程, 直到其他 9 个子集都分别作为训练子集和验证子集。最后, 合成 10 个子集的训练和验证误分率, 得到决策树的整体训练误分率和验证误分率(成功率)。

## 2 结果与分析

### 2.1 海雾预报决策树的建立

以汕头、珠海和湛江的海雾事件作为目标变量, 以预报参考点(A, B, C 点)的海洋气象要素作为预报变量, 采用 CART 方法进行分类分析, 得到决策树相对成本与节点数的关系(图 2)。本文中采用的错分惩罚系数为 1, 相对成本与误分率具有同样的意义。

由图 2 可见, 若无限地持续对样本进行分类, 汕头、珠海和湛江决策树都可产生众多的节点, 节点数可分别增长至 105, 132 和 230。但过少或过多的节点数, 并不意味着决策树整体误分率变小。当汕头的决策树节点数在 1~4 之间时, 误分率达到 55%~80%; 当内部节点数为 5 时, 误分率最小(49.5%); 而当节点数超过 5 以后, 整体误分率又重新增大。同时, 珠海决策树的内部节点数为 6 时, 误分率最小(47.7%); 当湛江决策树的内部节点数为 10 时, 误分率最小(46.1%)。

根据 CART 方法的剪枝原理, 以误分率作为指标分别对汕头、珠海和湛江决策树进行剪枝处理, 得到不仅整体误分率最小, 同时节点数也最少的最佳决策树。汕头、珠海和湛江最佳决策树的内部节点

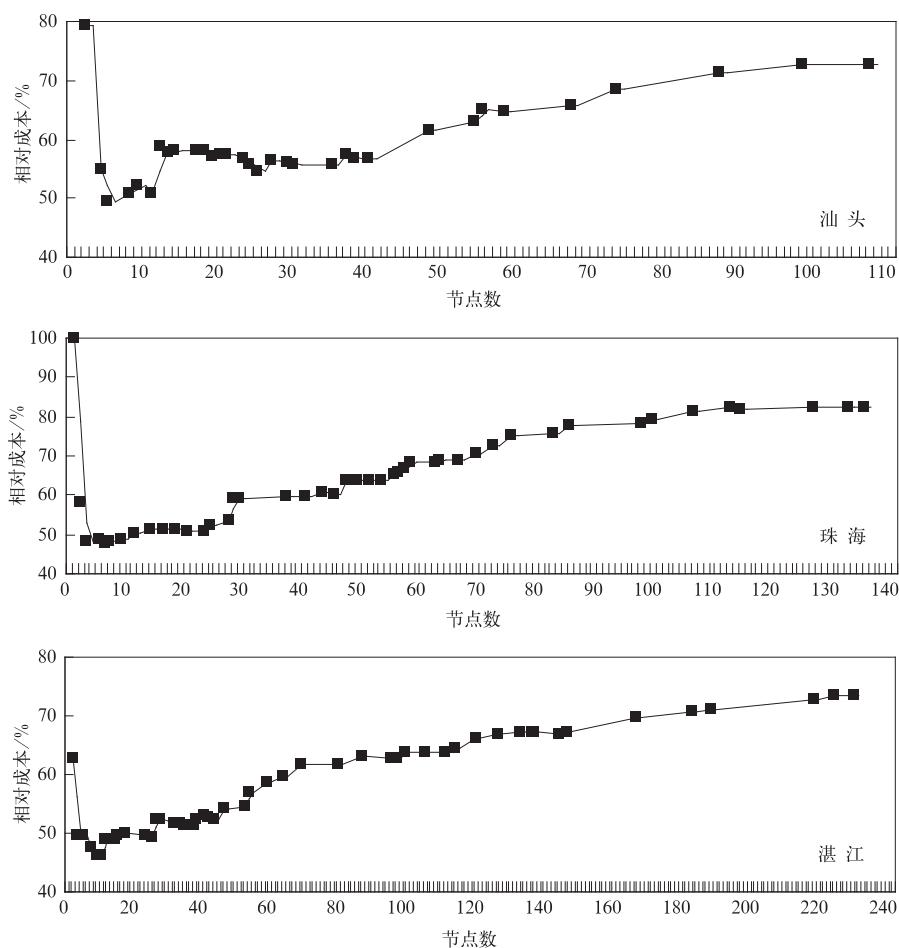


图2 汕头、珠海和湛江的决策树节点数与相对成本关系

Fig. 2 The relationship between relative costs and the number of nodes for the decision trees of Shantou, Zhuhai and Zhanjiang

数分别为5,6和10,包含的分类规则分别为9,11和17(图略)。

## 2.2 预报准确性检验

在样本数量较多的条件下,检验决策树分类/预测准确性的最简单方法是将样本集分为训练和测试两个子集,分别进行检验:训练子集用于建立决策

树,而验证子集用于检验决策树的分类/预测性能。由于资料的限制,3个站的分类样本分别只有2510,4613组和3833组,有雾样本仅有110,133组和313组;分类样本的总数不多,有雾的样本更少。本文采用10次交叉检验法对3个站决策树的分类/预测准确性进行检验,结果见表2。在此,误分率同样也可

表2 决策树的预报准确率检验结果

Table 2 The testing results for the classification/forecasting accuracy of sea fog decision trees

站点	类别	样本数	训练误分率/%	验证误分率/%	训练成功率/%	验证成功率/%
汕头	0	2267	31.8	32.2	68.2	67.8
	1	110	13.7	17.3	86.3	83.7
珠海	0	4480	18.9	21.4	81.1	78.6
	1	133	18.8	26.3	81.2	73.7
湛江	0	5031	26.8	28.5	74.2	71.5
	1	412	13.4	17.6	86.6	82.4

以表示为成功率。

由表 2 可知,对于无雾样本(类别 0),汕头、珠海和湛江的训练和验证成功率都比较低,3 个站的训练成功率分别为 68.2%,81.1%和 74.2%;验证成功率只有 67.8%,78.6%和 71.5%。本研究的主要目的是对海雾形成前的海洋气象条件进行分类分析,进而建立起海雾的决策树预报模型。因此,关注的对象是有雾类别的分类准确性,而无雾样本的误分对海雾预报的影响不大。从表 2 中有雾样本(类别 1)的检验结果可见:汕头、珠海和湛江最佳决策树无论训练成功率,还是验证成功率都远低于无雾样本。3 个站的训练成功率分别为 85.3%,81.2%和 86.6%,而验证成功率达到 83.7%,73.7%和 82.4%。上述检验结果表明:最佳决策树有较好的分类/预测性能,其准确率可达到 73%~83%。

在本文中,由于预报变量在时间上比目标变量提前了 24 h,因此,决策树分类规则可作为预报海雾的判别规则。根据决策树的分类规则及其排列顺序,可得到基于逐级逻辑判别原则的海雾预报模型。由上述检验结果也可知,采用 CART 方法建立的海雾决策树预报模型,对不同地域海雾预报的准确率总体上达到 73%以上,可满足海雾业务预报的需求。

### 2.3 海雾判别流程

在计算机条件许可的前提下,决策树预报模型中的判别变量和阈值可作为海雾预报的判别指标,对相应的海洋气象条件进行逐级逻辑判别,并得到预报结论。但是,3 个站的海雾预报模型中分别涉及到 9,11 和 17 个预报判别规则,难以在业务工作中直接做有雾/无雾判别。

由于决策树的建立过程是一个不断搜索最佳分类变量和阈值,以取得最小误分率的过程。因此,对于判别规则来说,高层节点上的判别变量和阈值比低层节点上的判别变量和阈值更有价值<sup>[18, 22]</sup>。根据判别规则在决策树中的排列顺序,分析内部节点的分类效率(即内部节点中包含的样本数相对于总样本数的比例),舍弃缺乏实际应用意义的小概率事件,可进一步建立适用于海雾业务预报的判别流程(图 3~5)。由图 3~5 可见,相对于海雾决策树预报模型,预报判别流程的更为简洁、实用。在一般情况下,采用判别流程可方便地进行有雾/无雾判别。

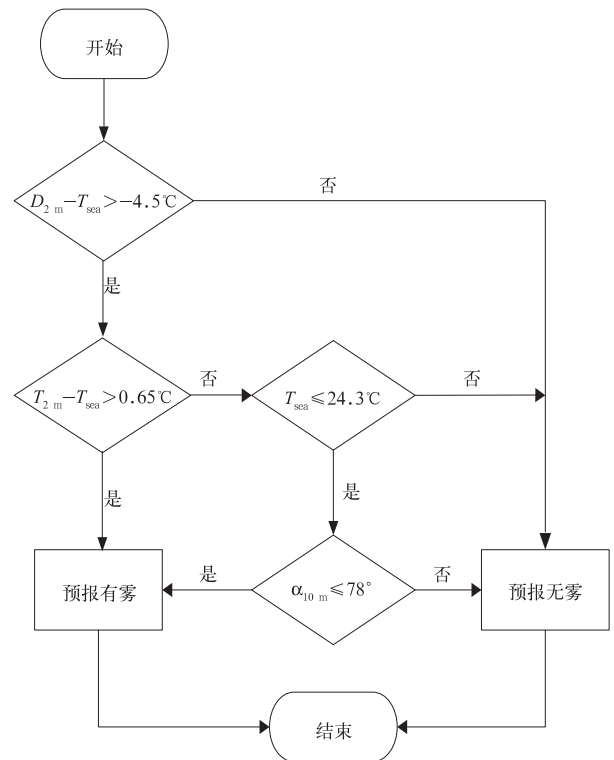


图 3 汕头海雾判别流程图  
Fig. 3 The decision-making procedure of sea fog forecasting for Shantou

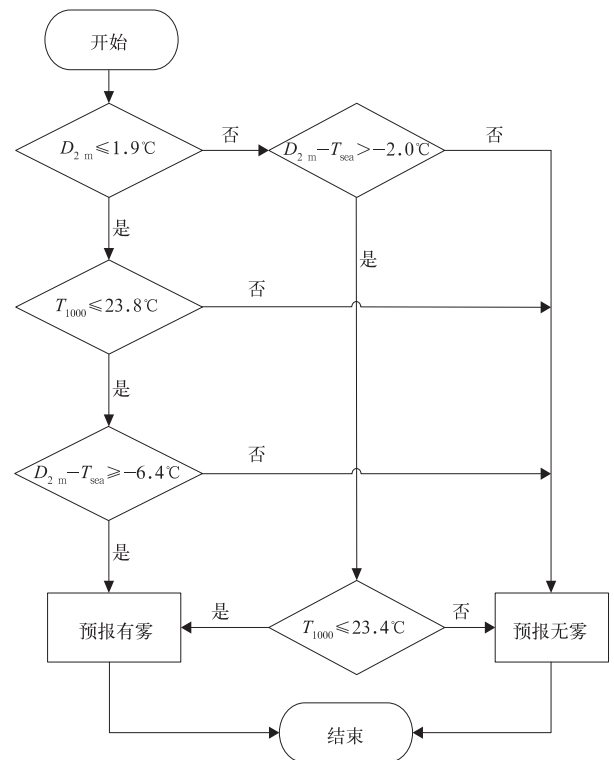


图 4 珠海海雾判别流程图  
Fig. 4 The decision-making procedure of sea fog forecasting for Zhuhai

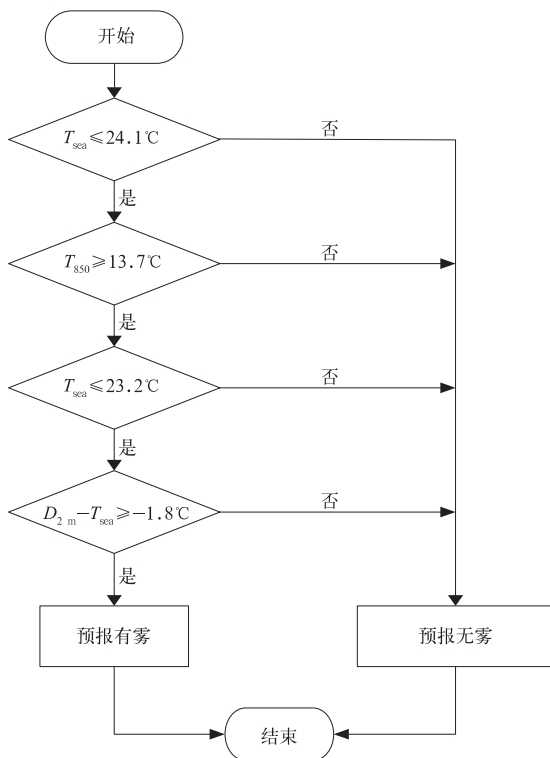


图5 湛江海雾判别流程图  
Fig. 5 The decision-making procedure of sea fog forecasting for Zhanjiang

### 3 海雾判别流程的物理意义

我国近海的海雾基本上都属于平流冷却雾,即暖湿空气平流到冷海面上,通过与冷海面的湍流热量交换,冷却达到饱和而形成的雾。有利于平流冷却雾形成的海洋气象条件包括:近海面层水汽充沛;海气温差的范围一般为 $-0.1\sim 5^{\circ}\text{C}$ ,并且海表温度不超过 $25^{\circ}\text{C}$ <sup>[1]</sup>;适宜的风向将暖湿空气输送到沿岸的冷海面上,以及稳定的大气层结等<sup>[23-24]</sup>。一些研究也认为,由于天气系统动力和热力结构不同,海洋大气边界层中的水汽和稳定度条件存在一定的差异;在某些情况下,雾顶长波辐射冷却等对平流冷却雾的发展和维持也起到一定的作用<sup>[25-26]</sup>,但近海面空气与冷海面的湍流交换过程仍然是平流冷却雾形成的最重要物理过程<sup>[5]</sup>。

由汕头的海雾判别流程(图3)可见,除风向条件外,海雾的第一判别规则是预报参考点A点来自海上的空气必须为湿空气,水汽含量需超过一定的指标( $D_{2\text{m}} - T_{\text{sea}} > -4.5^{\circ}\text{C}$ );第二,海雾形成前海气温差必须存在一定的差异( $T_{2\text{m}} - T_{\text{sea}} > 0.65^{\circ}\text{C}$ ),即

空气在平流过程中需要通过与冷海面的热量交换而冷却。当满足上述两个条件时,即可预报24 h后海雾将影响该地区。另一种情况,当来自海上的空气为湿空气( $D_{2\text{m}} - T_{\text{sea}} > -4.5^{\circ}\text{C}$ ),但海气温差较小( $T_{2\text{m}} - T_{\text{sea}} \leq 0.65^{\circ}\text{C}$ )时,还需满足海表温度较低的条件( $T_{\text{sea}} \leq 24.3^{\circ}\text{C}$ ),并且此时为来自东北方向的气流( $\alpha_{10\text{m}} \leq 78^{\circ}$ )。在满足上述条件的前提下,也可判断24 h后将有海雾生成。一些观测研究也显示,汕头地区的特殊地理位置和海岸线走向,海雾既可出现在地面为偏南风的条件下,也可在东北或偏东气流中形成<sup>[27]</sup>。可见,海雾判别流程具有明确的物理意义,不仅能够很好地反映水汽与海表冷却条件在平流冷却雾形成过程中的重要性,同时也反映了来自台湾海峡的气流与来自南海的气流,在动力和热力特性上可能存在的差异。

由图4~5也可知,珠海和湛江的海雾判别流程与汕头存在一定的差异,但两者都强调了水汽条件、海面冷却作用在海雾形成过程中的重要性。由广东省沿岸海岸线跨度可知,3个站之间的距离可达到800 km左右,判别流程的差异很可能与海雾形成的天气气候背景有关。过去的研究已表明:广东沿岸不同地域的海雾无论在其盛行期上,还是在控制天气系统上都有一定的差异<sup>[28]</sup>。雷州半岛以2—3月雾的频率最大,从雷州半岛向东,雾期却明显推迟;珠江口的海雾盛行于3—4月;而粤东的雾比较集中于4月<sup>[1]</sup>。2—3月广东沿海仍处于冬季风强盛期,海雾多发的原因与出海变性高压回流的暖湿气流有关;4月,冬季风对广东沿海的影响已经减弱,海雾形成的原因与低纬热带天气系统外围环流的输送条件有密切关系<sup>[28]</sup>。不同地域的海雾形成于不同的水汽、稳定度以及海表温度分布条件下,有可能是海雾判别流程存在一定差异的主要原因。

### 4 小结

利用2000—2008年1—5月广东沿岸汕头、珠海和湛江3个地面观测站的海雾历史资料和NCEP/NCAR FNL再分析资料,本文探讨了采用CART方法建立海雾预报模型的可行性;并根据现有的海雾理论认识,讨论了海雾决策树预报规则所包含的物理意义。显然,本项工作还存在海雾样本偏少,预报参考点选址缺乏客观标准等的缺陷,建立起一个真正适用于广东沿岸的海雾决策树预报方



法,还需要开展进一步的分析研究工作。初步的分析结果,归纳如下:

1) 采用 CART 方法建立汕头、珠海和湛江的海雾决策树预报模型,由 10 次交叉检验的结果可知:预报准确率分别达到 83.7%, 73.7% 和 82.4%, 即对广东沿岸不同地域的海雾预报准确率总体可达到 73% 以上。

2) 海雾决策树预报模型具有逻辑关系清晰的优点,根据决策树预报模型建立的海雾判别流程,可在业务工作中直接用于有雾/无雾判别。

3) 海雾预报的判别流程有明确的物理意义,能够较好地反映水汽与海表冷却条件在平流冷却雾形成过程中的重要性。

4) CART 方法计算过程简单、分类准确性高,可作为海雾业务预报的有效建模工具。

### 参 考 文 献

- [1] 王彬华. 海雾. 北京: 海洋出版社, 1983.
- [2] 刘小宁, 张洪政, 李庆祥, 等. 我国大雾的气候特征及变化初步解释. 应用气象学报, 2005, 16(2): 220-230.
- [3] 张苏平, 鲍献文. 近十年中国海雾研究进展. 中国海洋大学学报, 2008, 38: 359-366.
- [4] 张春桂, 蔡义勇, 张加春. MODIS 遥感数据在我国台湾海峡海雾监测中的应用. 应用气象学报, 2009, 20(1): 8-16.
- [5] Lewis J M, Koracin D, Redmond K T. Sea fog research in the United Kingdom and United States, a historical essay including outlook. *Bull Amer Meteor Soc*, 2004, 75: 395-408.
- [6] Edson J, Crawford T, Crescenti J, et al. The coupled boundary layers and air-sea transfer experiment in low winds. *Bull Amer Meteor Soc*, 2007, 88: 341-356.
- [7] Taylor W. Mesoscale Case Study: Project Haar. Met Office Met. O. 11 Tech, Note 239, 1987.
- [8] 傅刚, 张涛, 周发琇, 等. 一次黄海海雾的三维数值模拟研究. 青岛海洋大学学报(自然科学版), 2002, 32: 859-867.
- [9] Gao Shanhong, Lin Hang, Shen Biao, et al. A heavy sea fog event over the Yellow Sea in March 2005: Analysis and numerical modeling. *Adv Atmos Sci*, 2007, 24: 65-81.
- [10] 周发琇. 海雾: 第四讲 海雾预报. 海洋预报, 1986, 6: 71-78.
- [11] 胡基福, 郭可彩, 鄢利农. 应用模式输出统计作海雾出现判别预报. 青岛海洋大学学报, 1996, 4: 439-445.
- [12] 王厚广, 曲维政. 青岛地区的海雾预报. 海洋预报, 1997, 14(3): 53-57.
- [13] 于润玲, 王亚男, 李永平. 黄海和东海海雾短期客观预报业务系统. 大气科学研究与应用, 2007(2): 28-36.
- [14] 刘科峰, 张韧, 江海英, 等. 模糊逻辑仿真建模及其在青岛海雾分析中的应用. 海洋湖沼通报, 2004(4): 17-25.
- [15] 朱云清, 于世永. 青岛近海海雾的数学预报方法. 海洋预报, 1991, 8(3): 68-71.
- [16] 李法然, 周之栩, 陈卫锋, 等. 湖州市大雾天气的成因分析及预报研究. 应用气象学报, 2004, 16(6): 794-803.
- [17] 梅珏. 人工神经网络在辐射雾预报中的应用. 应用气象学报, 1999, 10(4): 511-512.
- [18] Steinberg D, Colla P. CART: Tree-structured Nonparametric Data Analysis. San Diego, CA, Salford Systems, 1995.
- [19] Lewis D M. Forecasting Advective Sea Fog with the Use of Classification and Regression Tree Analyses for Kunsan Air Base. AFIT/GM/ENP/04-08, Physics Graduate School of Engineering and Management, Air Force Institute of Technology, Air University, 2004.
- [20] Wantuch F. Visibility and fog forecasting based on decision tree method. *IDOJARAS*, 2003, 105: 29-38.
- [21] 梁军, 张胜军, 隋洪起, 等. 大连地区大雾特征. 应用气象学报, 2009, 20(1): 28-35.
- [22] Breiman L, Friedman J H, Olshen R A, et al. Classification Regression Trees. Wadsworth, Belmont, CA, 1984.
- [23] 周发琇, 刘龙太. 长江口及济州岛邻近海域综合调查报告(第七节 海雾). 山东海洋学院学报, 1986, 16(1): 115-131.
- [24] Zhang Suping, Xie Shangping, Liu Qinyu, et al. Seasonal variations of Huanghai sea fog: Observations and mechanisms. *J Climate*, 2009, 22: 6758-6772.
- [25] Huang Jian, Wang Xin, Zhou Wen, et al. The characteristics of sea fog with different airflow over the Huanghai Sea in boreal spring. *Acta Oceanologica Sinica*, 2010, 29(4): 3-12.
- [26] 黄健, 王斌, 周发琇, 等. 华南沿海暖海雾过程中的湍流热量交换特征. 大气科学, 2010, 34(4): 715-725.
- [27] 张朝锋. 粤东海区海雾的气候特征分析. 广东气象, 2002, 2: 2-21.
- [28] 林钢, 卢山, 薛登智. 华南沿岸雾的气候特征及影响系统. 广东省气象台研究技术报告, 2000.

## Decision Tree Forecasting Models of Sea Fog for the Coast of Guangdong Province

Huang Jian<sup>1)</sup> Huang Huijun<sup>1)</sup> Huang Minhui<sup>2)</sup> Xue Dengzhi<sup>2)</sup> Mao Weikang<sup>1)</sup> Bai Yujie<sup>3)</sup>

<sup>1)</sup> (*Institute of Tropical and Marine Meteorology, CMA, Guangzhou 510080*)

<sup>2)</sup> (*Guangzhou Central Meteorological Observatory, Guangzhou 510080*)

<sup>3)</sup> (*Training Center of Guangdong Provincial Meteorological Bureau, Guangzhou 510080*)

### Abstract

Sea fog is a phenomenon of water vapor condensation or sublimation in marine atmospheric boundary layer and is also one of the main disastrous weathers on the coast of Guangdong Province in spring. However, there is no suitable method for operational sea fog forecasting in Guangdong due to the complexity of physical processes involved in the formation of sea fog. Therefore, historical sea fog reports from Shantou, Zhuhai and Zhanjiang surface meteorological observation and NCEP/NCAR FNL reanalysis for the period of 2000—2008 are analyzed to explore the feasibility of sea fog forecasting with a 24-hour lead time. The relationship between marine atmospheric conditions and sea fog events is examined by Classification and Regression Trees (CART), employing the NCEP/NCAR reanalysis data 24 hours before the sea fog events. Then, the decision tree models for sea fog forecasting are developed based on results of classification analysis. Finally, the physical significance of the forecasting rules is discussed based on existing theoretical knowledge on sea fog.

The validation results by 10 cross-validation show that the forecasting accuracy of sea fog decision tree models developed by CART can reach 83.7%, 73.7% and 82.4% respectively for Shantou, Zhuhai and Zhanjiang on the coast of Guangdong Province. It can be interpreted or understood easily due to the clear logical relationship. The decision-making procedure can be developed and used directly to make fog/no-fog identification in operational sea fog forecasting with clear physical meanings. It also reflects the importance of the water vapor and the cooling effect of cold sea surface in the formation of advective cooling fog well. Simple calculation processes and relatively high classification accuracy make the CART an effective tool to develop sea fog forecasting model.

**Key words:** sea fog on the coast of Guangdong Province; Classification and Regression Trees (CART); decision tree mode; decision-making procedure