

结合FY-4A卫星及随机森林的日间沿海海雾识别模型的研究

耿丹¹, 刘婷婷², 李超³

(1.江苏省气象信息中心,江苏南京210041;2.江苏省气象服务中心,江苏南京210041;3.江苏省气象台,江苏南京210041)

摘要: 利用2019年8月—2021年7月期间的FY-4A卫星数据,结合江苏及周边地区自动气象站同期的能见度观测数据,建立了包括有海雾时次及非海雾的卫星像素集,利用随机森林算法构建了海雾识别模型,实现对江苏及周边区域的日间海雾识别。检验结果表明:与基于阈值法的海雾识别模型相比,训练得到的随机森林海雾识别模型具有较高的识别精度,该模型平均命中率、平均临界成功指数和平均误报率分别为83.46%、79.46%和5.7%,均优于阈值法的结果。两种识别模型对2021年4月12日发生在黄渤海区域海雾天气个例识别结果的对比表明,随机森林海雾识别模型能够更好地识别出发生海雾的区域。

关键词: FY-4A卫星;能见度观测数据;卫星像素集;随机森林;日间海雾识别

中图分类号: P732 **文献标识码:** A **文章编号:** 1003-0239(2022)03-0083-11

1 引言

海雾是海上中低层大气层水汽凝结造成的天气现象,凝结水汽或冰晶积聚形成的海雾会造成海上的水平能见度降至1 000 m以下。海上大雾一般分平流雾和辐射雾等多种类型,其中平流雾对人类日常危害性最大,严重影响海上运输、海产养殖和渔业捕捞等行业。

随着海上各行业的蓬勃发展,海雾造成的损失越来越大。由于上海雾监测点有限,常规观测和船载观测基本无法实现对海雾区域大范围和长时间的实时监测^[1]。卫星遥感技术的发展,使人类通过卫星遥感影像对海上大雾的监测成为可能。郑新江^[2]结合地球同步气象卫星(Geostationary Meteorology Satellite, GMS)资料分析了黄海海雾生成及演变的特征;Ellrod^[3]通过对地静止环境工作卫星(Geostationary Operational Environmental Satellite, GOES)的双通道红外资料对沿海夜间海雾和低云进行分析研究;鲍献文等^[4]使用GMS-5和美国国家

海洋和大气管理局(National Oceanic and Atmospheric Administration, NOAA)卫星遥感资料,分析出海雾在卫星遥感资料的光谱和辐射特征特点,实现对云和雾的识别及分离的定量分析;何月等^[5]通过多用途运输卫星(Multifunctional Transport SATellites, MTSAT)卫星遥感影像,采用分级判识太阳高度角阈值和大雾指数的方式反演出浙江海上大雾的每小时发生情况;Shang等^[6]结合“葵花8号”卫星中红外与长波红外通道亮温差与植被和雪覆盖指数等建立了日间陆雾识别模型;张培等^[7]将星载双波长偏振Mie散射激光雷达(Cloud-Aerosol Lidar with Orthogonal Polarization, CALIOP)数据获得的样本点用于“葵花8号”卫星对日间海雾通道及阈值选择的研究,实现了一种日间海雾的监测方法;衣立^[8]和Wang等^[9]通过不同方式不断提升阈值法海雾识别模型(简称“阈值法”)在海雾识别中的精度;张春桂等^[10]使用中分辨率成像光谱仪(MODerate-Resolution Imaging Spectroradiometer, MODIS)卫星数据分析出不同下垫面的可见光和红外辐射的特

收稿日期: 2021-11-30; 修回日期: 2021-12-27。

基金项目: 江苏省气象局青年基金课题(KQ202004)。

作者简介: 耿丹(1985-),男,工程师,硕士,主要从事航空及海洋服务产品监测预警研究。E-mail: 494541088@qq.com

通讯作者: 刘婷婷(1990-),女,工程师,本科,主要从事交通及健康等气象服务产品研究。E-mail: 853234916@qq.com

征特点,确定海雾识别阈值,建立日间海雾识别模型;孙艺等^[11]借助MTSAT可见光卫星云图、“葵花8号”可见光卫星云图和韩国气象厅(Korea Meteorological Administration, KMA)地面天气图对控制海雾产生的天气形势进行了分析,进一步讨论了各天气型下黄海海雾的高度特征;于海鹏等^[12]使用GOES9卫星的可见光云图和地面探空站资料对海雾天气进行识别。

阈值法是通过分析云雾在遥感辐射的差异来实现对海雾的监测,最难的是如何选择合适的阈值,而且阈值法不能充分使用各个波段遥感数据内容。随着机器学习在卫星遥感领域的不断应用,能够充分利用卫星遥感影像提供的各波段数据内容。Kim等^[13]通过决策树算法结合“葵花8号”和地球静止水色卫星(Geostationary Ocean Color Imager, GOCI)遥感数据的方法实现了海雾识别,并利用卫星遥感数据与朝鲜3个岛屿能见度观测值匹配进行模型训练及验证。Shin等^[14]基于通信、海洋和气象卫星(Communication, Ocean and Meteorological Satellite, COMS)红外通道亮温差,使用无监督学习方法实现对朝鲜半岛西部海雾区域的识别。许赞等^[15]利用随机森林算法结合卫星遥感影像对云、雪和雾进行分类识别,并通过增加两次检测减少有效区域的错检率。姜红等^[16]通过归一化差值沙尘指数和随机森林算法等3种方法,利用“风云四号”(FY-4A)卫星遥感影像实现对塔里木盆地沙尘情况的监测研究,通过试验检验分析可知随机森林模型和卷积神经网络(Convolutional Neural Networks, CNN)模型都具有较强的沙尘监测能力。张环宇等^[17]基于卫星数据的中红外到热红外波段遥感影像数据和ERA5水汽再分析产品等数据,结合随机森林算法实现对晴空大气可降水量的反演,该算法可有效提升大气可降水量的精度。柳青青等^[18]结合随机森林算法,利用海表面盐度遥感机理和土壤湿度及海水盐度(Soil Moisture and Ocean Salinity, SMOS)卫星遥感数据盐度反演理论建立海表面盐度反演模型,能够大幅提高盐度反演精度。

本文利用机器学习中的随机森林算法,结合FY-4A卫星遥感影像对江苏及周边省份沿海日间海雾进行识别研究。相对于海雾反演的传统经验阈值算法,随机森林算法能处理高维度的数据(即具

有很多特征的数据),不用特征选择,减少人为经验的误差,能更客观地表现数据联系和结果。该研究成果将为今后FY-4A卫星数据在海雾监测业务中的应用提供重要依据。

2 资料选取与方法介绍

2.1 海雾天气过程个案获取

通过分析江苏及周边省份沿海城市发生大雾的预警信号,选取日间江苏及山东省沿海城市发生大雾天气60个个例(见表1),从中随机选取20个个例作为检验识别模型的检验个个例集,剩下40个个例作为构建随机森林海雾识别模型的训练个个例集。

2.2 卫星数据获取

作为我国自主研发的第二代地球静止气象卫星, FY-4A是一种三轴稳定地球静止气象卫星。除了常规的成像探测设备外,还配备能够大幅提升我国短临天气预报和应对极端气候变化等防灾减灾能力的设备——红外高光谱垂直探测仪和闪电探测设备。FY-4A多通道扫描成像辐射计(AGRI)具有6个可见/近红外波段、2个中波红外波段、2个水汽波段和4个长红外波段(见表2),这14个探测波段范围为0.45~13.8 μm,并涉及(0—3)4个不同级别的卫星产品。

为了获得满足沿海海雾识别所需的卫星遥感影像,需要将AGRI的扫描成像数据经过0级数据格式变换、定位和定标等处理生成1级高频次中国区域产品数据。该产品数据包括14个波段信息,无全圆盘观测时进行5 min中国区域观测,空间分辨率为4 km,符合沿海海雾动态监测时空分辨率要求。

本文使用4 km分辨率中国区域产品数据和地对地静止轨道(GEostationary Orbit, GEO)定标数据,通过Python中H5py功能包进行数据解析获得14个波段数据,结合GEO定标文件和遥感图像处理软件对这些波段数据进行几何校正,然后利用自动站时次和经纬度信息获取FY-4A卫星14个通道影像上对应的像素值,组成样本点像素组。

本文使用江苏省及山东省沿海及近海约40个自动气象站观测FY-4A的1级数据产品,空间分辨率为4 km,通过空间插值到站点。由于卫星数据在

表1 江苏省和山东省沿海城市60个大雾个例

Tab.1 60 cases of dense fog in the coastal cities of Jiangsu and Shandong provinces

序号	发生地点	大雾预警等级	预警发布时间
1	山东威海	黄色	2019年9月09日07时00分
2	山东烟台	黄色	2019年9月25日07时10分
3	江苏盐城响水	黄色	2019年11月22日07时30分
4	山东日照	黄色	2019年11月23日08时40分
5	山东烟台莱州	黄色	2019年12月09日08时05分
6	江苏盐城射阳	黄色	2020年1月06日08时30分
7	山东东营	黄色	2020年1月19日07时12分
8	山东青岛莱西	黄色	2020年1月23日09时00分
9	山东青岛即墨区	黄色	2020年2月12日08时10分
10	江苏南通启东	黄色	2020年2月14日06时54分
11	山东日照	黄色	2020年2月20日07时30分
12	山东东营	黄色	2020年2月21日09时15分
13	山东东营	黄色	2020年2月25日07时32分
14	山东烟台	黄色	2020年3月08日14时20分
15	江苏盐城滨海	黄色	2020年3月26日08时18分
16	山东威海	黄色	2020年3月26日13时00分
17	江苏南通启东	黄色	2020年5月11日05时53分
18	江苏南通启东	黄色	2020年5月16日06时08分
19	山东烟台	黄色	2020年5月17日10时30分
20	江苏南通如东	黄色	2020年5月23日05时55分
21	山东烟台	黄色	2020年5月23日14时35分
22	山东青岛	黄色	2020年5月31日08时25分
23	江苏盐城滨海	黄色	2020年6月11日05时50分
24	山东烟台	黄色	2020年6月12日06时15分
25	山东威海	黄色	2020年6月24日08时30分
26	江苏连云港赣榆	黄色	2020年6月27日06时19分
27	山东日照	黄色	2020年6月28日05时42分
28	山东青岛	黄色	2020年7月06日16时15分
29	山东威海	黄色	2020年7月07日05时50分
30	山东威海	黄色	2020年7月08日06时50分
31	山东青岛	黄色	2020年7月19日17时46分
32	山东青岛	黄色	2020年7月30日16时35分
33	山东威海	黄色	2020年8月08日06时10分
34	江苏盐城响水	黄色	2020年8月28日05时52分
35	山东威海	黄色	2021年1月21日08时30分
36	山东青岛	黄色	2021年1月25日08时10分
37	江苏南通如东	黄色	2021年1月26日08时35分
38	江苏盐城大丰	黄色	2021年1月27日07时59分

表1 (续)
Tab.1 (Continued)

序号	发生地点	大雾预警等级	预警发布时间
39	山东青岛	黄色	2021年3月11日13时30分
40	山东威海	黄色	2021年3月21日13时00分
41	山东烟台牟平	黄色	2021年3月26日14时35分
42	江苏连云港赣榆	黄色	2021年3月28日07时44分
43	山东烟台	黄色	2021年4月12日11时10分
44	山东青岛	黄色	2021年4月23日17时00分
45	山东烟台海阳	黄色	2021年5月12日06时28分
46	山东日照	黄色	2021年5月12日17时10分
47	山东青岛	黄色	2021年5月14日16时14分
48	山东青岛	黄色	2021年5月15日08时40分
49	山东威海	黄色	2021年6月01日16时00分
50	山东日照	黄色	2021年6月08日14时25分
51	山东青岛	黄色	2021年6月12日16时45分
52	山东威海	黄色	2021年6月15日16时00分
53	山东烟台	黄色	2021年6月16日06时30分
54	山东青岛	黄色	2021年6月17日09时40分
55	山东日照	黄色	2021年6月25日08时45分
56	山东青岛	黄色	2021年6月25日16时15分
57	山东日照	黄色	2021年6月27日05时50分
58	山东青岛	黄色	2021年6月28日06时23分
59	山东威海荣成	黄色	2021年7月05日07时30分
60	山东烟台长岛	黄色	2021年7月14日06时10分

表2 FY-4A 多通道扫描成像辐射计主要技术参数

Tab.2 Main technical parameters of FY-4A multi-channel scanning imaging radiometer

波段	中心波长/ μm	波长范围/ μm	空间分辨率/ μm	主要用途
通道01	0.47	0.45~0.49	1	昼间云、沙尘和气溶胶观测
通道02	0.65	0.55~0.75	0.5	昼间云、沙尘、积雪等观测
通道03	0.825	0.75~0.90	1	云、气溶胶、植被和海洋特性检测
通道04	1.375	1.36~1.39	2	冰晶粒子、卷云观测
通道05	1.61	1.58~1.64	2	低云/雪识别和水云/冰云判识
通道06	2.225	2.10~2.35	2~4	卷云、气溶胶, 粒子大小观测
通道07	3.75	3.5~4.0(H)	2~4	对火点高温、白天的强太阳反射信号
通道08	3.75	3.5~4.0(L)	4	低云/雾的监测
通道09	6.25	5.8~6.7	4	对流层高层的水汽分布特征
通道10	7.10	6.9~7.3	4	对流层中层的水汽分布特征
通道11	8.50	8.0~9.0	4	昼夜云顶微物理状态相态、沙尘信息
通道12	10.80	10.3~11.3	4	地球表面和云顶温度
通道13	12.00	11.5~12.5	4	大气窗区边缘, 有弱吸收
通道14	13.50	13.2~13.8	4	大气中云、对流层中低层及地表信息

无全圆盘观测时的时间周期为5 min,而自动站能见度观测数据时间周期为5 min,这样每个卫星数据时间时次都能够获得自动站能见度观测数据,因此本文获取了全天165次中国区观测卫星像素。

气象大数据云平台简称为“天擎”,作为国省共建的气象部门核心的业务支撑系统,该系统能够提供实时及历史各类气象数据。本文通过“天擎”系统获取所涉及的自动站能见度数据。

本文通过分析海雾天气过程个例发生海雾的大致区域及时间段,从“天擎”系统查询大致区域及时间段内所涉及自动站能见度值首次小于1 000 m的观测时次。若查询该时次后连续2个时次(5 min/次)能见度值都小于1 000 m,则确认该时次为有效海雾出现时次;若后续连续2个时次能见度值不都小于1 000 m,则继续查询,直到查询超过该个例大致时间段终止本次查询。

当确认所涉及自动站的有效海雾出现时次,通过“天擎”系统查询有效海雾出现时次之后能见度值首次大于1 000 m出现时次(5 min/次)。若该时次后续连续2个时次(5 min/次)能见度值都大于1 000 m,则确认该时次为有效海雾消散时次;若后续连续2个时次不都大于1 000 m,则继续查询,直到查询超过该个例大致时间段终止本次查询。

获取涉及自动站有效海雾出现时次一有效海雾消散时次之间每个观测时次,剔除能见度值大于1 000 m的观测时次(5 min/次),同时剔除无法获得中国区产品数据的观测时次(5 min/次),获得涉及自动站的有效海雾时次数据集,然后对有效海雾时次和经纬度获取FY-4A卫星14个通道影像上对应的像素值,形成样本点像素组,然后由该个例所有涉及自动站发生海雾的样本点像素组组成有效海雾卫星像素集。

获取涉及自动站有效海雾出现时次之前和有效海雾消散之后一段时间(不少于1 h)每个观测时次,通过剔除能见度值小于1 000 m的观测时次,同时剔除无法获得中国区产品数据的观测时次(5 min/次),获得涉及自动站非海雾时次数据集,然后对应涉及自动站非海雾时次和经纬度获取FY-4A卫星14个通道影像上相对应的像素值,形成样本点像素组,然后由该个例所有涉及自动站未发生海雾的样本点像素组组成非海雾卫星像素集。

2.3 方法介绍

利用阈值法和随机森林算法作为海雾识别方法,结合FY-4A中AGRI的L1级空间分辨率为4 km的数据,利用14个不同波段光谱值作为输入要素,对江苏及周边省份沿海海雾天气开展识别研究。

2.3.1 阈值法

首先,获取可见光通道数据后剔除陆地,再利用可见光通道数据计算得到表观反射率,基于此数据初步剔除海面;其次,由于雾顶温度与云顶温度存在差异,利用剔除陆地后的长波红外通道亮温做初步判断,进而剔除中高云;最后,结合双通道差值技术剔除低云进而识别日间沿海海雾。具体识别流程见图1。

(1)剔除海面:获取可见光通道数据后,利用海陆掩膜文件把陆地去除(剔除陆地),再将可见光通道数据除以太阳天顶角的余弦得到可见光通道的

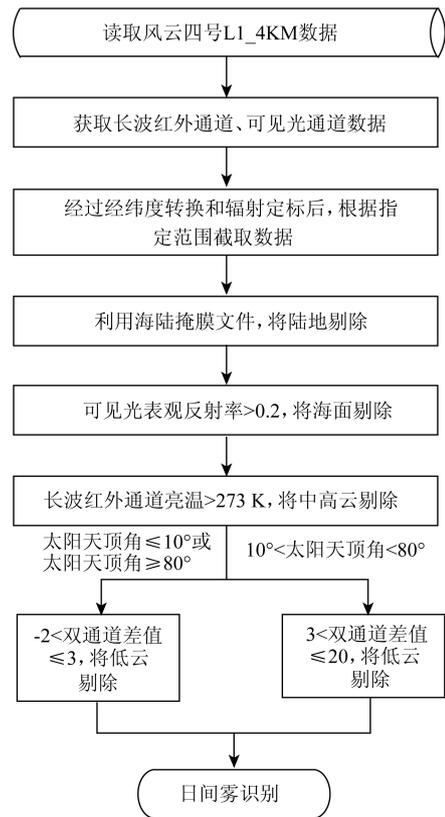


图1 基于阈值法日间海雾识别流程图

Fig.1 Flow chart of daytime sea fog recognition based on threshold method

表观反射率,其值大于0.2为云或雾,初步剔除海面。

(2)剔除中高云:由于雾顶温度与云顶温度存在差异,利用剔除陆地后的长波红外通道(13通道波长12 μm)亮温做初步判断,即大于273 K可剔除中高云。

(3)剔除低云:双通道差值指剔除陆地后的红外通道亮温差(14通道—12通道)。当太阳天顶角≤10°或太阳天顶角≥80°,−2<双通道差值≤3时,剔除低云;当10°<太阳天顶角<80°,3<双通道差值≤20,剔除低云。

2.3.2 随机森林海雾识别模型

随机森林模型是通过随机方式建立一个具有很多分类树的森林,且每个分类树之间没有关联。每当有一个新的样本输入随机森林模型中,每棵决策树都需要进行判断,通过投票方式得出最终分类结果(见图2)。

随机森林模型构造流程为:

(1)假设具有n个样本,进行有放回的随机选择n个样本(即每次随机选择一个样本,然后放回继续选择),使用这n个样本训练一个分类树。

(2)当每个样本的属性为M,在分类树的每个节点需要分裂时,随机从M个属性中选取m个属性(m<<M),采用某种策略在m个属性中选择1个属性作为节点分裂属性。

(3)分类树形成过程中每个节点都要按照步骤2来分裂(如果下一次该节点选出来的那个属性为刚刚父节点分裂时所用的属性,则该节点已到叶子

节点,不用继续分裂),这个过程到不能够再分裂为止,整个分类树形成过程中无需剪枝。

(4)按照步骤1—3建立n个分类树,如此构成1个随机森林模型。

利用随机森林算法确定最优参数n_estimators和m。n_estimators是指对原始数据集进行有放回抽样生成的子数据集个数,即决策树的个数。若n_estimators太小容易欠拟合,太大则不能显著地提升模型,所以n_estimators需要选择适中数值,默认值是100个。研究中分别选取设置100、200、300、400、500、600、700、800、900、1 000、1 100、1 200、1 300、1 400、1 500、1 600、1 700、1 800、1 900和2 000,子特征数m选取从2增大到9,通过超参数训练得到最优参数,即n_estimators=200,m=6,最优参数评估均方根误差(Root Mean Square Error, RMSE)误差最小。

利用Sklearn框架对随机森林算法进行表达,对于建模需要的图像数据集,首先获取为随机森林海雾识别模型(简称“随机森林模型”)建模所用的训练个例集,通过计算获得个例的有效海雾卫星像素集和非海雾卫星像素集,以此组成随机森林模型建模所用的标识后卫星像素集,其中有效海雾卫星像素集有8 356个样本点像素组,非海雾卫星像素集有52 346个样本点像素组。将标识后卫星像素集按照8:2随机分为训练样本点像素组集和测试样本点像素组集,以随机森林函数对所有训练样本点像素组集进行随机森林分类训练,逐步建立随机森林模型。

3 结果与分析

为了分析阈值法和随机森林模型的效果,利用事先选取的20个江苏及山东沿海城市大雾发生个例的检验个例集,在每个个例涉及自动站观测时次,分别从海雾发生前、发生中和消散后选取一定数量的卫星遥感影像,组成该个例的检验输入样本集,同时记录对应所涉及自动站观测时次,通过查询“天擎”系统获得该个例所涉及自动站能见度实测值,组成该个例检验样本点集。

利用海雾识别模型逐一对个例的检验输入样本集进行海雾雾区识别,获得海雾雾区信息,利用

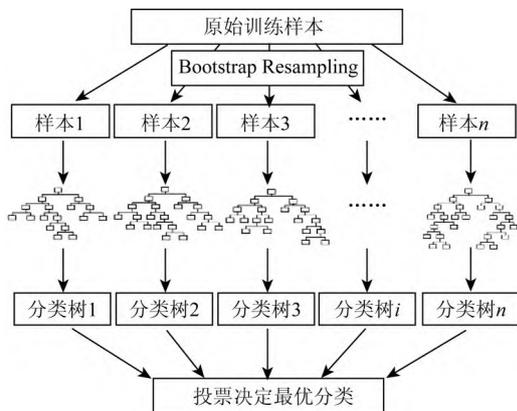


图2 随机森林模型结构示意图

Fig.2 Schematic diagram of random forest structure

该时次样本所涉及自动站观测时次和经纬度信息获取对应位置的海雾信息(有或无),并获得个例检验样本点集中该时次所涉及自动站的实测值,如果实测能见度值小于1 000 m,则为有海雾,如果大于等于1 000 m,则为无海雾。统计该个例卫星识别结果和自动站实测结果都有海雾的样本点个数、卫星识别结果有海雾而自动站实测结果没有海雾的样本点个数以及卫星识别结果没有海雾而自动站结果有海雾的样本点个数。

通过POD检验方法检测不同海雾识别模型的准确性,POD检验公式如下:

$$POD = \frac{N_H}{N_H + N_M} \times 100\%$$

$$FAR = \frac{N_F}{N_H + N_F} \times 100\%$$

$$CSI = \frac{N_H}{N_H + N_M + N_F} \times 100\%$$

式中,POD(Probability Of Detection)为命中率,FAR(False Alarm Rate)为误报率,CSI(Critical Success Index)为临界成功指数; N_H 为卫星识别结果和自动站实测结果都有海雾的样本点个数; N_M 为卫星识别结果有海雾而自动站实测结果没有海雾的样本点个数; N_F 为卫星识别结果没有海雾而自动站结果有

海雾的样本点个数。

3.1 阈值法对海雾识别有效性

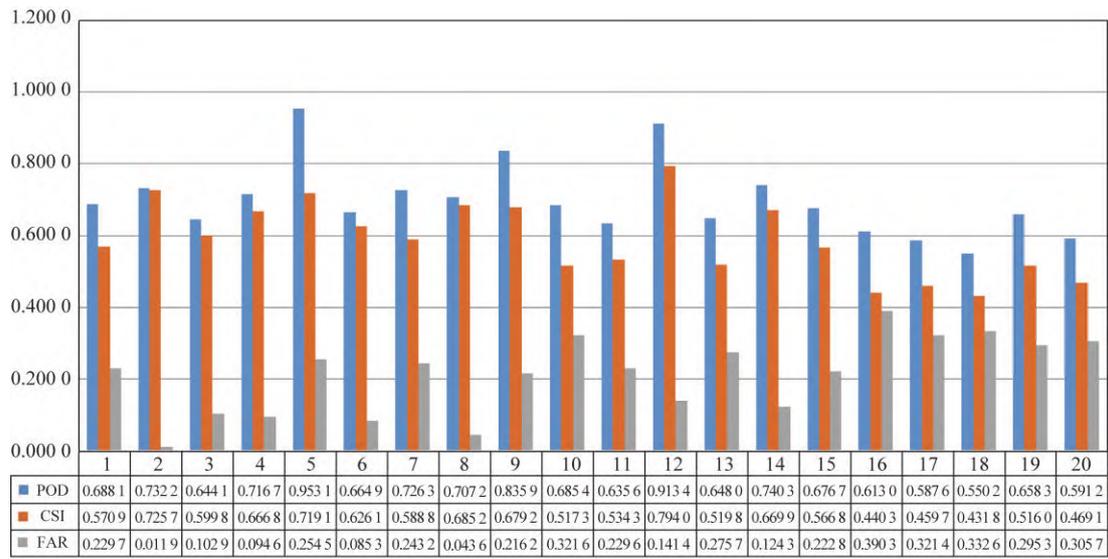
利用阈值法识别模型对每个个例的检验输入样本集进行海雾雾区识别,统计结果见图3。20个检验个例中,POD最大值为0.953 1,CSI最大值为0.794 0,FAR最大值为0.390 3;平均POD为0.698 4,平均CSI为0.589 0,平均FAR为0.212 1。综上说明,基于阈值法FY-4A卫星沿海海雾识别模型能够有效地反映实况海雾的分布情况。

3.2 随机森林模型对海雾识别有效性

利用随机森林识别模型对每个个例的检验输入样本集进行海雾雾区识别,统计结果见图4。20个检验个例中,POD最大值为0.932 7,CSI最大值为0.871 9,FAR最大值为0.143 3;平均POD为0.834 6,平均CSI为0.794 6,平均FAR为0.057 0。综上说明,基于随机森林FY-4A卫星沿海海雾识别模型能够更精确地反映实况海雾的分布情况。

3.3 不同海雾识别模型对比分析

通过对两种海雾识别模型的对比分析,我们可以看出,相比阈值法,随机森林模型在平均POD和



样本个例

图3 基于阈值法FY-4A卫星沿海海雾识别模型检验结果图

Fig.3 Test results of FY-4A satellite coastal sea fog recognition model based on threshold method

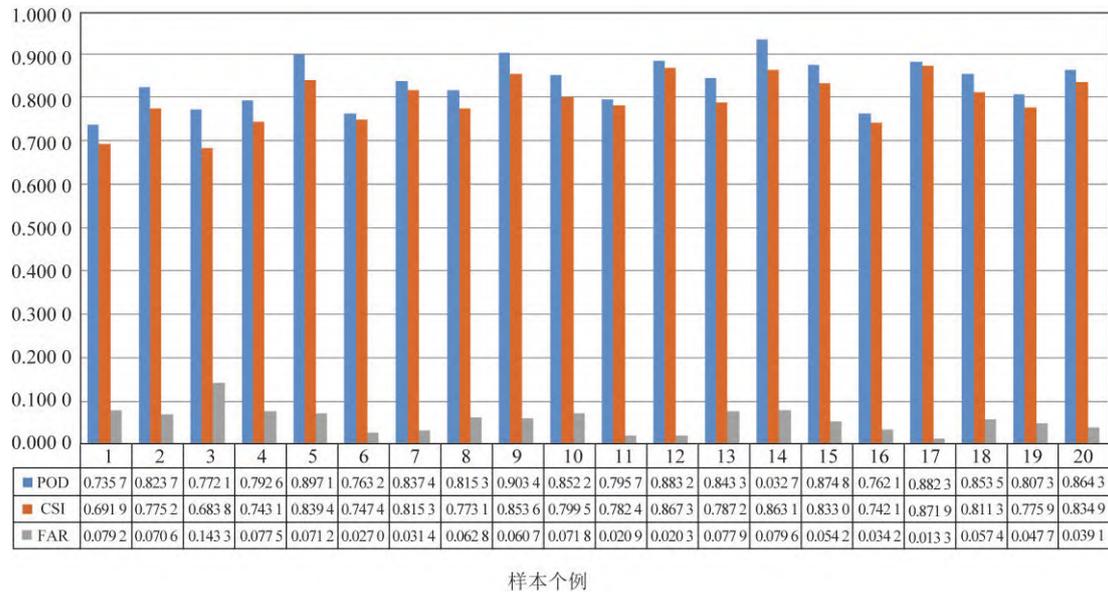


图4 基于随机森林FY-4A卫星沿海海雾识别模型检验结果图
 Fig.4 Test results of FY-4A satellite coastal sea fog recognition model based on random forest

平均CSI具有较大的提升,同时平均FAR大幅降低,这说明随机森林模型具有更高的精准性。另外,与阈值法中海雾识别效果忽高忽低相比,随机森林模型具有更好的稳定性,对于识别难度较大的海雾,随机森林模型也能够有较好的效果。

为了更好地揭示随机森林模型的性能,本文选取2021年4月12日黄渤海区域发生的海雾天气过程进行分析。我们分别利用这两种海雾识别模型对10:30(北京时,下同)时刻卫星遥感影像进行海雾雾区识别。

图5分别为FY-4A卫星华东区域3个红外通道亮度温度图以及经过随机森林模型和阈值法识别后的海雾雾区图。红外通道3.75 μm、7.33 μm和10.8 μm亮度温度对雾的识别具有重要作用,确实能够从图5a—c的亮温颜色看到与识别海雾雾区具有很大重叠,同时可以发现随机森林算法识别雾区更贴合(绿色为识别出的海雾雾区)。为了检验本次海雾雾区识别的实际效果,选取了所涉区域6个自动站作为检验样本点,通过查询“天擎”系统获得的该时次6个自动站能见度值都小于1 000 m,应该判

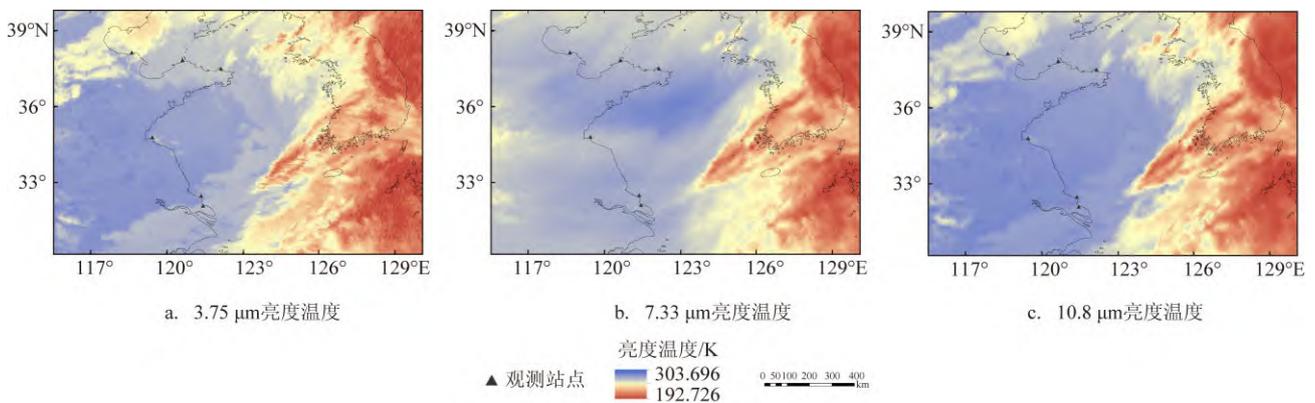


图5 2021年4月12日10:30时刻海雾识别图
 Fig.5 Sea fog identification map at 10:30(UTC+8)on April 12, 2021

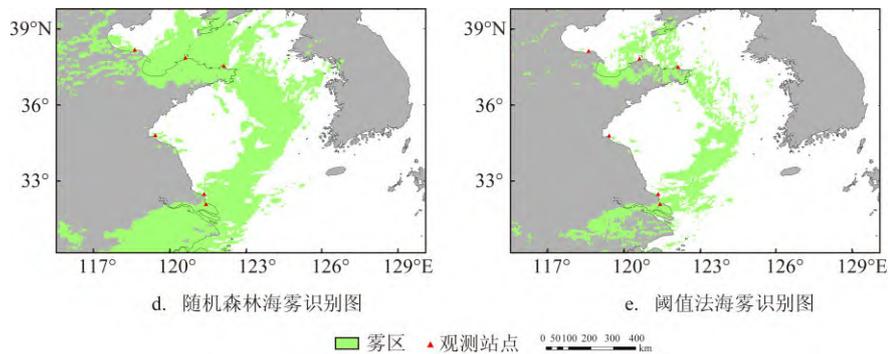


图5 (续)

Fig.5 (Continued)

定为有海雾发生。通过对比图5d和5e发现,阈值法未识别到1个自动站海雾发生点,通过分析本次海雾发生过程,可知2021年4月12日上午该区域海雾正在逐步生成,该时次海雾雾区正在逐步变大,对那些刚刚达标生成海雾雾区的识别难度大幅度增加,这可能是造成阈值法未识别到的原因,而随机森林模型因训练样本点较充分,能够准确识别到这个海雾雾区,因此说明随机森林模型具有更精准的识别效果。

4 结论与讨论

本文选取2019年8月—2021年7月江苏省及周边省份60个大雾天气个例,分别用阈值法和随机森林算法结合FY-4A卫星AGRI中4 km分辨率的14个波段数据对研究区域的沿海海雾雾区进行识别。结论如下:

(1) 阈值法海雾识别模型对海雾具有一定的识别能力,但对不同时期和不同阶段的海雾天气卫星遥感影像,需要取不同阈值范围,才能更好地识别出海雾区域。

(2) 通过40个个例的训练个例集建立的随机森林海雾识别模型,具有较高的精准性,该模型的参数配置合理。

(3) 分别用阈值法和随机森林模型对检验个例集的20个个例进行海雾雾区识别,对比阈值法,随机森林模型具有更精准的识别能力和更稳定的识别性能,对于识别难度较大的海雾,随机森林模型表现更加优异。

本文尝试利用阈值法和随机森林两种算法来实现海雾识别,随机森林算法在海雾识别方面具有较大的应用潜力,相比传统的阈值法,效果提高显著。虽然随机森林模型在训练时取得较高的精度,但是在实际识别中,当外在条件复杂或者遇到异物同波谱时,会出现错误识别。今后如果要将海雾识别结果投入到业务应用中,在训练建模时,必须提供更加多样的海雾天气个例,增加模型的容错性,其次,必须将随机森林算法与其他算法相结合,建立更加健壮的模型,提高模型的性能,在不同复杂条件下对海雾精准识别,从而实现FY-4A气象卫星海雾识别的业务化。

参考文献:

- [1] 王宏斌, 张志薇, 刘端阳, 等. 基于葵花8号新一代静止气象卫星的夜间雾识别[J]. 高原气象, 2018, 37(6): 1749-1764.
Wang H B, Zhang Z W, Liu D Y, et al. Detection of fog at night by using the new geostationary satellite Himawari-8[J]. Plateau Meteorology, 2018, 37(6): 1749-1764.
- [2] 郑新江. 黄海海雾的卫星云图特征分析[J]. 气象, 1988, 14(6): 7-9, 65.
Zheng X J. On satellite imagery features of sea fogs over the Yellow Sea[J]. Meteorological Monthly, 1988, 14(6): 7-9, 65.
- [3] Ellrod G P. Advances in the detection and analysis of fog at night using GOES multispectral infrared imagery[J]. Weather and Forecasting, 1995, 10(3): 606-619.
- [4] 鲍献文, 王鑫, 孙立潭, 等. 卫星遥感全天候监测海雾技术与应用[J]. 高技术通讯, 2005, 15(1): 101-106.
Bao X W, Wang X, Sun L T, et al. The weatherproof detection system of sea fog by remote sensing and its applications[J]. Chinese High Technology Letters, 2005, 15(1): 101-106.
- [5] 何月, 张小伟, 杜惠良, 等. 利用静止气象卫星监测浙江海上大雾

- [J]. 遥感技术与应用, 2015, 30(3): 599-606.
- He Y, Zhang X W, Du H L, et al. Monitoring sea fog of Zhejiang from geostationary meteorological satellite data[J]. Remote Sensing Technology and Application, 2015, 30(3): 599-606.
- [6] Shang H Z, Chen L F, Letu H, et al. Development of a daytime cloud and haze detection algorithm for Himawari-8 satellite measurements over central and eastern China[J]. Journal of Geophysical Research: Atmospheres, 2017, 122(6): 3528-3543.
- [7] 张培, 吴东. 基于Himawari-8数据的日间海雾检测方法[J]. 大气与环境光学学报, 2019, 14(3): 211-220.
- Zhang P, Wu D. Daytime sea fog detection method using Himawari-8 data[J]. Journal of Atmospheric and Environmental Optics, 2019, 14(3): 211-220.
- [8] 衣立. 黄海海雾/层云的空间分布及云底高度遥感方法研究[D]. 青岛: 中国海洋大学, 2015.
- Yi L. Spatio-temporal detection of sea fog/stratus and cloud base height over yellow sea with satellite data-A feasibility study[D]. Qingdao: Ocean University of China, 2015.
- [9] Wang B, Dong L L, Zhao M, et al. An infrared maritime target detection algorithm applicable to heavy sea fog[J]. Infrared Physics & Technology, 2015, 71: 56-62.
- [10] 张春桂, 林炳青. 基于FY-2E卫星数据的福建沿海海雾遥感监测[J]. 国土资源遥感, 2018, 30(1): 7-13.
- Zhang C G, Lin B Q. Application of FY-2E data to remote sensing monitoring of sea fog in Fujian coastal region[J]. Remote Sensing for Land & Resources, 2018, 30(1): 7-13.
- [11] 孙艺, 杨悦, 甄晴. CALIPSO卫星资料的春夏季黄海海雾高度特征分析[J]. 海洋预报, 2020, 37(3): 54-61.
- Sun Y, Yang Y, Zhen Q. The characteristics of the top height of sea fog over the Yellow Sea in spring and summer based on CALIPSO satellite data[J]. Marine Forecasts, 2020, 37(3): 54-61.
- [12] 于海鹏, 田付友. GOES9云图在黄海海雾区域识别中的应用[J]. 海洋预报, 2010, 27(3): 23-29.
- Yu H P, Tian F Y. Application of GOES9 cloud images in sea fog in the Yellow Sea[J]. Marine Forecasts, 2010, 27(3): 23-29.
- [13] Kim D, Park M S, Park Y J, et al. Geostationary ocean color imager (GOCI) marine fog detection in combination with Himawari-8 based on the decision tree[J]. Remote Sensing, 2020, 12(1): 149, doi: 10.3390/rs12010149.
- [14] Shin D, Kim J H. A new application of unsupervised learning to nighttime sea fog detection[J]. Asia-Pacific Journal of Atmospheric Sciences, 2018, 54(4): 527-544.
- [15] 许赞, 许艾文. 基于随机森林的遥感影像云雪雾分类检测[J]. 国土资源遥感, 2021, 33(1): 96-101.
- Xu Y, Xu A W. Classification and detection of cloud, snow and fog in remote sensing images based on random forest[J]. Remote Sensing for Land & Resources, 2021, 33(1): 96-101.
- [16] 姜红, 何清, 曾晓青, 等. 基于随机森林和卷积神经网络的FY-4A号卫星沙尘监测研究[J]. 高原气象, 2021, 40(3): 680-689.
- Jiang H, He Q, Zeng X Q, et al. Sand and dust monitoring using FY-4A satellite data based on the random forests and convolutional neural networks[J]. Plateau Meteorology, 2021, 40(3): 680-689.
- [17] 张环宇, 唐伯惠. 融合物理机理与随机森林算法的FY-4A AGRI数据晴空大气可降水量遥感反演[J]. 遥感学报, 2021, 25(8): 1836-1847.
- Zhang H Y, Tang B H. Remote sensing retrieval of total precipitable water under clear-sky atmosphere from FY-4A AGRI data by combining physical mechanism and random forest algorithm[J]. National Remote Sensing Bulletin, 2021, 25(8): 1836-1847.
- [18] 柳青青, 孟朔羽, 徐茗, 等. 随机森林反演卫星遥感海表面盐度研究[J]. 武汉大学学报·信息科学版, 2021, doi: 10.13203/j.whugis20210153.
- Liu Q Q, Meng S Y, Xu M, et al. Satellite sea surface salinity retrieval using random forest model[J]. Geomatics and Information Science of Wuhan University, 2021, doi: 10.13203/j.whugis20210153.

Research on a daytime sea fog identification model based on FY-4A satellite data and random forest algorithm

GENG Dan¹, LIU Tingting², LI Chao³

(1. Jiangsu meteorological information center, Nanjing 210041 China; 2. Jiangsu meteorological service center, Nanjing 210041 China; 3. Jiangsu Meteorological Observatory, Nanjing 210041 China)

Abstract: Using the FY-4A satellite data from August 2019 to July 2021 combined with the visibility observation data of automatic weather stations in Jiangsu province and surrounding areas in the same period, a satellite pixel set including sea fog and non-sea fog is established, and the daytime sea fog identification in Jiangsu and

surrounding areas is realized based on a sea fog recognition model that is constructed using random forest algorithm. The validation results show that the random forest sea fog recognition model trained in this paper has higher recognition accuracy compared with the sea fog recognition model based on threshold method, and the average hit rate, average critical success index and average false positive rate of the model are 83.46%, 79.46% and 5.7%, respectively. At the same time, by comparing the identification results of the two identification models for a sea fog weather case in the Yellow Sea and Bohai Sea on April 12, 2021, it shows that the random forest sea fog identification model can better identify the sea fog area.

Key words: FY-4A satellite; visibility observation data; satellite pixel set; random forest; daytime sea fog identification