夏进一,官莉.2024.基于随机森林算法优选FY-4B AGRI通道订正雨量产品研究[J].暴雨灾害,43(5):598-606. XIA Jinyi, GUAN Li. 2024. The study on precipitation products corrected by optimal selection FY-4B AGRI channel based on random forest algorithm [J]. Torrential Rain and Disasters,43(5):598-606 (in Chinese). doi:10.12406/byzh.2023-194

基于随机森林算法优选FY-4B AGRI通道订正雨量产品研究

夏进一,官莉
(南京信息工程大学中国气象局气溶胶与云降水重点开放实验室,南京 210044)

摘　要:为提高卫星反演地面定量降水产品的精度,基于2022年6月1日—7月31日FY-4B静止轨道辐射成像仪AGRI(Advanced Geosynchronous Radiation Imager)中国区域一级辐射观测产品、二级云检测产品、二级定量降水估计业务产品及国家站雨量观测,利用随机森林算法分别建立了白天和夜间的降雨有无判断模型和降雨量反演模型,评估了AGRI一级辐射观测15个通道重要性,筛选白天和夜间适合判断和反演地面雨量的优选通道并建立各自优选通道判断和反演模型,最后依据各自优选通道模型反演2022年8月的降水并与FY-4B AGRI的二级定量降水估计业务产品进行检验对比,结果表明:(1) 白天中波红外通道7和8重要性较低,白天判断和反演模型不使用通道7和8,夜间判断和反演模型使用所有红外通道,反演的降水量与国家站雨量观测最接近;(2) 白天和夜间优选通道判断模型对有无降水的判断均优于FY-4B降水业务产品,尤其是命中率大大提高,但两种优选通道雨量反演模型反演雨量都偏多,白天反演雨量的准确率高于夜间;(3) 白天和夜间两种优选通道雨量反演模型反演雨量的误差均小于FY-4B降水业务产品,降水业务产品和两种优选通道模型反演结果均高估小雨雨量、低估大雨雨量,降水强度越强低估越严重;(4) 对于优选通道反演模型不同量级降水的反演,除夜间暴雨外,白天和夜间其他量级降水的均方根误差均较降水业务产品有所提升,其中小到中雨的精度提升最明显。

关键词:FY-4B AGRI; 通道选择; 随机森林算法; 地面降水
中图法分类号:P405　　**文献标志码**:A　　**DOI**:10.12406/byzh.2023-194

The study on precipitation products corrected by optimal selection FY-4B AGRI channel based on random forest algorithm

XIA Jinyi, GUAN Li

(*China Meteorological Administration Aerosol-Cloud and Precipitation Key Laboratory, Nanjing University of Information Science and Technology, Nanjing* 210044)

Abstract: To improve the accuracy of quantitative precipitation products derived from regional satellite observations, the level 1 radiation observation products, level 2 cloud detection products, and level 2 quantitative precipitation estimation operational products from the FY-4B geostationary radiation imager AGRI (Advanced Geosynchronous Radiation Imager), and national rain gauge observations collected between June 1 and July 31, 2022, in China were utilized in this study. The random forest algorithm was employed to establish separate models for predicting the presence or absence of rainfall and estimating rainfall amounts during the day and at night. The importance of 15 channels from the AGRI level 1 radiation observations was evaluated, and suitable channels for determining and estimating ground rainfall during the day and night were selected. Judgment and retrieval models based on these optimal channels were then created. Finally, using the models based on their respective optimal channels, precipitation for August 2022 was estimated and compared with the FY-4B AGRI level 2 quantitative precipitation estimation operational products. The results are as follows: (1) Channels 7 and 8 in the mid-wave infrared range were of lower importance during the day, and the daytime judgment and retrieval models did not use these channels. However, the nighttime models employed all infrared channels, and the retrieval precipitation closely matched the national gauge observations. (2) The optimal channel judg-

收稿日期: 2023-09-18;**定稿日期**:2024-03-18
资助项目: 国家自然科学基金项目(41975028)
第一作者: 夏进一,主要从事气象卫星资料应用及二次产品开发。E-mail: 202212050016@nuist.edu.cn
通信作者: 官莉,主要从事气象卫星资料应用及二次产品开发、资料同化研究。E-mail: liguan@nuist.edu.cn
© Editorial Office of *Torrential Rain and Disasters*. OA under CC BY-NC-ND 4.0

ment models for both day and night outperformed the FY-4B precipitation operational products in identifying whether precipitation occurred, particularly showing a significant improvement in hit rates. Nevertheless, rainfall retrieval from both optimal channel retrieval models tended to overestimate rainfall amounts, with daytime retrieval precipitation being more accurate than nighttime. (3) The errors of rainfall retrieval from both daytime and nighttime optimal channel models were both lower than those from the FY-4B precipitation operational products. Both the precipitation operational products and the two optimal channel model estimates tended to overestimate light rainfall and underestimate heavy rainfall, with the underestimation becoming more severe as precipitation intensity increased. (4) For the retrieval of different precipitation magnitudes using optimal channel models, improvements in root mean square errors were observed for nearly all levels of rainfall except for nighttime heavy rain, with the most significant accuracy enhancement seen in the small to moderate rainfall range.

Key words: FY-4B AGRI; channel selection; random forest algorithm; surface rainfall

引　言

降水量是人们最关注的气象要素之一，其时空分布对气象、水文、生态以及经济、农业和其他相关学科研究有重要意义。传统的地面雨量站测量降水最准确可信，但由于中国的复杂地形和不同地区经济水平的差异，雨量站分布密度东西部差异很大，不能准确反映大尺度降水的时空特征(傅云飞，2019)。另外，由于降水在空间和时间上存在极大的不确定性，且地面雨量站和地基雷达测量范围有限，加之空间分布不均匀，在洋面和无人区几乎没有地面观测，使用插值的方法得到的大面积雨量和实际降水差别很大(Castro et al.，2014)。因此，如何得到准确且空间均匀的降水量是开展降水相关研究的首要问题，也是难点。

静止气象卫星在地球上空观测地球和大气可不受地理和自然条件限制，可实现大范围、全过程监测云系的发展演变，具有覆盖范围广、时空分辨率高的特点，能有效克服观测站网观测降水的缺陷(廖捷等，2013；刘瑜等，2017)。我国新一代风云四号静止气象卫星采用概率密度匹配方法生成定量降水估计产品，即通过统计在一定时空窗中的匹配星载被动微波降水和红外亮温的直方图和累积直方图，得到亮温到降水的传递特征参数(游然，2018)。此方法只用到了卫星的红外亮温进行降水反演，未考虑到可见光通道反照率对降水反演的作用。卫星反演降水产品的准确率仍然很低，需要进行订正以改善降水产品质量。

近年来，机器学习算法逐渐成为利用卫星多通道数据改善降水信息的有力工具。Bellerby等(2000)将地球静止业务环境卫星(Geostationary Operational Environmental Satellites)多光谱图像信息作为预测变量输入至神经网络进行训练以估计降水。Ma等(2018)基于葵花8号静止气象卫星成像仪AHI(Advanced Himawari Imager)多通道观测亮温及亮温差数据，用梯度提升树算法分别在三个时间段建立了地面降水反演子模型。这些算法突破了传统阈值法的单一性，将降水反演算法带领到新的思路和方法上。

FY-4B AGRI(Advanced Geosynchronous Radiation Imager)一级辐射观测包括15个通道产品，由于部分通道在白天时接收到的辐射信息较为复杂，全部15个通道能否用于地面降水反演值得商榷，目前还没有针对FY-4B静止轨道辐射成像仪(Advanced Geosynchronous Radiation Imager，AGRI)的15个通道是否适用于反演地面降水的研究。相较于其他机器学习模型，随机森林模型可给出每个变量的重要性，这为AGRI反演地面降水的通道筛选提供条件。

本文基于FY-4B静止轨道辐射成像仪AGRI一级辐射观测产品、二级云检测产品、二级定量降雨估计业务产品及国家站雨量观测，分别建立了白天、夜间基于随机森林算法的降雨有无判断模型和雨量反演模型，再根据AGRI的通道重要性评估筛选适合反演地面降雨量的优选通道，并重新建立白天和夜间优选通道降雨判断和雨量反演模型，最后以优选通道降雨判断和雨量反演模型反演2022年8月的雨量并检验精度，并与FY-4B AGRI的二级业务降雨产品进行了评估比较，以期提升卫星降雨产品的精度，为其他研究提供更精细准确率更高的卫星降水产品。

1 资料与方法

1.1 资料介绍

本文使用的资料包括FY-4B静止轨道辐射成像仪AGRI中国区域一级辐射观测产品、二级云检测产品、二级定量降水估计业务产品及国家站雨量观测，时间为2022年6月1日—8月31日。

1.1.1 FY4B AGRI资料介绍

静止气象卫星风云四号B星(FY-4B)于2021年6月3日成功发射，半年后投入业务试运行，搭载在其上的先进的静止轨道辐射成像仪AGRI具有很高的时空分辨率。本文使用资料共三种，分别为FY-4B AGRI全圆盘4 km一级辐射观测产品、二级云检测产品CLM (Cloud Mask)和二级定量降水估计业务产品QPE (Quantitative Precipitation Estimation)。本文仅使用全圆盘资料中的中国区域部分。

AGRI一级辐射观测产品共设置15个通道，表1列出了AGRI各通道的主要参数(AGRI仪器介绍来自http://www.nsmc.org.cn/nsmc/cn/instrument/AGRI.html)。通道权重函数代表了不同波段不同高度大气对卫星仪器的辐射贡献，峰值可以表示哪个层次的贡献是主要的(即有效信息层)(尹若莹等，2019)。其中前2个通道(0.47μm和0.65 μm)为可见光窗区通道，3—6(0.825 μm、1.379 μm和2.225 μm)为近红外通道，7—8为短波红外通道(中心波长均为3.75 μm，但是空间分辨率不一样)，该通道可同时接收目标物发射的红外辐射和反射的太阳辐射，9—11分别为权重函数峰值高度在对流层不同高度的水汽吸收通道，13—15(10.8 μm、12 μm和13.3 μm)为长波红外窗区通道。表1中1—6为通道观测的反照率值，无量纲，7—15通道为观测的亮温值，单位：K。

表1 FY-4B AGRI 15个通道参数及用途列表

Table 1 List of the parameters and uses of 15 channels from FY-4B AGRI

通道	中心波长/μm	带宽/μm	空间分辨率/km	主要观测目标与用途
1	0.47	0.45~0.49	1	小粒子气溶胶，真彩色图像合成
2	0.65	0.55~0.75	0.5	植被，图像导航配准，恒星观测
3	0.825	0.75~0.90	1	植被，水面上空气溶胶
4	1.379	1.371~1.386	2	卷云
5	1.61	1.58~1.64	2	低云/雪识别，水云/冰云判识
6	2.225	2.10~2.35	2	卷云、气溶胶，粒子大小
7	3.75	3.50~4.00 (high)	2	云等高反照率目标，火点
8	3.75	3.50~4.00 (low)	4	低反照率目标，地表
9	6.25	5.80~6.70	4	高层水汽(670 hPa)
10	6.95	6.75~7.15	4	中层水汽(750 hPa)
11	7.42	7.24~7.60	4	低层水汽(850 hPa)
12	8.55	8.3~8.8	4	云
13	10.80	10.30~11.30	4	云、地表温度等
14	12.00	11.50~12.50	4	云、总水汽量，地表温度
15	13.3	13.00~13.60	4	云、水汽

注：9-11通道括号内为权重函数峰值高度。

二级云检测产品可提供4类像素级云检测结果，包含晴空、可能晴空、可能云、及云；二级定量降水估计业务产品可以提供时空分辨率较高的同区域降水动态变化信息。上述资料的时间范围为2022年6月1日—8月31日，空间分辨率均为4 km，时间分辨率为1 h，数据来源于风云卫星遥感数据服务网(https://satellite.nsmc.org.cn/portalsite/default.aspx)。

1.1.2 *其他资料介绍及数据匹配*

中国区域共有2 164个地面气象站，使用和上述资料相同时间段国家站雨量观测小时累积降水量来构建训练样本并检验模型反演精度，数据来源于中国气象数据网。将同时间与地面站点距离最近的AGRI观测视场与国家站雨量观测位置进行时间、空间匹配，将定量降水估计业务产品与国家站观测降水量组成比较对。

1.2 随机森林算法介绍

随机森林算法通过集成学习的Bagging思想将多棵树集成，它的基本单元就是决策树(Breiman.，1999)。建立决策树时，从原始训练样本中有放回地随机抽取N组自变量和因变量数据，形成新的训练样本集；从所有自变量中不重复地随机抽取m个变量，依次用抽中的变量将因变量数据分成两份，并计算每种分割法得到的子集内部的纯净程度，使用纯净程度最高的分割法所用的变量来分割数据，完成该节点上的决策。反复二分数据使决策树最大限度生长，直到分割停止，完成单棵决策树的建立。以上步骤重复N_{tree}次构建包含N_{tree}颗决策树的随机森林模型。在本文中，使用训练好的模型进行预测时，向模型输入15个通道的观测数据，每棵决策树单独进行预测，分类树型将所有树的判断结果中出现次数最多的类别作为因变量类别，即有无降水；回归树类模型则取所有树输出结果的平均值作为最终输出结果，即具体的降水量(Quesada-Ruiz et al.，2022)。

随机森林算法可以根据纯净度的增加程度对每个输入自变量的重要性进行评价。分类树类型中自变量的重要性由基尼系数的平均降低程度来衡量，若某个变量分割后子集与总体样本的基尼系数之差的绝对值越大，说明该变量的分割效果越好，即重要性越高。基尼系数计算公式如下

$$G(D)=1-\sum_{k=1}^{K}\left(\frac{|C_k|}{|D|}\right)^2 \tag{1}$$

其中 C_k 表示样本总数 D 中属于第 k 类的样本子集，K 为类的个数。回归树类型中自变量的重要性由方差来衡量，分割后子集的方差越小，说明分割的效果越好，即重要性越高。基于随机森林这个特有的功能，在初步训练结束后，对AGRI 15个通道的判断和反演地面降水的重要性进行排序。保留重要性高的通道，剔除重要性低的通道，利用剩余的通道重新训练模型，并测试模型的准确度是否提升。

1.3 检验方法

使用空报率(R_{FA})、漏报率(R_{MA})和命中率(P_{OD})评估降水有无的准确率(Chen et al, 2020)

$$R_{\mathrm{FA}}=\frac{N_B}{N_{\mathrm{A}}+N_{\mathrm{B}}} \tag{2}$$

$$R_{\mathrm{MA}}=\frac{N_{\mathrm{C}}}{N_{\mathrm{A}}+N_{\mathrm{C}}} \tag{3}$$

$$P_{\mathrm{OD}}=\frac{N_{\mathrm{A}}}{N_{\mathrm{A}}+N_{\mathrm{C}}} \tag{4}$$

式中，N_A定义为卫星定量降水业务产品判断为有降水的正确站数，即卫星降水业务产品和站点观测均有降水的站数；N_B为空报站数，即卫星降水业务产品有降水但观测无降水的站数；N_C为漏报站数，即卫星降水业务产品无降水但观测有降水的站数，N_D表示卫星降水业务产品和观测均无降水的站数。

对卫星降水业务产品和观测均有降水的样本，统计两种降水量的平均偏差(M_B)、平均绝对误差(M_{AE})、均方根误差(R_{MSE})，评估卫星反演降水量的误差(韩芙蓉等，2023)，计算公式如下

$$M_{\mathrm{B}}=\frac{\sum_{i=1}^{n}(p_i-o_i)}{n} \tag{5}$$

$$M_{\mathrm{AE}}=\frac{\sum_{i=1}^{n}\left|p_i-o_i\right|}{n} \tag{6}$$

$$R_{\mathrm{MSE}}=\sqrt{\frac{\sum_{i=1}^{n}(p_i-o_i)^2}{n}} \tag{7}$$

p_i为卫星定量降水估计业务产品的降水量或模型反演的降水量，o_i为地面站点观测的降水量，n为命中的样本总数。i代表某时刻的某一像元或某一个地面雨量站。

2 模型的建立与验证

在训练随机森林模型之前，首要任务是确定并制定适当的训练集、验证集和测试集。代表性的训练集能够构建出真实反映现实情况的模型，而有效的验证集则能发现和改进模型因训练集数据质量差等原因引起的过拟合或欠拟合的问题。i代表某时刻的某一像元或某一个地面雨量站。

2.1 训练与验证样本集

将FY-4B 2022年6月1日—7月31日中国区域地面雨量计观测和时空匹配的AGRI一级辐射观测产品对分成训练样本集与验证样本集，分别占80%和20%。由于算法分两步，先判断卫星视场是否有降水，对识别为降水的视场再反演地表降水，所以样本分为降水判断样本A和雨量反演样本B。由于AGRI前6个通道位于太阳辐射波段范围内，在夜间反照率值接近于0，因此需要分别建立白天与夜间模型。白天样本的时次取02—08时(世界时，下同)，夜间样本的时次取14—20时，中间时间段(即09—13时和21—次日01时)卫星观测范围内白天与黑夜均存在，根据反照率是否接近0将视场分别归入白天或夜间样本。两个月内得到白天训练与验证样本A共321 330组、夜间训练与验证样本A共242 283组；为保证样本的多样性和代表性，结合AGRI云检测产品和地面降水，样本A中包含有云有降水、有云无降水和无云无降水(晴空)的三种情况，分别各占样本量的三分之一。再将样本A中地面观测到有降水的样本挑出来，构成白天反演样本B(107 110组)和夜间反演样本B(80 761组)。

2.2 测试样本集

将FY-4B 2022年8月的AGRI一级辐射观测产品用于算法检验，与时间、空间匹配的地面雨量测值构成测试样本集。测试样本集同样分为白天测试样本集和夜间测试样本集。

2.3 模型建立

经大量测试发现，若基于AGRI一级辐射观测产品直接用随机森林回归算法估计降水量，与地面站雨量计观测相比误差较大，尤其会将大量的晴空视场反演出降水量，导致降水落区判断有误。因此，建模分两步：先建立降水有无判断模型(简称判断模型)，再建立降水量反演模型(简称反演模型)，模型流程如图1所示。白天样本使用全部15个通道观测的平均小时反照率和平均小时亮温，夜间样本选用9个红外通道的平均小时亮温(即通道7－15)作为输入因子。

白天判断模型基于白天样本A，自变量为AGRI 15个通道观测的反照率和亮温，夜间则基于夜间样本A，输入为9个AGRI红外通道的亮温。白天和夜间判断模型都把地面站是否有降水(有降水1和无降水0)作为因变量，由于因变量是分类变量，因此训练判断模型时采用的分类树类型的决策树，经测试，每个模型都设置分类树N_{tree}为300棵。

对降水有无判断模型判断为降水的样本进一步反演其地面降水量。白天样本B用于训练白天降雨量

反演模型，夜间样本B则用于建立夜间降雨量反演模型，两者都是把地面观测的降水量作为因变量(输出)。由于因变量是连续变量，因此训练反演模型时采用的回归树类型的决策树，同样经过测试每个模型设置300棵回归树(钟宇璐，2021)。

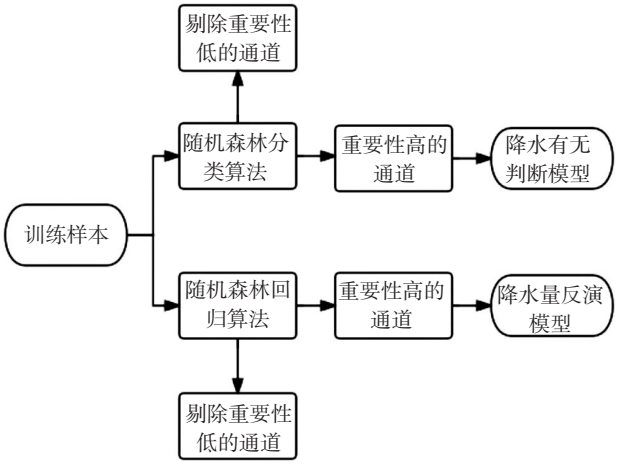


图1 随机森林模型建立流程图

Fig. 1 Random forest model development process diagram

2.4 AGRI优选通道筛选

针对上述初步建立好的四个模型，对每个输入通道的重要性进行评估，如图2所示。其中图2a、b分别为白天判断、反演模型15个通道的重要性图；图2c、d为夜间判断反演模型9个通道的重要性图。从图2a和b可看出白天模型中不论是降水有无判断还是反演降水量第7和8通道的重要性均较低，这是由于7和8通道的中心波长为3.75 μm，位于太阳短波辐射和地-气长波红外辐射的重合区，在白天接受到的辐射包括反射的太阳辐射和地面或云面发射的红外辐射两项，信息较为复杂，因此对白天模型的贡献量较小。而第1、4、5和6通道重要性较高，通道1中心波长为0.47 μm，太阳光谱在该波长辐射值最大，因此通道1对白天模型的贡献较大。通道4对冰晶粒子敏感，主要用于云粒子相态识别，卷云时反照率比较高、水云时较低，通常两层云时(即高层的冰云下面有暖云时)降水效率比较高，而只有高层卷云时地面往往没有降水。通道5和通道6中心波长分别为1.61 μm和2.225 μm，主要用于云粒子大小的识别，较小观测值往往对应大的云滴甚至雨滴，因此能反映云滴大小和云发展旺盛的程度。从图2c、d可看出，夜间模型各个通道贡献都很大，尤其是几个水汽吸收通道。夜间由于没有反射的太阳辐射的影响，通道7和8的重要性大大提高。FY-4B中这两个通道的空间分辨率均为2 km，但通道7对高温端敏感，因此，图像上较暖的晴空区和云区的灰度对比度增加，有利于判断是否会出现降水，所以图2c夜间降雨有无判断模型中通道7的贡献稍大于通道8。而地面能出现降水的云往往都是垂直发展比较高的，这时云顶亮温比较低，通道7在低温端对温度变化不敏感，但通道8对低温端敏感，因此对反演降水量的贡献是通道8高于通道7。

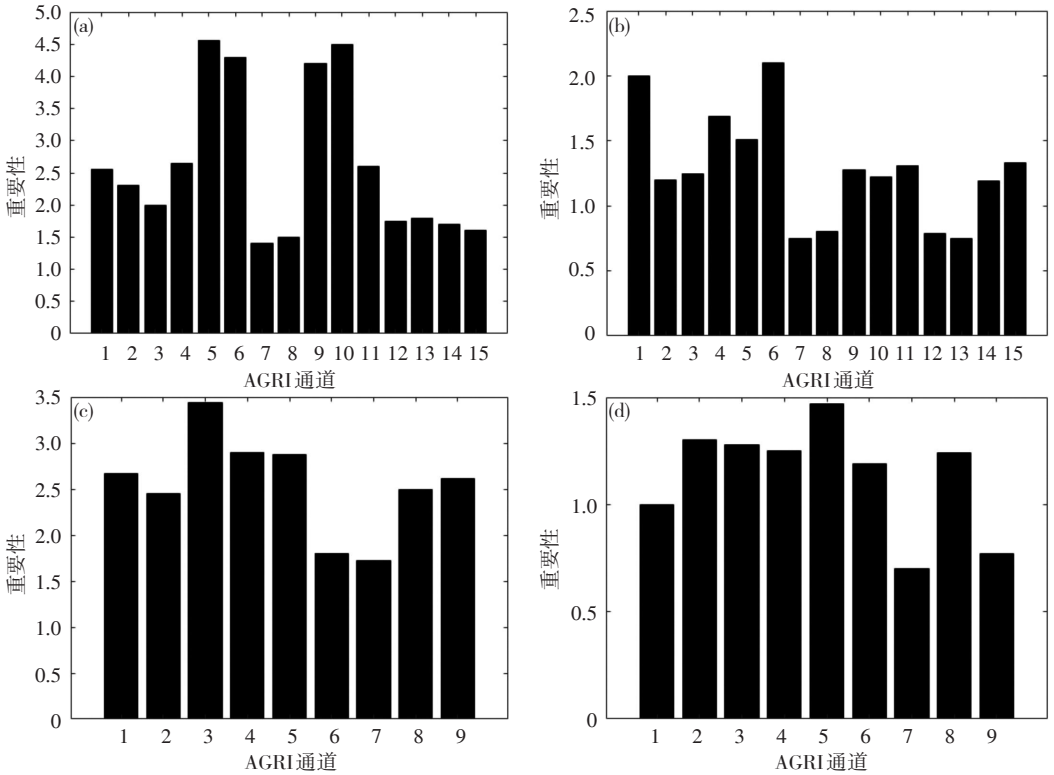


图2 白天(a、b)和夜间(c、d)降雨有无判断模型(a、c)、降雨量反演模型(b、d)不同输入通道的重要性直方图

Fig. 2 The histograms of the importance of different input channels for (a, b) daytime and (c, d) nighttime (a, c) rainfall judgment model and (b, d) rainfall retrieval model

鉴于中波红外通道7和8的特殊属性和白天较低的重要性，使用2022年8月1日—31日FY-4B AGRI一级辐射观测产品分别测试了这两个通道对判断模型和反演模型精度的影响。测试结果见表2和表3，表2为全部使用、分别弃用和全弃用7、8通道对降水有无判断准确度的统计，表3则为反演降水量的精度对比。

从表2可以看出，弃用通道7和8，白天判断模型结果准确率提高，表现在命中率提高、空报率和漏报率有所降低；而夜间判断模型不论是弃用通道7或8、还是同时弃用，模型的预测准确率均降低。表3同样展示，不使用通道7和8，白天反演的降水量平均误差和均方根误差均降低(即精度高)，夜间则使用全部9个通道的反演模型结果精度最高。这是由于白天时7和8通道接收的地气长波红外辐射中混杂了太阳短波辐射，白天反演模型增加这两个通道反而使得精度降低。而夜间没了反射太阳辐射的干扰，这两个窗区通道多增加的信息使得模型的判断和反演能力增强。因此，在实际应用中白天判断模型和反演模型不使用通道7和8，即用剩余的13个通道的数据参与模型训练；夜间判断模型和反演模型依旧采用包含7和8通道在内的9个通道训练网络。

表2 第7、8通道对降水有无判断模型的影响评估

Table 2 The impact assessment of channels 7 and 8 on the precipitation judgment model

模型名称	空报率	漏报率	命中率
白天降水有无判断模型(15个通道)	0.531 8	0.250 8	0.749 2
白天降水有无判断模型(弃用7通道)	0.531 1	0.196 5	0.803 5
白天降水有无判断模型(弃用8通道)	0.532 9	0.203 8	0.796 2
白天降水有无判断模型(弃用7、8通道)	0.532 4	0.190 5	0.809 5
夜间降水有无判断模型(9个通道)	0.586 5	0.505 8	0.494 2
夜间降水有无判断模型(弃用7通道)	0.583 1	0.574 4	0.415 6
夜间降水有无判断模型(弃用8通道)	0.590 7	0.587 3	0.422 7
夜间降水有无判断模型(弃用7、8通道)	0.587 4	0.610 6	0.389 4

表3 第7、8通道对降雨量反演模型的影响评估

Table 3 The impact assessment of channels 7 and 8 on the precipitation retrieval model

模型名称	平均误差	平均绝对误差	均方根误差
白天降水量反演模型(15个通道)	0.315 0	2.354 5	4.004 7
白天降水量反演模型(弃用7通道)	0.292 4	2.314 4	3.826 8
白天降水量反演模型(弃用8通道)	0.308 3	2.321 7	3.880 1
白天降水量反演模型(弃用7、8通道)	0.290 4	2.312 4	3.817 7
夜间降雨量反演模型(9个通道)	1.627 3	2.841 7	4.281 4
夜间降雨量反演模型(弃用7通道)	1.874 6	3.057 6	4.432 8
夜间降雨量反演模型(弃用8通道)	2.013 3	3.227 1	4.707 6
夜间降雨量反演模型(弃用7、8通道)	2.366 4	3.516 9	4.868 8

3 模型应用及检验

3.1 个例应用

为验证模型的普适性，随机挑选在中国区域存在较大范围降水的个例，以检验优选通道模型的订正效果。以2022年8月18日12—23时为例，该过程降水主要在华北至河套地区，华南部分地区也有降水，图3对比了中国区域降水观测和卫星相关的多种产品结果。可以看到，有云时可见光通道观测的反照率相对晴空区域较高，反照率越高对应云层越厚、甚至多层云，说明该区域降水概率越大，如图3d中黄色区域；低反照率的蓝色区域一般对应晴空。红外通道观测亮温对目标物的温度敏感，对流云中上升气流较强，使得云顶高度高，则对应亮温值低，表明该区域降水概率越大，如图3e中蓝色区域。对比发现优选通道判断模型判断的降水区域与图3e中观测亮温值较低的区域对应很好，但范围比图3d中对应高反照率区有所扩大，说明强对流云顶有一些高空扩散的卷云。从图3b中可看出优选通道判断模型判断的降水分布与大小与地面观测图3a接近，但存在一定程度的空报，使得雨带增宽，多是将高层卷云误判有降水。同样，AGRI二级定量降水业务产品也存在空报现象，使得雨带增宽。

3.2 独立样本检验

将地面站观测降水作为真值，用2022年8月整月

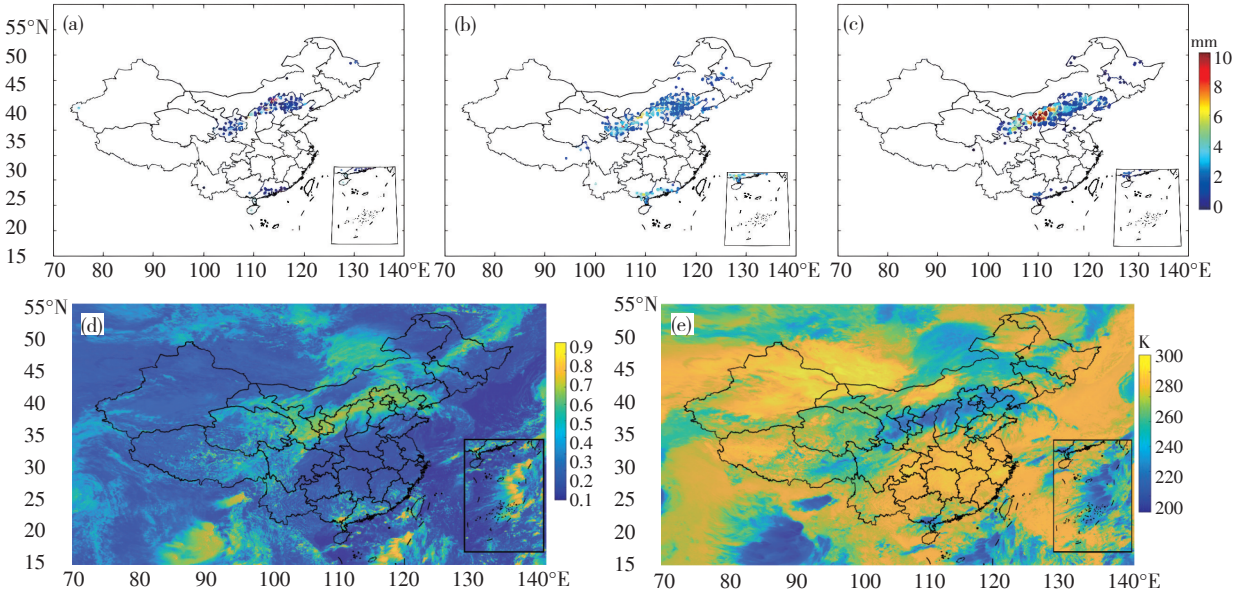


图3 2022年8月18日12时中国区域地面观测降水量(a)、随机森林模型反演降水量(b)、FY-4B AGRI业务产品降水量(c)、FY-4B AGRI 0.47 μm通道小时平均观测反照率(d)和FY-4B AGRI 10.8 μm通道小时平均观测亮温(e)分布

Fig. 3 The distribution of (a) surface observed precipitation in China, (b) random forest model retrieval precipitation, (c) precipitation from FY-4B AGRI operations, (d) hourly average observed albedo of the FY-4B AGRI 0.47 μm channel, (e) hourly average observed brightness temperature of the FY-4B AGRI 10.8 μm channel at 12:00 BT on August 18, 2022

AGRI一级辐射观测产品应用到降雨判断和雨量反演优选通道模型,并和同时段的FY-4B二级定量降水估计业务产品进行评估比较。一个月中FY-4B降水产品共匹配上634 052个白天样本、547 492个夜间样本。表4分白天和夜间给出了优选通道判断模型和AGRI二级定量降水估计业务产品对降水有无的判断准确率。

对比发现,白天优选通道判断模型结果的命中率较FY-4B定量降水估计业务产品提升了近35%;FY-4B定量降水估计业务产品的空报率较高,优选通道判断模型结果和业务产品比较,漏报率大大降低,同时空报率也稍有改善。夜间与白天类似,优选通道判断模型在降水有无的判断上均优于FY-4B的定量降水估计业务产品,尤其是命中率大大提高。优选通道判断模型结果白天的命中率比夜间高、空报率和漏报率较低,说明夜间优选通道判断模型对降水有无的判断准确率不如白天高,这是由于白天模型的解释变量中加入了可见光通道的反照率数据。当视场中有云时,卫星在可见光通道观测的反照率比晴空区域高,且云层越厚或者存在多层云时反照率越高,说明该视场存在降水的可能性越大。可见光通道的加入给模型提供了更多的判断信息,这导致白天优选通道判断模型相较于只依据红外通道判断的夜间模型准确率更高。

表4 2022年8月优选通道判断模型不同时段降水有无的对比检验

Table 4 The comparison and verification of precipitation presence or absence in different periods based on the optimal channels judgment model in August 2022

检验指标	白天		夜间	
	优选通道判断模型	FY-4B降水产品	优选通道判断模型	FY-4B降水产品
空报率	0.532 4	0.674 2	0.586 5	0.678 7
漏报率	0.190 5	0.636 7	0.505 8	0.705 9
命中率	0.809 5	0.363 3	0.494 2	0.294 1

表5列出了2022年8月优选通道雨量反演模型反演降水量和AGRI二级定量降水估计业务产品的误差,二者的平均偏差均为正,说明二者对雨量的估计都偏高。白天优选通道反演模型结果的平均偏差大于FY-4B定量降水估计业务产品,但平均绝对误差和均方根误差均小于定量降水估计业务产品,均方根误差减少超过2 mm,说明白天优选通道反演模型反演的降雨量与真值更接近。夜间两种产品的偏差同样为正值,且均比白天偏大,从绝对误差和均方根误差角度而言,反演产品精度较高。因此,无论是AGRI二级定量降水估计业务产品还是优选通道反演模型,白天反演降水量的准确率均高于夜间。

表5 2022年8月优选通道反演模型不同时段反演降水量的精度检验

Table 5 The accuracy verification of precipitation in different periods based on the optimal channels retrieval model in August 2022

检验指标	白天		夜间	
	优选通道反演模型	FY-4B降水产品	优选通道反演模型	FY-4B降水产品
平均误差	0.290 4	0.255 9	1.627 3	1.051 5
平均绝对误差	2.313 2	3.591 6	2.841 7	4.361 6
均方根误差	3.817 7	6.402 6	4.281 4	7.265 9

根据地面雨量站观测将小时降雨量划分不同强度等级(沈艳等，2013)：小雨(0.1~2.4 mm)、中雨(2.5~7.9 mm)、大雨(8~15.9 mm)和暴雨(≥16 mm)，分别统计了2022年8月FY-4B二级定量降水估计产品和优选通道反演模型反演雨量的平均误差(图4a)和均方根误差(图4b)。由图4a看出，不论业务产品还是反演产品，小雨时对雨量高估、而大雨时则低估，雨强越大低估越严重，雨量在中雨时与雨量计观测最接近。白天优选通道反演模型反演结果的均方根误差在各降水强度等级中是最小的，夜间优选通道反演模型也只是在暴雨时稍高，小到中雨时优选通道反演模型对定量降水估计业务产品精度的提高最明显，随着雨强的增加反演误差增大，说明优选通道反演模型反演结果的离散程度小于AGRI降水产品，反演结果更好。实际应用中，雨强较小时(小时雨量小于8 mm)本模型反演的降水量比较可靠，但雨强较大时(小时雨量大于8 mm)反演的降水量严重偏低，使用时需要慎重。

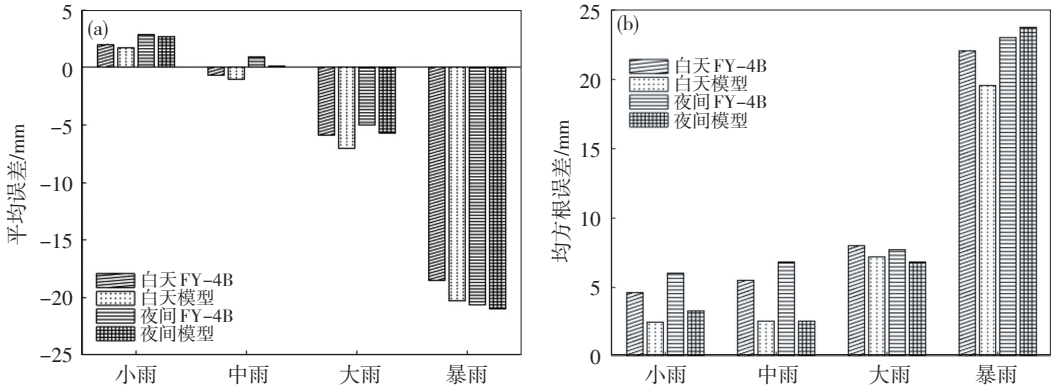


图4 2022年8月中国区域FY-4B降水估计业务产品和优选通道反演模型降雨量不同雨强的平均误差(a)和均方根误差(b)

Fig. 4 (a) The mean error and (b) mean square error of optimal channels retrieval model precipitation and FY-4B precipitation estimation service products in different precipitation intensities in August 2022 in China

4 结论和讨论

本文基于2022年6月1日至7月31日FY-4B AGRI一级观测辐射15个通道产品及时空匹配的国家站雨量，训练随机森林模型判断和反演地面降水量并优选适合白天和夜间的通道产品，建立判断和反演优选通道模型，并采用2022年8月的实况降水独立检验该模型反演结果的精度，和FY-4B卫星AGRI的二级业务降水产品进行比较，得到以下主要结论：

(1) 白天优选通道判断和反演模型不使用AGRI短波红外通道7和8、夜间优选通道判断和反演模型使用包括通道7和8在内的所有红外通道使得判断和反演的地表降水与国家站雨量观测最接近。

(2) 白天和夜间优选通道判断模型在降水有无的判断上均优于FY-4B二级定量降水估计业务产品，尤其是命中率大大提高。由于白天优选通道判断模型可见光和近红外通道的加入，使得对降水有无的判断准确率高于夜间优选通道判断模型。

(3) 以地面气象站逐小时降水资料作为真值，优选通道反演模型反演及定量降水估计业务产品雨量均普遍偏高，白天反演降水量的准确率均高于夜间；优选通道反演模型反演产品的绝对误差和均方根误差均小于FY-4B定量降水估计业务产品。

(4) 不论业务产品还是反演产品，小雨时对雨量高估、而大雨时则低估，雨强越大低估越严重。和夜间优选通道反演模型结果及白天、夜间定量降水估计业务产品比较，白天优选通道反演模型反演雨量的均方根误差在各种降水强度等级中最小，夜间模型也只是在暴雨时稍高，优选通道反演模型对业务产品精度的提高在小到中雨时最明显。

本模型反演雨量相对于地面站测量雨量存在偏差，可能涉及以下几个方面原因：(1) 随机森林模型的

不完善。在降水量反演训练样本中，小雨、中雨、大雨和暴雨的样本量非等比例分布，大雨和暴雨的样本量稀少，中雨占样本总量的比例很高，这导致小雨和中雨时反演结果较好，雨强越大时低估越严重。后续可通过训练样本中增加大雨和暴雨的样本数来提高强降水时的反演精度。当然，也包括一些随机森林关键模型参数的优化，如M_{try}、N_{tree}等。(2) AGRI一级辐射观测值存在观测误差、定标误差等。(3) AGRI仪器自身特性的局限。AGRI使用的是可见光和短波、长波红外通道，波长均比较短，对云没有穿透性，观测信息主要来自云顶，无法得知云中及以下的大气状态，因此反演降水会有误差，尤其是雨量比较大的强对流云。

参考文献(References)：

傅云飞. 2019.卫星主被动仪器遥感中国暴雨的研究进展[J].暴雨灾害,38(5): 554–563. Fu Y F. 2019. Satellite–borne active and passive instruments for remote sensing of heavy rain in China: A review [J]. Torrential Rain and Disasters,38(5): 554–563. doi:10.3969/j.issn.1004–9045.2019.05.0016

韩芙蓉,鹿翔,吴天贻,等.2023.多卫星融合降水产品对2015—2020年登陆浙江台风降水的监测能力评估[J].暴雨灾害,42(1):57–66. Han F R, Lu X, Wu T Y, et al. 2023. Evaluation of monitoring ability of the integrated multi–satellite retrievals for precipitation during typhoon landing in Zhejiang from 2015 to 2020 [J]. Torrential Rain and Disasters,42(1): 57–66 (in Chinese). doi:10.12406/byzh.2022–015

廖捷,徐宾,张洪政.2013.地面站点观测降水资料与CMORPH卫星反演降水产品融合的试验效果评估[J].热带气象学报,29(5):865–873. Liao J, Xu B, Zhang H Z. 2013. Evaluation of the experimental effect of the fusion of ground station observation precipitation data with CMORPH satellite inversion precipitation products [J]. Journal of Tropical Meteorology,29(5):865–873(in Chinese)

刘瑜,吴裕珍,冯志州,等.2017.多种卫星降水产品对中国极端降雨反演效果评估[J].热带地理,37(3):417–433. Liu Y, Wu Y Z, Feng Z Z, et al. 2017. Evaluation of the inversion effect of a variety of satellite precipitation products on extreme rainfall in China [J]. Tropical Geography,37(3):417–433 (in Chinese). doi:10.13284/j.cnki.rddl.002949.

沈艳,潘旸,宇婧婧,等.2013.中国区域小时降水量融合产品的质量评估[J].大气科学学报,36(1):37–46. SHEN Y, PAN Y, YU J J, et al. 2013. Quality assessment of hourly merged precipitation product over China [J]. Transactions of Atmospheric Sciences,36(1):37–46. (in Chinese)

杨轩, 曾燕, 邱新法, 等.2023. 基于机器学习算法的多源月尺度融合降水产品在中国区域的检验评估[J].暴雨灾害,42(5): 595–605. Yang X, Zeng Y, Qiu X F, Zhu X C. 2023. Examination and evaluation of multi–source monthly scale fusion precipitation product in China based on machine learning algorithm [J]. Torrential Rain and Disasters, 42(5):595–605. doi:10.12406/byzh.2023–006

游然.2018.卫星定量降水估计方法[C]//中国气象学会.第35届中国气象学会年会 S21 卫星气象与生态遥感,2018. 1. You R. 2018. Satellite quantitative precipitation estimation method [C]// The 35th Annual Meeting of the Chinese Meteorological Society S21 Satellite Meteorology and Ecological Remote Sensing,2018:1

尹若莹, 韩威, 高志球,等.2019.基于FY–4A卫星探测区域模式背景误差和观测误差估计的长波红外通道选择研究[J].气象学报,77(5): 898–910. Yin R Y, Han W, Gao Z Q, Wang Gen. 2019. A study on longwave infrared channel selection based on estimates of background errors and observation errors in the detection area of FY–4A [J]. Acta Meteorologica Sinica,77(5):898–910. doi:10.11676/qxxb2019.051

钟宇璐.2021.基于风云四号卫星AGRI观测用随机森林算法反演地面降水[D].南京:南京信息工程大学. Zhong Y L. 2021. Based on the observation of Fengyun–4 satellite AGRI, the random forest algorithm was used to invert the ground precipitation [D]. Nan Jing: Nanjing University of Information Science and Technology (in Chinese)

张强,赵煜飞,范邵华.2016.中国国家级气象台站小时降水数据集研制[J].暴雨灾害,35(2):182–186. Zhang Q, Zhao Y F, Fan S H. 2016. Development of hourly precipitation datasets for national meteorological stations in China [J]. Torrential Rain and Disasters,35(2):182–186. doi:10.3969/j.issn.1004–9045.2016.02.011

Bellerby T, Todd M, Kniveton D, et al. 2000. Rainfall estimation from a combination of TRMM precipitation radar and GOES multispectral satellite imagery through the use of an artificial neural network [J]. Journal of applied Meteorology,39(12):2115–2128. doi:10.1175/1520–0450(2001)040<2115:REFACO>2.0.CO;2

Breiman L.1999. Random Forests–Random Features [J]. Machine Learning.45(1): 5–32

Chen Y D, Shen J, Fan S Y, et al. 2020. Characteristics of Fengyun–4A satellite atmospheric motion vectors and their impacts on data assimilation[J]. Advances in Atmospheric Sciences,37(11):1222–1238. doi: 10.1007/s00376–020–0080–0

Castro M L, Giron á s J, Fern á ndez B. 2014. Spatial estimation of daily precipitation in regions with complex relief and scarce data using terrain orientation [J]. Journal of Hydrology,517(10–12):481–492. doi: 10.1016/j.jhydrol.2014.05.064

Ma L, Zhang G, Lu E. 2018. Using the gradient boosting decision tree to improve the delineation of hourly rain areas during the summer from advanced Himawari imager data [J]. Journal of Hydrometeorology,19(5): 761–776. doi:10.1175/JHM–D–17–0109.1

Quesada–Ruiz L C, Rodriguez–Galiano V F, Zurita–Milla R, et al. 2022. Area and feature guided regularised random forest: a novel method for predictive modelling of binary phenomena. The case of illegal landfill in Canary Island [J]. International Journal of Geographical Information Science,36(12):2473–2495. doi:10.1080/13658816.2022.2075879

（责任编辑　张　文）