文章编号：1672-8785(2023)07-0039-07

基于高光谱 GIIRS 红外亮温的大气三维风场反演研究

王　根[1]　袁　松[2]　叶　松[1]　谢　丰[2]　陈　娇[2]

(1. 巢湖学院电子工程学院，安徽 合肥　238000；
2. 安徽省气象台，安徽 合肥　230031)

摘　要：风场对于天气形势的演变和预报至关重要。基于风云四号 A 星干涉式大气垂直探测仪(GIIRS)中波通道资料和 ERA5 风场资料，采用 LightGBM 进行大气三维风场反演研究。首先，构建模型特征变量。GIIRS 通道最优选择采用二步特征选择法：(1)建立 GIIRS 通道黑名单；(2)采用置换特征重要性(Permutation Feature Importance，PFI)方法选择特征变量，在形成通道最优子集的基础上，构建含有时空信息的特征变量。其次，构建基于 LightGBM 的三维风场反演方法。最后，基于台风“利奇马”期间的 GIIRS 加密资料开展了 LightGBM 超参数优化和相关反演试验。结果表明，相对于 ERA5 风场资料，测试集中风场 U 和 V 分量的均方根误差(Root Mean Square Error，RMSE)分别小于 1 m/s 和 1.5 m/s。本文中的二步特征选择法能够实现 GIIRS 通道的动态最优选择。

关键词：FY-4A/GIIRS；大气风场反演；特征选择；LightGBM；台风“利奇马”

中图分类号：P407　　**文献标志码：**A　　**DOI：**10.3969/j.issn.1672-8785.2023.07.007

Retrieval of Atmospheric Three-Dimensional Wind Field Based on Hyperspectral GIIRS Infrared Brightness Temperature

WANG Gen[1], YUAN Song[2], YE Song[1], XIE Feng[2], CHEN Jiao[2]

(1. *School of Electronic Engineering, Chaohu University, Hefei 238000, China*;
2. *Anhui Meteorological Observatory, Hefei 230031, China*)

Abstract: Wind fields are crucial to the evolution and prediction of weather situations. Based on the medium wave channel data of GIIRS and the wind field data of ERA5, LightGBM is used to retrieve the three-dimensional atmospheric wind field in this paper. First, model feature variables are constructed. The two-step feature selection method is adopted for the optimal selection of GIIRS channels: (1) The blacklist of GIIRS chan-

收稿日期：2023-04-03

基金项目：安徽省高校杰出青年科研项目(2022AH020093)；安徽省重点研究与开发计划项目(2022h11020002)；安徽省自然科学基金项目(2108085QD183)；巢湖学院高层次人才科研启动经费项目(KYQD-202211)；巢湖学院学科建设质量提升工程项目(kj22zsys02)；国家自然科学基金项目(41805080)

作者简介：王根(1983-)，男，江苏泰州人，副高，博士，主要从事高光谱卫星资料同化、卫星估测降水、多源数据融合与人工智能可解释性等方面的研究。E-mail：203wanggen@163.com

nels is established; (2) Feature variables are selected by PFI method, and feature variables containing spatio-temporal information are constructed on the basis of forming optimal subsets of channels. Secondly, a three-dimensional wind field retrieval method based on LightGBM is constructed. Finally, LightGBM hyperparameter optimization and correlation retrieval experiments are carried out based on GIIRS encrypted data during Typhoon "Lekima". The experimental results show that RMSE of the U and V components of the wind field in the test set is less than 1 m/s and 1.5 m/s respectively, compared with the ERA5 wind field data. The two-step feature selection method in this paper can realize the dynamic optimal selection of GIIRS channels.

Key words: FY-4A/GIIRS; atmospheric wind field retrieval; feature selection; LightGBM; Typhoon "Lekima"

0 引言

大气是一个不断变化的动态系统[1]，而风场信息对于大气(天气)形势的演变和预报至关重要。从地面到大气层顶，风场可以量化分成多个层次，且每个层次的风速各不相同[2]。在低层大气中，风速随着气压层高度的增加而增大[3]。静止卫星资料能够实现大范围、快速和长期连续大气观测，在三维风场反演中具有显著的优势。

基于静止卫星观测资料，国内外学者已经开发了根据连续观测资料反演大气运动矢量的方法，称为“云导风”[4]。传统的云导风反演方法有基于三幅连续图像的特征检测法、跟踪法和图像处理中的光流法及相关方法的变体等[5]。孔德华等[6]提出了基于加速鲁棒特征图像匹配的云导风计算方法，其本质是改进了尺度不变特征变换的云导风计算方法。在多云情况下[7]，卫星可见光和长波红外资料可用于跟踪基于云运动的矢量；而在晴空情况下，水汽通道资料可用于跟踪水汽运动矢量的水分特征[8]。虽然上述方法对于风场反演取得了一定的效果，但其反演精度受高度分配的影响[8]。

风云四号 A 星搭载的 GIIRS 具有较高的时间分辨率，在特殊加密时期(比如有台风时)可达到 15 min。GIIRS 能对空间遥感地球温度、风场及大气成分的垂直分布实现大范围、快速和长期观测[9]。为了得到高频次的风场信息并解决高度分配问题，Ma Z 等[8]基于神经网络以及台风“玛莉亚”期间的 GIIRS 加密资料反演了大气风场信息。该研究将风场反演的结果与全球数据同化系统和 ERA5 风场资料对比后发现，对流层风场 U 分量和 V 分量的 RMSE 值均小于 2 m/s。

本文参考 Ma Z 等[8]的研究工作，基于 LightGBM[10]开展 FY-4A/GIIRS 中波红外通道亮温反演大气三维风场研究。

1 方法

1.1 基于特征重要性的通道最优选择

特征变量选择有助于降低数据维数和在模型开发之前提取信息量较大的特征变量，是机器学习建模中最重要的步骤之一。它可以将预测变量减少至几个重要变量，使模型更容易解释。在模型执行过程中，有些变量对提高模型预测精度的贡献可能不那么重要，也可能会降低模型的整体性能(如计算时效性等)，故需进行变量特征重要性分析。高光谱 GIIRS 通道多，这在应用其资料时较为重要。

本文中的 GIIRS 通道最优选择分两步：(1)建立 GIIRS 通道黑名单[11-12]；(2)采用 PFI 方法计算特征变量的重要程度，选取排序在前的特征变量，形成通道最优组合。

在建立通道黑名单时，本文采用 Tiros 业务垂直探测器辐射传输(RTTOV)模式[13]来模拟 FY-4A/GIIRS 中波通道亮温。背景场资料采用美国国家环境预报中心的最终全球资料同化系统资料，具体操作参考王根等[12]的研究工作。

在建立通道黑名单的基础上，本文采用 PFI 方法[14]计算特征变量的重要性。该方法的基本原理如下：首先打破目标和其中某个预测

因子之间的联系。其次，估算模型预测精度的下降程度。劣化程度越大，预测因子对提高模型的预测精度就越重要。最后，得到特征变量的重要性排序。

1.2　LightGBM

梯度提升是一种基于树的集成方法，通过组合弱模型进行预测。目前相对较新且快速的梯度提升方法主要有极端梯度提升(eXtreme Gradient Boosting，XGBoost)和 LightGBM。与 XGBoost 相比，微软公司提出的LightGBM[10]在性能和计算时间方面都有较大的改进。LightGBM 主要技术如下：(1)基于梯度的单边采样有助于选择信息量较大的观测数据；(2)利用高维数据通常具有的稀疏特性合并互斥特征。该稀疏性为设计数据接近无损降维提供了可能。因此，LightGBM 方法可能比较适用于高光谱红外探测器的多通道数据。反演方法的精度除了依赖于所选模型之外，也依赖于模型的超参数组合。本文基于均方误差优化了 LightGBM 模型的超参数。需要优化的超参数有学习率(learning_rate)、每棵树的叶子数量(num_leaves)、树的数量(n_estimators)。其它超参数采用模型的默认值。

2　数据集和预处理

本文将 FY-4A/GIIRS 中波通道亮温和欧洲中期天气预报中心(European Centre for Medium-Range Weather Forecasts，ECWMF)风场资料分别作为 LightGBM 模型的输入(特征变量、自变量)和输出(因变量)数据，分别构建了风场 U 分量和 V 分量的反演模型。

2.1　FY-4A/GIIRS 数据

FY-4A/GIIRS 是静止气象卫星携带的第一个高光谱红外大气垂直探测仪。GIIRS 的 1650 个通道覆盖 700～2250 cm^{-1} 的光谱区域，长波和中波通道数分别为 689 和 961。在高影响天气发生时，GIIRS 能够以高频次(如每 15 min)观测选定的需要监测的区域[8]。关于 GIIRS 的详细介绍可参考 Yang J 等[9]和 Yin R Y 等[15]的研究工作。本文使用的 FY-4A/GIIRS 资料来源于国家卫星气象中心官方网站(http://satellite.nsmc.org.cn/portalsite/default.aspx?currentculture=en-US)。

2.2　云检测产品

GIIRS 晴空和有云视场点的判识参考风云四号 A 星多通道扫描成像辐射计(AGRI)的全圆盘云检测产品(CLM)[16]。本文使用的 CLM 产品资料来源于国家卫星气象中心官方网站(http://satellite.nsmc.org.cn/portalsite/default.aspx?currentculture=en-US)。

2.3　ERA5 数据

本文将 ERA5 风场的 U 分量和 V 分量廓线数据作为 LightGBM 模型反演风场的输出值(因变量)和验证方法精度的基准。ERA5 作为 ECWMF 生成的第五代再分析产品，能够以 1 h 的时间分辨率和 0.25 的水平分辨率提供高精度的大气参数信息[17]。本文使用的 ERA5 风场资料来源于 ECWMF 官方网站(https://apps.ecmwf.int/datasets)。

2.4　数据预处理与试验数据

2.4.1　数据预处理

在风场 U 分量和 V 分量反演试验中，需对数据进行预处理，以提高数据的质量。本文采用“切趾”函数处理 FY-4A/GIIRS 观测资料[11]。通过“最近邻法”将 AGRI 的 CLM 产品插值至 GIIRS 视场点，以判断 GIIRS 视场点云量信息。本文参考文献[12]的研究工作，规定当云量小于 0.1 时，标记 GIIRS 视场点为晴空视场点。将此视场点的相关资料作为风场反演的样本。插值方法和过程可参考 Yin R Y 等[15]和 Zhang Q 等[18]的研究工作。

为了不引入其它误差，在构建机器学习模型样本时，以 ERA5 的分层(1～1000 hPa，共 37 层)为基准。取 FY-4A/GIIRS 和 ERA5 的整点数据，并将 ERA5 风场 U 分量和 V 分量插值至 GIIRS 视场点位置。

需要说明的是，借鉴 Ma Z 等[8]的研究成果，本文利用 GIIRS 入选通道子集的时空信息构建样本。采用选定的晴空视场点及周边四个

视场点亮温信息，以构建样本的“空间信息”。按照空间信息操作方式，加入待反演时次的前一时次的相关信息，以构建样本的“时间信息”。

2.4.2 试验数据

台风“利奇马”（国际编号：1909）是2019年太平洋台风季中的第9个风暴[19]。“利奇马”登陆时，中心附近最大风力为16级（52 m/s），使其成为1949年以来登陆浙江的第三强台风。台风“利奇马”共造成我国1402.4万人受灾，直接经济损失达500多亿元。本文的研究时间段为2019年8月9日0时至2019年8月9日15时（世界时），研究目标区域为12°N～50°N、98°E～161°E。

3 大气三维风场反演试验

接下来探讨FY-4A/GIIRS中波红外通道亮温反演风场U分量和V分量的可行性。首先进行LightGBM超参数优化分析；其次，分析影响反演精度的原因（归因分析）。训练和独立测试样本集分别有19319和4830个视场点数据。

图1所示为LightGBM不同超参数组合下训练和独立测试数据集风场U分量和V分量廓线反演的RMSE分布（单位为m/s）。选取超参数组合（num_leaves（40、50和60）、learning_rate（0.7、0.8和0.9）以及n_estimators（70、80和90））进行测试验证。为了更好地展示不同参数组合的U分量和V分量反演精度，图1仅给出了部分结果。

由图1可知，在不同num_leaves、learning_rate和n_estimators组合下，RMSE误差曲线显示出基本相同的变化规律。与其它超参数组合相比，三个超参数分别为60、0.7和90的组合下的风场U和V分量反演结果整体最优。在训练样本预测中，风场U分量和V分量整层（37层）RMSE分别小于0.0856 m/s和

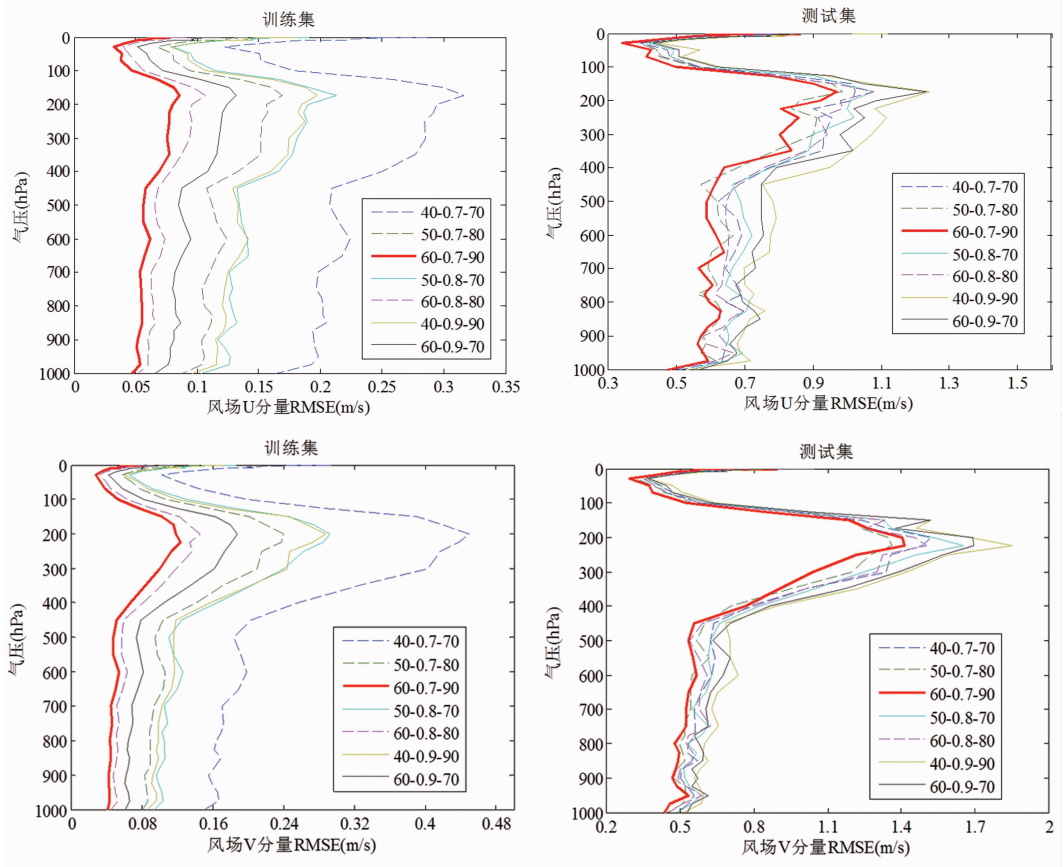

图1 LightGBM不同超参数组合下风场U和V分量反演精度对比

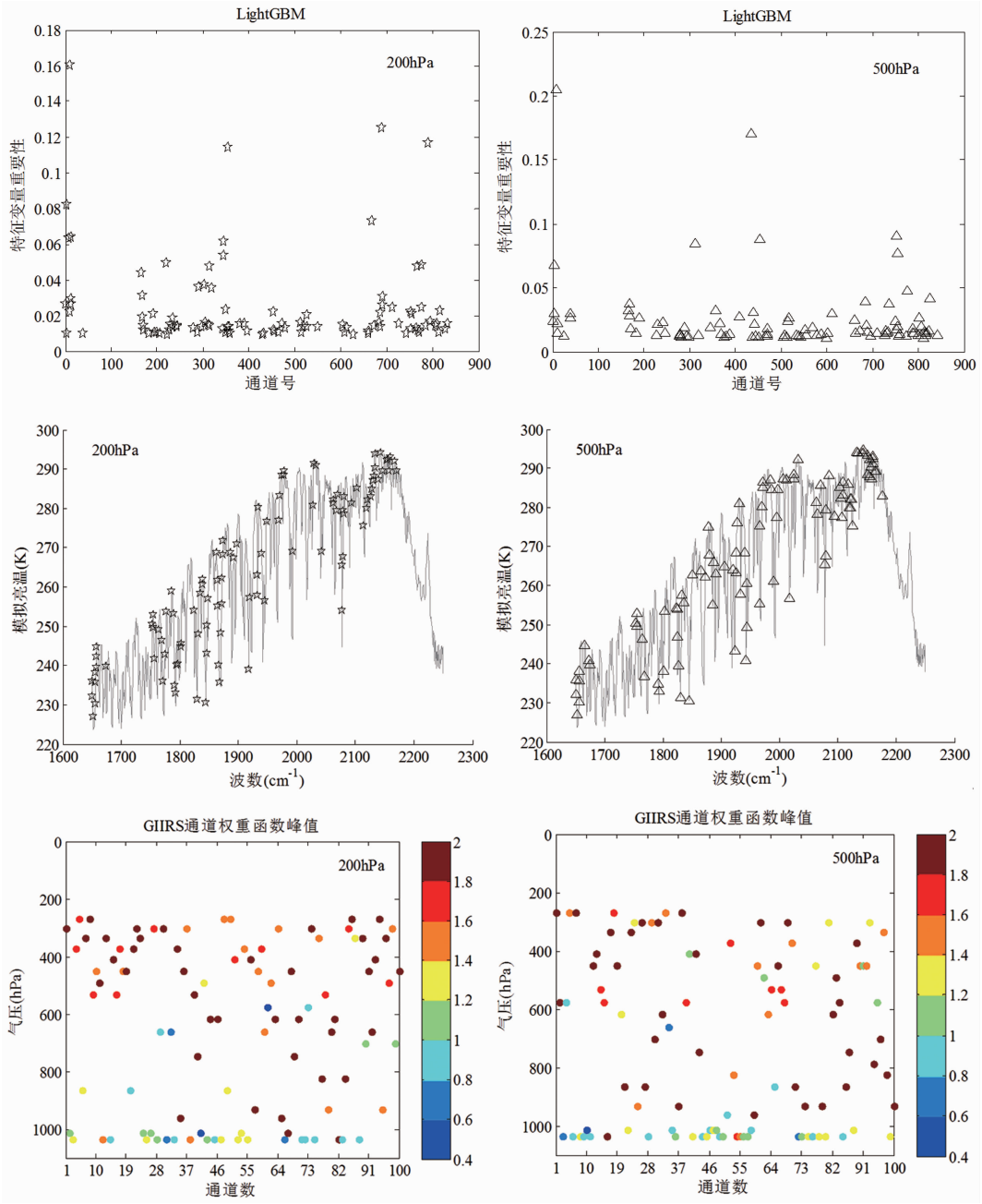

图 2 GIIRS 通道特征变量重要性、相关通道分布和权重函数峰值层分布

0.1235 m/s；在独立测试验证样本预测中，风场 U 分量和 V 分量整层(37 层)RMSE 分别小于 0.9677 m/s 和 1.4154 m/s。

与其它气压层相比，100～250 hPa 之间的风场 U 分量和 V 分量 RMSE 偏大。下面从 GIIRS 入选通道特征变量重要性和通道权重函数的角度分析其原因，并进一步分析典型气压层反演效果较好的原因。本文中的 FY-4A/GIIRS 通道权重函数和模拟亮温通过用 RTTOV 模式[13]计算中纬度夏季廓线得到[12]。

由于篇幅限制，图 2 所示为基于 LightGBM 模型的 200 hPa 和 500 hPa 风场 U 分量反演结果。图 2 给出了采用 PFI 得到的前 100 个特征变量重要性、相关通道分布和权重函数峰值层分布。

由图 2 可知，在反演 200 hPa 风场 U 分量时，特征重要性排名前 2 的是 GIIRS 中波通道 11 和 688，其通道权重函数峰值分别位于

267.10 hPa 和 962.26 hPa。在反演 500 hPa 风场 U 分量时，特征重要性排名前 2 的是 GIIRS 中波通道 9(236.28 hPa)和 434(532.58 hPa)。区别于熵减法[12]，此处采用的 PFI 方法可根据不同气压层或不同场景实现 GIIRS 通道的动态最优选择。

此外，在变量重要性排名前 100 的通道中，100～200 hPa 气压层里没有通道入选。与理想通道权重函数[11]相比，实际权重函数峰值层有一定的宽度[20]，表明在权重函数峰值层附近气压层的相关大气参数信息也能被探测。因此，本文中附近气压层的风场 U 分量和 V 分量也能被反演，但精度有待提高。

4 结束语

考虑到风场信息对于天气形势的演变和预报至关重要，而目前经典风场反演方法的精度受高度分配影响，本文基于 FY-4A/GIIRS 和 ERA5 再分析资料，采用 LightGBM 开展了大气三维风场反演研究。在建立 GIIRS 通道黑名单的基础上，采用 PFI 实现了不同气压层的 GIIRS 通道动态最优选择。在构建特征变量时，引入了 GIIRS 资料的时空信息。相对于 ERA5 风场资料，在测试集中风场 U 分量和 V 分量的 RMSE 分别小于 1 m/s 和 1.5 m/s。通过进一步分析可知，风场高层反演误差较大的主要原因是入选通道较少。

虽然本文中的个例和反演方法取得了一定的效果，但也存在不足之处。例如，虽然采用的是 GIIRS 加密资料，但作为机器学习算法的样本，资料还是偏少[8]。建立一个具有代表性的训练数据集对于 LightGBM 模型的应用非常重要。但由于处理大量数据的计算资源有限，本文只使用了有限的数据进行训练。因此，当该方法应用于另一种台风时，反演精度可能会降低。而本文基于以下假设：如果机器学习模型经过很好的训练，那么它可以从高时间分辨率的卫星观测中高效地反演大气风场，即可实现实时或近实时应用。所以后期的研究工作拟加入 GIIRS 长波通道资料，以提高所有气压层风场的反演精度，并进一步将反演的风场资料应用于高影响天气监测和预警中，以期更好地开展公共气象服务。

参考文献

[1] Kalnay E. Atmospheric modeling, data assimilation and predictability [M]. Cambridge: Cambridge University Press, 2003.

[2] Guillermo T S, Manel M R. Multi-layer wind velocity field visualization in infrared images of clouds for solar irradiance forecasting [J]. *Applied Energy*, 2021, **288**(1): 116656.

[3] 杨璐，陈敏，陈明轩，等. 高时空分辨率三维风场在强对流天气临近预报中的融合应用研究[J]. **气象学报**，2019，**77**(2)：243-255.

[4] Velden C, Daniels J, Stettner D, et al. Recent innovations in deriving tropospheric winds from meteorological satellites [J]. *Bulletin of the American Meteorological Society*, 2005, **86**(2): 205-224.

[5] Szantai A, Héas P, Mémin E. Comparison of atmospheric motion vectors and dense vector fields calculated from MSG images [C]. Beijing: 8th International Winds Workshop, 2006.

[6] 孔德华，张东，张卓，等. 基于加速鲁棒特征图像匹配的云导风计算方法 [J]. **海洋科学**，2022，**46**(9)：64-76.

[7] McNally A P. A note on the occurrence of cloud in meteorologically sensitive areas and the implications for advanced infrared sounders [J]. *Quarterly Journal of the Royal Meteorological Society*, 2002, **128** (585): 2551-2556.

[8] Ma Z, Li J, Han W, et al. Four-dimensional wind fields from geostationary hyperspectral infrared sounder radiance measurements with high temporal resolution [J]. *Geophysical Research Letters*, 2021, **48**(14): e2021GL093794.

[9] Yang J, Zhang Z, Wei C, et al. Introducing the new generation of Chinese geostationary weather satellites, Fengyun-4 [J]. *Bulletin of the American Meteorological Society*, 2017, **98**(8): 1637-1658.

[10] Ke G L, Meng Q, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree [C]. Long Beach: 31st Annual Conference on Neural Information Processing Systems, 2017.

[11] 王根. **卫星资料广义同化、大气科学中的数学反问题与人工智能应用——学习笔记** [M]. 北京：气象出版社，2020.

[12] 王根，邵立瑛，丁卫东，等. 高光谱 GIIRS 中波通道的最优选择及其对云检测的影响 [J]. **红外**，2021，**42**(7)：36-42.

[13] Saunders R, Hocking J, Turner E, et al. An update on the RTTOV fast radiative transfer model (currently at version 12) [J]. *Geoscientific Model Development*, 2018, **11**: 2717-2737.

[14] Altmann A, Tolosi L, Sander O, et al. Permutation importance: a corrected feature importance measure [J]. *Bioinformatics*, 2010, **26**(10): 1340-1347.

[15] Yin R Y, Han W, Gao Z Q, et al. The evaluation of FY4A's Geostationary Interferometric Infrared Sounder (GIIRS) longwave temperature sounding channels using the GRAPES global 4DVar [J]. *Quarterly Journal of the Royal Meteorological Society*, 2020, **146**(728): 1459-1476.

[16] Min M, Wu C Q, Li C, et al. Developing the science product algorithm testbed for Chinese next-generation geostationary meteorological satellites: Fengyun-4 series [J]. *Journal of Meteorological Research*, 2017, **31**(4): 708-719.

[17] Hersbach H, Bell B, Berrisford P, et al. The ERA5 global reanalysis [J]. *Quarterly Journal of the Royal Meteorological Society*, 2020, **146**(730): 1999-2049.

[18] Zhang Q, Yu Y, Zhang W M, et al. Cloud Detection from FY-4A's Geostationary Interferometric Infrared Sounder Using Machine Learning Approaches [J]. *Remote Sensing*, 2019, **11**(24): 3035.

[19] 王根，陈娇，戴娟，等. 风云四号红外高光谱 GIIRS 中波通道亮温偏差订正 [J]. **红外**，2021，**42**(5)：39-44.

[20] Joiner J, Brin E, Treadon R, et al. Effects of data selection and error specification on the assimilation of AIRS data [J]. *Quarterly Journal of the Royal Meteorological Society*, 2007, **133**(622): 181-196.