

文章编号: 1672-8785(2020)04-0041-08

# KNN不同距离度量对 FY-4A/AGRI 红外亮温反演降水的影响研究

王 根<sup>1,2,3</sup> 陆雅君<sup>2</sup> 王 悦<sup>2</sup> 吴瑞姣<sup>2</sup> 丁从慧<sup>2</sup>

(1. 中亚大气科学研究中心, 新疆乌鲁木齐 830002;

2. 安徽省气象台 强天气集合分析和预报重点实验室, 安徽合肥 230031;

3. 安徽省气象科学研究所 安徽省大气科学与卫星遥感重点实验室, 安徽合肥 230031)

**摘 要:** 静止卫星的高时空分辨率对高影响灾害性天气的监测和预报有显著优势。开展了基于风云四号 A 星多通道扫描成像辐射计(Advanced Geosynchronous Radiation Imager, AGRI)红外光谱亮温的台风降水反演研究。探讨了正则化反问题方法中 K-最近邻(K-Nearest Neighbor, KNN)不同距离度量对降水反演精度的影响。降水反演共分两步: 一是降水视场点识别, 主要基于训练字典样本, 利用 KNN 识别待反演的亮温“降水”和“非降水”信号; 二是降水视场点反演, 即在判识视场点有降水的基础上采用正则化反问题方法进行红外亮温降水反演。KNN 距离度量分别采用欧氏、标准化欧氏、马氏和布洛克距离。以台风“安比(2018)”为例, 开展了降水反演试验。试验表明, 反演结果与 GPM 的相似度较高, 且不同距离度量在反演“极端降水”时各有优势。

**关键词:** FY-4A/AGRI; 台风“安比”; 降水反演; 距离度量; 正则化反问题

**中图分类号:** P412.27 **文献标志码:** A **DOI:** 10.3969/j.issn.1672-8785.2020.04.007

## Research on Influence of Different Distance Measurements of KNN on Precipitation Retrieval by AGRI Infrared Bright Temperatures of FY-4A

WANG Gen<sup>1,2,3</sup>, LU Ya-jun<sup>2</sup>, WANG Yue<sup>2</sup>, WU Rui-jiao<sup>2</sup>, DING Cong-hui<sup>2</sup>

(1. Center of Central Asia Atmospheric Science Research, Urumqi 830002, China;

2. Anhui Meteorological Observatory, Key Lab of Strong Weather Analysis and Forecast, Hefei 230031, China;

3. Anhui Institute of Meteorological, Anhui Key Lab of Atmospheric Science and Satellite Remote Sensing, Hefei 230031, China)

**Abstract:** The high spatial and temporal resolution of geostationary satellite has significant advantages in mo-

**收稿日期:** 2020-03-10

**基金项目:** 中亚大气科学研究基金项目(CAAS202003); 国家自然科学基金项目(41805080); 上海台风研究基金项目(TFJJ201909); 安徽省气象局项目(KM201902)

**作者简介:** 王根(1983-), 男, 江苏泰州人, 博士, 副高, 主要从事卫星资料同化、正则化反问题与人工智能等方面的研究。E-mail: 203wanggen@163.com

monitoring and forecasting disastrous weather with high impact. A typhoon precipitation retrieval study based on infrared spectrum bright temperatures of the advanced geosynchronous radiation imager (AGRI) in FY-4A satellite is conducted. The influences on precipitation retrieval accuracy of different distance measurements of K-nearest neighbor (KNN) in the inverse problem of regularization method are discussed. The precipitation retrieval consists of two steps: (1) the identification of precipitation field of view, which uses KNN to identify the "precipitation" and "non-precipitation" signals of brightness temperatures to be retrieved mainly based on the training dictionary samples; (2) the retrieval of the precipitation field of view, which uses inverse problem of regularization method to retrieve the infrared brightness temperatures precipitation on the basis of identifying the precipitation field of view. KNN distance measurement adopts Euclidean, normalized Euclidean, Mahalanobis and Cityblock respectively. Taking typhoon "Ampil (2018)" as an example, the precipitation retrieval experiment is carried out. The experiment shows that the retrieval result has a high similarity with GPM, and different distance measurements have their own advantages in the retrieval of "extreme precipitation".

**Key words:** FY-4A/AGRI; typhoon Ampil; precipitation retrieval; distance measurement; inverse problem of regularization

## 0 引言

降水资料在气象监测和水文模式中至关重要,以生物多样性、生态环境和水文气象为导向的项目也需要高质量降水产品。在台风、暴雨和强对流等灾害性天气的发生和发展过程中,对降水的监测和预报也至关重要<sup>[1]</sup>。目前获取降水信息的方法主要包括地面雨量传感器、地基雷达、卫星观测和数值预报等。这些方法在观测原理、算法精度、产品的时空分辨率和覆盖范围等方面存在差异和互补<sup>[2]</sup>。其中,地面雨量传感器和雷达观测网的密度不够,较难监测和追踪暴雨等灾害性天气系统,更不足以捕捉台风降水的空间强度变化<sup>[1]</sup>。

国内外的相关研究表明,从静止气象卫星的光学传感器信号中可提取降水信息<sup>[3]</sup>。静止气象卫星获得的云顶亮温可揭示云存在和所处演变阶段的特征,能在一定程度上反映台风雨区分布<sup>[1]</sup>。由于卫星红外光谱无法穿透云,当云量达到一定厚度时,红外光谱信号就会“饱和”<sup>[4]</sup>。所以基于卫星红外资料的降水反演问题一直是研究热点和难点。目前,该研究可分为降水检测和降水反演两步来执行。孙绍辉等<sup>[5]</sup>利用查算表开展了日本 Himawari-8(H8)/AHI 红外亮温反演降水研究,结果表明,得到的查算表维数越多,

反演精度越高。Feidas H 等<sup>[6]</sup>建议从多光谱数据集中获取光学和微物理云参数以改善光学降水反演。Thies B 等<sup>[7]</sup>从 MSG/SEVIRI 数据中获取云特性,将雨区不同降水过程区域分开,并把它作为昼夜卫星反演降水方案的一部分。Tebbi M A 等<sup>[8]</sup>基于支持向量机和人工神经网络,利用 MSG/SEVIRI 红外资料建立了地中海南岸地区的降水监测和对流单体识别人工智能系统。Meike K 等<sup>[9]</sup>基于 MSG/SEVIRI 云顶高度、云顶温度、云相态和云水路径,利用随机森林得到了适用于中纬度地区的 24 h 降水反演。Hirose H 等<sup>[10]</sup>提出了一种基于红外资料随机森林机器学习的降水反演算法。他们将 H8/AHI 亮温作为输入数据,并将全球降水测量雷达的观测数据作为输出数据以训练样本。Ebtehaj A M 等<sup>[11]</sup>提出将正则化反问题用于 TRMM 资料反演降水,得到了较高的反演精度。Wang G 等<sup>[1]</sup>在 Ebtehaj A M 等<sup>[11]</sup>的研究基础上,采用正则化反问题方法开展了 AGRI 红外亮温反演台风“玛莉亚(2018)”降水研究,并基于贝叶斯模型平均法度量不同通道亮温对降水反演目标函数的贡献率。王根等<sup>[12]</sup>以安徽区域为例,经研究发现:当有降水发生时,H8/AHI 通道 7~16 的亮温梯度均有显著变化。另外,他们还开展了对流

和台风降水反演研究。

由于 Ebtehaj A M 等<sup>[11]</sup>、Wang G 等<sup>[1]</sup>和王根等<sup>[12]</sup>均基于欧几里得距离度量得到与待反演视场点亮温较近的通道亮温组合, 本文在其研究基础上, 探讨了正则化反问题降水反演中 KNN 不同距离度量(欧氏、标准化欧氏、马氏和布洛克距离)对台风“安比(2018)”降水反演的影响。该研究旨在为后期构建具有降水特征的距离度量函数做铺垫, 以期提高降水反演精度, 从而更好地将 FY-4A/AGRI 资料用于降水监测, 进一步提高气象防灾减灾能力。

## 1 基于正则化反问题方法的卫星红外光谱亮温反演降水

假设每个视场点的卫星观测亮温和降水分别记为  $\mathbf{y}=(y_1, y_2, \dots, y_{n_c})^T$  和  $\mathbf{x}$ , 其中  $n_c$  表示卫星通道数。卫星亮温与降水之间的关系可表示为<sup>[1, 11-12]</sup>

$$\mathbf{y}=H(\mathbf{x})+\mathbf{v} \quad (1)$$

式中,  $H(\cdot)$ 表示投影关系;  $\mathbf{v} \in R^n$ 表示误差。

本文参考 Ebtehaj A M 等<sup>[11]</sup>、Wang G 等<sup>[1]</sup>和王根等<sup>[12]</sup>的研究成果, 将亮温和降水字典定义为  $\Gamma=\{(\mathbf{b}_i, \mathbf{r}_i)\}_{i=1}^M$ 。其中, 通道亮温和降水字典原子分别表示为  $\mathbf{b}_i=[b_{1i}, b_{2i}, \dots, b_{n_i}]^T \in R^n$  和  $\mathbf{r}_i=[r_{1i}, r_{2i}, \dots, r_{n_i}]^T \in R^n$ ; 亮温和降水字典分别记为矩阵  $\mathbf{B}=[\mathbf{b}_1 | \dots | \mathbf{b}_M] \in R^{n \times M}$  和  $\mathbf{R}=[\mathbf{r}_1 | \dots | \mathbf{r}_M] \in R^{n \times M}$ 。

正则化反问题方法反演降水分为两步:

(1)判断待反演视场点是否存在降水。基于最优“原子”组合, 定义 KNN 邻域大小为  $K_1$ , 检测概率为  $p$ 。如果  $K_1$  个视场点出现降水的总概率小于  $p$ , 则将此待反演视场点降水记为 0, 并将其标记为“无降水”, 否则就把视场点标记为“有降水”。

(2)降水反演。这一步重点研究 KNN 中不同距离度量对降水反演精度的影响。在将降水视场点标记为“有降水”时, 由最优“原子”组合拟合待反演的观测亮温, 得到原子系

数  $\mathbf{c}$ , 则有

$$\mathbf{y}=\mathbf{B}_s \mathbf{c}+\mathbf{v} \quad (2)$$

式中,  $\mathbf{B}_s$  表示最优“原子”构成的字典子集;  $\mathbf{v}$  表示观测误差;  $\mathbf{c}$  表示原子系数。

通过分析, 将问题转换为求解原子系数  $\mathbf{c}$ 。文中采用 Ebtehaj A M 等<sup>[11]</sup>提出的算法, 将降水反演模型定义如下:

$$\begin{cases} \hat{\mathbf{c}}=\operatorname{argmin}_{\mathbf{c}} \frac{1}{2}\|\mathbf{W}^{1/2}(\mathbf{y}-\mathbf{B}_s \mathbf{c})\|_2^2+\lambda_1\|\mathbf{c}\|_1+\lambda_2\|\mathbf{c}\|_2^2 \\ \mathbf{c} \geq 0, \mathbf{1}^T \mathbf{c}=1 \end{cases} \quad (3)$$

式中,  $\mathbf{W}$  表示通道观测亮温对目标函数的贡献率。Wang G 等<sup>[1]</sup>和王根等<sup>[12]</sup>基于贝叶斯模型平均估计了  $\mathbf{W}$  值。本文将各个通道权重赋为 1。  $\lambda_1$  和  $\lambda_2$  表示正则化参数。

将式(3)中的原子系数  $\mathbf{c}$  应用于降水字典子集  $\mathbf{R}_s$  后得到  $\mathbf{x}$ 。  $\mathbf{x}$  即为降水反演值, 其表达式如下:

$$\mathbf{x} \approx \mathbf{R}_s \mathbf{c} \quad (4)$$

式中,  $\mathbf{R}_s$  表示与  $\mathbf{B}_s$  对应的  $\mathbf{R}=[\mathbf{r}_1 | \dots | \mathbf{r}_M] \in R^{n \times M}$  字典子集。

## 2 资料与台风背景介绍

### 2.1 资料信息

风云四号 A 星于 2016 年 12 月 11 日发射成功, 其携带的 AGRI 共有 14 个通道<sup>[13]</sup>。由 Wang G 等<sup>[1]</sup>的前期统计分析可知, 当降水发生时 AGRI 红外通道 8~14 的亮温梯度均有显著变化。本文参考此结论, 仅将这七个红外通道的亮温用于降水反演。采用 1 h 时间分辨率和 4 km 空间分辨率的 AGRI 资料(来源于国家卫星气象中心网站)。

本文采用全球降水量测量 GPM IMERG<sup>[14]</sup>的降水数据。该数据来源于美国国家宇航局(National Aeronautics and Space Administration, NASA), 其空间覆盖范围为  $89.95^\circ\text{S} \sim 89.95^\circ\text{N}$  和  $179.95^\circ\text{W} \sim 179.95^\circ\text{E}$ , 分辨率为  $0.1^\circ \times 0.1^\circ$ , 时间分辨率为 30 min。

在构建历史训练样本时, 参考 Wang G 等<sup>[1]</sup>的工作, 采用 KNN 法将 FY-4A/AGRI 各

通道亮温插到 GPM 视场点, 并将其作为本文的反演降水数据源。

## 2.2 台风“安比(2018)”背景介绍

强热带风暴“安比”(国际编号为 1810)是 2018 年太平洋台风季中的第 10 个风暴。2018 年 7 月 18 日 8 时, 台风“安比”在西北太平洋洋面生成。2018 年 7 月 22 日 12 时 30 分, 台风“安比”中心在我国上海市崇明岛沿海登陆。登陆时中心附近的最大风力为 10 级(28 m/s), 中心最低气压为 982 hPa。它成为 1949 年以来第 3 个直接登陆上海的台风, 导致飞机延误、轮渡停航、列车停运。

本文使用了“安比(2018)”整个台风过程中的 GPM 降水总量(单位为 mm/30 min)。时间覆盖范围为从 2018 年 7 月 18 日 8 时至 2018 年 7 月 24 日 23 时(<http://typhoon.nmc.cn/web.html> 网站), 区域覆盖范围为 12°N~43°N 和 110°E~150°E。台风“安比(2018)”发生、发展和消亡的运动轨迹(黑线)以及该时间段的 GPM 降水量累加数据见图 1。

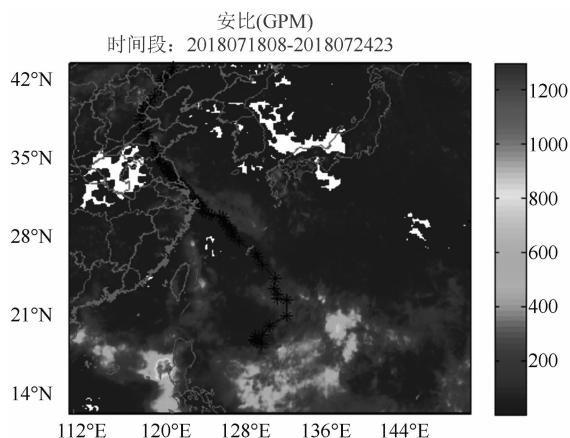


图 1 台风“安比(2018)”的运动轨迹和整个台风生命周期内的 GPM 降水总量分布

## 3 FY-4A/AGRI 红外亮温反演降水试验结果分析与比较

### 3.1 精度评估度量方法

本文采用的降水反演精度评估量化方法有皮尔森相关系数(Pearson product-moment correlation coefficient, 文中记为“Corr”)、结构相

似性(Structural SIMilarity, SSIM)、峰值信噪比(Peak Signal to Noise Ratio, PSNR)、命中率(Probability of Detection, POD)、虚警率(False-Alarm Ratio, FAR)和临界成功指数(Critical Success Index, CSI)。具体公式介绍可参考 Wang G 等<sup>[1]</sup>的工作。其中, POD、FAR 和 CSI 降水阈值定义为 5.0 mm/h。

### 3.2 基于 KNN 不同距离度量的台风“安比(2018)”降水反演试验

由于本文主要探讨 KNN 不同距离度量对降水反演精度的影响, 采用前一时次的资料作为训练样本来反演下一时次的降水。因篇幅限制, 本文仅给出了 2018 年 7 月 22 日 12 时和 23 时的降水反演结果。各时次的 AGRI 云图、真实 GPM 降水、基于不同距离(欧氏、标准化欧氏、马氏和布洛克距离)度量的 AGRI 反演降水分布以及 4 个不同距离度量反演降水与 GPM 相关性最大和最小的散点分布如图 2 和图 3 所示。

结合图 2 和表 1 可知, 采用不同距离度量反演的台风“安比(2018)”降水形态结构与 GPM 较为接近, 说明本文基于正则化反问题反演降水的方法是可行的。由 SSIM 分析可知, 基于“欧氏距离”能反演出台风“极端”强降水; 由 Corr 分析可知, 基于“布洛克距离”反演的降水与 GPM 真实降水的相关性最大。从图 2 也可看出, 不同的度量距离会影响降水视场点的判识。有些距离度量方法将某视场点识别为“降水”, 而另一些方法则将其识别为“非降水”。

结合图 3 和表 2, 由 SSIM 可知, 基于“标准化欧氏距离”能反演出台风此时刻的“极端”强降水; 由 Corr 可知, 基于“欧氏距离”反演的降水与 GPM 真实降水的相关性较大。

## 4 总结与展望

文中探讨了正则化反问题反演方法中 KNN 不同距离度量对 FY-4/AGRI 资料反演降水的影响, 并以台风“安比(2018)”为例开展



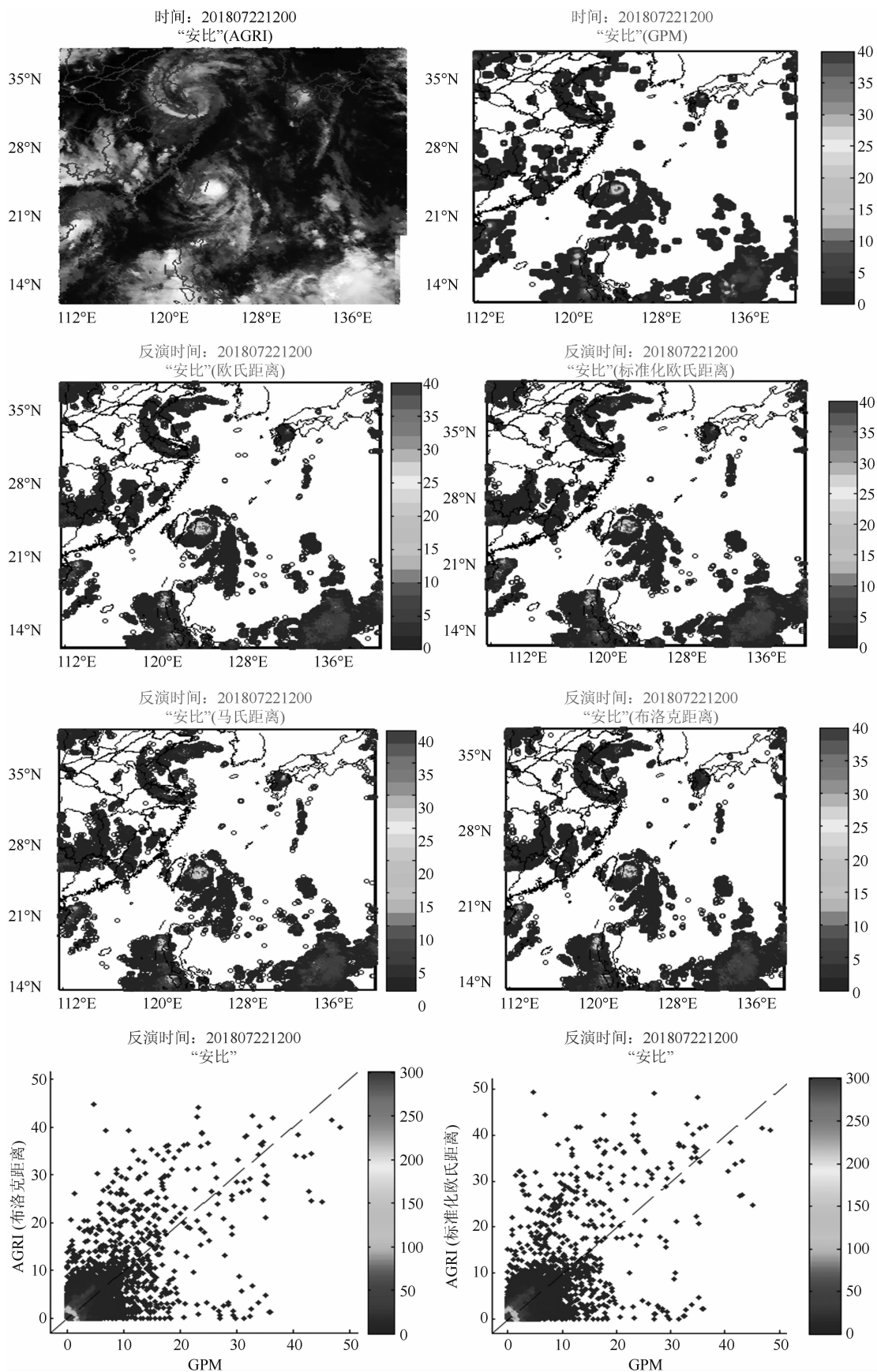


图 2 2018 年 7 月 22 日 12 时 KNN 不同距离度量反演台风“安比(2018)”降水分布

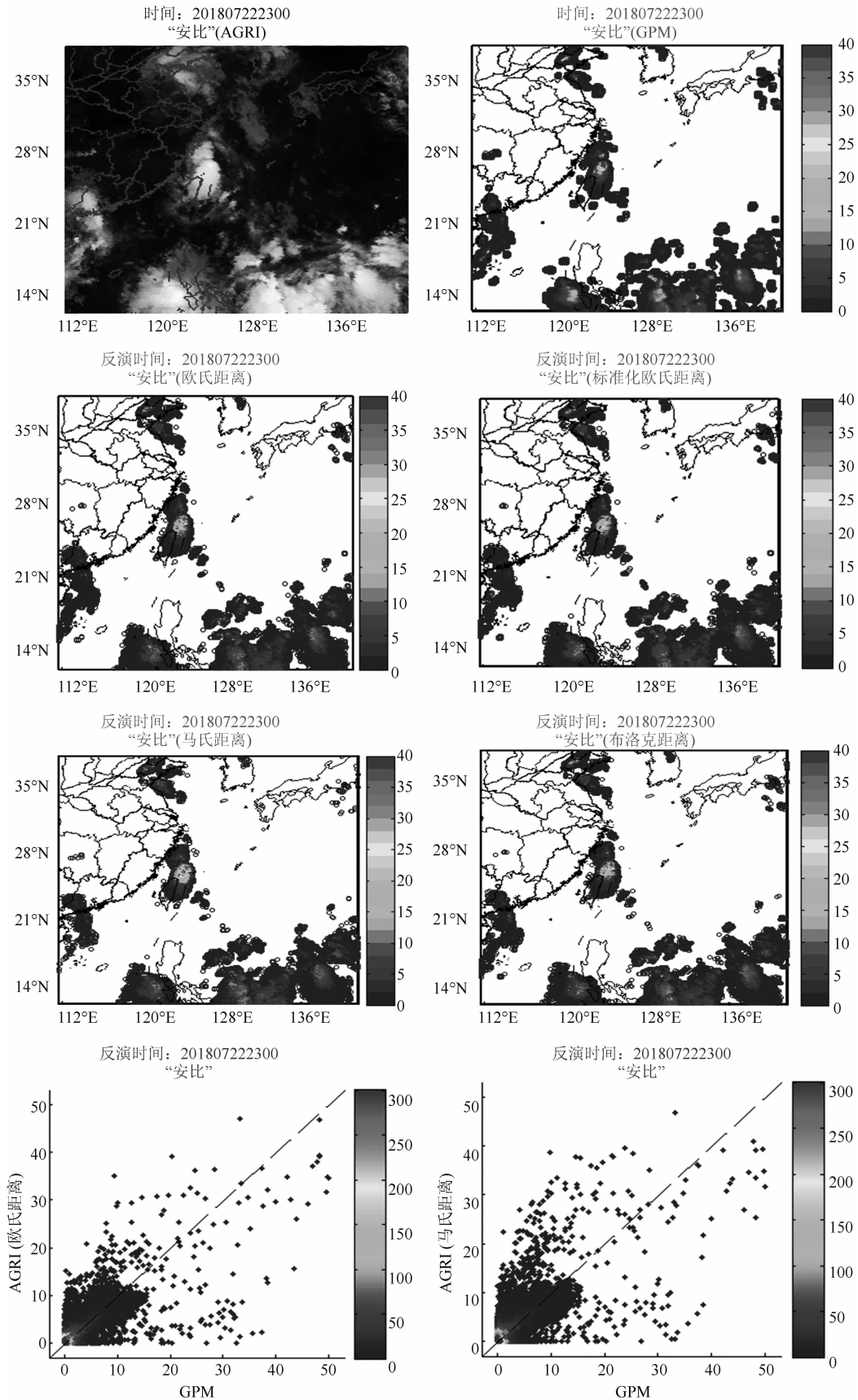


图3 2018年7月22日23时KNN不同距离度量反演台风“安比(2018)”降水分布

表 1 2018 年 7 月 22 日 12 时 KNN 不同距离度量反演台风“安比(2018)”降水精度分析

	Corr	SSIM	PSNR	POD	FAR	CSI
欧氏距离	0.7541	0.9298	32.1315	0.9944	0.0093	0.9852
标准化欧氏距离	0.7531	0.9284	31.8014	0.9943	0.0096	0.9848
马氏距离	0.7581	0.9281	32.4531	0.9943	0.0092	0.9852
布洛克距离	0.7702	0.9294	32.5827	0.9943	0.0091	0.9852

表 2 2018 年 7 月 22 日 23 时 KNN 不同距离度量反演台风“安比(2018)”降水精度分析

	Corr	SSIM	PSNR	POD	FAR	CSI
欧氏距离	0.8184	0.9397	34.1361	0.9943	0.0057	0.9886
标准化欧氏距离	0.8172	0.9417	34.1559	0.9940	0.0052	0.9889
马氏距离	0.7823	0.9316	32.7014	0.9951	0.0093	0.9859
布洛克距离	0.8141	0.9374	34.0402	0.9939	0.0058	0.9881

了相关试验。主要结论如下:

(1) 基于 KNN 不同距离度量的 AGRI 反演降水结构形态与 GPM 真实降水较为接近, 相关性较好。

(2) KNN 不同距离度量不仅影响视场点有无降水的判识, 也影响降水反演精度。

(3) 由台风“安比(2018)”降水反演试验可知, 基于“标准化欧氏距离”整体上能得到较高的降水反演精度, 为 KNN 识别亮温降水信号提供了参考。

本文仅初步开展了降水反演算法应用研究, 聚焦 KNN 不同距离度量对降水反演精度的影响。然而统计分析资料的时间尺度不够, 开展的个例研究较少, 历史字典样本库的代表性也不够。后期将开展大量的实际个例研究, 以构建符合降水特征的距离度量函数, 从而实现 FY-4A/AGRI 反演降水成果在降水监测、预警和预报中的业务化应用, 使其更好地服务于强对流和台风降水等公共气象工作。

## 致谢

感谢美国国家宇航局 NASA 和国家卫星气象中心免费在线提供 GPM IMERG 和 FY-4A/AGRI 数据。

## 参考文献

[1] Wang G, Wang K F, Han W, et al. Typhoon

maria precipitation retrieval and evolution based on the infrared brightness temperature of the Feng-Yun 4A/Advanced Geosynchronous Radiation Imager [J]. *Advances in Meteorology*, 2020, **2020**: 4245037.

[2] 王圆圆, 郭微, 李贵才, 等. 基于广义加性模型估算 1979–2014 年三峡库区降水及其特征分析 [J]. *地理学报*, 2017, **72**(7): 1207–1220.

[3] Min M, Bai C, Guo J P, et al. Estimating summertime precipitation from Himawari-8 and global forecast system based on machine learning [J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2018, **57**(5): 2557–2570.

[4] 王根, 唐飞, 刘晓蓓, 等. M-估计法广义变分同化 FY-3B/IRAS 通道亮温 [J]. *遥感学报*, 2017, **21**(1): 52–61.

[5] 孙绍辉, 李万彪, 黄亦鹏. 利用 Himawari-8 卫星红外图像反演降雨 [J]. *北京大学学报(自然科学版)*, 2019, **55**(2): 215–226.

[6] Feidas H, Giannakos A. Identifying precipitating clouds in Greece using multispectral infrared Meteosat Second Generation satellite data [J]. *Theoretical and Applied Climatology*, 2011, **104**(1–2): 25–42.

[7] Thies B, Nauss T, Bendix J. Discriminating raining from non-raining clouds at mid-latitudes using meteosat second generation daytime data [J]. *Atmospheric Chemistry and Physics*, 2008, **8**(9):

- 2341–2349.
- [8] Tebbi M A, Haddad B. Artificial intelligence systems for rainy areas detection and convective cells' delineation for the south shore of Mediterranean Sea during day and nighttime using MSG satellite images [J]. *Atmospheric Research*, 2016, **178**: 380–392.
- [9] Meike K, Appelhans T, Thies B, et al. Precipitation Estimates from MSG SEVIRI Daytime, Nighttime, and Twilight Data with Random Forests [J]. *Journal of Applied Meteorology and Climatology*, 2014, **53**(11): 2457–2480.
- [10] Hirose H, Shige S, Yamamoto M K, et al. High temporal rainfall estimations from Himawari-8 multi-band observations using the random-forest machine-learning method [J]. *Journal of the Meteorological Society of Japan*, 2019, **97**(3): 689–710.
- [11] Ebtahaj A M, Bras R L, Foufoula-Georgiou E. Shrunk Locally Linear Embedding for Passive Microwave Retrieval of Precipitation [J], *IEEE Transactions on Geoscience and Remote Sensing*, 2015, **53**(7): 3720–3736.
- [12] 王根, 王东勇, 吴蓉. Himawari-8/AHI 红外光谱资料降水信号识别与反演初步应用研究 [J]. *红外与毫米波学报*, 2020, **39**(2): 251–262.
- [13] Yang J, Zhang Z, Wei C, et al. Introducing the new generation of Chinese geostationary weather satellites—FengYun 4 (FY-4) [J]. *Bulletin of the American Meteorological Society*, 2017, **98**(8): 1637–1658.
- [14] Sounak K B, Chandrasekar V. Cross-validation of observations between the GPM dual-frequency precipitation radar and ground based dual-polarization radars [J]. *Remote Sensing*, 2018, **10**(11): 1773.