

王玥, 谢涛, 李建, 等. 星载 GNSS-R 融冰期海冰密集度反演研究[J]. 海洋学报, 2024, 46(5): 127–136, doi:10.12284/hyxb2024026
Wang Yue, Xie Tao, Li Jian, et al. Sea ice concentration retrieval using spaceborne GNSS-R during the melting period[J]. Haiyang Xuebao, 2024, 46(5): 127–136, doi:10.12284/hyxb2024026

星载 GNSS-R 融冰期海冰密集度反演研究

王玥¹, 谢涛^{1,2,3,4*}, 李建^{1,3,4}, 张雪红^{1,3,4}, 白淑英^{1,3,4}, 王明华^{1,3,4}

(1. 南京信息工程大学 遥感与测绘工程学院, 江苏 南京 210044; 2. 青岛海洋科技中心区域海洋动力学与数值模拟功能实验室, 青岛 山东 266200; 3. 自然资源部遥感导航一体化应用工程技术创新中心, 江苏 南京 210044; 4. 江苏省协同精密导航定位与智能应用工程研究中心, 江苏 南京 210044)

摘要: 针对北极融冰期的海冰密集度反演, 并改善全球导航卫星系统反射测量 (GNSS-R) 对海水的海冰密集度高估问题, 本文提出一种利用机器学习算法生成高时空分辨率的融冰期海冰密集度估算方法, 提取 GNSS-R 时延多普勒图 (DDM) 的特征参数, 并结合海表温度数据建立 LightGBM 模型, 将反演结果与参考海冰密集度值进行相关性分析和评估。本文的模型结果与 OSI SAF 的海冰密集度产品显示出较好的一致性, 相关系数、平均绝对误差和均方根误差分别为 0.965、0.061 和 0.090。该方法能够实现北极海冰边缘区的海冰密集度高精度估计。

关键词: GNSS-R; DDM; 融冰期; 海冰密集度; LightGBM; 北极

中图分类号: P715.6

文献标志码: A

文章编号: 0253-4193(2024)05-0127-10

1 引言

海冰密集度 (Sea Ice Concentration, SIC) 是给定区域海冰覆盖面积的比例, 是北极气候研究的重要参数之一。海冰影响大气和海洋之间的热交换, 了解海冰覆盖变化对气候变化研究至关重要。在全球变暖的背景下, 北极海冰融化期变长, 海冰边缘区扩大, 海冰覆盖在海冰边缘区呈现出更为显著的空间和时间变化, 监测海冰边缘区成为极地海冰研究的热点和难点^[1]。夏季北冰洋海冰的减少也提高了北极航道的通航潜力^[2], 准确获取北极融冰期高时空分辨率的海冰密集度, 有助于了解北极海冰的时空分布, 对于冰区航运具有重大意义。

冰冻圈环境恶劣, 卫星遥感是监测极地地区的最有效方法, 而传统海冰探测方法存在一些局限和缺点。光学传感器的成像质量易受到云层的影响, 监测

区域受限; 被动微波和散射计数据最广泛用于海冰密集度估计, 但是空间分辨率较低, 海冰边缘区的误差大; 而主动微波遥感空间分辨率高, 在海冰边缘区表现较好, 但时间分辨率较低^[3]。全球导航卫星系统反射测量 (GNSS-Reflectometry, GNSS-R) 作为一种新兴的遥感技术, 可实现全天候大范围监测, 来自不同卫星星座的丰富信号源使得它具有高空间和时间分辨率的优势, 是进行海冰密集度时空监测的有效手段。星载 GNSS-R 接收器是一种新型无源微波传感器, 以双基地雷达的形式配备^[4], 不需要发射信号, 测量方法成本低。GNSS-R 信号工作在 L 波段, L 波段波长较长, 对冰层具有更强的穿透力。

星载 GNSS-R 技术已成功应用于海冰探测。Yan 等通过神经网络^[5]、卷积神经网络^[6] 和支持向量机回归^[7] 对 TechDemoSat-1 卫星的时延多普勒图 (Delay-Doppler Mapping, DDM) 进行海冰检测和海冰密集度

收稿日期: 2023-05-30; 修订日期: 2023-11-21。

基金项目: 国家自然科学基金项目 (42176180); 国家重点研发计划项目 (2021YFC2803302)。

作者简介: 王玥 (1998—), 女, 江苏省南通市人, 主要从事 GNSS-R 海洋遥感研究。E-mail: wy2805@outlook.com

* 通信作者: 谢涛 (1973—), 男, 湖南省张家界市人, 教授, 主要从事海洋微波遥感研究。E-mail: xietao@nuist.edu.cn

反演,展现了时延多普勒图反演海冰密集度的能力,其中基于 SVR 的方案显示出更高的准确性,相关系数为 0.94,均方根误差为 0.15。Llaveria 等^[8]提出了一种使用 FSSCat 任务 CAT-5/a 卫星 MWR 和 GNSS-R 组合数据生成海冰密集度和海冰范围地图的神经网络方法,结合土地覆盖率和肤温数据生成的海冰密集度地图的平均绝对误差在南极洋为 3.57%,在北冰洋为 1.95%。Zhu 等^[9]利用 TechDemoSat-1 卫星 GNSS-R 数据,首次使用差分延迟波形的右缘波形求和(REWS)开发了一种用于反演海冰密集度的地球物理模式函数(GMF),在北半球可以实现 11.78%的均方根误差和 1.67%的偏差,在南半球可以实现 12.10%的均方根误差和 1.94%的偏差。上述研究主要使用的是极地冷季样本,海冰边缘区较少,大部分区域海冰密集度较高,反演难度比暖季样本低,海冰边缘区的 GNSS-R 反演效果难以突出。目前还缺乏应用 GNSS-R 技术针对融冰期海冰的研究,而监测融冰期海冰密集度对极地气候、生态和航运意义重大。北极的融冰期一般在 7 月、8 月,融冰过程中部分海冰表面会出现融化池,且海冰类型增多,海冰表面粗糙度也会有差异,对海冰密集度的反演会更加具有挑战。

时延多普勒图(DDM)特征参数(如 DDM 平均值、前沿斜率)被广泛应用于对风速^[10-11]的反演研究中,许多学者也利用多个 DDM 特征参数对海冰进行了研究,将 DDM 特征与海冰条件联系起来,大量研究^[12-15]表明利用从时延多普勒图得到的特征参数对海面粗糙度敏感,可以精确地检测海冰。但是目前大多数研究是利用 DDM 特征参数进行海冰和海水的分类,没有单独研究这些参数与海冰密集度的关系,且对海冰密集度的反演比海冰海水分类困难。

本文旨在探索 GNSS-R 在北极融冰期的应用,提出一种基于 GNSS-R DDM 特征参数的海冰密集度反演方法,与海冰密集度产品进行相关性分析。引入先进高效的 LightGBM 机器学习算法,同时考虑海表温度数据加入模型来提高海冰边缘区的海冰密集度反演精度,减少对平静海水的错估情况。

2 数据

2.1 星载 GNSS-R 数据

本文主要研究数据是英国 TechDemoSat-1(TDS-1)卫星的 GNSS-R 1B 级数据^[16],数据格式为 NetCDF,该数据可在 MERRBys 网站(www.merrbys.co.uk)申请,然后在 <ftp.merrbys.co.uk> 下载。TDS-1 卫星发射于 2014 年 7 月,2018 年停止服务。该星运行在准太阳同步轨

道上,轨道高度约 635 km,倾角为 98.4°,可以检测到极地高纬度的反射信号。

TDS-1 卫星搭载的 GNSS-R 载荷 SGR-ReSI 捕获的反射 GNSS 信号被处理成时延多普勒图(DDM)。每个 DDM 的原始尺寸为 128 个时延像素 × 20 个多普勒像素,时延和多普勒的分辨率分别为 250 ns(即 1/4 个 C/A 码片的长度)和 500 Hz。DDM 的元数据里包含每个 DDM 的镜面反射点信息,镜面反射点为地球表面上的点,是 GNSS 导航卫星发射的信号经过地表反射后到达接收机路径延迟最短的点。GNSS-R 信号以镜面反射点处的散射分量为主。当反射面如海冰表面平静光滑时,GNSS 信号产生相干反射。相干反射区为围绕镜面反射点的第一菲涅尔区,该区域是接收的反射信号能量主要贡献区域,约 1 km × 7 km,将该区域的大小视为本研究的空间分辨率。

2.2 再分析资料

欧洲中期天气预报中心 ECMWF 第五代再分析资料的海表温度(Sea Surface Temperature, SST)作为本研究的辅助数据。从 2007 年 9 月至今,ERA5 的海表温度基于 OSTIA 数据集得到,时间分辨率为 1 h,空间分辨率是 0.25°,在高纬度地区分辨率约为 10 km。SST 是海面附近海水的温度,单位为开尔文(K),本文减去了 273.15 转换为摄氏度(°C)。Yao 等^[17]对该数据进行了评估,ERA5 海温值与 NDBC 浮标测量值基本一致,具有 -0.04°C 偏差和 4.60°C 标准差。

2.3 参考海冰产品

欧洲气象卫星组织(EUMETSAT)的 OSI SAF 海冰密集度日产品的海冰密集度作为本研究参考的海冰密集度真值,假设海冰密集度在 1 天内无变化。OSI SAF 3 级产品(OSI-401-d)基于 SSMIS(Special Sensor Microwave Imager Sounder)微波遥感资料得到,空间分辨率为 10 km。

3 方法

3.1 DDM 预处理与特征参数计算

在提取 DDM 特征参数之前,先对 DDM 进行了质量控制与预处理。根据镜面反射点的位置,提取北极 70°N 以北的数据并利用 GSHHS 海岸线数据过滤陆地和海岸线附近 50 km 以内的 DDM,从而减少陆地散射的影响;噪声影响严重的 DDM 无法正确显示反射表面的特性,因此利用峰值信噪比作为数据质量指标,将低于 0 dB 的 DDM 数据排除,并剔除含直射信号的 DDM。

图 1 展示了 3 个预处理后的 DDM 和对应的归一

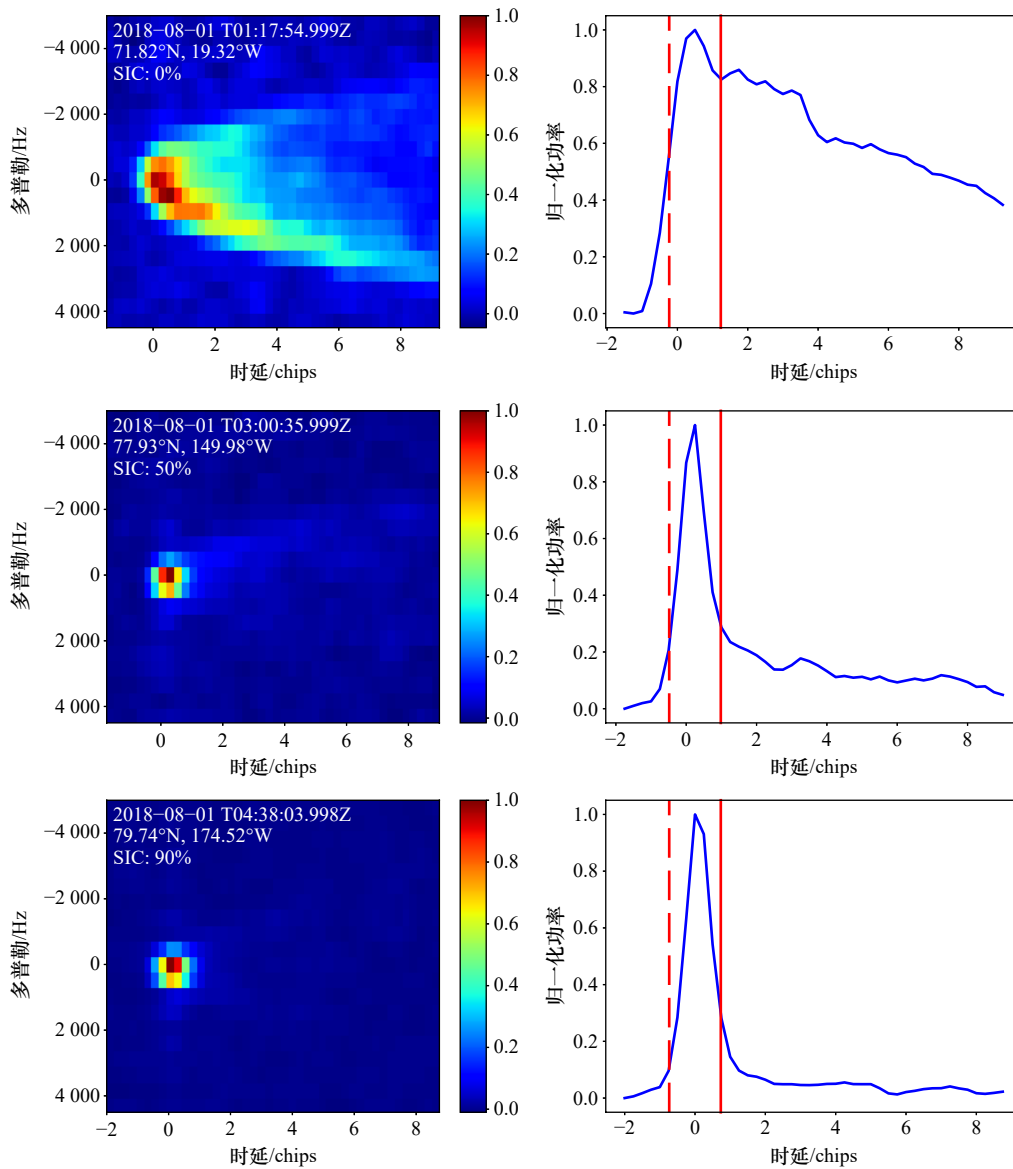


图1 0%、50%和90% SIC DDM和对应的归一化积分延迟波形

Fig. 1 DDM for 0%, 50% and 90% SIC DDM and corresponding normalized integrated delay waveforms

化IDW,所选的DDM实际海冰密集度分别为0%、50%和90%,图中还标注了收集时间和位置,波形图中红色虚线和实线分别经过峰值点前0.75 chips处和后0.75 chips处。DDM描述了观测表面的散射功率,DDM的扩散程度由表面粗糙度决定,海冰密集度不同,相应DDM的扩散程度通常不同,因此利用DDM特征参数描述DDM的功率变化和扩散程度,实现对海冰密集度的反演。本文采取了以下8种DDM特征参数,充分利用DDM的特征参数保证海冰密集度的反演精度。为了减少天线和仪器系统噪声变化引起的噪声波动的影响,使用的是减去底噪的DDM并且只使用了DDM峰值点周围-2~8.75 chips的时延范围。DDM底噪去除方法是先计算DDM的噪声底值,再用原始DDM减去该噪声底值,然后根据DDM的峰

值功率进行归一化。噪声底值 N 为DDM无信号区的散射功率平均值,本文将无信号区定义为DDM前5个时延像素×所有多普勒像素的矩形区域^[18]。

(1)等效反射率 Γ :假设信号由近镜面方向上的表面散射的相干分量支配,等效反射率 Γ 由以下雷达方程^[19]提取:

$$\Gamma(\theta) = P_s \frac{(4\pi)^2 (R_t + R_r)^2}{\lambda^2 G' G'' P_t}, \quad (1)$$

式中, P_s 和 P_t 分别表示接收和发射功率,计算 P_s 时减去了噪声功率; θ 表示镜面反射点处的入射角; λ 表示信号波长(19 cm); G' 和 G'' 分别表示镜面反射点处发射器和接收器的天线增益; R_t 和 R_r 分别是发射器和接收器到镜面反射点的距离,根据发射器、接收器和镜面反射点的位置计算。海冰的反射率远大于海水, Γ

可以表征海面的相干反射特性,与海冰密集度有关。

(2)时延多普勒图平均值(Delay-Doppler Map Average, DDMA):是在以归一化 DDM 峰值为中心的指定时延多普勒窗口上计算的平均散射功率^[20]。计算时需要先定位到 DDM 的峰值点所在像素位置,DDMA 计算公式如下:

$$DDMA = \frac{1}{15} \sum_{i=m-2}^{m+2} \sum_{j=n-1}^{n+1} DDM(t_i, f_j), \quad (2)$$

式中, t_i 和 f_j 分别代表时延和多普勒频移坐标值; m 和 n 分别代表峰值点的时延坐标和多普勒频移坐标。在计算的时延多普勒窗口内,海水面粗糙会使得散射功率扩散程度高,而海冰的信号功率更集中,所以海冰的 DDMA 往往小于海水。

(3)信噪比(Signal-to-Noise Ratio, SNR):散射功率与噪声底值的比值,计算公式如下:

$$SNR = 10 \times \lg(DDMA/DDM_{noise}), \quad (3)$$

式中, DDMA 为上一步计算的时延多普勒图平均值, DDM_{noise} 为噪声底值。一般来说,海冰的状态是相对稳定的,而海水表面容易受到海风的影响,反射便会是扩散的,因此海冰上散射信号的信噪比通常高于海水^[21]。

(4)峰值功率比(Peak Power Proportion, PP):峰值功率比是峰值周围 5 时延像素×3 多普勒像素的总功率,与整个 DDM 的功率之比^[22]。

$$PP = \frac{\sum_{i=m-2}^{m+2} \sum_{j=n-1}^{n+1} DDM(t_i, f_j)}{\sum_{i=m-8}^{m+35} \sum_{j=1}^{20} DDM(t_i, f_j)}, \quad (4)$$

式中, t_i 和 f_j 分别代表时延和多普勒频移坐标值; m 和 n 分别代表峰值点的时延坐标和多普勒频移坐标。分母中-8、35 代表裁切后的 DDM 的时延范围,即峰值周围-2~8.75 chips, 1 时延像素 = 0.25 chips, 1、20 代表多普勒范围,由图 1 可见本文使用了 20 行多普勒像素。PP 对相干信号敏感,也与海冰密集度有密切联系,高密度下的 PP 值较大。

(5)后沿斜率(Trailing Edge Slope, TES):后沿斜率表示积分延迟波形(Integrated Delay Waveform, IDW)的最大值与峰值后 0.75 chips 处的值之间斜率的绝对值,它会随信号的相干分量变化而变化。如图 1 所示,通常海冰密集度越高,坡度越陡即 TES 值越大。IDW 是多普勒域上 DDM 的积分,考虑了所有多普勒频移在相同延迟下的散射功率,热噪声可变性降低,并且可以更好地观察到非相干散射功率^[23]。

(6)前沿斜率(Leading Edge Slope, LES):前沿斜

率表示积分延迟波形 IDW 的最大值与峰值前 0.75 chips 处的值之间的斜率,与后沿斜率类似,都表征反射表面的粗糙度,与散射面的相干性或非相干性直接相关。

(7)峰度(Kurtosis, K):DDM 散射功率的统计参数,在本研究里是计算 IDW 的峰度。一般情况下,海冰的峰度是远大于海水的,密集度越高,功率越集中于镜面反射点处,峰度也就越大。

(8)偏度(Skewness, SK):与峰度类似,是计算的 IDW 偏度,表示 IDW 的偏斜方向和程度。从图 1 可以明显看出海水的 IDW 为左偏态,因为功率向右扩展程度高;而海冰的 IDW 为右偏态,SK 值一般会随海冰密集度的增高而变大。

3.2 数据匹配

GNSS-R DDM 数据与 ERA5 再分析资料和参考海冰密集度数据进行时空匹配。考虑到参考海冰密集度产品的分辨率,本文直接采用了镜面反射点周围的空间覆盖,这样所使用到的数据也都具有相似的空间分辨率。因此,根据 DDM 采集时间和镜面反射点位置,利用双线性空间插值可以确定对应的海表温度和海冰密集度值。其中,海表温度的匹配时间是距 DDM 采集时间最近的整点时间,最大时间误差小于 30 min。最后将 DDM 的特征参数与海表温度构成特征参数集。

3.3 LightGBM 模型构建

LightGBM(Light Gradient Boosting Machine)是一种基于树的机器学习方法,由 Microsoft Research 提供并于 2017 年发布。LightGBM 回归算法属于决策树的集成模型,其通过带有深度限制的按叶子生长(leaf-wise)算法构建决策树,面对非线性数据效果往往很好。LightGBM 还有 3 个重要的改进技术:独立特征合并(Exclusive Feature Bundling, EFB)、直方图算法(Histogram-based Algorithm, HA)以及梯度单边采样(Gradient-based One-Side Sampling, GOSS)。EFB 通过捆绑互斥的特征减少特征数量。HA 利用离散化的特征构造直方图,从而寻找到最优分割点。GOSS 优先选取大梯度样本并随机选取一部分小梯度的样本。这些技术使得模型具有训练速度快、泛化能力强等优势。模型架构图如图 2 所示,该算法的更多细节可见相关论文^[24-25]。

本研究采用 LightGBM 模型算法来反演海冰密集度,训练过程中以 8:2 的比例将样本随机划分为训练集与测试集,将弱学习器指定为梯度提升决策树,调整模型参数对模型进行优化。本研究提出的模型自变量与因变量之间的映射关系可表达为

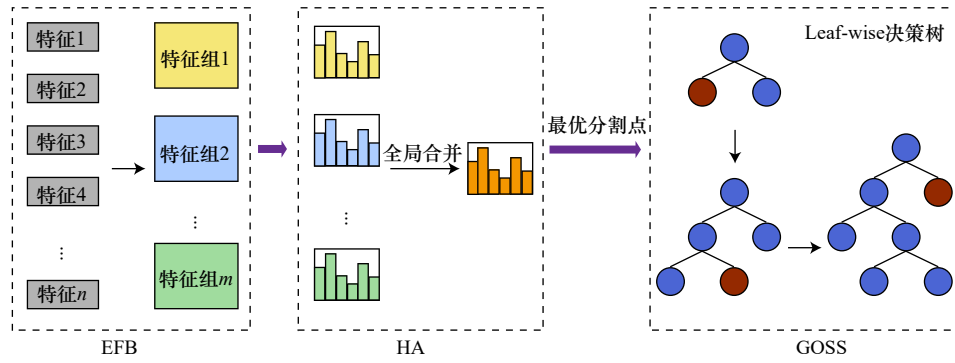


图2 LightGBM 模型架构
Fig. 2 Framework of the LightGBM model

$$SIC = \text{LightGBM}(\Gamma, DDMA, SNR, PP, TES, LES, K, SK, SST). \quad (5)$$

采用的评估标准为相关系数 R 、平均绝对误差 (Mean Absolute Error, MAE) 和均方根误差 (Root Mean Squared Error, RMSE), 它们的计算公式为

$$R = \frac{\text{Cov}(\hat{Y}, Y)}{\sigma_{\hat{Y}}\sigma_Y}, \quad (6)$$

$$MAE = \frac{1}{n} \sum_{i=1}^n |\hat{y}_i - y_i|, \quad (7)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}, \quad (8)$$

式中, \hat{y}_i 表示第 i 次海冰密集度预测值; y_i 表示相应的参考海冰密集度值; n 代表样本数量。

4 结果与讨论

使用 TDS-1 卫星 2018 年 8 月 8 月 1-5 日收集于北极 70°N 以北的数据, 按照 3.3 节 LightGBM 模型构建

方法来建立海冰密集度反演模型。总样本有 15 162 个 DDM, DDM 采集时间处于北极海冰融化阶段, 其中海冰边缘区的 DDM 有 7 029 个。用于验证的数据集是另外的 3 个夏季时间段 (2018 年 7 月 20-22 日、8 月 9-10 日、8 月 15-16 日) 的数据, 共 25 138 个 DDM。

4.1 海冰密集度反演模型训练结果和评估

先探索 GNSS-R 技术单独反演北极夏季海冰密集度的能力, 只使用 DDM 特征参数和海冰密集度真值建模。图 3a 显示了测试样本的基于 GNSS-R 观测值的 LightGBM 模型反演结果与 OSISAF 的海冰密集度真值之间的比较。可以看出, DDM 特征参数反演海冰密集度结果与 OSISAF 的海冰密集度大体上接近, R 为 0.875, MAE 为 0.120, RMSE 为 0.171。有少量估计结果与其相差较大, 尤其是在纯海水区域有较多高估值, 可能是海面比较平静导致相干反射, 并且有研究表明, 当海况和风速较低时, 靠近冰边缘或部分被冰覆盖区域的海冰密集度倾向于被高估^[5]。另外在

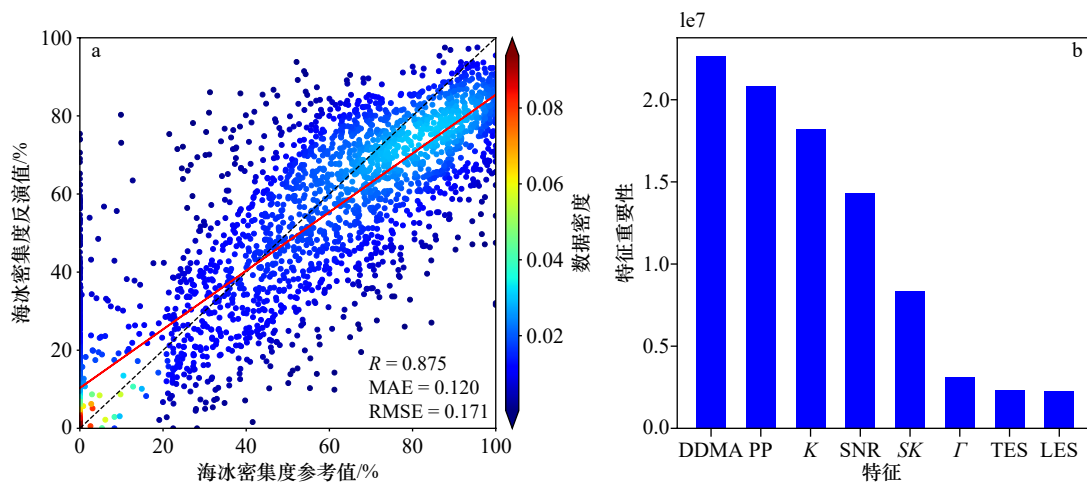


图3 测试集基于 DDM 特征参数的反演结果与 OSISAF 海冰密集度对比散点图(a)和基于 DDM 特征参数的 LightGBM 特征重要性排序结果(b)

Fig. 3 Scatter plot of retrieval results based on DDM observables against OSISAF sea ice concentration from the test samples (a) and LightGBM feature importance ranking results based on DDM observables (b)

较高密集度的区域模型倾向于低估,这可能是因为冰层过厚,L波段无法完全穿透冰层且海冰表面更粗糙,也有OSISAF产品误差的影响。

图3b显示了LightGBM模型的DDM特征参数的排序结果,是根据决策树构建过程中特征被分割的总增益来排序的。其中,DDMA的信息增益最高,DDMA反映了海表面的散射功率大小,说明DDM峰值附近的时延多普勒区域对海面粗糙度敏感,DDMA在构建该海冰密集度反演模型时起了较大的作用,是一个对于估计海冰密集度很重要的GNSS-R参数。为证

明反演过程中每个DDM特征参数的必要性,表1统计了不同的特征参数组合反演得到的R、MAE和RMSE。由表1可见,第9组使用了所有的DDM特征参数取得了最佳的结果,相关性最强且误差最小。前8组的统计结果表明缺少任何一项参数都会造成反演的误差变大,其中,SNR在该模型中表现出的影响最为明显,忽略SNR的模型结果与OSISAF产品的相关性下降了5.8%,MAE和RMSE也是9组中增大最多的。因此,本研究综合使用这8个特征参数以提高海冰密集度的反演精度。

表1 不同特征参数组合的反演误差统计表

Table 1 Error Statistics for Different DDM observables Combinations

序号	特征参数组合	R	MAE	RMSE
1	DDMA + SNR + PP + TES + LES + K + SK (无I)	0.867	0.124	0.176
2	I + SNR + PP + TES + LES + K + SK (无DDMA)	0.859	0.128	0.181
3	I + DDMA + PP + TES + LES + K + SK (无SNR)	0.817	0.146	0.204
4	I + DDMA + SNR + TES + LES + K + SK (无PP)	0.851	0.132	0.186
5	I + DDMA + SNR + PP + LES + K + SK (无TES)	0.871	0.122	0.174
6	I + DDMA + SNR + PP + TES + K + SK (无LES)	0.871	0.122	0.174
7	I + DDMA + SNR + PP + TES + LES + SK (无K)	0.866	0.124	0.177
8	I + DDMA + SNR + PP + TES + LES + K (无SK)	0.866	0.125	0.177
9	I + DDMA + SNR + PP + TES + LES + K + SK	0.875	0.120	0.171

北极夏季海冰融化,海冰边缘区增多,热力学参数在海冰变化中起很大作用,影响海气交换、热通量。海洋表层温度SST与海冰密集度存在紧密联系,SST也会引起DDM特征参数的变化,对反演结果造

成影响。接下来引入SST参数,分析比较单独使用辅助参数及其结合DDM参数的海冰密集度模型估计性能。

图4a显示了与OSISAF的海冰密集度真值比较,

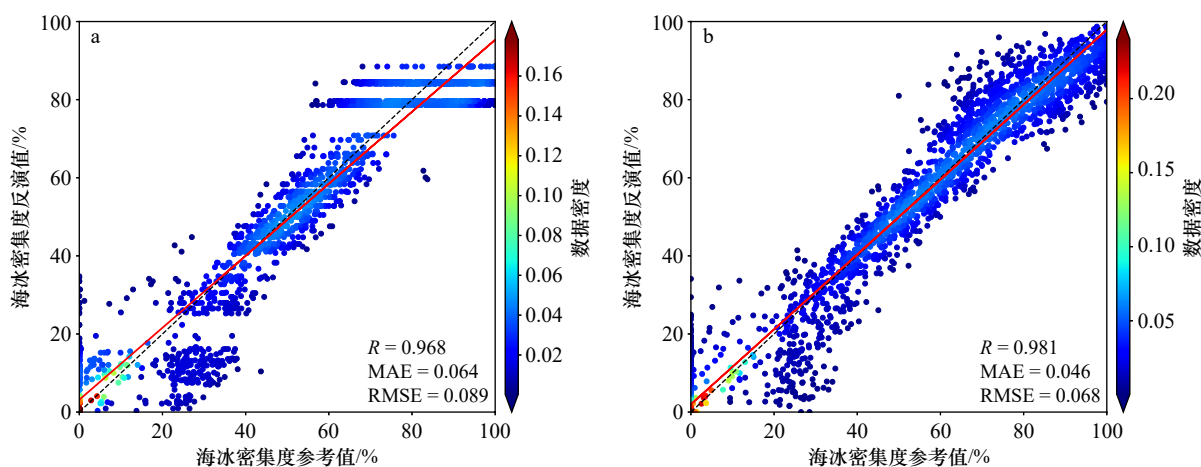


图4 测试集海冰密集度反演结果与OSISAF海冰密集度对比散点图

Fig. 4 Scatter plot of retrieval results against OSISAF sea ice concentration from the test samples

a. 基于SST; b. 基于DDM特征参数、SST

a. Based on SST; b. based on DDM observables and SST

基于 SST 的模型反演结果。SST 与海冰密集度显著相关, R 为 0.968, 较基于 DDM 特征参数的模型提高了 9.3%; RMSE 和 MAE 分别为 0.089 和 0.064, 各减少了 0.082 和 0.056。在 40%~60% 的海冰密集度区间, 基于 SST 的模型反演结果很好, 由于冰面和水面的 SST 值差异明显, 海冰边缘区冰的占比越多, SST 值一般就越低, SST 与海冰密集度呈负相关, 所以 SST 能有效反演海冰边缘区的海冰密集度。但是当海冰密集度在 70% 以上时, 由于冰区冰面温度变化很小, 反演结果倾向于稳定的较低值。另外在 20%~40% 的低密集度区有较多偏低的估计结果, 可能是这部分样本较少机器学习训练不充分或是受到融池的热量影响, 也有可能是 ERA5 再分析资料在极地的 SST 值存在热偏的误差情况导致。

图 4b 显示了 DDM 特征参数和 SST 组合的特征参数集作为模型的输入反演海冰密集度的结果。基于 DDM 观测值与 SST 的参数组合模型是所有训练模型中最好的, 特别是对海冰边缘区 (15%~80% 的海冰密集度) 的海冰密集度反演。与参考的海冰密集度值相关性很好, 相关系数达到了 0.981, RMSE 为 0.068, MAE 为 0.046, 可见 SST 的作用显著。与基于 GNSS-R 观测值的模型相比, RMSE 和 MAE 分别减少了 0.103 和 0.074, 结果表明了反演的海冰密集度准确度较高, 在海冰边缘区具有较好的预测能力, 也证明了模型结

果对 DDM 的依赖性。模型 70%~100% 的海冰密集度结果更加准确, 说明加入 DDM 观测值能改变 SST 在高海冰密集度反演效果不好的情况。DDM 在高海冰密集度低估, SST 难以反演高海冰密集度, 然而两者的结合能够取得很好的结果。

4.2 模型验证结果和评估

使用 2018 年 7 月 20-22 日、8 月 9-10 日、8 月 15-16 日 3 个时间段的数据验证 LightGBM 模型的预测能力。本研究分成了 3 种不同的海冰密集度区域, 0%~15% 的低密集度区、15%~80% 的海冰边缘区以及 80% 以上的高密集度区。图 4 是应用整个验证集对比基于 GNSS-R 观测值的模型和基于 DDM 特征参数、SST 的模型在不同海冰密集度区域与海冰密集度差值的箱线图, 中位线代表数据从小到大排序中间的数, 海冰密集度差值由真值减去反演值得到。图 5 反映了两个模型海冰密集度差异随海冰密集度大小的变化情况, 可见模型加入 SST 后, 不同海冰密集度区的反演精度都得到了提高, 在低密集度区能够显著改善只使用 DDM 特征参数引起的高估问题, 在海冰边缘区考虑 SST 的影响反演更加精准, 结果数据分布更加集中, 误差更小, 在高密集度区也有一定的改进效果, 异常值大大减少。其中加入了 SST 的模型在海冰边缘区验证结果的 R 、RMSE 和 MAE 分别为 0.901、0.096 和 0.073。

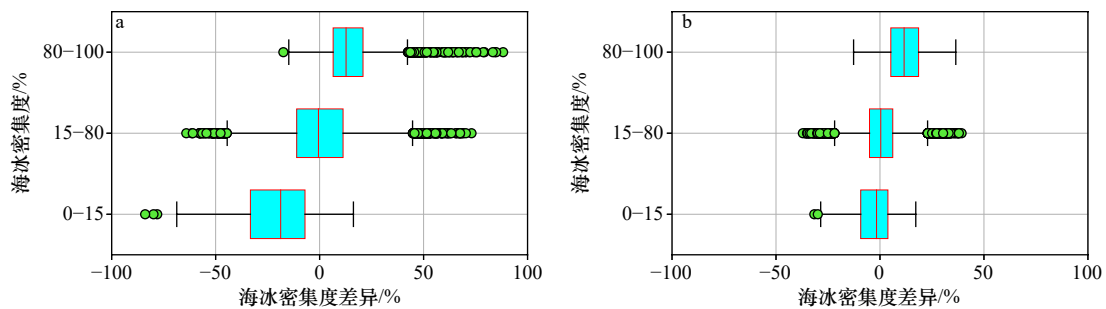


图 5 整个验证集不同海冰密集度区间与反演误差箱线图

Fig. 5 Box plots of different sea ice concentration intervals and retrieval error using the entire validation set

a. 基于 DDM 特征参数的模型; b. 基于 DDM 特征参数、SST 的模型

a. The model based on DDM observables; b. the model based on DDM observables and SST

对基于 DDM 特征参数、SST 的模型在不同时间段进行验证, 验证结果如图 6 所示。总体上看, 3 个验证结果的相关系数都达到了 94% 以上, 与训练结果差别都不大, 平均下来, 总体相关系数为 0.965, RMSE 为 0.061, MAE 为 0.090, 说明反演结果与 OSI SAF 产品的海冰密集度具有较好的一致性。海冰密集度较高时, SST 不能发挥优势, 反演的海冰密集度值有部分偏低, 其中距训练样本时间最近的验证结果最佳。验

证结果表明本文所提出的方案可以适用于北极夏季的海冰密集度反演, 能大大提高对海冰边缘区的反演精度, 基于 DDM 特征参数、SST 的模型应用于这 3 个不同的数据集中也能有较强的适应能力, 验证结果良好, 可见该模型具有一定的泛化性。

这里进一步进行 GNSS-R 的镜面反射点检验, 以 2018 年 8 月 9 日为例, 图 7 展示了以 OSI-SAF 的海冰密集度产品为底图的 GNSS-R 镜面反射点处海冰密集

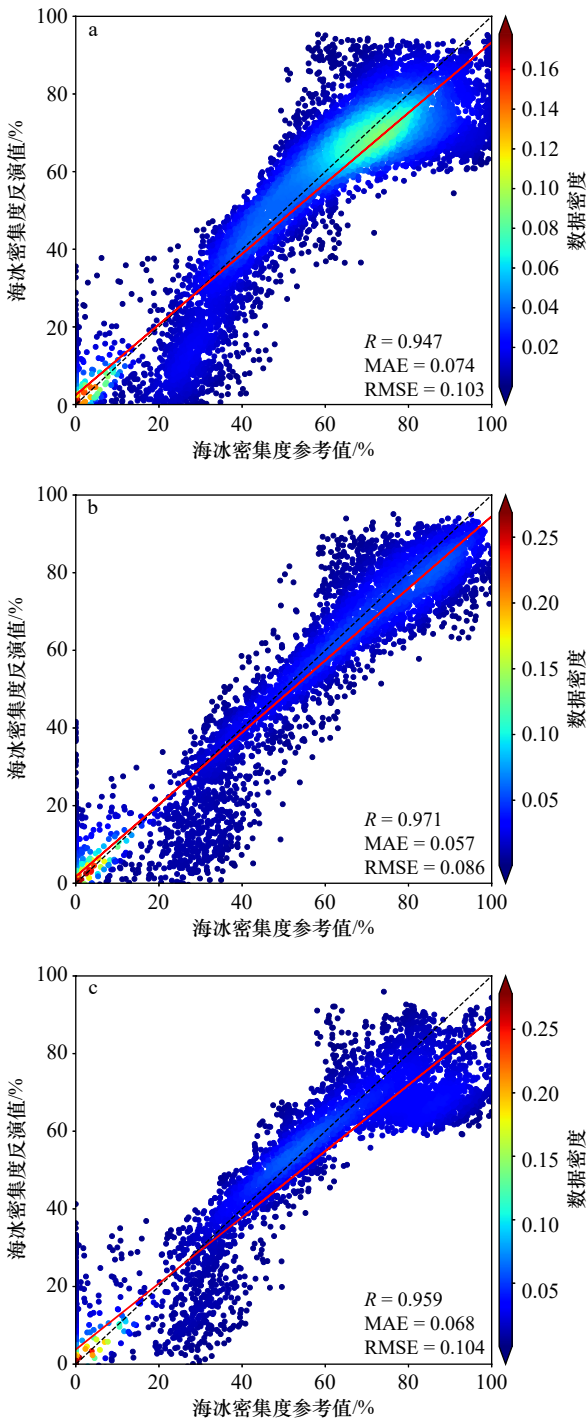


图 6 基于 DDM 特征参数和 SST 的反演结果与 OSISAF 海冰密集度对比散点图

Fig. 6 Scatter plot of retrieval results based on DDM observable against OSISAF sea ice concentration

a. 2018 年 7 月 20-22 日; b. 2018 年 8 月 9-10 日;
c. 2018 年 8 月 15-16 日

a. July 20-22, 2018; b. August 9-10, 2018; c. August 15-16, 2018

度估计结果, 右边颜色条表示 OSISAF 海冰密集度值, 下边的颜色条表示使用模型反演的海冰密集度值。根据 GNSS-R 镜面反射点的地面轨迹, 将模型的海冰

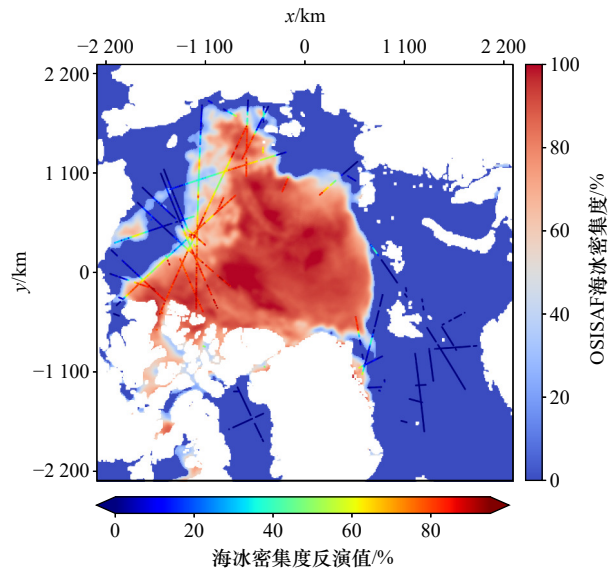


图 7 2018 年 8 月 9 日模型反演结果与 OSISAF 海冰密集度地图比较

Fig. 7 Comparison of model retrieval results with OSISAF sea ice concentration map on August 9th, 2018

密集度估计结果映射到 OSISAF 海冰密集度图上, 可以发现模型能够很好地识别冰水过渡区域, 较为精确地检测到海冰边缘, 海冰边缘区密集度的轻微变化也能清晰反映出来, 与 OSISAF 产品较为相似。

5 结论

本文通过提取 GNSS-R 时延多普勒图的特征参数, 结合 ERA5 再分析资料的海表温度, 与 OSISAF 海冰密集度产品训练来构建 LightGBM 机器学习回归模型, 研究星载 GNSS-R 多参数与北极融冰期海冰密集度的相关性。研究的目的是设计出适用于北极融冰期的高时空分辨率海冰密集度反演模型, 提高海冰边缘区反演精度。本文提出的基于 GNSS-R 特征参数、SST 的模型采用 LightGBM 算法, 在北极区域运行速度快且泛化能力强, 得到了与 OSISAF 海冰密集度值 96% 以上的相关性, 平均绝对误差为 0.061, 均方根误差为 0.090, 优于大多数其他北极冬季的研究结果。

本研究发现 GNSS-R 多参数(等效反射率 Γ 、时延多普勒图平均值 DDMA、信噪比 SNR、峰值功率比 PP、后沿斜率 TES、前沿斜率 LES、峰度 K 、偏度 SK)可以用于估计海冰密集度; SST 在海冰边缘区的反演效果很好, 但在高密度区难以估计; DDM 特征参数结合海表温度进行北极融冰期海冰密集度的反演取得了最佳的结果, 实验证明了在 GNSS-R 反演海冰密集度时加入 SST 这一因素可以提高反演性能, 面对平

静海水和高密集度冰区的错估问题也能够极大地缓解。后续研究可以引入风速参数来研究风条件的影响,并扩大样本数量提高模型的代表性。

目前研究领域内对海冰密集度的反演使用的DDM空间覆盖一般是 $125\text{ km} \times 125\text{ km}$,而本文提出的方法使用GNSS-R特征参数来反演海冰密集度,可

以提供更高的空间分辨率。相比于其他的遥感方法,也具有覆盖范围广、成本低和海冰边缘区反演精度高等优势。期望本文提出的新方法可以更好地监测北极海冰边缘区的变化,为北极融冰期的冰区航行提供海冰密集度的业务预报,对北极交通运输、极地考察等人类活动具有重要的现实价值。

参考文献:

- [1] Li Xiaoming, Sun Yan, Zhang Qiang. Extraction of sea ice cover by sentinel-1 SAR based on support vector machine with unsupervised generation of training datas[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2021, 59(4): 3040–3053.
- [2] Chen Jinlei, Kang Shichang, You Qinglong, et al. Projected changes in sea ice and the navigability of the Arctic Passages under global warming of 2 °C and 3 °C[J]. *Anthropocene*, 2022, 40: 100349.
- [3] 谢涛, 赵立. 海冰密集度卫星遥感反演研究进展 [J]. *海洋科学进展*, 2022, 40(3): 351–366.
Xie Tao, Zhao Li. Advances in sea ice concentration retrieval based on satellite remote sensing[J]. *Advances in Marine Science*, 2022, 40(3): 351–366.
- [4] Zavorotny V U, Gleason S, Cardellach E, et al. Tutorial on remote sensing using GNSS bistatic radar of opportunity[J]. *IEEE Geoscience and Remote Sensing Magazine*, 2014, 2(4): 8–45.
- [5] Yan Qingyun, Huang Weimin, Moloney C. Neural networks based sea ice detection and concentration retrieval from GNSS-R delay-Doppler maps[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 10(8): 3789–3798.
- [6] Yan Qingyun, Huang Weimin. Sea ice sensing from GNSS-R data using convolutional neural networks[J]. *IEEE Geoscience and Remote Sensing Letters*, 2018, 15(10): 1510–1514.
- [7] Yan Qingyun, Huang Weimin. Sea ice concentration estimation from TechDemoSat-1 data using support vector regression[C]//2019 IEEE Radar Conference (RadarConf). Boston: IEEE, 2019: 1–6.
- [8] Llaveria D, Munoz-Martin J F, Herbert C, et al. Sea ice concentration and sea ice extent mapping with l-band microwave radiometry and GNSS-R data from the FFSCat mission using neural networks[J]. *Remote Sensing*, 2021, 13(6): 1139.
- [9] Zhu Yongchao, Tao T, Zou J, et al. Spaceborne GNSS reflectometry for retrieving sea ice concentration using TDS-1 data[J]. *IEEE Geoscience and Remote Sensing Letters*, 2021, 18(4): 612–616.
- [10] Guo Wenfei, Du Hao, Guo Chi, et al. Information fusion for GNSS-R wind speed retrieval using statistically modified convolutional neural network[J]. *Remote Sensing of Environment*, 2022, 272: 112934.
- [11] Liu Hongsu, Jin Shuanggen, Yan Qingyun. Evaluation of the ocean surface wind speed change following the super typhoon from spaceborne GNSS-reflectometry[J]. *Remote Sensing*, 2020, 12(12): 2034.
- [12] Yan Qingyun, Huang Weimin. Spaceborne GNSS-R sea ice detection using delay-Doppler maps: First results from the U. K. Tech-DemoSat-1 mission[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2016, 9(10): 4795–4801.
- [13] Cartwright J, Banks C J, Srokosz M. Sea ice detection using GNSS-R data from TechDemoSat-1[J]. *Journal of Geophysical Research: Oceans*, 2019, 124(8): 5801–5810.
- [14] Rodriguez-Alvarez N, Holt B, Jaruwatanadilok S, et al. An Arctic sea ice multi-step classification based on GNSS-R data from the TDS-1 mission[J]. *Remote Sensing of Environment*, 2019, 230: 111202.
- [15] 邵连军, 胡磊, 李冰, 等. 基于 CART 决策树的星载 GNSS-R 海冰检测方法 [J]. *海洋测绘*, 2021, 41(1): 70–74.
Shao Lianjun, Hu Lei, Li Bing, et al. Sea ice detection using spaceborne GNSS-R data by CART decision tree[J]. *Hydrographic Surveying and Charting*, 2021, 41(1): 70–74.
- [16] Jales P, Unwin M. MERRByS product manual: GNSS reflectometry on TDS-1 with the SGR-ReSI[R]. Guildford: Surrey Satellite Technol. Ld., 2015.
- [17] Yao Ling, Lu Jiaying, Xia Xiaolin, et al. Evaluation of the ERA5 sea surface temperature around the Pacific and the Atlantic[J]. *IEEE Access*, 2021, 9: 12067–12073.
- [18] Zhu Yongchao, Yu Kegen, Zou Jingui, et al. Sea ice detection based on differential delay-Doppler maps from UK TechDemoSat-1[J]. *Sensors*, 2017, 17(7): 1614.
- [19] Santi E, Clarizia M P, Comite D, et al. Detecting fire disturbances in forests by using GNSS reflectometry and machine learning: a case study in Angola[J]. *Remote Sensing of Environment*, 2022, 270: 112878.
- [20] Clarizia M P, Ruf C S, Jales P, et al. Spaceborne GNSS-R minimum variance wind speed estimator[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2014, 52(11): 6829–6843.
- [21] Zhu Yongchao, Tao Tingye, Yu Kegen, et al. Sensing sea ice based on Doppler spread analysis of spaceborne GNSS-R data[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2020, 13: 217–226.
- [22] Yin Cong, Xia Junming, Huang Feixiong, et al. Sea ice detection with FY3E GNOS II GNSS reflectometry[C]//2021 IEEE Specialist

- Meeting on Reflectometry using GNSS and other Signals of Opportunity (GNSS+R). Beijing: IEEE, 2021: 36–38.
- [23] Alonso-Arroyo A, Zavorotny V U, Camps A. Sea ice detection using U. K. TDS-1 GNSS-R data[J]. *IEEE Transactions on Geoscience and Remote Sensing*, 2017, 55(9): 4989–5001.
- [24] Ke Guolin, Meng Qi, Finley T, et al. LightGBM: A highly efficient gradient boosting decision tree[C]//Proceedings of the 31st International Conference on Neural Information Processing Systems. Long Beach: Curran Associates Inc., 2017: 3149–3157.
- [25] Chen Jiajia, Shen Huanfeng, Li Xinghua, et al. Ground-level ozone estimation based on geo-intelligent machine learning by fusing in-situ observations, remote sensing data, and model simulation data[J]. *International Journal of Applied Earth Observation and Geoinformation*, 2022, 112: 102955.

Sea ice concentration retrieval using spaceborne GNSS-R during the melting period

Wang Yue¹, Xie Tao^{1,2,3,4}, Li Jian^{1,3,4}, Zhang Xuehong^{1,3,4}, Bai Shuying^{1,3,4}, Wang Minghua^{1,3,4}

(1. School of Remote Sensing and Geomatics Engineering, Nanjing University of Information Science and Technology, Nanjing 210044, China; 2. Laboratory for Regional Oceanography and Numerical Modeling, Qingdao Marine Science and Technology Center, 266200, Qingdao, Shandong Province, China; 3. Technology Innovation Center for Integration Applications in Remote Sensing and Navigation, Ministry of Natural Resources, Nanjing 210044, China; 4. Jiangsu Province Engineering Research Center of Collaborative Navigation/Positioning and Smart Application, Nanjing 210044, China)

Abstract: In this paper, a high spatial-temporal resolution sea ice concentration estimation method for the Arctic melting season is proposed, aiming to improve the overestimation of sea ice concentration in seawater by the Global Navigation Satellite System-Reflectometry (GNSS-R). The method utilizes machine learning algorithms to extract feature parameters from the Delay Doppler Maps (DDM) obtained through GNSS-R and combines them with sea surface temperature data to establish a LightGBM model. The inversion results are then subjected to correlation analysis and evaluation against reference sea ice concentration values. The model's performance is compared with the sea ice concentration product from OSI SAF, demonstrating good consistency, with correlation coefficient, mean absolute error, and root mean square error being 0.965, 0.061, and 0.090, respectively. This approach enables high-precision estimation of sea ice concentration in the Arctic marginal ice zone.

Key words: GNSS-R; DDM; melting season; sea ice concentration; LightGBM; Arctic