

张立,佟海昕,周菲. 基于随机森林算法的岸基 GNSS-R 海面风速反演[J]. 陕西气象, 2024(4):72-76.

文章编号:1006-4354(2024)04-0072-05

# 基于随机森林算法的岸基 GNSS-R 海面风速反演

张立<sup>1,2</sup>, 佟海昕<sup>1,2</sup>, 周菲<sup>1,2</sup>

(1. 山东省气象防灾减灾重点实验室, 济南 250031; 2. 东营市气象局, 山东东营 257000)

**摘要:**GNSS-R 技术可应用于海面风速反演,但传统的反演方法存在计算量大、反演精度不高等缺点。利用随机森林算法,对东营沿海东青五号石油平台上的岸基 GNSS-R 数据进行海面风速反演,结果显示:训练集的决定系数  $R^2$  为 0.94,均方根误差为 1.02 m/s;测试集的  $R^2$  为 0.84,均方根误差为 1.55 m/s。与传统方法相比,随机森林算法反演效果大幅提升,可应用于 GNSS-R 海面风速反演中。岸基 GNSS-R 反演海面风速中最重要的特征因子是风向和相关时间,不同风力等级下的反演结果也存在明显差异。

**关键词:**GNSS-R;随机森林;反演;海面风速

**中图分类号:**P228.4:TP181

**文献标识码:**A

全球导航卫星定位系统反射信号(GNSS-reflectometry,GNSS-R)<sup>[1]</sup>是全球卫星导航定位系统(global navigation satellite system,GNSS)的一个分支。自1993年 Martin-Neria 提出 PARIS (passive reflectometry and interferometry system)<sup>[2]</sup>概念后,GNSS-R 技术凭借低成本、全天候、高时空分辨率、覆盖范围广等优点<sup>[3-4]</sup>,在海面测高<sup>[5]</sup>、测风<sup>[6]</sup>、海冰探测<sup>[7]</sup>、土壤湿度测量<sup>[8]</sup>及海面溢油探测<sup>[9]</sup>等方面都有所应用。

海面在风的作用下会引起波浪,从而引起海面粗糙度的变化,海风大小与海面的粗糙程度直接相关,而海面粗糙度又会直接影响 GNSS 卫星的反射信号。传统的 GNSS-R 风速反演方法主要包括波形匹配方法<sup>[10]</sup>和经验函数法。其中,波形匹配方法是直接将实测时延功率曲线与理论曲线进行匹配来反演海面风速,其原理是通过波形图像对比进行反演,需要建立仿真波形数据库,计

算量大;经验函数法依靠函数模型,但用于反演海面风速的因子少,往往导致反演精度不高。近年来科研人员也开展了有理函数模型<sup>[11]</sup>、自适应 CDF 匹配<sup>[12]</sup>等反演方法研究,反演精度和准确率较传统方法均有提升。机器学习已在气象行业中的降水预测<sup>[13]</sup>、海雾预报<sup>[14]</sup>、气候预测<sup>[15]</sup>、雷电预警<sup>[16]</sup>等领域开展了广泛应用。本文引入机器学习中的随机森林算法进行海面风速反演。该算法相比波形匹配方法模型简单,不需要建立庞大的数据库;相比经验函数法,可以建立多个特征因子与海面风速的关系,从而提升反演效果。主要应用岸基 GNSS-R 数据进行反演。

## 1 反演方法

### 1.1 GNSS-R 反演海面风速原理

GNSS-R 设备的原始观测数据是直射信号和反射信号的复数序列,同时包含观测时间、信噪比、卫星高度角、卫星号、卫星系统(Beidou/GPS)等

**收稿日期:**2023-04-17

**作者简介:**张立(1977—),男,汉族,山东潍坊人,学士,高级工程师,主要研究方向为低能见度天气预报机理和海上大风等灾害性天气预报。

**通信作者:**佟海昕(1987—),男,满族,吉林通化人,学士,工程师,主要研究方向为雷电灾害防御技术和黄河三角洲生态环境演变机理。

**基金项目:**山东省气象局科研项目(2020sdqxm12,2022sdqzx14)

参数。假设海表面高度遵循高斯分布,GNSS 反射信号复数相关值的自相关函数近似为高斯函数<sup>[17]</sup>,

$$R(\Delta t) \approx A(h_{\text{swh}}, l_z, \theta, G_r) e^{-\frac{(\pi h_{\text{swh}} \sin \theta \Delta t)^2}{2(\lambda r_z)^2}} \quad (1)$$

式中: $R(\Delta t)$ 为时间序列的自相关函数; $h_{\text{swh}}$ 为海面有效波高; $l_z$ 为海面相关长度; $\theta$ 为卫星高度角; $G_r$ 为接收天线增益; $\lambda$ 为导航信号波长; $r_z$ 为海面相关时间,近似可表示为 $ah_{\text{swh}} + b$ ,其中 $a$ 和 $b$ 是未决系数。相关时间定义为自相关函数值下降至 $1/e$ 对应的时间。根据式(1)可得相关时间为

$$\tau_{\text{icf}} = \frac{\sqrt{2}\lambda}{\pi \sin \theta} \left( a + \frac{b}{h_{\text{swh}}} \right) \quad (2)$$

基于 Pierson-Moskowitz 海浪谱,成熟海域有效波高和风速( $U_{10}$ )的关系<sup>[18]</sup>为

$$h_{\text{swh}} = 0.0235 U_{10}^2 \quad (3)$$

假设岸基场景下卫星的高度角 $\theta$ 稳定不变,则相关时间和风速的关系可近似表示为

$$\tau_{\text{icf}} = a + b/U_{10}^2 \quad (4)$$

公式(4)为相关时间和海面风速的经验函数,其中 $a$ 和 $b$ 需要通过实验同位测量确定,可通过最小二乘法拟合得到。当 $a, b$ 这两个未决系数确定后,可利用反射信号计算得到相关时间,再利用公式(4)的逆函数 $U_{10} = \sqrt{b/(\tau_{\text{icf}} - a)}$ 反演得到风速。可以看出,经验函数的海面风速反演仅应用了相关时间一个特征因子。

## 1.2 随机森林反演风速流程

随机森林是决策树的集合,是将多个决策树的预测组合为一个模型。该算法可以将多棵树组合为一个集合模型,从而有效降低决策树等柔性模型的高方差,避免过拟合产生。利用随机森林反演风速,是在传统经验函数的基础上,应用包含相关时间在内的更多因子,利用随机森林算法建立与海面风速的关系,达到反演海面风速的目的。具体流程如图1所示。

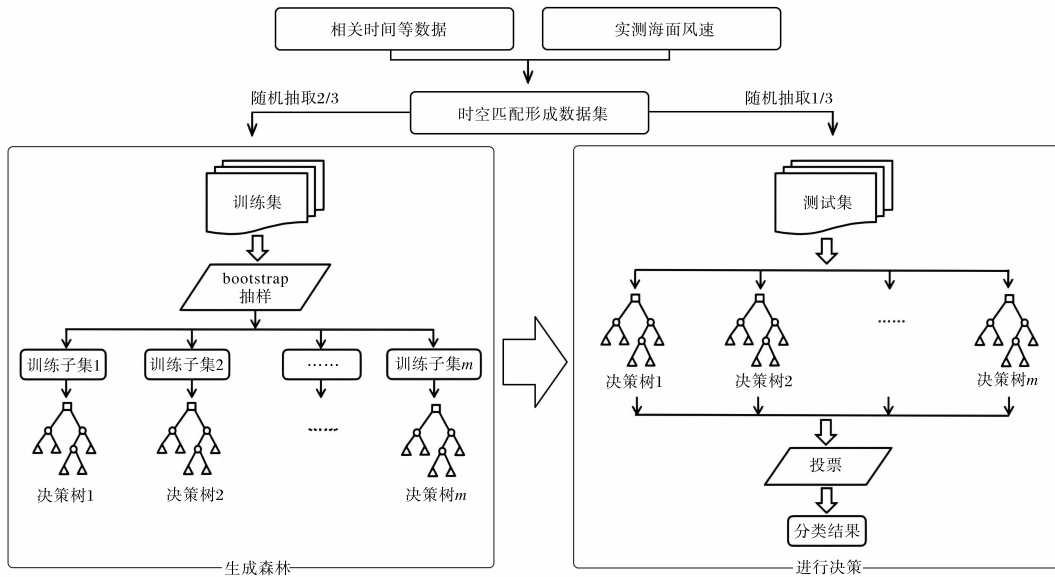


图1 随机森林反演风速流程图

## 2 反演数据

基于上述反演海面风速方法,进行了真实海面上 GNSS-R 观测实验,计算得到相关时间,与实况风速进行拟合分析。反演所用数据为岸基接收机数据。

### 2.1 数据获取场景

2021年10月26日—11月20日在山东省东

营市沿海青东5号石油平台开展 GNSS-R 反演的岸基连续观测实验。实验海域水深8~20 m,石油平台距岸边约500 m,高度约20 m,天线高度约50 m。直射天线使用3 dB全向天线,极化方式为右旋圆极化,垂直朝上放置,用于接收卫星直射信号;反射天线为四单元阵列式天线,天线增益12 dB,波束角 $\pm 20^\circ$ ,极化方式为左旋圆极化。

天线朝向为 205°, 天线仰角 45°, 信号采集器支持 GPS、北斗双系统的信号采集。本次实验经历 25 d, 累计获得 596 h 的复数相关值数据。同处架设一套云帆 XFY3-1 型风向风速仪, 取得实验同期实况数据。

### 2.2 特征因子选择和处理

GNSS-R 观测设备的原始数据按照 1.1 中方法处理后, 得到的相关时间作为特征因子, 单位是 ms, 数据时间间隔为 5 min。在此基础上, 加入观测时间、信噪比、卫星高度角、卫星号、卫星系统 (Beidou/GPS) 等参数作为特征因子。同时, 为区别向岸风和离岸风产生海浪的差异, 将风向作为特征因子加入反演。为了消除量纲影响, 采用离

差标准化方法对相关时间、信噪比、卫星高度角、卫星号、卫星系统 (Beidou/GPS)、风向、观测时间等 7 个特征因子进行了归一化处理。

## 3 反演结果分析

### 3.1 准确性

利用随机森林将 7 个特征因子加入反演, 反演过程将 2/3 的样本作为训练集, 1/3 的样本作为测试集, 结果见图 2。可以看出, 训练集的决定系数  $R^2$  为 0.94, 均方根误差为 1.02 m/s; 测试集的  $R^2$  为 0.84, 均方根误差为 1.55 m/s。同期观测数据利用经验函数也进行了反演, 其  $R^2$  为 0.06。利用随机森林反演的准确性比经验函数大幅提升。

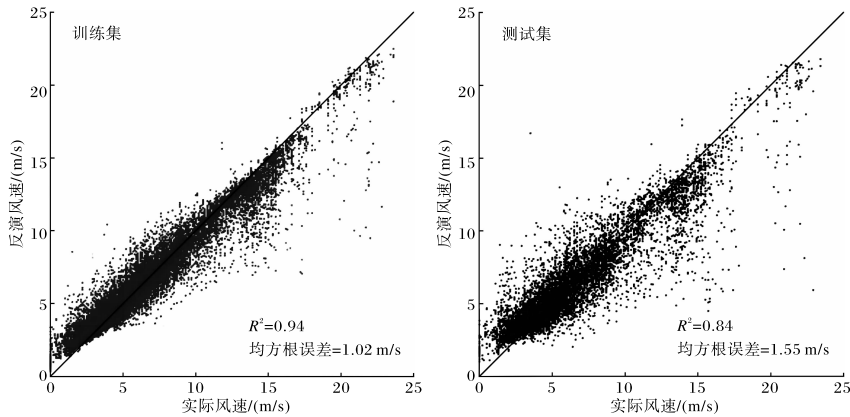


图 2 随机森林反演海面平均风速结果散点图(基于 2021-10-26—11-20 东营市沿海岸基 GNSS-R 数据, 下同)

### 3.2 特征因子重要性

在随机森林中, 每个决策树都只取一部分特征因子和数据样本建模, 这样可以避免过拟合。在构建多个决策树之后, 可以对每个特征因子的重要性进行排序, 以确定最有用的特征因子。在计算中, 首先以每个特征因子对最终预测贡献的样本比例确定该特征因子的相对重要性, 然后通过相对重要性与纯度减少相结合得到特征因子的重要性。每个特征因子的重要性, 实际就是每个特征因子在随机森林中每棵树上的贡献值。重要性分析结果以数值显示。

按照上述方法对 7 个特征因子进行重要性分析, 结果见图 3。可以看出, 风向的重要性最大, 相关时间其次, 这是对反演风速影响最大的 2 个

特征因子。之后依次是卫星高度角、信噪比、时间、卫星号和卫星类别。

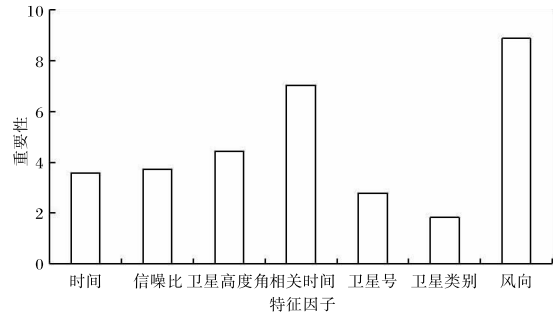


图 3 海面风速反演特征因子的重要性

### 3.3 不同风力等级样本反演结果

将样本按实况的风力等级进行分类, 利用随机森林分别反演并计算均方根误差, 结果见图 4。

可以看出:当实况风力为 3 级时均方根误差最小,为 0.5 m/s,反演效果最好;风力在 2~5 级时均方根误差小于 1 m/s,反演效果较好;风力在 0~1 级或 6 级以上时均方根误差均大于 1 m/s,反演效果不佳;当风力达到 7 级以上时均方根误差随风速增大而明显增大,达到 2 m/s 以上,反演效果相对较差。这也说明,基于岸基 GNSS-R 反演海面风速,反演效果受实况风速影响大。

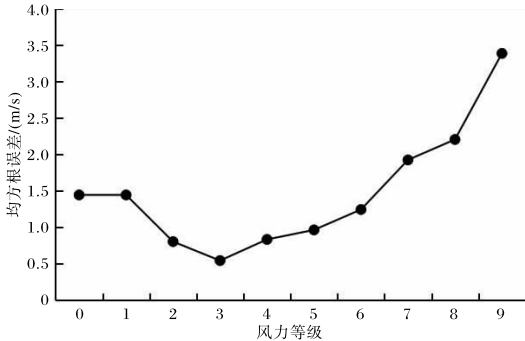


图 4 不同风力等级样本反演结果的均方根误差

### 3.4 卫星可用性

本实验接收了 32 颗卫星的 GNSS-R 数据,其中北斗卫星 8 颗, GPS 卫星 24 颗。针对反演海面风速的实验, GNSS-R 数据中的相关时间最为重要。分别将 32 颗卫星的相关时间与海面风速进行反演,得到每颗卫星反演结果的  $R^2$  (如图 5)。计算后可得北斗卫星  $R^2$  的平均值为 0.316, GPS 卫星  $R^2$  的平均值为 0.397,可见 GPS 卫星数据反演结果稍好于北斗卫星数据反演结果。

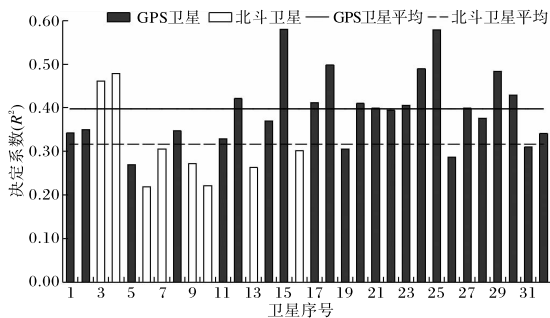


图 5 反演实验中 32 颗卫星数据的决定系数 ( $R^2$ )

## 4 结论和讨论

(1) 本文介绍了 GNSS-R 反演海面风速的原理,说明了传统反演方法的局限性,并采用随机森林算法解决反演因子单一问题。结果表明,基于

随机森林算法利用岸基 GNSS-R 数据反演海面风速的效果明显优于传统反演方法,具有应用价值。

(2) 本文使用了相关时间、信噪比等 7 个特征因子进行反演,如果要继续提升反演效果,需要更深入挖掘特征因子。在特征因子的重要性分析中,风向的重要性排位第一。因实验地点距离岸边仅约 500 m,导致向岸风和离岸风产生的波浪差异明显,故风向对反演结果影响大。对于距离岸边较远处的海面,风向因子的影响需做进一步研究。

(3) 利用 GNSS-R 反演海面风速,反演效果与实况风速有关。当实况风力等级在 2~5 级时反演效果较好,小于 2 级或者大于 5 级时反演效果均变差,特别是风力大于 7 级后反演效果明显变差。这也体现了基于 GNSS-R 反演海面风速的局限性。

(4) GPS 卫星数据反演结果稍好于北斗卫星数据反演结果仅限于本次实验研究。

### 参考文献:

- [1] 吕帆,修春娣,王峰,等. GNSS-R 海面风场反演模型仿真分析[J]. 导航定位学报,2018,6(3):87-91.
- [2] MARTIN-NEIRA M A. Passive reflectometry and interferometry system(PARIS): application to ocean altimetry [J]. ESA Journal,1993,17(4):331-335.
- [3] 万玮,陈秀万,彭学峰,等. GNSS 遥感研究与应用进展和展望[J]. 遥感学报,2016,20(5):858-874.
- [4] 胡媛,顾世森,刘卫,等. 星载 GNSS-R 监测海面目标技术进展[J]. 全球定位系统,2023,48(1):125-132.
- [5] 胡媛,陈行杨,顾旺旺,等. GNSS-R 海面测高现状及其常用方法研究进展[J]. 全球定位系统,2020,45(3):96-103.
- [6] 袁国良,张卫峰,卫豪杰. 基于 GNSS-R 的反演海面风速技术的研究[J]. 微型机与应用,2017,36(13):88-90.
- [7] 朱勇超. 星载 GNSS-R 海冰检测与海冰密集度反演方法研究[J]. 测绘学报,2020,49(12):1643.
- [8] 洪学宝,张波,阮宏梁,等. 基于相关功率修正的地基 GNSS-R 土壤湿度反演[J]. 北京航空航天大学

- 学报,2021,47(8):1558-1564.
- [9] 贾紫樱,张波,吴军,等. 岸基 GNSS-R 海上溢油探测方法[J]. 北京航空航天大学学报,2018,44(2):383-390.
- [10] 白伟华. GNSS-R 海洋遥感技术研究[D]. 北京:中国科学院研究生院(空间科学与应用研究中心),2008.
- [11] 刘原华,魏伟,牛新亮. 基于有理函数模型的 GNSS-R 岸基海面风速反演算法[J]. 现代电子技术,2019,42(1):65-67.
- [12] 杜皓,郭文飞,郭迟,等. 针对 GNSS-R 海面风速反演的自适应 CDF 匹配方法[J]. 武汉大学学报(信息科学版),2021,46(12):1924-1931.
- [13] 蒋薇,刘芸芸,陈鹏,等. 利用深度神经网络和先兆信号的江苏夏季降水客观预测方法[J]. 气象学报,2021,79(6):1035-1048.
- [14] 高松,徐江玲,刘桂艳,等. 基于机器学习的青岛市区近岸海雾集成预报方法[J]. 海洋科学,2021,45(3):33-42.
- [15] 贺圣平,王会军,李华,等. 机器学习的原理及其在气候预测中的潜在应用[J]. 大气科学学报,2021,44(1):26-38.
- [16] 刘海兵,卢秋芳. GMDH 神经网络在雷电预警中的应用[J]. 陕西气象,2023(5):74-79.
- [17] 王峰,杨东凯,李杰,等. 岸基多星多参数北斗 GEO 反射信号海面风速反演[J]. 北京理工大学学报,2022,42(9):961-968.
- [18] ANDERSON C, MARKLIN J, GOMMENGINGER C. Study of the impact of sea state on nadir looking and side looking microwave backscatter [R]. [S. l.]:[s. n.],2000.