

兰友国, 林明森, 张有广, 等. HY-2B 卫星和模式风速数据随机误差及相关性分析[J]. 海洋学报, 2023, 45(10): 183–194, doi:10.12284/hyxb2023131

Lan Youguo, Lin Mingsen, Zhang Youguang, et al. Analysis for random error and correlation of HY-2B satellite and model wind speed data[J]. Haiyang Xuebao, 2023, 45(10): 183–194, doi:10.12284/hyxb2023131

# HY-2B 卫星和模式风速数据随机误差及相关性分析

兰友国<sup>1,2</sup>, 林明森<sup>2\*</sup>, 张有广<sup>2</sup>, 杨典<sup>2</sup>

(1. 中国海洋大学 信息科学与工程学院, 山东 青岛 266100; 2. 国家卫星海洋应用中心, 北京 100081)

**摘要:** 当利用三配准 (Triple Collocation, TC) 方法进行误差分析时, 系统间随机误差 (简称误差) 不相关是一个重要前提假设, 而在实际应用中不同系统误差常存在相关性, 基于最小二乘法扩展配准 (Extended Collocation, EC) 方法能够在误差相关性存在情况进行误差分析, 但对于误差弱相关性情况不能够准确估计误差的标准差。为此本文提出利用 3 个误差独立系统对第四个系统进行误差估计的方法, 同时考虑误差相关性和表征误差, 在误差弱相关情况下能更精确估计系统误差的标准差。本文根据 HY-2B 卫星 3 个载荷风速观测数据集随机误差相互独立特点, 利用扩展三配准 (Extended Triple Collocation, ETC) 方法计算得到散射计、辐射计和高度计 3 个载荷风速产品误差的标准差分别为 0.600 m/s、0.742 m/s 和 0.533 m/s; 再对 ECMWF 再分析数据集 ERA5 的风速产品误差及相关性进行估计, 计算出 ERA5 再分析风速产品随机误差的标准差为 0.810 m/s, HY-2B 卫星散射计风速产品和 ERA5 再分析风速产品误差相关性为 0.231, HY-2B 卫星辐射计风速产品和 ERA5 再分析风速产品误差相关性为 0.105。本文提出利用已知 3 个误差独立数据集对第四个数据集误差及相关性进行估计的方法, 实现了在误差弱相关情况下对系统误差的标准差更为准确的估计, 有助于在同化和融合中更好地使用这些数据。

**关键词:** HY-2B 卫星; 海表面风速; 扩展三配准方法; 随机误差; 误差相关性

中图分类号: P717

文献标志码: A

文章编号: 0253-4193(2023)10-0183-12

## 1 引言

海表面风通过驱动海洋环流促进大气和海洋之间相互作用, 因此海面风场是研究海洋环境及变化的重要物理参数, 对海表面风的观测有助于飓风等海洋灾害预报和预防, 促进全球海气相互作用的研究<sup>[1]</sup>。HY-2B 卫星搭载的 3 个载荷都能对海表面风进行观测, 并且已经业务化运行, 对观测风速随机误差 (简称误差) 进行评估有助于推进 HY-2B 风场产品同化和融合等研究和应用, 进而提高海洋环境的预报精度。

在误差分析时, 通常需要将观测数据与模式数据

或浮标观测数据进行比较, 由于模式数据和浮标观测数据自身也存在误差, 这种直接比较方法得到误差精度存在偏差, 针对此问题, 1998 年 Stoffelen<sup>[2]</sup> 首先提出了三配准 (Triple Collocation, TC) 的方法, 但是没有考虑到误差间存在相关性的影响。随后 TC 方法应用于海面风场、海面温度和有效波高等海洋物理量误差分析, 当然 TC 方法也应用于地球环境观测其他领域, 特别是在土壤湿度观测领域。随着地球观测数据源越来越多, 往往需要对 3 个及以上数据集进行分析, 而被分析的部分数据集随机误差常存在相关性, 而误差相关性存在对误差精确估计有着重要影响。

收稿日期: 2023-05-06; 修订日期: 2023-06-19。

作者简介: 兰友国 (1974—), 男, 福建省福州市人, 正高级工程师, 研究方向为海洋信息探测处理。E-mail: lyg@mail.nsoas.org.cn

\* 通信作者: 林明森 (1963—), 男, 福建省莆田市人, 研究员, 研究方向为微波遥感。E-mail: mslin@mail.nsoas.org.cn

基于 TC 误差分析方法在发展中,由对 3 个数据集分析扩展到对多个数据集分析,也逐渐考虑误差相关性对误差分析精度的影响。Scipal 等<sup>[3]</sup>在 2008 年提出一种对误差仿射模型系数进行定标通用方法,并对全球土壤湿度数据误差进行估计。Vogelzang 等<sup>[4-5]</sup>于 2011 年利用 TC 方法对高分辨率的 ASCAT 风场质量评估开展研究,利用谱分析的方法来计算表征误差,计算得到 ASCAT 分辨率 12.5 km 海表面风场  $U$  分量和  $V$  分量误差精度为 0.7 m/s 和 0.8 m/s。Zwieback 等<sup>[6]</sup>于 2012 年研究 TC 方法统计特性和对违反假设的敏感度,利用协方差矩阵分析方法将配准数据集的数量扩展到 3 个以上。Pierdicca 等<sup>[7]</sup>于 2015 年提出一种基于最小二乘法 QC(Quadruple Collocation)方法对 4 组数据集进行估计,但是没有考虑不同数据集间相关性;Pan 等<sup>[8]</sup>于 2015 年利用最小二乘法在希尔伯特空间开展多元配准问题研究,对误差进行分组,允许组内数据集误差存在相关性,并利用土壤湿度的数据集进行验证;Gruber 等<sup>[9]</sup>于 2016 年提出基于最小二乘法扩展配准(Extended Collocation, EC)方法,对任意多个数据集误差进行估计,允许处理有限几个数据集的相关性问题,并用仿真数据和土壤湿度的数据集进行验证。McColl 等<sup>[10]</sup>于 2014 年利用基于协方差扩展三配准(Extended Triple Collocation, ETC)方法计算 ASCAT、ECMWF 模式风场和浮标现场观测数据集的随机误差,同时进一步推导出了 3 组数据集相对于观测物理量真值的相关系数,随后 Konings 等<sup>[11]</sup>利用仿真数据开展了相应验证研究。Lin 等<sup>[12]</sup>2015 年利用 TC 方法对浮标、ASCAT 和模式风场在高变化情况数据质量进行分析,提出一种在高变化情况下风场表征误差估计的方法。Abdalla 和 De Chiara<sup>[13]</sup>于 2017 年利用 TC 方法对 ECMWF 的预报数据和再分析数据、ASCAT-A、ASCAT-B 和 Jason-2 的风速数据进行分析,利用预报提前时间越长,散射计和模式预报数据误差相关性逐渐减小到 0 的分析结果,直接估计出 ASCAT 和 Jason-2 随机误差,并进一步求解出再分析数据集 ERA-Interim(ECMWF Reanalysis-Interim)风速数据和 ASCAT-B 数据误差相关系数。Vogelzang 和 Stoffelen<sup>[14-15]</sup>于 2021 年和 2022 年对海表面风场的 4 组数据和 5 组数据情况进行 EC 分析,同时考虑不同系统误差相关性以及表征误差存在情况,利用对数对协方差方程进行变换来简化协方差方程求解。2022 年吕思睿等<sup>[16]</sup>利用三配准方法对多源散射计和

辐射计风速与 ERA5 和浮标的等效力风进行分析,定义风速误差横向对比的指示因子,实现多源数据的相对误差估计。

以往研究虽然已经考虑误差相关性,但在弱误差相关性情况下的误差估计还未被研究过,这是本文研究的重点问题。本文将以 ECMWF 再分析数据 ERA5 的风场数据和 HY-2B 3 个载荷风场观测数据共 4 个数据集开展研究。

## 2 方法

本节简单对基于 TC 方法发展而来的 ETC 方法和 EC 方法进行介绍,着眼于利用这几种方法对不同观测系统的误差相关性的分析,TC 方法最基本的假设是不同观测系统误差是不相关的,但是在实际分析中不同观测系统的误差经常存在相关性,因此需要在误差估计中加以考虑。

在分析问题时要注意区分误差相关性与观测的相关性<sup>[13]</sup>。观测相关性指的观测值与被观测物理量相关程度,观测相关性越大,意味着观测值与被观测物理量越相关,因此越大越好。误差相关性是两个系统随机误差相关程度,如果误差相关性存在,则观测数据集误差协方差不等于 0,误差相关性主要由 3 种原因导致:(1)两个测量系统具有相同测量原理;(2)一个系统的数据用于标定第二个系统的反演算法;(3)一个系统同化另外一个系统测量数据。

### 2.1 误差模型

TC 方法是由 Stoffelen<sup>[2]</sup>首先提出,对 3 个误差独立观测系统进行误差分析,并假定一阶线性近似能够满足误差估计要求,采用模型见下式:

$$X_i = \beta_i(t + \varepsilon_i), \quad i = 1, 2, 3, \quad (1)$$

式中,  $t$  表示要观测物理量真值;  $X_i$  表示第  $i$  个系统观测值;  $\varepsilon_i$  表示随机误差。在误差求解过程中要任意先选定一个系统作为参考系统,然后对另外两个系统进行标定,也就是将系数  $\beta_i$  变换到 1,再对误差进行求解。后续一些研究通常采用仿射模型<sup>[6,10]</sup>,这两种模型等效,但是仿射模型更为通用,仿射模型见下式:

$$X_i = X'_i + \varepsilon_i = \alpha_i + \beta_i t + \varepsilon_i, \quad i = 1, 2, 3, \quad (2)$$

式中,  $t$  表示要观测物理量真值;  $X_i$  表示第  $i$  个系统观测值,  $X'_i$  表示定标后值,如果将  $X_i$  表示成  $X'_i$  可以给计算上带来方便。  $X_i$ ,  $X'_i$ ,  $t$ ,  $\varepsilon_i$  都是随机变量,  $\alpha$ 、 $\beta$  分别为最小二乘估计的截距和斜率。基于 TC 方法有 3 个前

提假设<sup>[2,10]</sup>: (1) 随机误差  $\varepsilon_i$  和被观测物理量  $t$  不相关, 即协方差  $Cov(t, \varepsilon_i) = 0$ ; (2) 如果不考虑表征误差, 不同系统随机误差  $\varepsilon_i$ ,  $i = 1, 2, 3$  不相关, 即协方差  $Cov(\varepsilon_i, \varepsilon_j) = 0$ ,  $i \neq j$ ; (3) 随机误差  $\varepsilon_i$ ,  $i = 1, 2, 3$  是无偏的, 即均值  $E(\varepsilon_i) = 0$ 。

## 2.2 ETC 方法

### 2.2.1 不考虑误差相关性

ETC 方法是由 McColl 等<sup>[10]</sup>于 2014 年提出, 采用仿射误差模型 (见式 (2)), 不需要提前将系数  $\beta$  变换到 1, 而可以通过方差直接进行求解, 利用协方差直接推导出绝对误差和相关系数, 不同系统测量观测数据协方差可以表示为

$$\begin{aligned} Cov(X_i, X_j) &= E(X_i X_j) - E(X_i)E(X_j) \\ &= \beta_i \beta_j \sigma_t^2 + \beta_i Cov(t, \varepsilon_j) + \beta_j Cov(t, \varepsilon_i) + Cov(\varepsilon_i, \varepsilon_j). \end{aligned} \quad (3)$$

根据模型式 (2), 引入一个新变量  $\theta_i = \beta_i \sigma_t$ , 我们可以得到

$$Q_{ij} = \begin{cases} \theta_i \theta_j, & i \neq j, \\ \theta_i^2 + \sigma_{\varepsilon_i}^2, & i = j. \end{cases} \quad (4)$$

对于 3 个观测系统情况, 正好有 6 个方程和 6 个未知数, 就可以求解得到相应误差的标准差:

$$\sigma_\varepsilon = \begin{bmatrix} \sqrt{Q_{11} - \frac{Q_{12}Q_{13}}{Q_{23}}} \\ \sqrt{Q_{22} - \frac{Q_{12}Q_{23}}{Q_{13}}} \\ \sqrt{Q_{33} - \frac{Q_{13}Q_{23}}{Q_{12}}} \end{bmatrix}. \quad (5)$$

对于相关系数, 根据最小二乘法估计, 模型斜率  $\beta_i$  可以表示为

$$\beta_i = \rho_{t, x_i} \frac{\sqrt{Q_{ii}}}{\sigma_t}, \quad (6)$$

则相关系数进一步表示为

$$\rho_{t, x_i} = \frac{\beta_i \sigma_t}{\sqrt{Q_{ii}}} = \frac{\beta_i \sigma_t}{\sqrt{\beta_i^2 \sigma_t^2 + \sigma_{\varepsilon_i}^2}}, \quad (7)$$

进一步利用协方差公式可以推导得到

$$\rho_{t, x_i} = \pm \begin{bmatrix} \sqrt{\frac{Q_{12}Q_{13}}{Q_{11}Q_{23}}} \\ \text{sign}(Q_{13}Q_{23}) \sqrt{\frac{Q_{12}Q_{23}}{Q_{22}Q_{13}}} \\ \text{sign}(Q_{12}Q_{23}) \sqrt{\frac{Q_{13}Q_{23}}{Q_{33}Q_{12}}} \end{bmatrix}. \quad (8)$$

### 2.2.2 误差相关性存在情况

如果 3 个系统中有部分系统误差的标准差已知, 那么可以对部分系统误差存在相关性情况进行求解。误差相关性和表征误差都表现为协方差不等于 0, 在数学上不存在区别。2.2.1 节公式都是没有考虑表征误差和误差相关性, 如果考虑表征误差和误差相

关性存在情况, 也就是存在  $Cov(\varepsilon_i, \varepsilon_j) \neq 0$ ,  $i \neq j$ , 那么在计算时候就需要从协方差中减掉误差的协方差再进行计算, 即  $Q_{ij} - Cov(\varepsilon_i, \varepsilon_j)$ 。假设有 3 个观测系统, 用 1、2、3 来表示, 本文对两种情况进行公式推导: (1) 只有一对系统间误差存在相关性; (2) 有两对系统间误差存在相关性。

第一种情况, 只有一对系统间误差存在相关性: (1) 系统 1 误差与系统 2 误差存在相关性, 相应协方差用  $C$  表示; (2) 系统 2 与系统 3 之间存在表征误差, 用  $Y$  表示; (3) 系统 1 误差与系统 3 误差不存在相关性。根据式 (5) 可以得到误差的标准差的表达式:

$$\sigma_\varepsilon = \begin{bmatrix} \sqrt{Q_{11} - \frac{(Q_{12}-C)Q_{13}}{Q_{23}-Y}} \\ \sqrt{Q_{22} - \frac{(Q_{12}-C)(Q_{23}-Y)}{Q_{13}}} \\ \sqrt{Q_{33} - \frac{Q_{13}(Q_{23}-Y)}{Q_{12}-C}} \end{bmatrix}. \quad (9)$$

如果系统 2、系统 3 误差的标准差已知, 那么可求解误差协方差和表征误差:

$$C = Q_{12} - Q_{13} \sqrt{\frac{Q_{22} - \sigma_2^2}{Q_{33} - \sigma_3^2}}, \quad (10)$$

$$Y = Q_{23} - \sqrt{(Q_{33} - \sigma_3^2)(Q_{22} - \sigma_2^2)}. \quad (11)$$

第二种情况, 有两对系统间误差存在相关性: (1) 系统 1 误差与系统 2 误差存在相关性, 相应协方差用  $C$  表示; (2) 系统 2 与系统 3 之间存在表征误差, 用  $Y$  表示; (3) 系统 1 误差与系统 3 误差也存在相关性, 相应协方差用  $D$  表示。这样在系统 1、系统 2 和系统 3 误差的标准差都已知情况下, 可以求解协方差  $C$ 、 $D$  和表征误差  $Y$ , 在这种情况下式 (5) 变为

$$\sigma_\varepsilon = \begin{bmatrix} \sqrt{Q_{11} - \frac{(Q_{12}-C)(Q_{13}-D)}{Q_{23}-Y}} \\ \sqrt{Q_{22} - \frac{(Q_{12}-C)(Q_{23}-Y)}{Q_{13}-D}} \\ \sqrt{Q_{33} - \frac{(Q_{13}-D)(Q_{23}-Y)}{Q_{12}-C}} \end{bmatrix}, \quad (12)$$

进一步可以推导  $C$ 、 $D$  表达式为

$$C = Q_{12} - \sqrt{(Q_{11} - \sigma_1^2)(Q_{22} - \sigma_2^2)}, \quad (13)$$

$$D = Q_{13} - \sqrt{(Q_{11} - \sigma_1^2)(Q_{33} - \sigma_3^2)}, \quad (14)$$

这种情况下推导出来的  $Y$  的表达式与式 (11) 相同。

## 2.3 EC 方法

EC 方法是基于最小二乘法对 4 个及以上观测系统的数据分析, 并且允许其中一对或者两对系统存在

误差相关性<sup>[6,9,14]</sup>。在前面 ETC 公式基础上进行推导, 假设现在有 4 个系统  $X_i, X_j, X_k, X_l$ , 其中系统  $X_i, X_j$  的误差存在相关性, 即  $Cov(\varepsilon_i, \varepsilon_j) \neq 0$ , 则由式(4)变为

$$Q_{ij} = \begin{cases} \theta_i \theta_j + \sigma_{\varepsilon_i \varepsilon_j}, & i \neq j, \sigma_{\varepsilon_i \varepsilon_j} \neq 0, \\ \theta_i^2 + \sigma_{\varepsilon_i}^2, & i = j, \end{cases} \quad (15)$$

$$\frac{Q_{ij} Q_{ik}}{Q_{jk}} = \beta_i^2 \sigma_i^2, \quad \forall i, j, k \text{ 当 } \sigma_{\varepsilon_i \varepsilon_j} = \sigma_{\varepsilon_i \varepsilon_k} = \sigma_{\varepsilon_j \varepsilon_k} = 0, \\ \frac{Q_{ik} Q_{jl}}{Q_{kl}} = \beta_i \beta_j \sigma_i^2, \quad \forall i, j, k, l \text{ 当 } \sigma_{\varepsilon_i \varepsilon_k} = \sigma_{\varepsilon_j \varepsilon_l} = \sigma_{\varepsilon_k \varepsilon_l} = 0, \quad (16)$$

将式(15)和式(16)写成矩阵形式:

$$y = \begin{bmatrix} \sigma_i^2 \\ \sigma_{ij} \\ \frac{Q_{ij} Q_{ik}}{Q_{jk}} \\ \frac{Q_{ij} Q_{kl}}{Q_{jl}} \\ \frac{Q_{ij} Q_{kl}}{Q_{jl}} \end{bmatrix}, \quad A = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix}, \quad x = \begin{bmatrix} \beta_i^2 \sigma_i^2 \\ \beta_i \beta_j \sigma_i^2 \\ \sigma_{\varepsilon_i}^2 \\ \sigma_{\varepsilon_i \varepsilon_j} \end{bmatrix}, \quad (17)$$

$$y = A \cdot x. \quad (18)$$

### 3 数据

HY-2B 卫星 2018 年发射, 星上搭载的散射计、辐射计和高度计能够对海表面风场进行同步观测, HY-2B 数据连续稳定, 并与后续发射的 HY-2C 和 HY-2D 卫星组成中国首个海洋动力环境卫星观测星座<sup>[17]</sup>。得益于 HY-2B 卫星 3 个载荷时空上同步观测, 因此能够在较短时间内获得大量时空配准数据, 本研究选取 2021 年 7 月到 2021 年 11 月 1742 轨 HY-2B 卫星 3 个载荷二级产品及同一时间段的 ECMWF 的 ERA5 再分析数据用于分析。HY-2B 3 个载荷 1742 轨数据在空间上为全球覆盖数据, 但为了减小极地海冰影响, 将南北纬度大于 65° 数据进行掩膜剔除。

#### 3.1 HY-2B 卫星散射计数据

HY-2B 卫星的散射计工作在 Ku 频段(13.256 GHz), 采用双极化(HH 和 VV)笔形圆锥扫描体制, 两个波束在不同固定入射角不同方位对海表面进行观测。在进行风场反演时, 将观测区域沿轨道方向和垂直轨道方向划分成 25 km × 25 km 观测面元, 通过地球物理函数 NSCAT-4, 利用极大似然法(MLE)对每个面元内多个方位观测后向散射系数进行反演, 利用圆中数滤波法进行模糊解去除得到 L2B 级产品。由于在反演过程中采用 ECMWF 预报数据作为背景初始场, 因此在分析中需要考虑散射计的反演风场误差和 ECMWF 再分析数据误差的相关性。

#### 3.2 HY-2B 卫星辐射计数据

HY-2B 卫星的辐射计一共配置 5 个工作频率: 6.6 GHz、10.7 GHz、18.7 GHz、23.8 GHz、37.0 GHz, 分

别采用 VH、VH、VH、VV 和 VH 极化工作方式, 扫描刈幅大于 1 600 km。HY-2B 辐射计采用多元线性回归法进行海面风速的反演, 也就是建立海面风速和多个通道亮温的线性关系模型, 通过统计回归算法求出最优系数解。目前业务分发的 L2B 级产品, 其中包括 Res0、Res6、Res10、Res18 4 种分辨率, Res0 为原始分辨率, Res6、Res10、Res18 分别为将亮温重采样到 6.925 GHz、10.7 GHz、18.7 GHz 对应频率的产品。而 Res0 中包括两种风速产品, 一种正常风速产品 SSW, 最高风速不超过 23 m/s; 另外一种为高风速试验产品 HSSW, 最高风速高达 60 m/s。用于本文分析的是 Res0 正常风速产品 SSW, 空间分辨率为 40 km × 50 km。

#### 3.3 HY-2B 卫星高度计数据

HY-2B 卫星的高度计有 13.58 GHz 和 5.25 GHz 两个工作频率, 脉冲足迹小于 2 km。HY-2B 卫星的高度计利用 Gourrion 等<sup>[18]</sup>提出双参数模型进行风速反演, 根据高度计观测到后向散射系数与海面风速和海况相关, 因此在风速反演函数中引入有效波高, 利用多层感知神经网络求解模型系数。用于本文分析的是 HY-2B 卫星高度计临时地球物理数据(Interim Geophysical Data Records, IGDR), IGDR 是利用 MOE 定轨数据和波形重构等方法得到的未经校正数据产品, 数据中主要包括了有效波高、海面风速、海面高度及用于计算海面高度所需的相关校正参数。IGDR 风速产品星下点分辨率约为星下点 7 km。

#### 3.4 ERA5 风场数据

ECMWF 再分析数据集 ERA5 的海表面风场用于本文研究, ERA5 再分析数据特点是高时间分辨率和空间分辨率, 在时间分辨率为逐小时, 空间分辨率已经从 ERA-Interim 的 79 km 提升到 31 km<sup>[16,19]</sup>。数据是从 Copernicus Climate Data Store 上获取逐小时空间分辨率为 0.25° × 0.25° 东西风和南北风的数据, 再计算出风速。

### 4 数据预处理

本文利用 ETC 和 EC 的方法对 HY-2B 卫星散射计、辐射计和高度计 3 个载荷风速产品以及 ECMWF 再分析产品进行误差结构的分析, 为了清楚和方便对这些数据集进行标识, 利用 4 个英文大写字母分别代表 4 个数据集:

- (1)S: HY-2B 卫星散射计数据集;
- (2)R: HY-2B 卫星辐射计数据集;
- (3)A: HY-2B 卫星高度计数据集;
- (4)E: ECMWF-ERA5 再分析数据集;

(5) ESAR: 表示 E、S、A 和 R 这 4 个数据集的一组排列;

(6) E/S/A/R: 表示 E、S、A 和 R 这 4 个数据集组合情况。

ESAR 表示包含 ECMWF-ERA5 再分析数据、HY-2B 卫星散射计数据、HY-2B 卫星高度计数据和 HY-2B 卫星辐射计数据 4 个数据集在内一组配对数据, 并且是按 E、S、A 和 R 这样顺序排列, 这是由于 EC 方法在分析时与数据集排列顺序相关。

#### 4.1 误差相关性定性分析

以前几乎所有的研究都是针对不同观测系统不同时间和空间数据进行配准分析, 在进行配准时不可避免会引入解释性误差<sup>[2]</sup>, 为了避免空间配准误差经常将不同数据集配准到一个统一空间网格上, 时间选取尽量接近, 但是还是会或多或少引入一些误差。HY-2B 卫星搭载微波散射计、微波辐射计和雷达高度计, 同时对相同区域的海表面风场进行独立的测量, 因此时空配准误差基本不存在。微波散射计和高度计虽然都是利用后向散射系数进行观测, 但是观测方式和反演方法不同, 两个载荷反演风场误差相关性很小基本可以忽略; 主动雷达微波观测和辐射计观测机理不同, 并且辐射计在反演过程没有引入背景场, 因此观测误差相关性也不存在。总而言之, 根据分析可以认为 HY-2B 卫星 3 个载荷风速产品数据观测上是相互独立, 误差间不存在相关性。另外 HY-2B 卫星散射计在反演过程中使用 ECMWF 预报数据作为背景场, 而且 HY-2B 卫星散射计数据也进入到 ECMWF 同化系统中, 所以认为 ECMWF 再分析数据误差和 HY-2B 卫星散射计风速数据集误差存在相关性。以上两点作为先验知识, 用于后面分析结果的验证, 也是本文误差相关性分析的出发点。

#### 4.2 数据配准

用于本文研究 HY-2B 卫星的散射计、辐射计和高度计 3 个载荷的风速产品在时空上是同步观测的, 但是分辨率不一样, 散射计空间分辨率为 25 km, 辐射计空间分辨率为 40~50 km, 高度计分辨率为星下点 7 km。为了减小表征误差影响, 通常将数据配准到一个  $0.5^\circ \times 0.5^\circ$  经纬度空间网格上<sup>[3]</sup>, 因此本文将落在配准公共网格内数据通过反向距离加权平均方法将数据配准到公共网格上。在配准过程中, 采用 HY-2B 辐射计降雨标识来标识降雨网格, 只要网格内有一个点受到降雨污染就标记该网格受到降雨污染。ECMWF 风场数据的空间分辨率为  $0.25^\circ \times 0.25^\circ$ , 时间分辨率为 1 h, 通过时空线性差值方法将数据配

准到  $0.5^\circ \times 0.5^\circ$  经纬度网格上, 时间点配准到 HY-2B 观测时间。经过配准得到包含 4 个数据集的 297 700 对配准数据。

#### 4.3 数据预处理及异常值去除

在使用传统 TC 方法进行误差分析时, 通常通过交互迭代方式进行模型系数(式(2)中的  $\alpha$ 、 $\beta$ )定标和异常值的剔除, 虽然 ETC 不需要提前对模型系数定标, 但还是需要对异常值进行剔除, 因为所有基于 TC 分析方法对异常值都非常敏感。第一步, 先利用 HY-2B 卫星 3 个载荷产品质量标识进行数据初步筛选, 剔除掉受陆地与海冰影响的面元数据、反演不成功面元数据和南北纬度大于  $65^\circ$  极地数据; 第二步把受到降雨影响的数据剔除, 根据 HY-2B 卫星辐射计降雨标记进行剔除, 总共剔除掉 83 727 个点, 剔除了近 21.8% 数据; 第三步, 假定 ECMWF 再分析风速数据异常值可以忽略, 以此作为参考, 将 HY-2B 卫星 3 个载荷观测值分别减去 ECMWF 对应风速值, 然后将数据大于 4 个标准差的数据剔除掉, 总共剔除 2 041 个点, 只剔除不到 1% 的数据, 最后得到 297 700 对配准数据。可以看出在剔除降雨污染数据后, 数据异常点已经非常少了。

从图 1(Hexbin 图, 也称二维直方图, 它展示了每个小六边形中观测点的数量)可以看出配准数据完成了大部分异常值剔除。

## 5 结果分析

### 5.1 配准数据概率密度分布

Stoffelen<sup>[2]</sup>认为对风场  $U$ 、 $V$  分量进行误差分析要比对风速和风向误差分析更为方便, 因为风场  $U$ 、 $V$  分量成近似高斯分布, 但是高斯的分布不是 TC 方法的强制要求, 只是更便于分析。另外一方面风速观测数据本身也近似对称(7 m/s)分布(图 2), HY-2B 卫星 3 个载荷观测风速与 ECMWF 风速差近似服从高斯分布(图 3), HY-2B 卫星 3 个载荷风速观测相互差值近似服从高斯分布, 因此在一定程度上我们可以认为这几个风速数据集的误差也近似服从高斯分布。

从图 3 中可以看出 HY-2B 卫星高度计测量的风速总体上比 ECMWF 偏小, 总体上 HY-2B 卫星散射计风速比 ECMWF 偏大, HY-2B 卫星辐射计风速与 ECMWF 再分析风速最为接近; 从图 4 可以看出散射计和高度计风速相差最小, 而辐射计与高度计风速相差最大。

### 5.2 基于 EC 的误差及相关性分析

本文采用 2016 年 Gruber 等<sup>[9]</sup>, Vogelzang 等<sup>[14]</sup>提

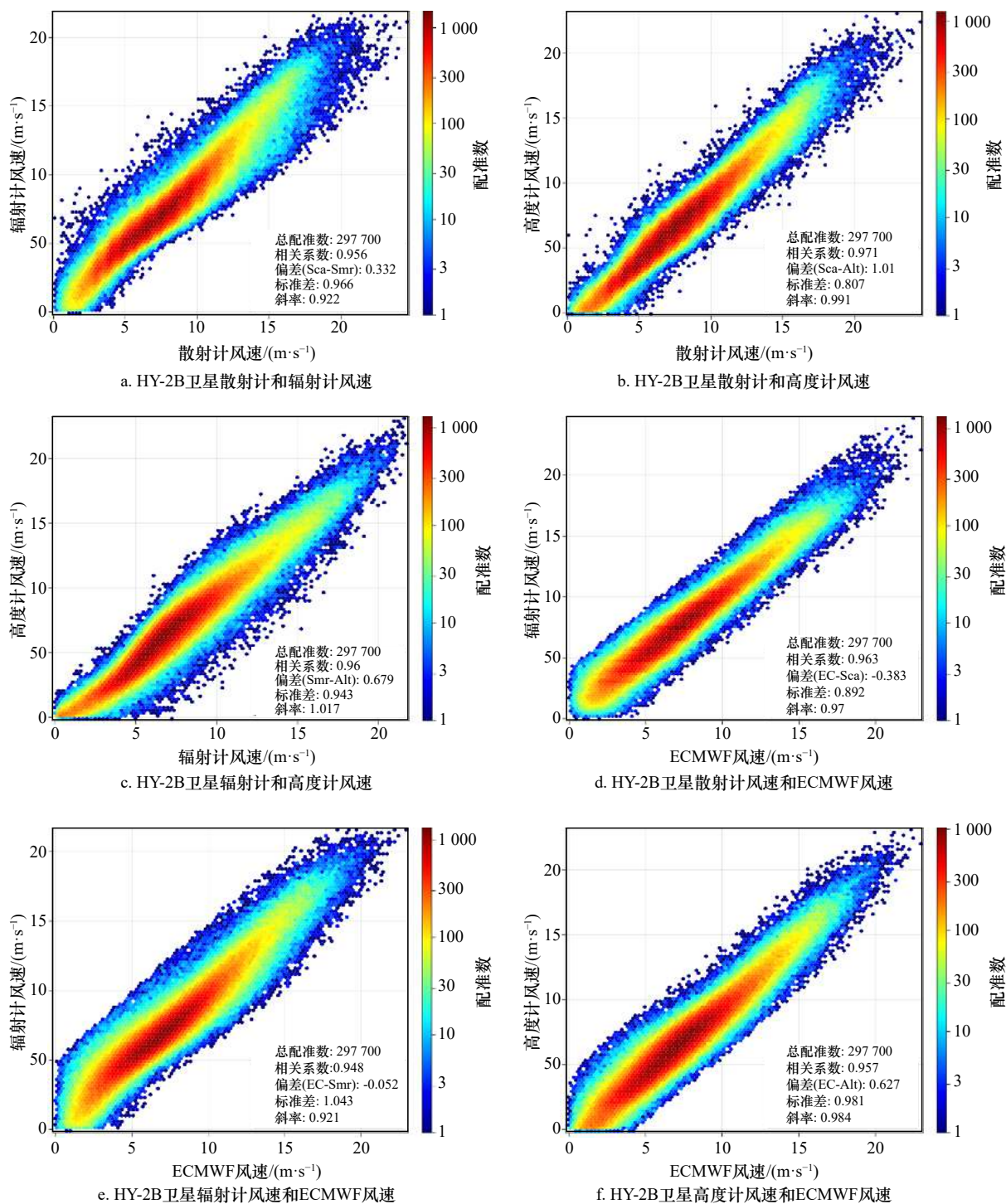


图1 散射计、辐射计、高度计、ECMWF 4个数据集风速比较

Fig. 1 Comparison of wind speeds across four datasets of S/R/A/E

出的EC方法对E、S、R和A这4个数据集进行分析。根据前面对E、S、R、A数据误差相关性的定性分析,认为S、R、A这3个数据集误差相互独立;由于散射计在反演过程中使用ERA5风场作为背景场,所以E误差与S误差存在相关性,其他误差相关情况未知,因此对以下3种情况进行分析:(1)只有E误差与S误差存在相关性;(2)E误差与S误差存在

相关性,同时E误差与R误差也存在相关性;(3)E误差与S误差存在相关性,同时E误差与A误差存在相关性。

本文只给出情况(2)的公式推导,根据式(17)代入ESAR这种排列情况,同时考虑E误差与S误差存在相关性、E误差与R误差存在相关性情况,进一步推导得到式(19):

$$y = \begin{bmatrix} Q_{ee} \\ Q_{ss} \\ Q_{aa} \\ Q_{rr} \\ Q_{es} \\ Q_{er} \\ \frac{Q_{ea}Q_{er}}{Q_{ar}} \\ \frac{Q_{sa}Q_{sr}}{Q_{ar}} \\ \frac{Q_{sa}Q_{sr}}{Q_{ar}} \\ \frac{Q_{er}Q_{ar}}{Q_{ea}} \\ \frac{Q_{sr}Q_{ar}}{Q_{ea}} \\ \frac{Q_{sa}Q_{sr}}{Q_{ar}} \\ \frac{Q_{ea}Q_{sr}}{Q_{ar}} \\ \frac{Q_{ea}Q_{sr}}{Q_{sa}} \end{bmatrix}, A = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, x = \begin{bmatrix} \beta_e^2 \sigma_t^2 \\ \beta_s^2 \sigma_t^2 \\ \beta_a^2 \sigma_t^2 \\ \beta_r^2 \sigma_t^2 \\ \beta_e \beta_s \sigma_t^2 \\ \beta_e \beta_r \sigma_t^2 \\ \sigma_{\varepsilon_e}^2 \\ \sigma_{\varepsilon_s}^2 \\ \sigma_{\varepsilon_a}^2 \\ \sigma_{\varepsilon_r}^2 \\ \sigma_{\varepsilon_e \varepsilon_s} \\ \sigma_{\varepsilon_e \varepsilon_r} \end{bmatrix}. \quad (19)$$

对给定的参数矩阵  $A$ , 用最小二乘法就进行求解:

$$\hat{x} = (A^T A)^{-1} A^T y, \quad (20)$$

从表 1 可以看出在 3 种情况中,  $Cov(\varepsilon_e, \varepsilon_s)$  不等于

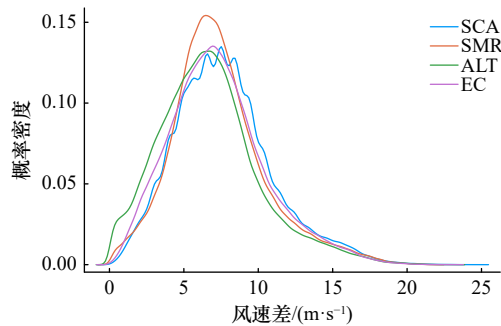


图 2 配准风速数据概率密度分布

Fig. 2 Probability density distribution of collocated wind

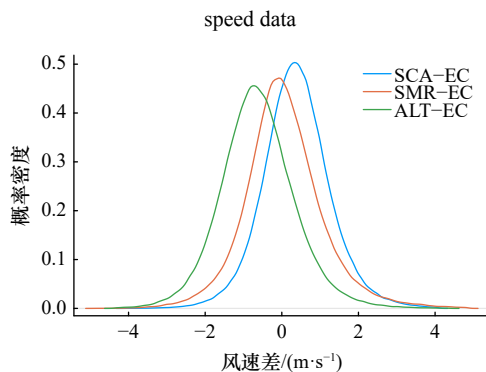


图 3 HY-2B 卫星风速与 ECMWF 风速差的概率密度分布

Fig. 3 Probability density distribution of wind speed bias between HY-2B satellite and ECMWF

0, 可以判断 E 误差与 S 误差存在相关性, 3 种情况下  $Cov(\varepsilon_e, \varepsilon_s)$  大小不同, 这意味着不同数据集误差相关性是相互影响的;  $Cov(\varepsilon_e, \varepsilon_r)$  值很小, 说明 E 误差与 R 误差存在微弱的相关性;  $Cov(\varepsilon_e, \varepsilon_a)$  为负值可以解释为 E 误差与 A 误差相关性不存在, 因为很难想象 E 误差增大会导致 A 误差减小机制存在的可能。与此相类似相关研究也认为 ECMWF 再分析风场误差与 Jason-2 高度计风速误差不存在相关性<sup>[13]</sup>。虽然在第三种情况中  $Cov(\varepsilon_e, \varepsilon_a)$  为负值, EC 方法也给出了 E 误差的标准差的估计与前面两种估计值是一致的, 也说明 EC 方法健壮性。

从表 1 可以看出这 3 种情况计算得到误差的协方差值有较明显不同, 却对误差的标准差估计几乎没

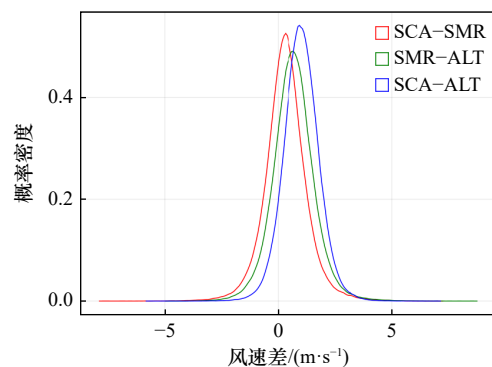


图 4 HY-2B 卫星 3 个风速产品相互差值概率密度分布

Fig. 4 Probability density distribution of wind speed bias between three products of HY-2B satellite

表 1 不同误差相关性假设条件下误差及相关性 (EC)

Table 1 Error and correlation under different assumption (EC)

	$Cov(\varepsilon_e \varepsilon_s) \neq 0$	$Cov(\varepsilon_e \varepsilon_s) \neq 0$ $Cov(\varepsilon_e \varepsilon_r) \neq 0$	$Cov(\varepsilon_e \varepsilon_s) \neq 0$ $Cov(\varepsilon_e \varepsilon_a) \neq 0$
$\sigma_{\varepsilon_e}$	0.769	0.769	0.769
$\sigma_{\varepsilon_s}$	0.600	0.600	0.600
$\sigma_{\varepsilon_r}$	0.721	0.721	0.721
$\sigma_{\varepsilon_a}$	0.565	0.565	0.565
$\sigma_t^2$	10.160	10.160	10.161
$Cov(\varepsilon_e \varepsilon_s)$	0.079	0.113	0.046
$Cov(\varepsilon_e \varepsilon_r)$	0.0	0.063	0.0
$Cov(\varepsilon_e \varepsilon_a)$	0.0	0.0	-0.068

有影响,也许是因为误差相关性较小缘故,因此误差估计结果上可能存在一些偏差。

5.3 基于 ETC 误差及相关性分析

如果有 3 个系统完全满足 TC 前提假设,那么可以利用这 3 个系统作为基准去对第四个系统进行分析,并允许第四个系统与这 3 个系统间存在误差相关性,进一步精确求解误差的标准差和相应的协方差。当我们利用 ETC 对 E、S、R、A 这 4 个数据集进行分析时,根据 4.1 节误差相关性定性分析结果,S、R、A 这 4 个数据集相互独立,这样先利用 ETC 方法求解 S、R、A 这 3 个数据集误差的标准差;然后以这 3 个误差独立的数据集作为基准,进一步对 ECMWF 再分析数据集进行误差及相关性的分析,求解 E 误差以及 E 与 S、R、A 的协方差。

5.3.1 误差相关性不存在情况

当不考虑表征误差和误差相关性时,可以利用 2.2 节介绍的 ETC 方法求解数据集误差的标准差,并且不需要考虑参考系统选择,这是 ETC 方法相对传统 TC 方法一个方便的地方,但是作为被测量物理量的  $\sigma_t^2$  则与参考系统选择有关。另外一方面 ETC 方法跟配对数据集排列顺序无关,因此表 2 每一列给出的是几个字母所代表数据集所有排列情况。

从表 2 可以看出不同组合计算出系统误差的标准差值存在差异,这可能是由于 E 与 S、R、A 某些数据集存在相关性导致。

5.3.2 误差相关性存在情况

从 5.3.1 节分析可以看出利用 EC 方法进行误差相关性分析时,虽然协方差不等于 0,但是对误差的标准差计算结果没有影响。为此本文基于 ETC 方法,利用已知 HY-2B 卫星 3 个载荷观测数据集不存在误差相关性假设和定性分析,通过 ETC 先求解得

表 2 散射计、辐射计、高度计、ECMWF 4 个数据集不同排列情况下的误差 (ETC)

Table 2 Error of different permutation of four datasets of E/S/R/A (ETC)

	S/R/A(SR, RS, SA, AS, RA, AR 6种排列情况)	E/S/A(ES, SE, EA, AE, SA, AS 6种排列情况)	E/R/A(ER, RE, EA, AE, RA, AR 6种排列情况)	E/S/R(ES, SE, ER, RE, SR, RS 6种排列情况)
$\sigma_{\varepsilon_e}$	不计算	0.739	0.769	0.693
$\sigma_{\varepsilon_s}$	0.600	0.495	不计算	0.560
$\sigma_{\varepsilon_r}$	0.742	不计算	0.595	0.770
$\sigma_{\varepsilon_a}$	0.533	0.635	0.699	不计算
$\sigma_t^2$	10.550	10.206	10.753 5	10.271

到 S、R、A 误差,然后对第四个数据集 E 的误差及相关性进行分析和估计。

根据 EC 方法分析结果(表 1),E 误差与 S 误差存在相关性,E 误差与 R 误差存在相关性,E 误差和 A 误差不存在相关性。因此先利用式(10)和式(11)求解 ESA(E 与 S 误差存在相关性)和 ERA(E 与 R 误差存在相关性)两种情况下误差协方差和表征误差,进一步可以求出 E 误差的标准差;然后把 E 误差的标准差当作已知量代入式(13)和式(14)求解 ESR 情况下误差协方差和表征误差。

当利用式(10)和式(11)求解 EAS 和 EAR 两种情况,计算结果 C 为负数,由于不存在 ECMWF 再分析数据集误差增加而导致 HY-2B 卫星高度计观测风速减小机制,因此 E 误差和 A 误差不存在相关性。利用公式(13)和公式(14)对 ERS 情况进行求解,结果与 ESR 情况完全一样,这从公式(12)和公式(14)对称性也可以直接得出。从表 3 可以看出计算得到表征误差 Y 很小,基本可以忽略不计。

表 3 3 个数据集不同排列情况误差的协方差及表征误差

Table 3 Error covariance and representative error of different permutation for three datasets

	ESA	EAS	ERA	EAR	ESR/ERS
$\sigma_{\varepsilon_e}$	0.810	0.659	0.810	0.725	0.810 (known)
C	$Cov(\varepsilon_e \varepsilon_s)$ : 0.113	$Cov(\varepsilon_e \varepsilon_a)$ : -0.115	$Cov(\varepsilon_e \varepsilon_r)$ : 0.063	$Cov(\varepsilon_e \varepsilon_a)$ : -0.068	$Cov(\varepsilon_e \varepsilon_s)$ : 0.113
D	不计算	不计算	不计算	不计算	$Cov(\varepsilon_e \varepsilon_r)$ : 0.063
Y	$3.04 \times 10^7$	$3.04 \times 10^7$	$5.93 \times 10^8$	$5.93 \times 10^8$	$3.61 \times 10^7$

从表 3 可以看出在考虑误差相关性的前提下,在 ESA 和 ERA 排列情况下计算误差协方差结果与在 ESR 情况下计算误差协方差结果是一致的。当同时考虑 E 与 S 误差相关性和 E 与 R 误差相关性情况下,

计算得到 E 误差的标准差也是一致的。最后求解得到 E 误差的标准差, 也就是  $\sigma_e = 0.810$  要比 EC 方法估计结果略大。

### 5.3.3 E、S、R 和 A 误差及协方差估计置信区间

Bootstrap 是由 Efron 和 Tibshirani<sup>[20]</sup> 于 1979 年提出一种非参数统计学习方法, 通过对总体样本进行重采样, 进而对总体估计量的分布特性进行统计推断。由于不同样本对 E、S、R 和 A 误差及协方差估计会有不同, 参照参考文献 [13, 21] 做法利用 Bootstrap 方法对被风速误差的标准差的方差及置信区间进行估计。从 297 700 对配准数据中进行可置换重采样,

每 100 000 次重采样的样本作为一组样本, 总共采集 1 000 组样本数据集, 在每组数据中利用 ETC 方法进行误差的标准差和协方差的计算, 然后对 1 000 组数据集利用 Bootstrap 方法计算误差的标准差及协方差分布及置信区间。

从图 5 和图 6 可以看出, 利用基于 ETC 方法对 4 个数据集进行误差和相关性分析的结果呈大致正态分布, 而且利用 Bootstrap 方法求解的均值与利用总体样本估计结果基本上是一致的, 并且 95% 置信区间相对集中, 因此本文基于 ETC 分析方法估计结果是可靠的。

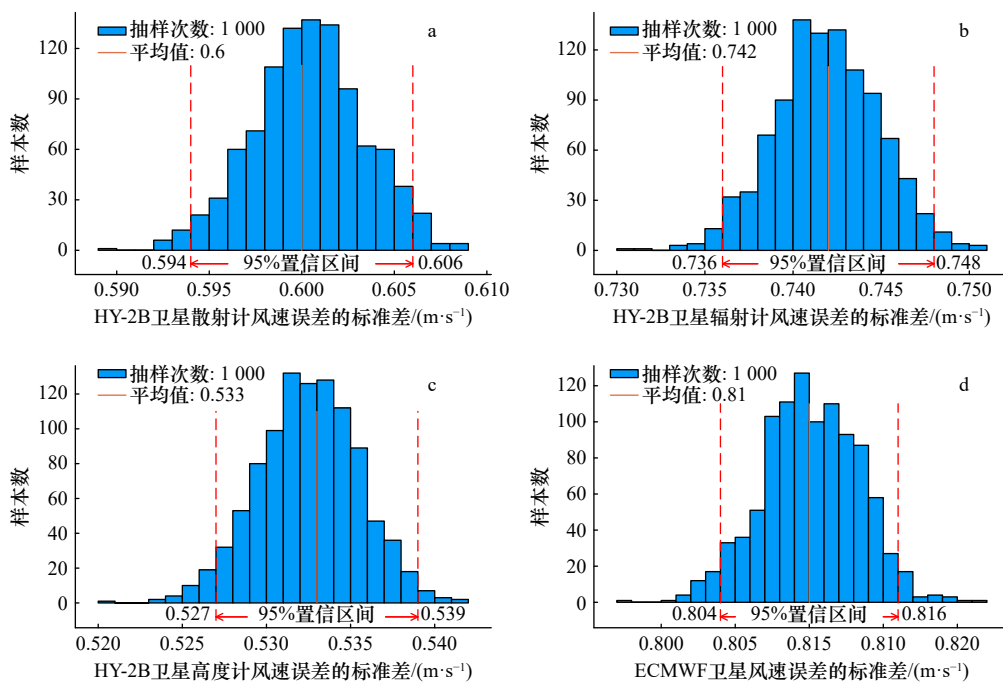


图 5 散射计、辐射计、高度计、ECMWF 4 个数据集误差的标准差分布

Fig. 5 Error SD distributions of four dataset of S/R/A/E

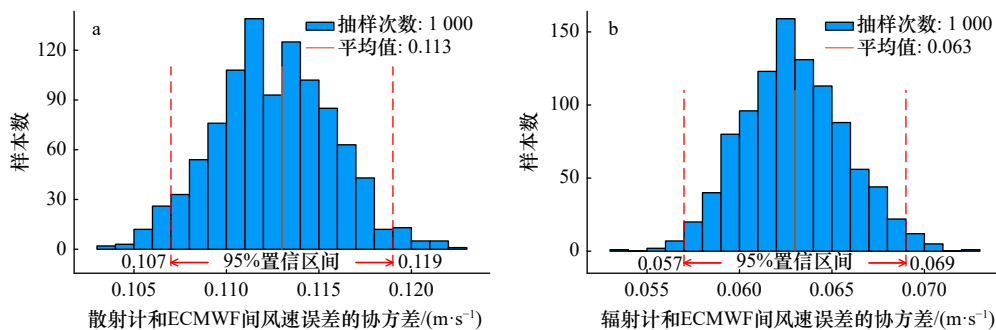


图 6 误差的协方差分布

Fig. 6 Distribution of error covariance

## 6 讨论

本文讨论的数据集随机误差是由观测系统中诸

多随机因素和地球物理函数不完备性导致的, 由于不同数据集分辨率不一样, 高分辨率观测到信号在与低分辨率数据比较时表现为误差, 在三配准方法中通常

表示为表征误差<sup>[2]</sup>。为了减少表征误差影响, 本文将数据统一配准到  $25^\circ \times 25^\circ$  经纬度空间网格, 空间分辨率大约为 50 km, 因此本文讨论随机误差是在空间分辨率为 50 km 情况下所估计的随机误差。

从表 1 结果可以看出 EC 方法在弱误差相关性时, 忽略或者低估了误差相关性影响, 本文提出的利用 3 个误差独立数据集对第四个数据集误差及相关性分析方法, 能够在弱误差相关情况下精确估计误差的标准差。5.4.3 节利用 Bootstrap 方法分析得到 E、S、R 和 A 数据集误差的标准差的 95% 置信区间为  $(0.810 \pm 0.006)$  m/s、 $(0.600 \pm 0.006)$  m/s、 $(0.742 \pm 0.006)$  m/s、 $(0.533 \pm 0.006)$  m/s; 从图 6 可以看出 E 数据集误差与 S 数据集误差协方差的 95% 置信区间为  $(0.113 \pm 0.006)$  m/s, E 数据集误差与 R 数据集误差协方差的 95% 置信区间为  $(0.063 \pm 0.006)$  m/s。

EC 方法利用最小二乘法直接对 4 个数据集进行求解, 但是并不是所有误差相关性组合都可以求解, 我们根据 E 误差与 S 误差存在相关性先验知识对 3 种情况进行估计: (1)  $Cov(\varepsilon_e, \varepsilon_s) \neq 0$ ; (2)  $Cov(\varepsilon_e, \varepsilon_r) \neq 0$ ,  $Cov(\varepsilon_e, \varepsilon_a) \neq 0$ ; (3)  $Cov(\varepsilon_e, \varepsilon_s) \neq 0$ ,  $Cov(\varepsilon_e, \varepsilon_a) \neq 0$ , 虽然 3 种情况下求解得到误差的协方差大小存在差异, 但是计算得到误差的标准差不存在明显差异, 也就是较小误差的协方差对误差的标准差估计没有影响, 因此利用 EC 方法对误差的标准差估计也存在偏差。在第三种情况中计算得到  $(Cov(\varepsilon_e, \varepsilon_a) = -0.068) < 0$ , 也就是 E 误差和 A 误差存在微小负相关, 意味着 E 误差增大会导致 A 误差的减小; 然而 E 和 A 的误差是随机误差, 而在现实测量中随机误差相互叠加, 不会因为一个系统观测的随机误差增大而导致另外一个系统随机误差的减小, 因此随机误差负相关机制不存在, 所以可以判断 E 误差和 A 误差相关性不存在。在第二种情况中, 计算得到  $(Cov(\varepsilon_e, \varepsilon_r) = 0.063) > 0$ , 这意味着 E 误差和 R 误差之间存在微弱的相关。

从表 1 结果可以看出 EC 方法计算结果能够定性和定量确定误差相关性, 但是没能够反映误差弱相关性对误差的标准差影响。我们利用改进 ETC 方法, 考虑误差相关性和表征误差影响, 对误差的标准差和误差协方差进行求解。求解得到 E 误差与 S 误差之间协方差为 0.113, E 误差与 R 误差之间协方差为 0.063, 结果与 EC 方法计算结果一致(表 1 第二列)。求解得到 E 误差的标准差为 0.810, 要比 EC 方法计算结果略大, 这是由于此方法能够反映误差弱相关性对误差的标准差估计的影响。2017 年 Abdalla 和 De Chiara<sup>[13]</sup> 在利用 Jason-2、ASCAT-A/B 和 ECMWF 的模式数据进

行风速误差分析得到再分析数据 ERA-Interim 风速误差的标准差约为 0.9 m/s, 而本文利用 ETC 方法分析得到 ERA5 再分析数据误差为 0.810 m/s, 这是由于 ERA5 风速的标准差比 ERA-Interim 有大幅度提升。这也说明如果有 3 个完全满足 TC 假设观测系统, 可以反过来对模式再分析数据进行精度评估。吕思睿等<sup>[16]</sup> 在三配准方法中将一个系统误差与参考系统误差比值称为固有误差, 并计算得到 HY-2B 卫星散射计固有误差为 0.73, 根据本文前面分析结果可知, HY-2B 卫星散射计固有误差为  $0.6/0.81 = 0.74$ , 这说明利用三配准方法进行误差估计有很好的的一致性。

从表 2 结果可以看出即使微小的误差协方差对误差方差也存在影响, 因此在分析中考虑这种影响能够得到更为准确结果, 换句话说, 误差相关性将影响到误差的标准差估计的准确性。利用表 3 计算得到总体样本协方差结果, 进一步计算得到 E 与 S 误差之间相关系数为 0.231, E 与 R 误差之间相关系数为 0.105, 相应的协方差都落在利用 Bootstrap 方法估计的 95% 置信区间内(图 6), 因此我们对协方差的估计结果是可靠的。相关系数结果显示 E 误差和 S 误差之间是弱相关, E 误差和 R 误差之间相关性则更小。

从图 1 可以看出辐射计风速相对于 ERA5 风速呈现出一定非线性, 这可能由于反演算法和模型不完善导致在低风速区域辐射计风速估计较 ERA5 风速偏大, 这种非线性可能会导致误差相关性估计出现偏差, 进一步定量分析需要进行非线性建模, 可以作为未来一个研究方向。

本文基于 ETC 方法针对不同系统随机误差存在相关性情况下进行误差分析, 对系统误差的标准差进行精确估计, 求解结果可以作为不同数据集融合权重, 精度越高权重越大; 也可以作为数据同化阈值, 因为如果被同化数据波动太大将导致同化系统较大的扰动。

## 7 结论

在利用基于 TC 方法进行误差分析时, 误差相关性对误差的标准差估计结果有重要影响。本文提出利用 3 个误差相互独立数据集对第四个数据集进行误差及相关性分析的方法, 能够在误差弱相关情况下进行准确求解。利用该方法, 本文对 HY-2B 卫星散射计、辐射计和高度计 3 个载荷和 ERA5 再分析风场数据风速观测数据进行分析, 精确求解得到误差的标准差分别为 0.600 m/s、0.742 m/s、0.533 m/s 和 0.810 m/s, 利用 Bootstrap 方法进行估计得到 95% 置信区间都围绕均值  $\pm 0.006$  m/s 范围内变化。本文提出的方法可以

用于模式数据误差的标准差精确地估计,也有助于更好地将这些数据用于同化和融合。

第一,本文提出方法假定有3个数据集完全满足TC假设前提条件,而这只能通过专家知识根据经验进行估计,具有一定主观性。第二,本文没有就数据质量对误差精度影响进行分析。第三,由于HY-2B卫星辐射计常规风速产品最高风速只有23 m/s,分析数据中高风速样本都作为异常值被剔除掉,因此风

速大于23 m/s样本未包含到分析中。第四,本文分析得到HY-2B卫星辐射计观测风速误差与ECMF再分析数据的风速误差存在微弱相关性,需要进一步探讨其相关性来源。散射计沿垂直轨道方向观测几何差异较大,对误差会有较大影响,海表温度也对风速反演有影响,后续可以继续开展误差的标准差的空间分布和季节变化的研究,这将有助于更好地同化和利用这些数据。

## 参考文献:

- [1] Martin S. 海洋遥感导论(修改版)[M]. 蒋兴伟,译. 北京:海洋出版社,2017.  
Martin S. An Introduction to Ocean Remote Sensing[M]. Jiang Xingwei, trans. Beijing: China Ocean Press, 2017.
- [2] Stoffelen A. Toward the true near-surface wind speed: error modeling and calibration using triple collocation[J]. *Journal of Geophysical Research: Oceans*, 1998, 103(C4): 7755–7766.
- [3] Scipal K, Holmes T, De Jeu R, et al. A possible solution for the problem of estimating the error structure of global soil moisture data sets[J]. *Geophysical Research Letters*, 2008, 35(24): L24403.
- [4] Vogelzang J, Stoffelen A, Verhoef A, et al. On the quality of high-resolution scatterometer winds[J]. *Journal of Geophysical Research: Oceans*, 2011, 116(C10): C10033.
- [5] Vogelzang J, King G P, Stoffelen A. Spatial variances of wind fields and their relation to second-order structure functions and spectra[J]. *Journal of Geophysical Research: Oceans*, 2015, 120(2): 1048–1064.
- [6] Zwieback S, Scipal K, Dorigo W, et al. Structural and statistical properties of the collocation technique for error characterization[J]. *Non-linear Processes in Geophysics*, 2012, 19(1): 69–80.
- [7] Pierdicca N, Fascetti F, Pulvirenti L, et al. Quadruple collocation analysis for soil moisture product assessment[J]. *IEEE Geoscience and Remote Sensing Letters*, 2015, 12(8): 1595–1599.
- [8] Pan Ming, Fisher C K, Chaney N W, et al. Triple collocation: beyond three estimates and separation of structural/non-structural errors[J]. *Remote Sensing of Environment*, 2015, 171: 299–310.
- [9] Gruber A, Su C H, Crow W T, et al. Estimating error cross-correlations in soil moisture data sets using extended collocation analysis[J]. *Journal of Geophysical Research: Atmospheres*, 2016, 121(3): 1208–1219.
- [10] McColl K A, Vogelzang J, Konings A G, et al. Extended triple collocation: Estimating errors and correlation coefficients with respect to an unknown target[J]. *Geophysical Research Letters*, 2014, 41(17): 6229–6236.
- [11] Konings A G, Mccoll K A, Alemohammad S H, et al. Error characterization of similar products: Triple collocation with correlated errors[C]//AGU Fall Meeting Abstracts. Washington: AGU, 2014.
- [12] Lin Wenming, Portabella M, Stoffelen A, et al. ASCAT wind quality under high subcell wind variability conditions[J]. *Journal of Geophysical Research: Oceans*, 2015, 120(8): 5804–5819.
- [13] Abdalla S, De Chiara G. Estimating random errors of scatterometer, altimeter, and model wind speed data[J]. *IEEE Journal of Selected Topics in Applied Earth Observations and Remote Sensing*, 2017, 10(5): 2406–2414.
- [14] Vogelzang J, Stoffelen A. Quadruple collocation analysis of *in-situ*, scatterometer, and NWP winds[R/OL]. (2023–01–18) [2023–02–02]. *Oceanography*, 2021. <https://essopenarchive.org/doi/full/10.1002/essoar.10505872.1>.
- [15] Vogelzang J, Stoffelen A. On the accuracy and consistency of quintuple collocation analysis of in situ, scatterometer, and NWP Winds[J]. *Remote Sensing*, 2022, 14(18): 4552.
- [16] 吕思睿, 林文明, 邹巨洪, 等. 多源卫星遥感海面风速误差分析和交叉标定 [J]. *海洋学报*, 2023, 45(5): 118–128  
Lü Sirui, Lin Wenming, Zou Juhong, et al. Error quantification and cross calibration of sea surface wind speeds from multiple remote sensing satellites[J]. *Haiyang Xuebao*, 2023, 45(5): 118–128.
- [17] 林明森, 何贤强, 贾永君, 等. 中国海洋卫星遥感技术进展 [J]. *海洋学报*, 2019, 41(10): 99–112,  
Lin Mingsen, He Xianqiang, Jia Yongjun, et al. Advances in marine satellite remote sensing technology in China[J]. *Haiyang Xuebao*, 2019, 41(10): 99–112.
- [18] Gourrion J, Vandemark D, Bailey S, et al. A two-parameter wind speed algorithm for Ku-band altimeters[J]. *Journal of Atmospheric and Oceanic Technology*, 2002, 19(12): 2030–2048.
- [19] Molina M O, Gutiérrez C, Sánchez E. Comparison of ERA5 surface wind speed climatologies over Europe with observations from the HadISD dataset[J]. *International Journal of Climatology*, 2021, 41(10): 4864–4878.
- [20] Efron B, Tibshirani R J. An Introduction to the Bootstrap[M]. New York: Chapman and Hall/CRC, 1994.
- [21] Caires S, Sterl A. Validation of ocean wind and wave data using triple collocation[J]. *Journal of Geophysical Research: Oceans*, 2003, 108(C3): 3098.

## Analysis for random error and correlation of HY-2B satellite and model wind speed data

Lan Youguo<sup>1,2</sup>, Lin Mingsen<sup>2</sup>, Zhang Youguang<sup>2</sup>, Yang Dian<sup>2</sup>

(1. College of Information Science Engineering, Ocean University of China, Qingdao 266100, China; 2. National Satellite Ocean Application Service, Beijing 100081, China)

**Abstract:** Random errors between systems are not correlated is a necessary assumption for Triple Collocation (TC) analysis, but this assumption does not always hold in practice. The least squares-based Extended Collocation (EC) method can estimate random error in the presence of error correlation, but it cannot accurately estimate standard deviation (SD) of the random error as error correlation is weak. This paper proposes an error estimation method for the fourth system using three error-independent systems, which can estimate the SD of the system error more accurately in case of weak correlation by considering both error correlation and representative error. The SD of the errors of the scatterometer, radiometer and altimeter are 0.600 m/s, 0.742 m/s and 0.533 m/s respectively, as assumed that random errors of three HY-2B wind speed products are independent. The SD of error of ERA5 reanalysis wind speed is also estimated to be 0.810 m/s, the correlation coefficient of the errors of wind speed between HY-2B scatterometer and the ERA5 is 0.231, the correlation coefficient of the errors of wind speed between HY-2B radiometer and the ERA5 is 0.105. This paper proposes a method to estimate random errors and their correlation with the fourth dataset using three known error independent datasets, which achieves a more precise estimation for the SD of the random error in the case of weak correlation, and it helps to use these data better in assimilation and fusion.

**Key words:** HY-2B satellite; ocean surface wind speed; extended triple collocation; random error; error correlation