

球面质心泰森多边形网格海浪预报的向量化检验方法

谢明炎¹, 魏立新^{1,2}, 高元勇^{1,2}, 高志一^{1,2*}

(1. 国家海洋环境预报中心, 北京 100081; 2. 国家海洋环境预报中心 自然资源部海洋灾害预报技术重点实验室, 北京 100081)

摘 要: 提出一种适于并行计算设备的向量化检验方法。采用特征半径近似判别法选取搜索域内观测点平均值作为实测值, 基于广播、索引等向量化方法及海浪平稳性条件实现模式和观测资料时间和空间的匹配, 并将待检区域等分为若干块分别计算以免内存溢出。本算法可在中央处理器(CPU)和图形处理器(GPU)上运行, 也适用于其它类型的模式网格。为评估算法实际效果, 以有限体积海浪模式产生的球面质心泰森多边形网格预报产品和卫星沿轨道遥感产品分别作为待检和参考数据, 结果表明 CPU 和 GPU 的计算结果一致; GPU 的计算效率较 CPU 提高 30 倍; 区块大小对效率提升有一定影响, 需根据实际算例调试优化。

关键词: 精度检验; GPU 加速; 向量化; 高性能检验

中图分类号: P731.33 **文献标识码:** A **文章编号:** 1003-0239(2024)02-0034-10

0 引言

精度检验是评估海浪数值模式预报技能和精度性能的重要步骤, 包括大面检验和单点检验等形式, 检验要素包括有效波高 (Significant Wave Height, SWH) 和平均波周期 (Mean Wave Period, MWP) 等。大面检验是以卫星遥感资料为依据, 对洋面或海域数值模拟结果进行的全场检验, 常用于横向比较不同机构的数值预报技术水平。

以往由于观测和遥感数据量有限, 大面检验算法的效率并未引起人们关注。学者们通常将沿轨道数据进行时空双线性插值, 生成新的规则网格数据集与模式资料进行匹配^[1-3]; 或从卫星轨迹重合的预报模型中提取数据并进行一定时空区间内平滑, 再逐对判断观测点和模式网格之间的相对位置关系^[2,4]。然而, 随着观测和模拟数据量的爆发性增长, 以及各种复杂非结构网格的广泛应用, 传统的匹配算法较为耗时, 提升大面检验的计算效率也成为业务上关心的问题。例如, DODET 等^[5]发现 1 000 次匹配可以保证观测和模式数据的检验误差水平

趋于稳定, 通过稀疏匹配数量来降低计算量。目前, 国内外业务化海洋动力卫星数量增长迅速, 包括 Jason 系列 (2/3)、ERS/Envisat^[6]、海洋二号 (HY-2) 系列 (B/C/D)^[7]、中法海洋卫星 (The China France Oceanography SATellite, CFOSAT)^[8] 等; 我国也于 2018—2021 年先后发射了 6 颗卫星, 观测数据量增长 4 倍。国际上主要预报机构的全球海浪数值预报系统的分辨率普遍从 25 km 提高到 5~10 km, 数值模拟数据量增加超过 6 倍。此外, 非结构网格在全球海浪数值预报系统中盛行, 特别是三角形网格^[9-10]和球面质心泰森多边形 (Spherical Centroidal Voronoi Tessellations, SCVT)^[11] 的出现, 进一步增加了检验计算的复杂度。粗略估计大面检验计算量较 5 年前增长了 24 倍, 预计未来还将继续快速增长。因此, 研制高性能大面检验算法才能跟上高性能数值预报和海洋遥感领域的发展。

本文提出一种适用于并行计算设备的 SCVT 网格向量化检验方法。该方法将高性能数据分析 (High-Performance Data Analysis, HPDA) 的核心思想融入, 基于广播、索引、数组合并等向量化操作构

收稿日期: 2023-10-05。

基金项目: 国家重点研发计划 (2023YFC3107800)。

作者简介: 谢明炎 (1999-), 女, 硕士在读, 主要从事海浪数值和智能模拟研究。E-mail: x_mingyan@126.com

*通信作者: 高志一 (1980-), 男, 高级工程师, 博士, 主要从事海浪理论与小尺度海气相互作用研究。E-mail: gaozy@nmefc.cn

建二维距离矩阵,可实现搜索域内 SCVT 网格的空间匹配并剔除近岸数据。在实际操作中,可将待检区域进行分块以免存储溢出。新方法可在中央处理器(Central Processing Unit, CPU)和图形处理器(Graphics Processing Unit, GPU)上运行,适用于不同类型卫星数据的大面检验计算,并可迁移到六边形 SCVT 网格外的其他网格预报资料中,如三角形网格、四边形网格等。为了评估该检验算法的实用效果,采用海洋二号 B 卫星(HY-2B)和 CFOSAT 卫星沿轨道数据对西太平洋有限体积海浪模式(Finite Volume Wave Model, FVWAM)产生的 SCVT 网格预报产品检验结果进行横向比较。

1 数据和方法

1.1 数值模拟数据

有限体积海浪模式 FVWAM 是国家海洋环境预报中心最新研制的第三代海浪谱数值模式,其突出特征是在动力框架中首次采用了 SCVT 计算网格和有限体积平流算法。该数值产品存储格式为 NetCDF,包含每日逐时的数值模拟海浪要素场,如模拟时间、网格质心坐标、有效波高、平均周期等参数。除产品文件之外,模式附带一个 SCVT 网格信息文件,包括网格质心坐标、网格质心索引、网格顶点坐标、网格顶点索引、同单元格(某个六边形)顶点索引、海陆分布掩模等。2021 年全年 FVWAM 模拟数据文件为 365 个,每个模拟数据文件包含 24 个逐时(00~23 时,世界时,下同)的海浪要素模拟场。本文选取 2021 年 $100^{\circ}\sim 150^{\circ}\text{E}$, $0^{\circ}\sim 50^{\circ}\text{N}$ 的预报产品进行检验,空间分辨率为 10 km。

1.2 卫星遥感数据

卫星遥感资料包括 CFOSAT 和 HY-2B 的沿轨道数据,时间范围均为 2021—2022 年。CFOSAT 轨道高度为 521 km,空间分辨率为 25 km,原始资料来自 Aviso+ 网站提供的波谱仪(Surface Wave Investigation and Monitoring, SWIM)二级产品(Geophysical Data Record, GDR),包括有效波高数据和采用人工智能算法由箱限谱反演得到的平均周期资料^[12],文件数为 6 476 个。HY-2B 数据由国家海洋环境预报中心通过国家卫星海洋应用中心专线获

取,包括散射计和高度计二级产品(Interim Geophysical Data Record, IGDR)^[7],轨道高度为 973 km,倾角 99.34° ,空间分辨率为 25 km,文件数为 12 500 个(含重复推送文件),每个文件包含 1 条沿轨道有效波高遥感记录。

1.3 检验方法

本文针对大面检验中遇到的复杂网格时空匹配和数据量激增的问题,提出了一种沿轨道遥感资料对 SCVT 非结构网格预报数据进行精度检验的向量化新方法。向量化算法流程见图 1,其中 Tr、Gr 和 CL 分别指由星下点、SCVT 网格和海岸线信息构成的经纬度距离矩阵。检验步骤如下:首先进行数据质量控制,去除邻近陆地和岛屿待匹配数据;其次,将距离 SCVT 网格质心在特征半径内的卫星观测结果进行平均并作为该网格的实测值,基于向量化运算编写观测点—SCVT 质心的距离矩阵进行观测和模拟数据的空间匹配,并依据海浪平稳性条件将卫星数据对齐至整点和半点实现时间匹配;最后计算

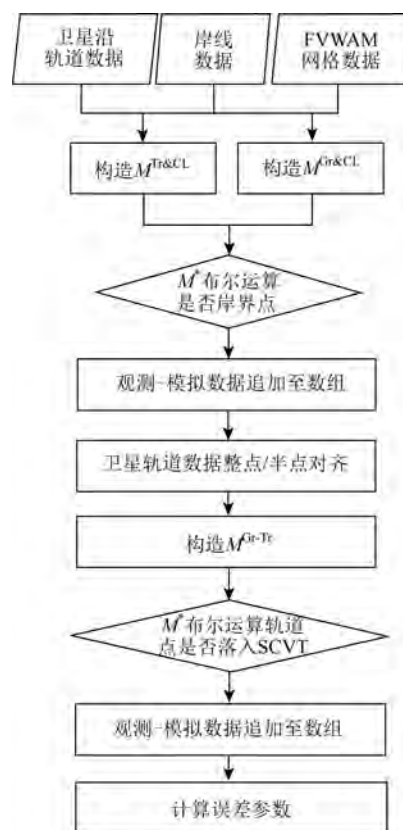


图1 向量化算法流程图

Fig.1 Vectorized algorithm flowchart

误差参数并绘图。

1.3.1 遥感-模拟数据时空匹配

为实现遥感-模拟数据的时空匹配,需将卫星轨道数据插值到SCVT网格上,然后根据输出时间间隔重整已插值的观测数据。时间匹配精确的做法是将整点的数值模拟结果插值到遥感时间点上,但计算代价较大。为简化匹配计算量,选取每小时的整点和半点数据作为时间检验点进行匹配。海

浪平稳性条件认为15~30 min内海浪场的统计特性不发生变化^[13],因此,可认为时间检验点前后15 min内的数值模拟和卫星遥感数据刻画了该时段的海浪状态,这样可以有效地节省时间插值成本。实际计算中,先对FVWAM输出的SCVT网格资料进行时间检验点补全,即将分辨率为1 h的模式数据加密为0.5 h,以整点(或半点)前后15 min内的遥感数据视为该时刻的实测结果(见图2)。

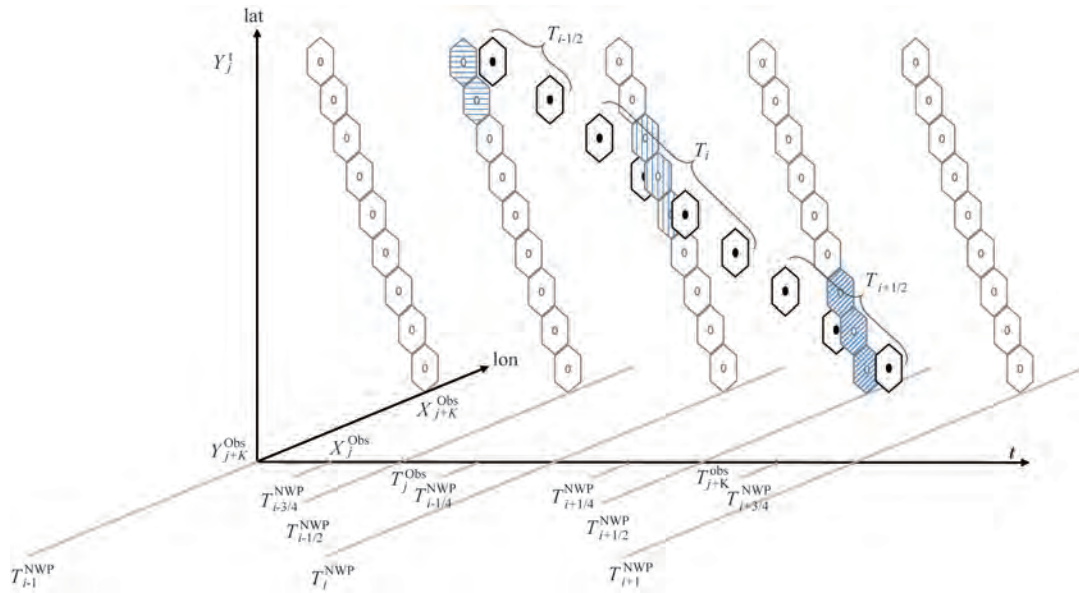
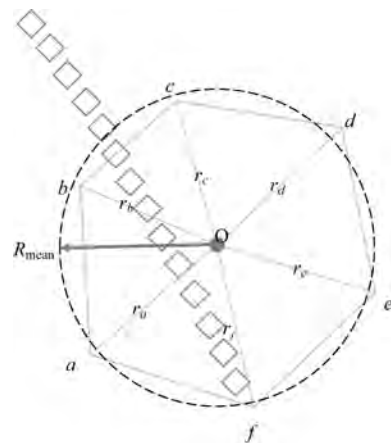


图2 SCVT网格数据时间序列插值补全

Fig.2 SCVT grid data time series interpolation completion

接下来对同一时间分组的遥感和模式数据进行空间匹配,即找到落入SCVT网格的卫星星下点作为数值计算格点的相邻观测点。尽管对于点与多边形网格的空间匹配问题已有广泛的研究,包括面积和判别法、夹角和判别法和射线法等^[14-15],但考虑到这些算法复杂且不利于提速增效^[16],本文基于SCVT网格的物理概念和网格参数,采用了高效的“特征半径近似判别法”实现遥感和数值模拟数据空间匹配。在此方法中,SCVT网格的范围可由以网格质心为圆心且面积与网格面积相同的圆形来近似表示,通过比较星下点*i*与SCVT网格*j*质心的距离以及网格自身特征半径的大小判断卫星点是否落入网格内(见图3)。六边形*a~f*为任一SCVT网格,穿过网格的四边形列为网格附近卫星星下点,特征半径 R_{mean} 取与网格单元面积 $S_j^0 = \sum_{n=1}^6 A_n^0$ 相



注:实六边形为任一SCVT网格,虚线圆为与SCVT网格面积相等的等效圆,四边形列代表SCVT网格附近卫星星下点

图3 “特征半径近似判别法”示意图

Fig.3 Schematic of the "feature radius approximation discrimination method"

等的圆的半径。计算公式为:

$$R_{\text{mean},j} = \sqrt{\frac{S_j^0}{\pi}} = \sqrt{\frac{\sum_{i=1}^6 A_n^0}{\pi}} \quad (1)$$

计算星下点 i 与 SCVT 网格 j 质心的距离,可先通过矩阵广播运算将 I 个 SCVT 网格坐标 $(lon_i^{\text{Gr}}, lat_i^{\text{Gr}})$ 和 J 个卫星轨道星下点坐标 $(lon_j^{\text{Tr}}, lat_j^{\text{Tr}})$ 分别整理为矩阵:

$$lon^{\text{Gr}} = \begin{bmatrix} lon_{1,1}^{\text{Gr}} & \cdots & lon_{1,I}^{\text{Gr}} \\ \vdots & \ddots & \vdots \\ lon_{J,1}^{\text{Gr}} & \cdots & lon_{J,I}^{\text{Gr}} \end{bmatrix} \quad (2)$$

$$lat^{\text{Gr}} = \begin{bmatrix} lat_{1,1}^{\text{Gr}} & \cdots & lat_{1,I}^{\text{Gr}} \\ \vdots & \ddots & \vdots \\ lat_{J,1}^{\text{Gr}} & \cdots & lat_{J,I}^{\text{Gr}} \end{bmatrix} \quad (3)$$

$$lon^{\text{Tr}} = \begin{bmatrix} lon_{1,1}^{\text{Tr}} & \cdots & lon_{1,I}^{\text{Tr}} \\ \vdots & \ddots & \vdots \\ lon_{J,1}^{\text{Tr}} & \cdots & lon_{J,I}^{\text{Tr}} \end{bmatrix} \quad (4)$$

$$lat^{\text{Tr}} = \begin{bmatrix} lat_{1,1}^{\text{Tr}} & \cdots & lat_{1,I}^{\text{Tr}} \\ \vdots & \ddots & \vdots \\ lat_{J,1}^{\text{Tr}} & \cdots & lat_{J,I}^{\text{Tr}} \end{bmatrix} \quad (5)$$

对上述坐标进行矩阵运算得到表示逐对的 SCVT 网格格点和卫星轨道星下点之间球面距离的矩阵 D^{GrTr} :

$$D^{\text{GrTr}} = \begin{bmatrix} d_{1,1} & \cdots & d_{1,I} \\ \vdots & \ddots & \vdots \\ d_{J,1} & \cdots & d_{J,I} \end{bmatrix} \quad (6)$$

式中: d_{ji} 为 SCVT 网格格点 i 和卫星轨道星下点 j 之间的球面距离,计算公式为:

$$d_{ji} = R_{\text{earth}} \arccos [\cos(lat_j^{\text{Tr}}) \cos(lat_i^{\text{Gr}}) \cos(lat_j^{\text{Tr}} - lat_i^{\text{Gr}}) + \sin(lat_j^{\text{Tr}}) \sin(lat_i^{\text{Gr}})] \quad (7)$$

式中: R_{earth} 表示地球半径。

通过广播运算将 I 个等效圆半径 r_i 向量扩展为判据矩阵 r^{Th} :

$$r^{\text{Th}} = \begin{bmatrix} r_{1,1} & \cdots & r_{1,I} \\ \vdots & \ddots & \vdots \\ r_{J,1} & \cdots & r_{J,I} \end{bmatrix} \quad (8)$$

利用矩阵元素的布尔运算:

$$D^{\text{GrTr}} < r^{\text{Th}} \quad (9)$$

筛选出位于 SCVT 网格内部的星下点并记录其索引,通过索引运算将这些点保留至样本集完成空间匹配。落入同一 SCVT 网格的卫星星下点数据被认为是同步观测所得,可取平均值减小噪声误差。由于模式和卫星数据量巨大,计算球面距离的矩阵 D^{GrTr} 容易超出 GPU 显存的单次容量限制,因此在实际检验中,为避免显存溢出,可将地理区域划分为若干小块逐片完成。如果计算机存在多个 GPU 可用,可考虑将任务分配到多个 GPU 上并行计算,进一步加速处理过程。

该方法具备通用性,适合于多种不同类型的网格。针对各网格自身的特点,搜索域的形状和范围也会发生变化。对于 SCVT 六边形网格,物理量的计算位置在其质心上(见图 4a),搜索域可直接选取为自身网格大小,然后将落在网格里的所有星下点的平均值作为该点匹配值。对于三角形和四边形网格,物理量的计算位置位于其顶点上,此时用于质心网格的方法不再适用,需要根据自身尺寸人为选取适当的搜索域进行匹配运算。一种方法是选取与网格面积相等的矩形搜索域(见图 4b),另一种方法是以网格边长为半径构建一个更大的圆形搜索域(见图 4c)。

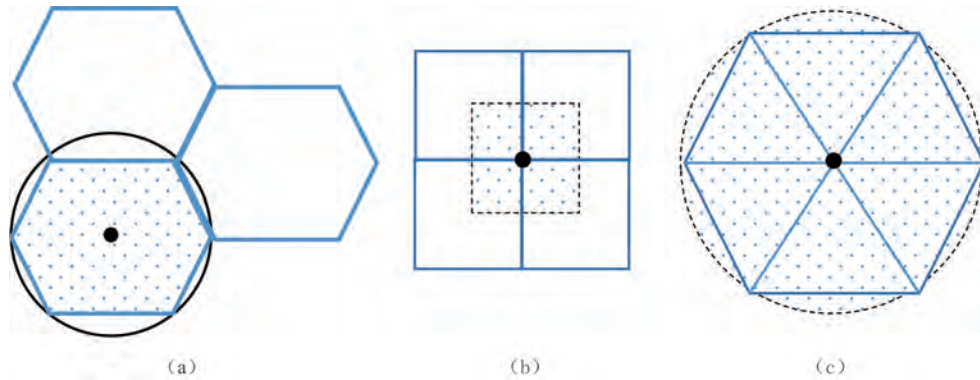


图4 SCVT 网格(a)、矩形网格(b)和三角形网格搜索域(c)

Fig.4 SCVT grid (a), rectangular grid (b), and triangular grid search domains (c)

1.3.2 数据质量控制方法

遥感模拟资料匹配之前,需组织和整理数据以确保计算的准确性。首先,将辅助信息(如波高、掩膜、水深等)转换为一维格式,按“时间-经纬度-物理量”的方式排列;其次,设置边界条件和阈值筛选出目标区域内 10 m 水深以下的海洋点,并根据 3-sigma 质量控制法删除超出范围的异观测值;最后,制作一维海岸线数据集,计算卫星星下点或 SCVT 网格格点与海岸线的最小距离来剔除内陆、边界和近岸点。

本文设计了一种适合并行计算的向量化算法来优化星下点与海岸线间的距离计算问题(见图 5)。该方法通过矩阵广播运算将 I 个 SCVT 网格或星下点格点坐标 $(lon_i^{GrTr}, lat_i^{GrTr})$ 和 K 个海岸线坐标 (lon_k^{CL}, lat_k^{CL}) 分别整理为适合并行计算的球面距离的矩阵 D^{GrTrCL} 。公式为:

$$D^{GrTrCL} = \begin{bmatrix} d_{1,1} & \cdots & d_{1,I} \\ \vdots & \ddots & \vdots \\ d_{K,1} & \cdots & d_{K,I} \end{bmatrix} \quad (10)$$

式中: $d_{k,i}$ 为 SCVT 网格格点 i 和等海岸线格点 k 之间的球面距离。 $d_{k,i}$ 的计算方法为:

$$d_{k,i} = R_{\text{earth}} \arccos [\cos(lat_k^{CL}) \cos(lat_i^{GrTr}) \cos(lat_k^{CL} - lat_i^{GrTr}) + \sin(lat_k^{CL}) \sin(lat_i^{GrTr})] \quad (11)$$

式中: R_{earth} 表示地球半径。

一般认为在近岸 20~50 km 内模式和观测数据质量下降严重^[17]。本检验选择距离 50 km 为过滤范围,利用矩阵元素的布尔运算剔除出近岸点:

$$D^{GrCL} < 50 \text{ km}, D^{TrCL} < 50 \text{ km} \quad (12)$$

若球面距离矩阵超过 GPU 显存单次容纳大小,同样可切分出若干小片区域进行预处理计算,分片大小因实际数据量和目标区域而定。

1.3.3 精度评价指标

对于时间和空间匹配的 SCVT 网格预报数据和 HY-2B 高度计数据,采用相关系数(R)、偏差(Bias)、均方根误差(Root Mean Square Error, RMSE)、平均绝对误差(Mean Absolute Error, MAE)和散射系数(Scattering Index, SI) 5 个统计指标进行比较,结果以散点图、Taylor 图和统计结果表格进行呈现。指标定义如下:

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2 \sum_{i=1}^n (y_i - \bar{y})^2}} \quad (13)$$

$$\text{Bias} = \bar{y} - \bar{x} \quad (14)$$

$$\text{RMSE} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})} \quad (15)$$

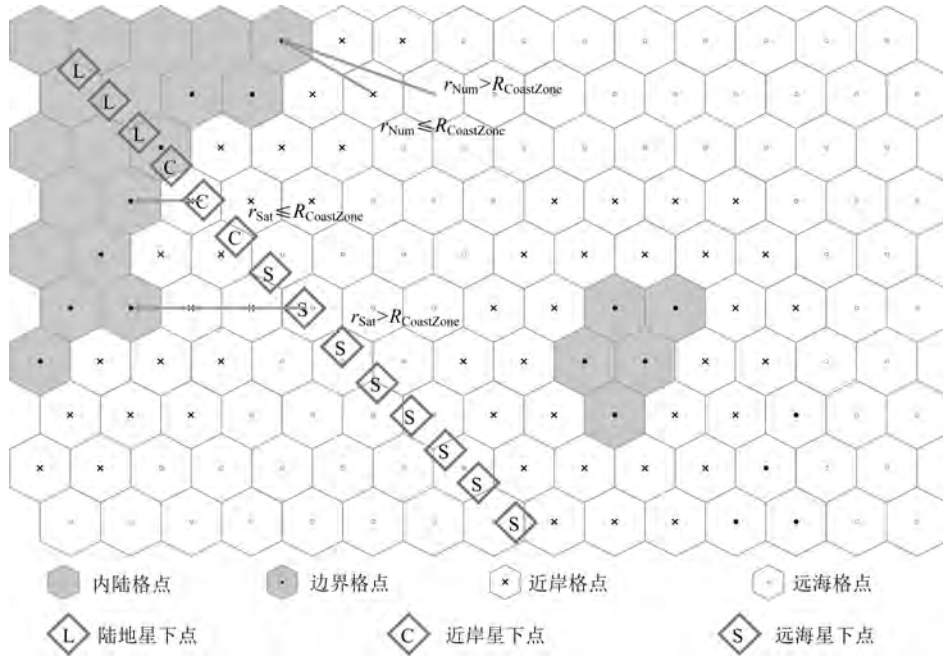


图5 近岸数据过滤示意图

Fig.5 Schematic of nearshore data filtering

$$MAE = \frac{\sum_{i=1}^n |y_i - x_i|}{n} \quad (16)$$

$$SI = \frac{RMSE}{\bar{x}} \quad (17)$$

式中: x_i 代表遥感观测数据; y_i 代表SCVT模拟数据; \bar{x} 和 \bar{y} 是分别是观测数据和模拟数据的平均值; n 为匹配的总样本数;SI是波浪模型比较的标准度量,表示误差标准差对平均值的贡献,已用于业务预报中^[18]。

2 开发和测试环境

本文的向量化检验算法是在Python环境下使用Numpy库(网址:<https://numpy.org/>)和Cupy库(网址:<https://docs.cupy.dev/en/stable/>)开发的。Numpy库提供了高性能的多维数组对象和数学函数,对高级向量扩展指令集(AVX)支持较好,适合

高度优化的CPU环境下科学计算和数学分析^[19]。CuPy是一个开源的Numpy兼容数组库,专为在NVIDIA的GPU上实现高性能计算而设计,其具有的多个CUDA核心可以实现更好的并行加速^[20]。对Numpy程序稍作语法修改即可实现CPU平台到GPU平台的转化。这两种工具的开发效率较CUDA环境下的Fortran开发更便捷。根据流程图(见图1),质量控制部分利用布尔运算实现边界和阈值筛选,使用广播和索引操作计算星下点—海岸线的距离,并将符合要求的数组合并;在时空匹配部分,首先通过索引、广播、数据重复和形状改变等操作计算特征半径,再基于海浪平稳性条件将观测数据对齐至整点和半点,利用广播操作和条件判断提取出落入SCVT网格特征半径以内的星下点数据,最后按照时间维将匹配的数组进行合并;为防止内存溢出,程序通过嵌套循环将待检区域分块计算。本文采用一台高性能图形工作站进行实验,软硬件配置见表1。

表1 检验计算环境软硬件配置

Tab.1 Verification calculation environment hardware and software configuration

硬件配置	软件配置
CPU: Intel Xeon Gold 6150(16C×2)	
GPU: Nvidia V100×2	OS: Linux4.18.0-338.el8.x86_64
RAM: DDR4 192 GB	ENV: CUDA 11.2, CUDNN 8.1.1.33, Python 3.8, Cupy 9.5.0, Numpy 1.21.2
HD: 25TB (RAID5)	

3 结果

3.1 FVWAM大面检验结果

3.1.1 CFOSAT有效波高及平均波周期检测

本文利用向量化匹配算法,基于2021年CFOSAT的SWH和MWP数据对FVWAM模式数据进行周期检验,比较算法在不同参数之间的使用情况,检验区域为西北太平洋(100°~150°E, 0°~50°N)。图5和图6分别展示了春(3—5月)、夏(6—8月)、秋(9—11月)、冬(12月—次年2月)4个季节中模式与观测资料SWH和MWP的散点图和泰勒图,表2为误差结果。CFOSAT与FVWAM的SWH基本一致,4个季节的相关系数均较高,在0.95以上;两者的RMSE均保持在0.2~0.4 m内,MAE <

0.26 m, SWH的误差冬季最大(RMSE和MAE分别为0.36 m和0.26 m),夏季最小(RMSE和MAE分别为0.23 m和0.16 m);两者的Bias均为负,表明FVWAM的有效波高数据略低于测量值。CFOSAT与FVWAM的MWP的全年相关性为86%,较SWH的模拟效果偏差,两者MWP的全年RMSE为0.86, MAE为0.63, SI较SWH的结果偏低,在0.2以下,数据更为集中。MWP的误差在夏季和冬季较小,秋季偏高,各月的Bias均为正,表明FVWAM模式对周期预报有系统性偏高。综合可见,本文提出的向量化检验方法对同一遥感资料的不同参数均适用, FVWAM的波高模拟较周期模拟更为精确,两者均有系统性偏差。

3.1.2 HY-2B有效波高检验

利用2021全年HY-2B卫星在西北太平洋区域

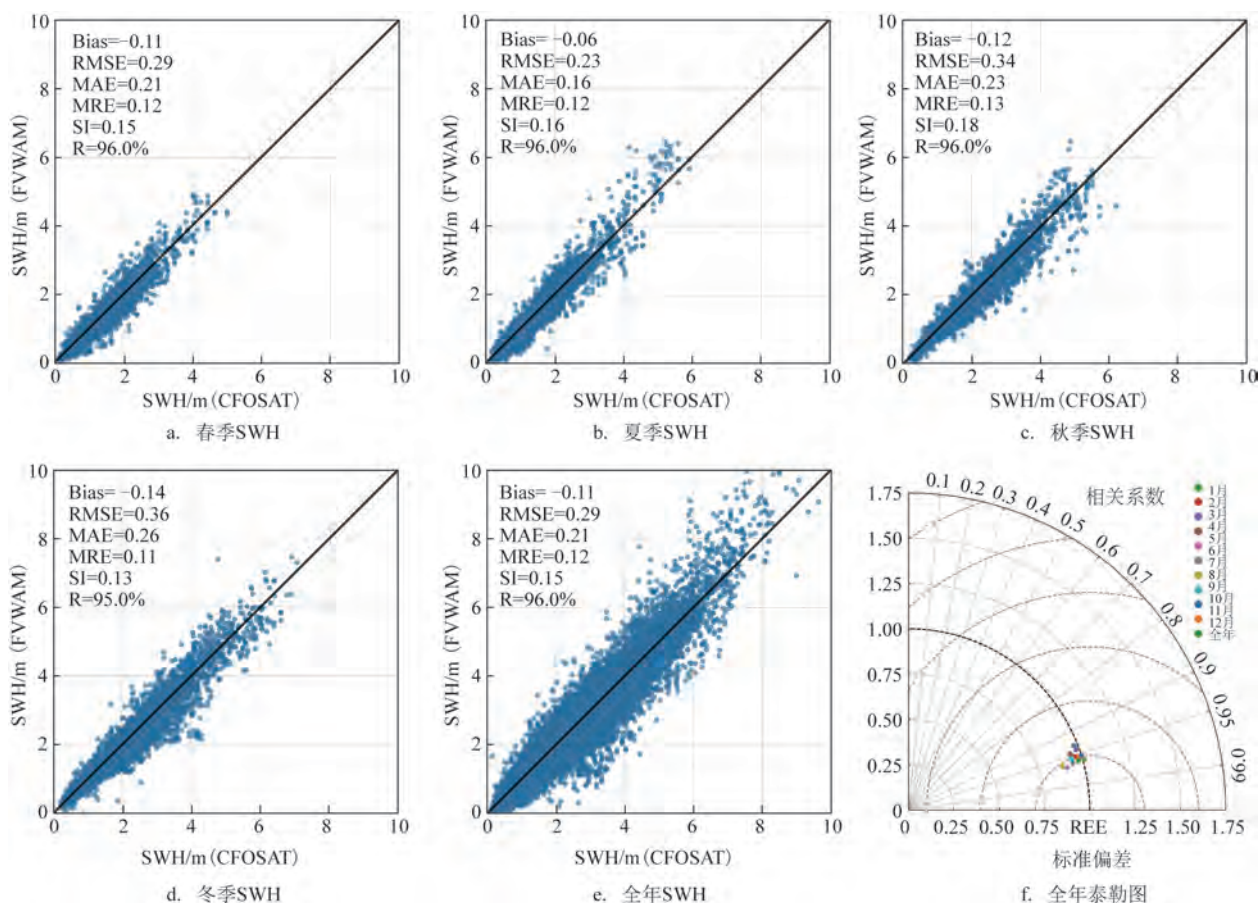


图6 2021年不同季节及全年CFOSAT卫星和FVWAM数据有效波高散点图和全年泰勒图

Fig.6 Scatterplots of significant wave heights for CFOSAT satellite and FVWAM data in different season and annually in 2021, as well as Taylor Diagram for the entire year

表2 2021年CFOSAT卫星与FVEAM数据有效波高、平均波周期误差表

Tab.2 Error table for significant wave height and mean wave period of CFOSAT Satellite and FVEAM data in 2021

时间	Bias		RMSE		MAE		SI		R	
	SWH/m	MWP/s	SWH/m	MWP/s	SWH/m	MWP/s	SWH/m	MWP/s	SWH/m	MWP/s
春	-0.11	0.36	0.29	0.91	0.21	0.65	0.15	0.13	0.96	0.86
夏	-0.06	0.23	0.23	0.80	0.16	0.61	0.16	0.13	0.96	0.83
秋	-0.12	0.39	0.34	0.93	0.23	0.68	0.18	0.13	0.96	0.86
冬	-0.14	0.37	0.36	0.8	0.26	0.60	0.13	0.10	0.95	0.84
全年	-0.11	0.34	0.29	0.86	0.21	0.63	0.15	0.12	0.96	0.86

的SWH数据对FVWAM进行精度检验。两者的有效波高较为一致,全年相关系数为0.88(见表3),其中秋、冬两季相关系数较高,为0.87,春、夏两季相关性较低,MAE为0.3~0.5 m, RMSE为0.5~0.7 m, SI普遍较低,在0.3 m以下。相较于CFOSAT卫星和FVWAM的SWH比较结果, HY-2B卫星与FVWAM

的观测-模拟相关性较低,这是由于本文使用的CFOSAT数据集对波高进行了滑动平均滤波,降低了随机噪声的影响,且滑动平均后的星下网格点和模式的单个网格点尺寸基本一致,进一步降低了表征误差。整体而言,本文所提出的向量化检验方法可适用于轨道性质完全不同的卫星资料参

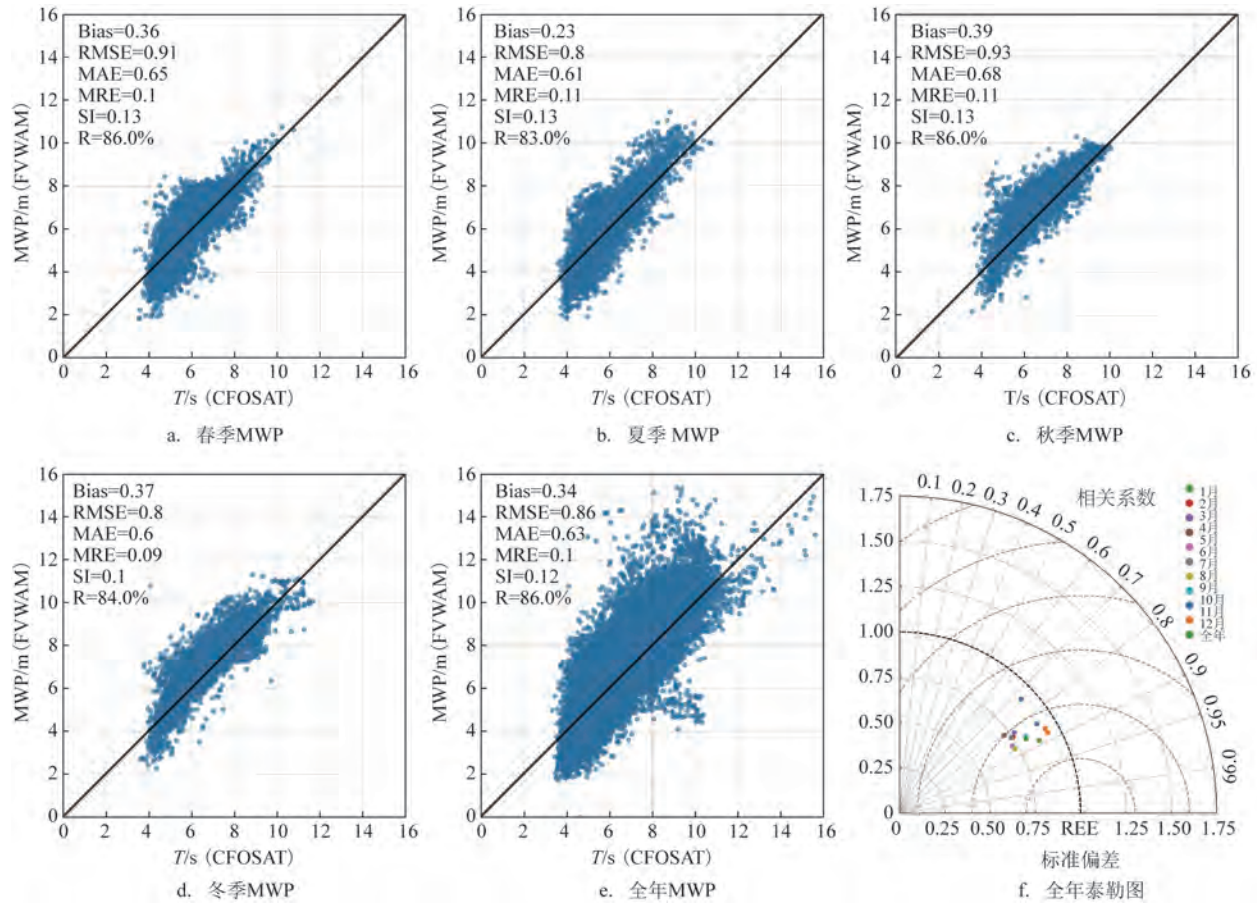


图7 2021年不同季节及全年CFOSAT卫星和FVWAM数据平均波周期散点图和全年泰勒图

Fig.7 Scatterplots of mean wave periods for CFOSAT Satellite and FVWAM Data in different season and annually in 2021, as well as Taylor Diagram for the entire year

表3 2021年HY-2B卫星与FVEAM数据有效波高误差表
Tab.3 Error table for significant wave height of HY-2B Satellite and FVEAM data in 2021

时间	匹配点数	Bias	RMSE	MAE	SI	R
春	64 742	-0.37	0.56	0.40	0.25	0.85
夏	64 876	-0.29	0.51	0.35	0.31	0.85
秋	64 488	-0.39	0.60	0.44	0.27	0.87
冬	61 250	-0.41	0.68	0.48	0.22	0.87
全年	255 356	-0.36	0.55	0.41	0.24	0.88

数检验。

3.2 效率对比

向量化检验算法适用于不同的硬件环境,包括中央处理器(CPU)和图形处理器(GPU)等。此外,算法的运行效率还受到矩阵分片大小的影响。为

全面评估算法在不同硬件和分块策略下的性能表现,本文分别利用NumPy库和CuPy库编写向量化检验程序,并在CPU和GPU上进行了8种矩阵分块方法的测试($5^{\circ} \times 5^{\circ}$ 、 $6^{\circ} \times 6^{\circ}$ 、 $7^{\circ} \times 7^{\circ}$ 、 $8^{\circ} \times 8^{\circ}$ 、 $9^{\circ} \times 9^{\circ}$ 、 $10^{\circ} \times 10^{\circ}$ 、 $5^{\circ} \times 10^{\circ}$ 和 $10^{\circ} \times 5^{\circ}$)。经计算(见表4),8种测试在CPU和GPU上的检验结果一致;不同分块策略在GPU环境下运行时间相对稳定,而在CPU平台上变化幅度较大。在CPU上6种正方形分块方式里 $6^{\circ} \times 6^{\circ}$ 运行机时较小,GPU上 $9^{\circ} \times 9^{\circ}$ 运行机时较小,纵向分块 $5^{\circ} \times 10^{\circ}$ 与横向分块 $10^{\circ} \times 5^{\circ}$ 的运行速度不尽相同。由此可见矩阵分块的运行机时与硬件特征有关,GPU上较大的矩阵分块更具有优势。不同的分块策略在GPU平台上执行匹配操作较CPU的速度提升幅度表现出较大差异,选择 $10^{\circ} \times 10^{\circ}$ 分块方案可实现最高的速度提升,提升速度为33.93倍。总的来说,本文提出的基于矩阵运

表 4 不同分块策略下的向量化匹配程序月平均运行时间

Tab.4 Monthly average execution time of vectorized matching program under different block strategies

片区尺寸	CPU 机时/s	GPU 机时/s	速率提升
5°×5°	42.090 3	5.381 3	7.82
6°×6°	38.669 3	3.745 7	10.32
7°×7°	43.811 7	3.173 6	13.81
8°×8°	47.340 0	2.618 3	18.08
9°×9°	52.831 9	2.115 8	24.97
10°×10°	73.996 1	2.181 0	33.93
5°×10°	44.809 1	3.068 7	14.60
10°×5°	44.637 5	3.295 8	13.54

算 SCVT 大面检验算法能够充分利用计算资源,在 GPU 上的计算效率较 CPU 可提高约 33 倍;区块大小对提升效率有一定影响,需根据实际算例调试优化。

4 结 论

针对大数据带来的计算量激增及复杂网格的时空匹配问题,本文提出了一种适用于并行计算设备的向量化大面检验方法,并将高性能数据分析的核心思想融入这一方法中。该方法将落入 SCVT 等效圆(与 SCVT 面积相等的圆)内的观测结果取平均值作为该网格实测值,采用广播、索引等向量化操作实现观测-模式资料的时空匹配,并对待检测区域进行分块避免内存溢出。该方法不仅适用于 SCVT 特定类型网格,也适用于三角形、四边形等其他网格模拟数据的大面检验中。结论如下:

①采用 2021 年 CFOSAT 卫星 SWH 和 MWP 数据对 FVWAM 模式产生的 SCVT 网格数据进行检验,发现向量化方法在 FVWAM 海浪模式的不同参数中均显示出较好的适应性,其中 FVWAM 有效波高模拟误差更小,全年相关系数在 0.92 以上。

②对于轨道性质完全不同的 HY-2B 卫星资料参数检验,SWH 预报结果和观测资料也显示出良好的一致性。

③分别将匹配算法在 CPU 和 GPU 上运行,实验结果表明两者的计算结果一致,GPU 的计算效率较 CPU 提高可达 33 倍。

④区域分块策略对效率提升也有一定影响,

GPU 上较大的矩阵分块更具有优势,需要根据实际情况优化和调试。

致谢:本项目得到“中国东盟蓝色伙伴关系建设”项目支持,在此表示感谢。

参考文献:

- [1] CHEN C T, ZHAO Y L, XING J Y, et al. Validation of the wave model SWAN against altimeter data from Jason-2 satellite[C]// 2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS). Fort Worth: IEEE, 2017: 3385-3388.
- [2] HARIKUMAR R, HITHIN N K, NAIR T M B, et al. Ocean state forecast along ship routes: evaluation using ESSO-INCOIS real-time ship-mounted wave height meter and satellite observations[J]. Journal of Atmospheric and Oceanic Technology, 2015, 32(11): 2211-2222.
- [3] 刘力铨, 付东洋, 邹巨洪, 等. 自主 HY-2B 卫星北极地区风场及有效波高产品交叉验证分析[J]. 遥感技术与应用, 2023, 38(1): 227-238.
- LIU L C, FU D Y, ZOU J H, et al. Cross-validation analysis of sea surface wind and significant wave height over the arctic from the autonomous HY-2B satellite[J]. Remote Sensing Technology and Application, 2023, 38(1): 227-238.
- [4] BHASKARAN P K, NAYAK S, BONTU S R, et al. Performance and validation of a coupled parallel ADCIRC-SWAN model for THANE cyclone in the Bay of Bengal[J]. Environmental Fluid Mechanics, 2013, 13(6): 601-623.
- [5] DODET G, ABDALLA S, ALDAY M, et al. Error characterization of significant wave heights in Multidecadal satellite altimeter product, model Hindcast, and in situ measurements using the triple collocation technique[J]. Journal of Atmospheric and Oceanic Technology, 2022, 39(7): 887-901.
- [6] AMANI M, MOGHIMI A, MIRMAZLOUMI S M, et al. Ocean remote sensing techniques and applications: a review (Part I)[J]. Water, 2022, 14(21): 3400.
- [7] 刘建强, 叶小敏, 兰友国. 我国海洋卫星遥感大数据及其应用服务[J]. 大数据, 2022, 8(2): 75-88.
- LIU J Q, YE X M, LAN Y G. Remote sensing big data from Chinese ocean satellites and its application service[J]. Big Data Research, 2022, 8(2): 75-88.
- [8] 刘晓燕, 陈莉, 邢建勇, 等. 海洋 2B 卫星和中法海洋卫星散射计海面风资料在台风“米娜”模拟中的一次应用[J]. 海洋预报, 2022, 39(3): 75-82.
- LIU X Y, CHEN L, XING J Y, et al. An application of scatterometer sea surface wind data of HY-2B satellite and China-France oceanography satellite in typhoon “Mitag” simulation[J]. Marine Forecasts, 2022, 39(3): 75-82.
- [9] SUN Z, LIU X L, CAI Z W, et al. A comparison of WAVEWATCH III grid models for a typical reef lagoon[C]//The 29th International

- Ocean and Polar Engineering Conference. Honolulu: ISOPE, 2019.
- [10] AMUNUGAMA M, SUZUYAMA K, MANAWASEKARA C, et al. Analysis of wind-waves with swan on structured mesh and unstructured mesh during the arrival of typhoon[J]. Journal of Japan Society of Civil Engineers, Ser. B3 (Ocean Engineering), 2019, 75(2): I_283-I_288.
- [11] SANTOS L F, PEIXOTO P S. Topography-based local spherical Voronoi grid refinement on classical and moist shallow-water finite-volume models[J]. Geoscientific Model Development, 2021, 14(11): 6919-6944.
- [12] JIANG H Y, SONG Y H, MIRONOV A, et al. Accurate mean wave period from SWIM instrument on-board CFOSAT[J]. Remote Sensing of Environment, 2022, 280: 113149.
- [13] ELGAR S. Waves in oceanic and coastal waters[J]. Oceanography, 2007, 20(3): 133-135.
- [14] 章磊, 何芬, 李鸿赞. 一种基于奇异射线法检测点在多边形内的方法[J]. 计算机应用研究, 2020, 37(S2): 133-135.
- ZHANG L, HE F, LI H Y. Detection method for point within polygon based on singular ray method[J]. Application Research of Computers, 2020, 37(S2): 133-135.
- [15] 于存光. 点在多边形内外的矩形区域射线法[J]. 牡丹江师范学院学报(自然科学版), 2013, 39(2): 1-2.
- YU C G. The ray method for determining whether a point lies inside or outside a polygonal area[J]. Journal of Mudanjiang Normal University (Natural Sciences Edition), 2013, 39(2): 1-2.
- [16] 吴坚, 郑康平, 王小椿. 一种检测点是否是多边形或多面体内的方法[J]. 小型微型计算机系统, 2003, 24(12): 2200-2203.
- WU J, ZHENG K P, WANG X C. A method for the decision of a point whether in or not in a polygon or a polyhedron[J]. Journal of Chinese Computer Systems, 2003, 24(12): 2200-2203.
- [17] DURRANT T H, GREENSLADE D J M, SIMMONDS I. Validation of Jason-1 and Envisat remotely sensed wave heights [J]. Journal of Atmospheric and Oceanic Technology, 2009, 26(1): 123-134.
- [18] PARK K A, WOO H J, LEE E Y, et al. Validation of significant wave height from satellite altimeter in the seas around Korea and error characteristics[J]. Korean Journal of Remote Sensing, 2013, 29(6): 631-644.
- [19] HARRIS C R, MILLMAN K J, VAN DER WALT S J, et al. Array programming with NumPy[J]. Nature, 2020, 585(7825): 357-362.
- [20] OKUTA R, UNNO Y, NISHINO D. CuPy: A NumPy-compatible library for NVIDIA GPU calculations[C]//31st Conference on Neural Information Processing Systems. Long Beach: NIPS, 2017.

Vectorized verification method for Spherical Centroid Voronoi Tessellation grid ocean wave forecasting

XIE Mingyan¹, WEI Lixin^{1,2}, GAO Yuanyong^{1,2}, GAO Zhiyi^{1,2*}

(1. National Marine Environmental Forecasting Center, Beijing 100081, China; 2. Key Laboratory of Marine Hazards Forecasting, National Marine Environmental Forecasting Center, Ministry of Natural Resources, Beijing 100081, China)

Abstract: This paper introduces a vectorized inspection method suitable for parallel computing devices. The method employs the feature radius approximate discrimination method to select the average value of observed points within the search domain as the measured value. It relies on vectorization techniques such as broadcasting, indexing, and the smoothness conditions of waves to achieve matching of patterns and observational data in both time and space. Additionally, the algorithm divides the inspection area into several blocks for separate calculations to prevent memory overflow. The algorithm can run on Central Processing Unit (CPU) and Graphics Processing Unit (GPU), and it is applicable to other types of pattern grids. To evaluate the practical effectiveness of the algorithm, Spherical Centroidal Voronoi Tessellations grid forecast products generated by the Finite Volume Wave Model and satellite along-track remote sensing products are used as the test and reference data, respectively. The results show consistency between CPU and GPU computations, with a 30-fold increase in efficiency on the GPU compared to the CPU. The block size has a certain impact on efficiency improvement, requiring tuning and optimization based on actual cases.

Key words: accuracy verification; GPU acceleration; vectorization; high-performance verification