

王辉赞,郭芑,倪钦彪,等. 基于密度峰值聚类的中尺度涡轨迹自动追踪方法[J]. 海洋学报, 2018, 40(8): 1—9, doi:10.3969/j.issn.0253-4193.2018.08.001

Wang Huizan, Guo Peng, Ni Qinbiao, et al. A CFSFDP clustering-based eddy trajectory tracking method[J]. Haiyang Xuebao, 2018, 40(8): 1—9, doi:10.3969/j.issn.0253-4193.2018.08.001

基于密度峰值聚类的中尺度涡轨迹自动追踪方法

王辉赞^{1,2}, 郭芑³, 倪钦彪^{2,4}, 李佳讯⁵

(1.国防科技大学 气象海洋学院, 湖南 长沙 410073; 2.国家海洋局第二海洋研究所 卫星海洋环境动力学国家重点实验室, 浙江 杭州 310012; 3.94587 部队, 江苏 连云港 222345; 4.厦门大学 近海海洋环境科学国家重点实验室, 福建 厦门 361005; 5. 海军海洋测绘研究所, 天津 300061)

摘要: 中尺度涡信息的提取包括涡旋的识别和轨迹追踪, 其自动识别与追踪对于基于海量数据的中尺度涡分析十分重要。传统涡旋轨迹自动追踪方法一般需要预先设定搜索半径的阈值, 存在一定的主观性。针对传统中尺度涡轨迹追踪方法存在的问题, 论文从聚类的角度出发, 提出基于密度峰值聚类算法实现对涡旋轨迹的自动追踪, 并以南海中尺度涡追踪为例, 将基于聚类的追踪算法与传统的相似度追踪算法进行比较分析。结果表明: (1) 基于密度峰值聚类算法, 可实现对海洋中尺度涡的自动追踪, 该算法涡旋追踪准确率优于传统相似度算法; (2) 该涡旋追踪算法对资料的完整性依赖度较低, 特别是对于存在部分缺损数据的情况仍能较准确追踪; (3) 该追踪算法克服了传统涡旋追踪算法需要预先设定搜索半径阈值的问题, 自适应性更强。

关键词: 中尺度涡; 轨迹追踪; 密度峰值聚类算法; 南海

中图分类号: P731.2

文献标志码: A

文章编号: 0253-4193(2018)08-0001-09

1 引言

海洋中尺度涡是一种重要的海洋中尺度现象, 是海洋物理环境的重要组成部分之一, 其典型的空间尺度为几十到几百千米, 时间尺度为几天到上百天。中尺度涡在海洋中几乎处处存在, 作为大尺度过程与小尺度过程的衔接, 其在能量传递中起着至关重要的作用。中尺度涡还对生物生产力、上层海洋生态环境、生物化学以及海洋声传播等都具有重要的影响。

目前, 随着卫星观测资料的出现, 卫星高度计数据相比常规的调查手段在观测覆盖范围和时间、空间分辨率方面具有较大优势, 基于卫星高度计资料可以

识别和分析中尺度涡, 且大大提高对中尺度涡的监测能力^[1-3]。正是由于高度计资料等海量数据被用于研究中尺度涡, 通过人工的方法分析海量的数据来识别涡旋已经不切实际, 因而在近十几年来, 多种涡旋识别和轨迹追踪的自动算法便应运而生。目前, 确定涡旋中心和大小(或边界)的常用方法有海面高度(SSH)法、Okubo-Weiss(OW)法、Winding-Angle(WA)法和 Vector Geometry(VG)等方法^[2,4-12]。待确定涡旋中心和大小之后, 就可以追踪涡旋的轨迹。

常用的涡旋轨迹追踪方法可以分为3种^[13]。第一种涡旋轨迹追踪方法是像素法^[14]: 将涡旋中心的网格点(即所谓的“像素”)标记为1, 其余网格点则标

收稿日期: 2017-06-10; 修订日期: 2017-10-30。

基金项目: 中国科学院战略性先导科技专项(A类)资助(XDA11010103); 国家自然科学基金(41706021, 41775053, 41206002); 国家海洋局第二海洋研究所专项资助(JG1416); 中国博士后科学基金(2014M551711); 江苏省自然科学基金(BK20151447)。

作者简介: 王辉赞(1983—), 男, 湖南省浏阳市人, 博士/讲师, 主要从事物理海洋学研究。E-mail: wanghuizan@126.com

记为 0。追踪轨迹时,搜索时间和空间(经向、纬向)上都最靠近的标记为 1 的网格点即可。该方法简单快速,适用于涡旋分离明显的海域,但在涡旋结构复杂的区域容易出错。第二种为距离法^[8]:假设 e_1 为 t_1 时的某个涡旋, e_2 为 t_2 时相同类型的涡旋, D 则是 e_1 中心到 e_2 中心的距离,那么, D 最短者构成同一轨迹。为了避免两条不同的轨迹连接在一起, D 需设定一个上限(即搜索半径)。因为涡旋活动于背景流中,所以该上限可近似取局地平均流流速和资料的时间分辨率之积。该方法对搜索半径的选择具有一定的依赖性。第三种为相似度法^[6,9],该方法是距离法的扩展。它基于一个由距离差、半径差、涡动能差和涡度差组成的无量纲相似参数距离,同一条轨迹上前后时刻的涡旋也不能超过一个阈值。该方法适用范围更广,总体效果不错,但相对复杂,且对阈值具有一定的依赖性和主观性。因此,如何在涡旋识别的基础上,克服传统涡旋追踪方法的不足,实现海洋中尺度涡轨迹的准确追踪,具有重要的意义。

涡旋轨迹追踪是将不同时刻具有相近特征、相近距离的涡旋认为是同一个涡旋,假设不同时刻组成的同一个涡旋是一类,那么涡旋轨迹追踪的本质可以看成聚类。值得一提的是,吴笛^[15]采用聚类方法对南海中尺度涡移动轨迹进行分析,但该方法是把一个完整生消的涡旋看成一个样本,对不同涡旋移动轨迹进行聚类,研究涡旋典型移动特征,而本文是将每个时刻识别出的单个涡旋看成一个样本,将同一个涡旋不同时刻特征聚类成同一个涡旋。南海作为一个半封闭深水海盆,有着较为复杂多样的海底地形,且受到黑潮和冬夏季交替季风的显著影响,导致南海的中尺度涡纷繁复杂^[9,16-18]。本文拟选择南海($5^{\circ} \sim 25^{\circ}\text{N}$, $105^{\circ} \sim 125^{\circ}\text{E}$)作为中尺度涡轨迹追踪的研究区域,基于卫星海洋海面高度计资料,拟从聚类的角度出发,采用 Rodriguez 和 Laio^[19]在 *Science* 杂志新近提出的一种快速搜索基于密度峰值聚类算法(Clustering by Fast Search and Find of Density Peaks, CFSFDP),实现海洋中尺度涡旋的自动追踪,并将其与传统的相似度法进行比较。

2 数据与方法

2.1 卫星高度计资料

本文所采用的卫星高度计资料(SLA)是来自于法国空间海洋局的 Archiving, Validation, and Interpretation of Satellite Oceanographic data (AVISO) 计

划。该资料融合了 T/P、Jason-1、Jason-2 和 Geosat Follow-On 等多颗轨道卫星的数据,其空间分辨率为 $(1/4)^{\circ} \times (1/4)^{\circ}$,时间分辨率为 1 d。

2.2 密度峰值聚类算法

常用的经典聚类算法(如 K 均值聚类算法)一般已知聚类中心和聚类数,再通过简单迭代的方法更新数据的聚类中心来进行聚类,但是由于其将每个点都聚类到距离最近的中心,这又会导致其不能检测非球面的数据分布。虽然传统的密度聚类算法对于任意形状分布的数据也可以进行分类,但它必须要通过指定一个密度阈值除去噪音点,对密度阈值依赖性较大。CFSFDP 聚类是基于密度的新聚类算法,由 Rodriguez 和 Laio^[19]在 *Science* 杂志发表提出,该方法可聚类非球形数据集,具有聚类速度快、实现简单等优点,目前得到了较为广泛的应用。

CFSFDP 聚类算法的基本中心思想:假设确定的聚类中心周围都是密度值比它低的点,同时这些密度值比它低的点距离该聚类中心的距离相比于其他聚类中心最小。

CFSFDP 聚类算法步骤如下:

(1) 指标计算。对于每一个样本点 i , 计算其两个指标: i 点的局部密度 ρ_i , i 点与所有高于 i 点密度的点之间距离最小值 δ_i 。这些指标仅依赖于数据点之间的距离 d_{ij} 。

i 点的局部密度 ρ_i 定义如下:

$$\rho_i = \sum_j \chi(d_{ij} - d_c), \quad (1)$$

当 $x < 0$, $\chi(x) = 1$; 否则, $\chi(x) = 0$ 。其中 d_c 为截断距离,默认其为所有样本点的相互距离由小到大排列占 2% 的位置距离数值。

最小距离 δ_i 定义如下:

$$\delta_i = \min_{j: \rho_j > \rho_i} (d_{ij}), \quad (2)$$

式中, δ_i 表示 i 与所有比 i 点密度高的点的最近距离。但是对于最大密度的点,其为所有样本点与样本点之间距离的最大值 $\delta_i = \max(d_{ij})$ 。

(2) 确定聚类中心和归类。不妨以图 1 为例说明。图 1 共有 28 个样本点(分别用 1~28 进行编号),计算所有样本点的密度值并按照由高到低排列,可以看出样本“1”表示密度最高的点。图 1b 表示图中每个点最小距离与局部密度的不同函数的图示,称为决策图,其展示了二维平面内的 28 个点分布,容易发现样本点 1 和点 10 的密度最大,可以将其作为聚类中心。图 1 可以发现“9”和“10”号点拥有相近的密

度值但是其距离值不同,这里“9”属于“1”号类别且比“9”密度高的其他点离它很近,然而比“10”密度高的临近点属于其他类别。“26”、“27”和“28”号点有一个相对较大的距离值,但是其密度值太小,这主要是因为它们是孤立点,我们可以通过给定的 δ_{\min} 和 ρ_{\min} 筛选出同时满足 $(\rho_i > \rho_{\min})$ 和 $(\delta_i > \delta_{\min})$ 条件的点作为聚类中心点。正如预期的那样,只有具有高 δ 和相对较高的 ρ 的点才可以确认为是聚类中心。因为点 26、27、28 是孤立的,所以有相对较高的 δ 值和低 ρ

值,它们也可以被看作是由单个点做成的聚类。当找到聚类中心之后,按照“剩余的每个样本点被归属到比它有更高密度的最近邻聚类中心所属类别,当前样本点的类别应该与高于当前样本点密度的最近的点的类别一致”的原则,指定剩下样本点的类别。CFS-FDP 聚类分配只需一步即可完成,不像其他算法要对目标函数进行迭代优化。详细算法可参考文献 Rodriguez 和 Laio^[19]。

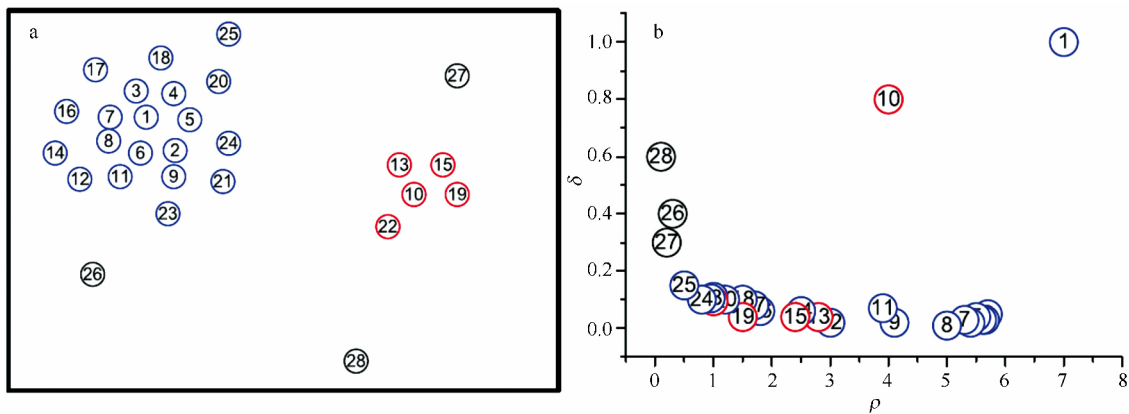


图1 二维 CFSFDP 算法

Fig.1 The CFSFDP algorithm in two dimensions

a. 样本点分布,数据点已经按照降密度排列, b. 对应的数据决策图,不同的颜色对应不同的聚类(引自 Rodriguez and Laio^[19])
a. Point distribution, data points are ranked in order of decreasing density, b. decision graph for the data in fig.1a, different colors correspond to different clusters (cited from Rodriguez and Laio^[19])

3 自动追踪技术

3.1 中尺度涡自动识别技术

本文在 Wang 等^[1]的中尺度涡识别方法的基础上,采用的是改进的基于 SLA 的海洋中尺度涡自动识别方法。具体步骤为:

(1) 针对某一日期 SLA 场,以 1 cm 为间隔提取 SLA 数据的所有等值线集合,从等值线集合中筛选出所有闭合等值线,找出最内圈闭合等值线,取几何中心为涡心;

(2) 筛选出包含涡心的最近最外圈闭合等值线,确定涡边;

(3) 比较涡心和涡边 SLA 值确定涡旋类型。当涡心 SLA 值大于涡边 SLA 值,为反气旋涡;当涡心 SLA 值小于涡边 SLA 值,为气旋涡。本文舍弃振幅小于 2 cm 的涡旋;

(4) 计算涡旋的各种属性:时间点(天)、涡心的位

置(经、纬度)、振幅、半径、动能、相对涡度。

3.2 中尺度涡自动追踪技术

将识别出涡旋并确定涡旋类型将涡旋轨迹进行聚类,具体步骤为:

(1) 粗聚类:首先根据涡旋类别将涡旋分成气旋涡和反气旋涡两大类,保证任意两个不同的轨迹集合之间的两条轨迹是不相关的;

(2) 精聚类:分别对气旋和反气旋两类涡旋进行聚类。由识别涡旋技术得到任意一天内所有涡旋的属性:时间点(天)、涡心的位置(经、纬度)、振幅、半径、动能、相对涡度。选取判别因子:时间点(天)、涡心的位置(经、纬度),采用 CFSFDP 算法对冷暖涡旋进行聚类。根据涡旋识别算法得到的已有一年内所有 m 个涡旋三维样本 $X = \{x_i\} (i=1, 2, 3)$ (即时间点(天)、涡心经度、涡心纬度),首先将时间点(天)进行转换,初始化样本集合(本文将时间变量乘以系数 0.05,使之与位置匹配),其目的是将具有不同单位的

数据匹配(也可以通过直接除以各变量的标准差实现)。待聚类轨迹集合为 n 条 $Y=\{X\}$, 由 2.2 节中 CFSFDP 算法介绍计算每个涡旋样本的两个指标: 局部密度 ρ_i 、高于 i 点密度的最小距离 δ_i , 得到图 2 所展示每个涡旋三维样本和的不同函数的图示(决策

图), 选择具有高 δ (本文 δ 取为 0.6) 和相对较高的 ρ 的点作为轨迹中心(对于有相对较高的 δ 值和低 ρ 值的样本点, 它们可以被看作是由单个点形成的类簇, 也就是异常点), 得到轨迹数和聚类中心样本点。

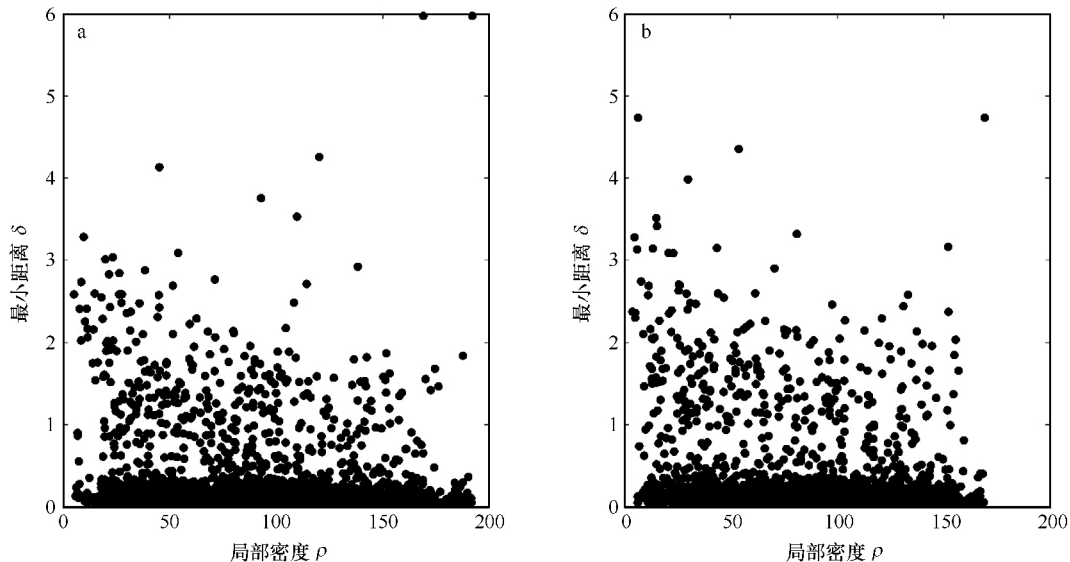


图 2 中尺度涡轨迹追踪聚类决策图

Fig.2 The decision graph for eddy trajectory tracking

a. 气旋涡; b. 反气旋涡

a. Cyclonic eddies; b. anticyclonic eddies

(3) 当聚类中心样本点确定之后, 剩下的涡旋样本点按照 2.2 节中 CFSFDP 算法介绍的原则划分到指定类别, 这样, 就得到了所有涡旋的轨迹集合。考虑到在处理 SLA 数据时可能产生的误差, 排除偶然性以及短暂持续的涡旋信号, 本文只统计大于等于 14 d 的涡旋轨迹, 将得到的涡旋的轨迹集合去除生命周期小于 14 d, 得到最终的轨迹集合。

4 中尺度涡自动追踪效果评估

如引言所述, 目前主要包括像素法、距离法和相似度法共 3 类中尺度涡旋自动追踪方法, 其中相似度法效果较好。为检验本文提出的基于密度峰值聚类算法 CFSFDP 实现对涡旋轨迹自动追踪方法的有效性, 将之与相似度追踪法比较从以下 3 个方面比较:

(1) 一个涡旋的生命周期被聚类在同一涡旋轨迹内的概率大小(指标 1);

(2) 一条聚类轨迹集合中的涡旋是同一涡旋的概率大小(指标 2);

(3) 对资料的完整性依赖度大小(指标 3)。

前两个指标是对聚类准确性的要求。其中, 第一个指标主要用于刻画同一个涡旋被追踪方法分为多个涡旋的可能情况, 即本应属于同一个涡旋被划分至不同轨迹集合; 第二个指标主要刻画追踪的同一条轨迹集合中出现多个涡旋的可能情况, 即本应属于不同涡旋的轨迹被划分至同一个轨迹集合。同时考虑前两个方面指标使轨迹划分结果评估更为准确。最后一个指标主要刻画追踪算法对资料的连续性要求。

为评价所采用的两种不同方法(传统相似度方法和本文 CFSFDP 聚类方法)对涡旋轨迹进行追踪的效果, 需要检验追踪轨迹结果是否与真实涡旋轨迹相符, 但由于目前尚缺乏公认的客观准确追踪方法, 因此本文采用与 Nencioli 等^[8]、Yi 等^[20]等文献相同的处理办法, 把专家人工判别追踪的涡旋轨迹作为真实涡旋轨迹。严格来说, 专家人工判别也有误差, 但由于专家人工判别能够结合专家经验判断, 所以本文参照前述文献做法以专家人工判别为标准(作为真实轨

迹),将两种不同自动判别方法(传统法和聚类法)追踪的轨迹与人工判别追踪的真实轨迹进行对比,统计3个指标大小,得出两种自动判别追踪方法的优劣。不妨以1996整年SLA逐日资料为样本对南海北部涡旋轨迹自动追踪技术进行验证,并举例说明。

4.1 长生命周期涡旋追踪

通过人工判别,1996年1月1日到6月22日吕宋岛西北侧存在一生命周期长达172d的冷涡(图3),图3a的7号涡旋,其生长到消亡时间长,而且历经春夏转换期间,涡旋数量多且容易变性,给其轨迹识别带来困难。我们分别采用CFSFDP聚类法和传统的相似度法对涡旋轨迹进行追踪,并把人工判别得到的作为真实涡旋,并与之对比追踪方法的优劣。这

样两种方法虽然得到都是同一个涡旋,但是由于不同方法差异,导致不同追踪方法的轨迹追踪结果的起始、终止位置都不一样。

应用涡旋自动追踪CFSFDP算法,对7号吕宋冷涡追踪后,得到轨迹如图3b所示,绿色点表示涡旋初生中心位置,黑色点表示涡旋消亡中心位置,红色点线表示涡旋生长到消亡中心位置。由图可以看出:CFSFDP算法对生命周期长的涡旋轨迹追踪有较高的效果。其涡旋中心运动如图4a所示,绿色点表示涡旋初生中心位置,黑色点表示涡旋消亡中心位置,红色点表示涡旋生长到消亡中心位置,黄色线段表示涡旋中心运动。

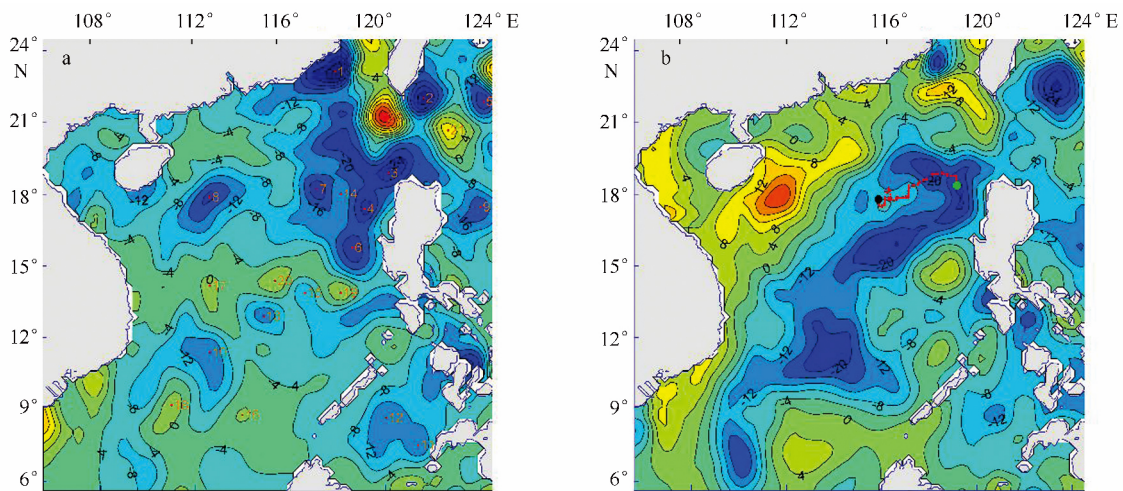


图3 吕宋涡旋轨迹追踪图

Fig.3 Trajectory tracking of Luzon cold eddy

- a.3月8日SLA场(7号涡旋为生命周期长达172d的吕宋冷涡), b.1月15日至6月8日CFSFDP追踪的吕宋冷涡轨迹,等值线代表6月8日的SLA场
- a. SLA field for the March 8 (eddy No.7 with life cycle up to 172 days is Luzon cold eddy), b. the trajectory of Luzon cold eddy from January 15 to June 8 tracking by the CFSFDP clustering, and the contour represents the SLA field on June 8

对比现在应用广泛的相似度涡旋追踪算法,其对吕宋冷涡追踪中心如图4b所示,绿色点表示涡旋初生中心位置,黑色点表示涡旋消亡中心位置,红色点表示涡旋生长到消亡中心位置。由图4可以看出:相似度算法对生命周期长的涡旋轨迹追踪效果较差,一个涡旋的生命周期被聚类在同一涡旋轨迹内的概率不到50%。经统计所有涡旋,基于CFSFDP算法,同一个涡旋的生命周期被聚类在同一涡旋轨迹内的概率能达到85%,而且一条聚类轨迹集合中的涡旋是同一涡旋的概率能够

达到95%以上。

4.2 资料缺损对涡旋轨迹追踪效果影响检验

对于传统追踪算法,其基本思想都是连续SLA场内涡旋对比进行追踪,一旦连续缺失资料,轨迹追踪就会中断。即使资料仅连续缺失3d,比如去除第50天至第52天(2月19日至2月21日)的SLA资料,传统追踪方法失效。然而,本文介绍的CFSFDP算法从聚类的角度出发仍然能够继续追踪。同样以1996年1月1日至6月22日吕宋冷涡为例,其在缺失第100天至第106天(4月9日至4月15日)的

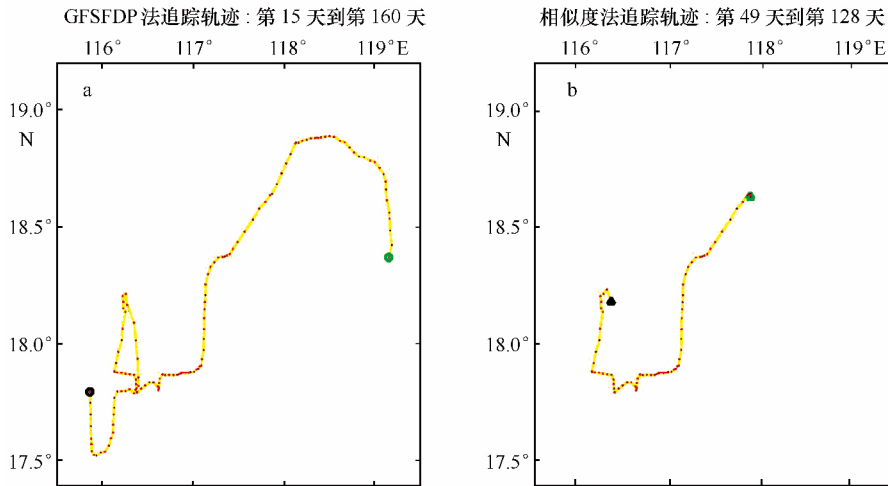


图 4 基于不同涡旋追踪方法追踪的吕宋冷涡中心运动图

Fig.4 Trajectory tracking of Luzon cold eddy center based on different tracking method

a.基于 CFSFDP 法, b.基于传统相似度法。其中 CFSFDP 法轨迹追踪得到的时间段为 1 月 15 日至 6 月 8 日;相似度法轨迹追踪得到的时间段为 2 月 18 日至 5 月 7 日

a. Based on CFSFDP method; b. based on the traditional similarity method. The time period of CFSFDP trajectory tracking is from January 15 to June 8, and the time period of similarity trajectory tracking is from February 18 to May 7

SLA 资料情况下,追踪结果如图 5a 所示,其追踪吕宋冷涡轨迹对应时间段为第 15 天至 119 天(1 月 15 日至 4 月 28 日),其追踪轨迹如图 5b 所示,绿色点表示涡旋初生中心位置,黑色点表示涡旋消亡中心位置,

红色点表示涡旋生长到消亡中心位置。对比图 5b 可以发现,矩形区域为缺失时间的涡旋,但 CFSFDP 算法仍然能够跳过缺失时间段继续对吕宋冷涡轨迹的追踪。

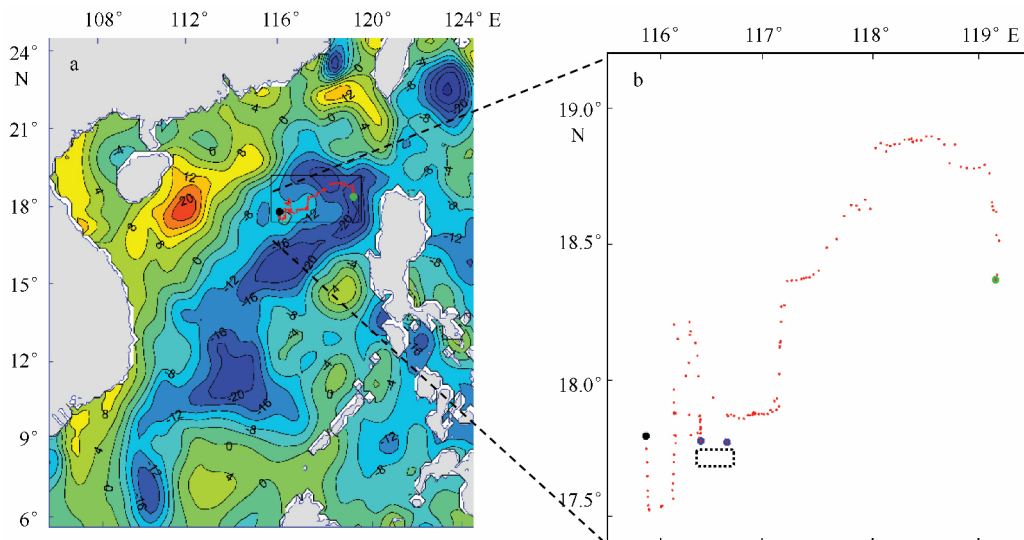


图 5 CFSFDP 聚类吕宋冷涡轨迹追踪图(缺失 7 d 数据)

Fig.5 Luzon cold eddy trajectory tracking based on CFSFDP clustering method (including seven missing days)

由以上案例可以看出,基于 CFSFDP 算法对南海北部涡旋轨迹自动追踪效果比传统算法较好,而且对资料的完整性依赖度较低。

基于 1996 全年数据统计发现(注:指标 3 为资料最多连续缺失天数),CFSFDP 算法的涡旋轨迹追踪在一定程度上优于传统算法如相似度算法结果(表

1), CFSFDP 算法对轨迹追踪的准确率大大提高, 降低了误判率、漏判率。同时, 通过对生命周期超过

50 d 涡旋与专家人工判别对比, CFSFDP 方法相比相似度方法具有优越性。

表 1 CFSFDP 算法与传统相似度算法涡旋追踪结果对比

Tab.1 Comparison of clustering methods for trajectory tracking between the CFSFDP method and the traditional similarity method

	轨迹总数	冷涡 轨迹数	暖涡 轨迹数	生命周期超过 50 d 涡旋数	指标 1	指标 2	指标 3
CFSFDP 算法	179	102	77	10	超过 80%	超过 95%	≥ 7 d
相似度算法	166	83	83	8	低于 70%	超过 95%	≤ 2 d

5 结论与讨论

由于涡旋轨迹追踪是将不同时刻具有相近特征、相近距离的涡旋认为是同一个涡旋, 追踪其运动轨迹, 假设不同时刻组成的同一个涡旋是一类, 那么涡旋轨迹追踪的本质就是聚类。基于此, 本文从聚类角度出发, 提出基于快速搜索发现密度峰值的 CFSFDP 聚类算法, 实现海洋中尺度涡旋的快速有效聚类, 涡旋追踪准确率明显优于传统算法。对生命周期长的涡旋轨迹追踪有较高的效果, 一个涡旋的生命周期被聚类在同一涡旋轨迹内的概率能达到 85%, 而且一条聚类轨迹集合中的涡旋是同一涡旋的概率能够达到 95% 以上; 相较于传统追踪算法适用性更强, 对于存在缺损数据的情况仍能准确追踪, 对资料的完整性依赖度低; 同时克服了传统算法中涡旋追踪影响范围参数阈值选择存在较大主观性的不足。

为检验本文提出的基于 CFSFDP 聚类的涡旋追踪在强流区的适应情况, 我们还以黑潮强流区为例, 进行了涡旋追踪实验(图略), 结果表明, 利用该方法也可以较好地追踪强流区涡旋。从理论上讲, 由于强流区的位置标准差大于非强流区, 所以相当于强流区的时间乘以的系数应该略小于非强流区的系数。

参考文献:

- [1] Wang Guihua, Su Jilian, Chu P C. Mesoscale eddies in the South China Sea observed with altimeter data[J]. Geophysical Research Letters, 2003, 30(21): 2121.
- [2] Chelton D B, Schlax M G, Samelson R M. Global observations of nonlinear mesoscale eddies[J]. Progress in Oceanography, 2011, 91(2): 167-216.
- [3] Li Jiaxun, Wang Guihua, Zhai Xiaoming. Observed cold filaments associated with mesoscale eddies in the South China Sea[J]. Journal of Geophysical Research: Oceans, 2017, 122(1): 762-770.
- [4] Okubo A. Horizontal dispersion of floatable particles in the vicinity of velocity singularities such as convergences[J]. Deep Sea Research and Oceanographic Abstracts, 1970, 17(3): 445-454.
- [5] Weiss J. The dynamics of enstrophy transfer in two-dimensional hydrodynamics[J]. Physica D: Nonlinear Phenomena, 1991, 48(2/3): 273-294.
- [6] Chaigneau A, Gizolme A, Grados C. Mesoscale eddies off Peru in altimeter records: Identification algorithms and eddy spatio-temporal patterns[J].

另外, 值得一提的是, 文中所使用的密度峰值聚类算法, 对于 N 个样本点, 由于在计算两个点之间距离过程中需要生成 $N \times N$ 的矩阵, 对于长时间序列(如 10 年)和大区域范围(如整个北太平洋)的涡旋进行轨迹追踪时会对内存有一定要求, 此时可以采用对样本点按照时间段或空间区域先分割成子块(相邻子块略有重叠)进行聚类追踪, 然后再合并轨迹的办法解决。

值得一提的是, 在本文研究过程中, 作者还对减法聚类(也是一种无需事先确定类别数的聚类方法)进行涡旋追踪聚类试验, 结果表明减法聚类得到的聚类数(即涡旋个数)明显偏少, 同时受参数设置影响大, 减法聚类虽然也是一种密度聚类方法, 但试验证明该方法不可行。

致谢: 感谢复旦大学王桂华教授对本文提出的宝贵意见; 感谢解放军理工大学气象海洋学院高睿、邓隆旺、高顶的有益讨论; 感谢审稿人的意见建议; 感谢专家参与涡旋轨迹人工判别追踪。高度计资料来自于法国空间海洋局的 Archiving, Validation, and Interpretation of Satellite Oceanographic data (AVISO), 在此致谢。

- Progress in Oceanography, 2008, 79(2/4): 106–119.
- [7] Sadarjoen I A, Post F H. Detection, quantification, and tracking of vortices using streamline geometry[J]. Computers & Graphics, 2000, 24(3): 333–341.
- [8] Nencioli F, Dong C M, Dickey T, et al. A vector geometry-based eddy detection algorithm and its application to a high-resolution numerical model product and high-frequency radar surface velocities in the Southern California Bight[J]. Journal of Atmospheric and Oceanic Technology, 2010, 27(3): 564–579.
- [9] Chen Gengxin, Hou Yijun, Chu Xiaoqing. Mesoscale eddies in the South China Sea: Mean properties, spatiotemporal variability, and impact on thermohaline structure[J]. Journal of Geophysical Research: Oceans, 2011, 116(C6): C06018.
- [10] Li Jiaxun, Zhang Ren, Jin Baogang. Eddy characteristics in the northern South China Sea as inferred from Lagrangian drifter data[J]. Ocean Science, 2011, 7(5): 661–669.
- [11] Dong Changming, Lin Xiayan, Liu Yu, et al. Three-dimensional oceanic eddy analysis in the Southern California Bight from a numerical product [J]. Journal of Geophysical Research: Oceans, 2012, 117(C7): C00H14.
- [12] 刘宇. 基于拉格朗日数据和欧拉数据的中尺度涡旋识别与分析[D]. 广州: 中国科学院南海海洋研究所, 2012.
Liu Yu. Mesoscale eddy detection and analysis from Lagrangian data and Eulerian data[D]. Guangzhou: South China Sea Institute of Oceanology, Chinese Academy of Sciences, 2012.
- [13] 倪钦彪. 吕宋海峡附近中尺度涡的统计特征和复合三维结构[D]. 厦门: 厦门大学, 2013.
Ni Qinbiao. Statistical characteristics and composite three-dimensional structures of mesoscale eddies near the Luzon Strait[D]. Xiamen: Xiamen University, 2013.
- [14] Henson S A, Thomas A C. A census of oceanic anticyclonic eddies in the Gulf of Alaska[J]. Deep-Sea Research Part I: Oceanographic Research Papers, 2008, 55(2): 163–176.
- [15] 吴笛. 南海中尺度涡移动轨迹聚类分析[D]. 北京: 中国科学院大学, 2015.
Wu Di. Trajectory clustering analysis of ocean eddies in the South China Sea[D]. Beijing: University of Chinese Academy of Sciences, 2015.
- [16] 王桂华, 苏纪兰, 齐义泉. 南海中尺度涡研究进展[J]. 地球科学进展, 2005, 20(8): 882–886.
Wang Guihua, Su Jilan, Qi Yiquan. Advances in studying mesoscale eddies in South China Sea[J]. Advances in Earth Science, 2005, 20(8): 882–886.
- [17] Xiu Peng, Chai Fei, Shi Lei, et al. A census of eddy activities in the South China Sea during 1993–2007[J]. Journal of Geophysical Research: Oceans, 2010, 115(C3): C03012.
- [18] 郑全安, 谢玲玲, 郑志文, 等. 南海中尺度涡研究进展[J]. 海洋科学进展, 2017, 35(2): 131–158.
Zheng Quanan, Xie Lingling, Zheng Zhiwen, et al. Progress in research of mesoscale eddies in the South China Sea[J]. Advances in Marine Science, 2017, 35(2): 131–158.
- [19] Rodriguez A, Laio A. Clustering by fast search and find of density peaks[J]. Science, 2014, 344(6191): 1492–1496.
- [20] Yi Jiawei, Du Y, He Z, et al. Enhancing the accuracy of automatic eddy detection and the capability of recognizing the multi-core structures from maps of sea level anomaly[J]. Ocean Science, 2014, 10(1): 39–48.

A CFSFDP clustering-based eddy trajectory tracking method

Wang Huizan^{1,2}, Guo Peng³, Ni Qinbiao^{2,4}, Li Jiaxun⁵

(1. Institute of Meteorology and Oceanography, National University of Defense Technology, Changsha 410073, China; 2. State Key Laboratory of Satellite Ocean Environment Dynamics, Second Institute of Oceanography, State Oceanic Administration, Hangzhou 310012, China; 3. Unit 94587, Lianyungang 222345, China; 4. State Key Laboratory of Marine Environmental Science, Xiamen University, Xiamen 361005, China; 5. Naval Institute of Hydrographic Surveying and Charting, Tianjin 300061, China)

Abstract: Information extraction of ocean mesoscale eddies includes the eddy identification and its trajectory tracking, both of which are very important for the research on mesoscale eddies based on the massive data. The traditional methods of trajectory tracking generally need to be set the threshold of the search radius beforehand, which could introduce a certain degree of subjectivity. To improve the existing problem of the traditional methods, an automatic tracking method of the mesoscale eddies is proposed in this study based on the Clustering by Fast Search and Find of Density Peaks (CFSFDP) from clustering point of view. Then it was compared with the traditional similarity algorithm by taking the South China Sea as a testbed. Our results show that: (1) Based on the CFSFDP

algorithm, the automatic tracking of mesoscale eddies is realized, and the accuracy is better than the traditional similarity method; (2) The proposed tracking algorithm is less dependent on the data integrity especially for the presence of partial missing data; (3) Our proposed tracking method has stronger adaptability, which overcomes the problem that the search radius need to be set beforehand in the traditional method.

Key words: mesoscale eddies; trajectory tracking; Clustering by Fast search and Find of Density peaks(CFSFDP); South China Sea