



国防科技大学
National University of Defense Technology



高分辨率涡旋建模融合温盐声数据集 (EFGO)
用户手册 (第一版)

国防科技大学

上海海洋大学

2025年12月 南京/上海

EFGO 数据集使用说明

高分辨率涡旋建模融合温盐声数据集 (Eddy Model Fusion Gradient Optimal Interpolation Data, EFGO) 的研制, 得到了国家自然科学基金面上项目“高时效高分辨率全球 Argo 数据产品智能系统研制” (42576203) 等项目的资助。

数据集制作主要基于研究团队首次提出的针对中尺度涡旋的三维温盐声速场重构模型和梯度依赖相关尺度最优插值客观分析方法等研究成果制作而成, 水平空间分辨率达 $1/16^\circ$ 、时间为“日”分辨率。突破了基于中尺度涡物理过程模型与数据驱动机器学习相结合的三维“温盐结构-声场特征”一致性重构和融合关键技术, 在充分挖掘历史观测信息基础上, 利用卫星观测的表观特征和针对性实时观测, 对温盐结构引入中尺度涡统一结构动力约束, 将海水密度垂直结构作为声速重构约束嵌入网络, 同时引入空间特征提取机制, 通过移位窗口对声速场高效编码, 更好地刻画了声速空间分布特征, 实现区域化三维温盐声场一体化重构。

数据集是首套考虑海洋涡旋特征的温盐声一致性高分辨率三维重构产品, 分为两个子数据集: 温盐数据子集 (EFGO.ST) 和声速数据子集 (EFGO.SS)。其中, EFGO.ST 数据集的开发和维护由国防科技大学前沿交叉学科学院和上海海洋大学共同完成; EFGO.SS 数据集的开发和维护由国防科技大学前沿交叉学科学院完成。本数据集可免费下载使用, 凡使用本数据集时, 请统一标注为“EFGO.ST”/“EFGO.SS”, 且按文献格式引用。

1、EFGO.ST数据集引用:

毛科峰, 张春玲, 陈希, 王鹏皓, 傅利福, 朱宇航, 范家慧, 2025. 高分辨率涡旋建模融合温盐声数据集 (EFGO) 用户手册EFGO.ST数据分册 (第一版), 国防科技大学, 29pp。

或者直接引用以下文献:

Penghao Wang, Kefeng Mao, Three-dimensional structure reconstruction of ocean mesoscale eddies based on physical process modeling and data-driven machine learning, *Ocean Modelling*, Volume 196, 2025, 102558.

Zhang Chun Ling, Xu Jian Ping, Bao Xian Wen, Wang Zhen Feng, An Effective

Method for Improving the Accuracy of Argo Objective Analysis. *Acta Oceanologica Sinica*, 32(7): 66-77, 2013。

2、EFGO.SS数据集引用:

李明, 毛科峰, 陈希, 李洪臣, 刘宇航, 王鹏皓, 朱宇航, 2025. 高分辨率涡旋建模融合温盐声数据集 (EFGO) 用户手册EFGO.SS数据分册 (第一版), 国防科技大学。

或者直接引用以下文献:

李洪臣, 李明, 陈希等. 基于机器学习和经验正交函数法的声速剖面快速重构[J]. *海洋与湖沼*, 2024.

李洪臣, 李明, 王鹏皓. 融合机器学习和物理模型的中尺度涡声速剖面重构[J]. *海洋学报*, 2025.

Li H CH, Liu Y H, Li M. A Deep Learning-based Reconstruction Model for 3D Sound Speed Field Combining Underwater Vertical Information [J]. *Frontiers in Marine Science*, 2025.

Li H CH, Li M. Integrating Machine Learning and Physical Models for Reconstructing Sound Speed Profiles in Mesoscale Eddies. *Journal of Oceanology and Limnology*, 2025.

由于时间仓促, 该数据产品难免会有不足之处, 欢迎广大用户提供宝贵意见或建议, 非常感谢!

联系人

毛科峰: 025-80802758, maokefeng@nudt.edu.cn

张春玲: 021-61900396, clzhang@shou.edu.cn

李明: 025-80802824, liming17c@nudt.edu.cn

刘增宏: 0571-81963197, zliu@sio.org.cn

第一部分 EFGO.ST 数据分册

目录

一、 EFGO.ST 数据集基本信息	1
二、 EFGO.ST 融合数据说明	2
三、 温盐融合建模简介	4
4.1 观测资料标准化	5
4.2 背景场的构建	5
4.3 涡旋内温盐剖面重构	6
4.4 多源数据融合	7
4.5 衍生要素计算	9
四、 EFGO.ST 数据集效果检验	9
5.1 相对分析误差	9
5.2 均方根误差	12
5.3 谱分析	15
5.4 实测剖面对比	17
5.5 涡旋结构对比	24
五、 EFGO.ST 数据产品总结	26

一、EFGO.ST 数据集基本信息

表 1.1 EFGO.ST 数据集基本信息

数据名称	Eddy Model Fusion Gradient Optimal Interpolation Data —Salinity and Temperature, EFGO.ST	
空间范围	15°N~35°N, 114°E~135°E, 0~5500m	
时间范围	2024 年 1 月 1 日至 2025 年 6 月 30 日	
时间分辨率	逐日	
空间分辨率	1/16°×1/16°	
数据格式	EFGO.ST_YYYYMMDD.nc	
环境要素 融合分析变量	<ul style="list-style-type: none"> ● temp (海水温度) ● terr (海水温度分析误差) ● salt (海水盐度) ● serr (海水盐度分析误差) ● u (海水东-西向密度流速) ● v (海水南-北向密度流速) ● SCD (海水表层声道) ● MLD (混合层深度) ● TBD (温跃层下界深度) ● TTG (温跃层强度) 	
nc 文件 存储变量	变量名 (全称)	要素 (维数)
	lon (Longitude)	经度 (337×1)
	lat (Latitude)	纬度 (321×1)
	time (Time)	时间: 距离 0000-01-01 的天数 (1×1)
	depth (Sea water depth)	垂向分层 (44×1)
	temp (Sea water temperature)	温度场 (321×337×44)
	terr (Sea water temperature error)	温度误差 (321×337×44)
	salt (Sea water salinity)	盐度场 (321×337×44)
	serr (Sea water salinity error)	盐度误差 (321×337×44)
	u (Sea water east-west density current velocity)	经向流 (321×337×44)
	v (Sea water south-north density current velocity)	纬向流 (321×337×44)
	SCD (Surface sound depth)	表层声道 (321×337)
	MLD (Mixed layer depth)	混合层深度 (321×337)
	TBD (Bottom depth of thermocline)	温跃层下界深度 (321×337)
TTG (Thermocline temperature gradient)	温跃层强度 (321×337)	

Eddy_center_lat (Eddy center latitude)	涡旋中心纬度 (NN×1)
Eddy_center_lon (Eddy center longitude)	涡旋中心经度 (NN×1)
Eddy_radius (Eddy radius)	涡旋半径 (NN×1)
Eddy_type (Eddy type)	涡旋类型 (1 代表气旋, -1 代表反气旋); (NN×1)
Ineddy_argo_id (Ineddy argo id)	涡旋重构是否使用了 Argo 数据 (1 代表采用, -1 代表没采用), (NN×1)
Eddy_boundary_lat (Eddy boundary latitude)	涡旋边界纬度; (NN×mm)
Eddy_boundary_lon (Eddy boundary longitude)	涡旋边界经度; (NN×mm)
N_boundary (Number of boundary profiles)	涡旋边界剖面的数量; (NN×1)
Ineddy_lat (Ineddy latitude)	涡旋内部的剖面纬度; (NN×nn)
Ineddy_lon (Ineddy longitude)	涡旋内部的剖面经度; (NN×nn)
N_ineddy (Number of ineddy profiles)	涡旋内部的剖面数量; (NN×1)
Ineddy_depth (Ineddy depth)	涡旋内部的深度剖面; (dd×1)
Ineddy_temp (Ineddy temperature)	涡旋内部的温度剖面; (NN×nn×dd)
Ineddy_salt (Ineddy salinity)	涡旋内部的盐度剖面; (NN×nn×dd)
注: NN 为当天涡旋个数, mm 为当天单个涡旋边界剖面数量最大值, nn 为当天单个涡旋内部剖面数量最大值, dd 为涡旋剖面深度的层数	

二、EFGO.ST 融合数据说明

EFGO.ST 数据集主要融合了观测增强的中尺度涡三维结构重构数据、卫星海表面观测和多源剖面观测数据。其中, 剖面观测资料包括 Argo、GTSP、WOD 等数据库中的温、盐度观测剖面, 涡旋三维结构重构数据包括中尺度涡内部温度和盐度剖面数据; 融合的海表面温度 (SST, 时间分辨率为天, 空间分辨率为 0.01°) 和海表面盐度 (SSS, 时间分辨率为天, 空间分辨率为 0.25°) 用于补充剖面观测中表层数据的缺失。具体数据信息如表 2.1 所示:

表 2.1 融合数据信息

名称	数据源	时间分辨率	空间分辨率	融合变量	备注
Argo	https://www.argo.org.cn/	天	散点	温度、盐度	处于涡心的剖面用于涡旋重构, 其余观测用于数据融合
GTSP	https://www.ncei.noaa.gov/products/global-temperature-and-salinity-profile-programme	天	散点	温度、盐度	剔除了重复的 Argo 剖面
WOD	https://www.ncei.noaa.gov/products/world-ocean-database		散点	温度、盐度	主要选取 CTD、XBT、XCTD、锚系浮标(MRB)、漂流浮标观测数据
观测增强的涡旋重构剖面数据	自研	天	格点	温度、盐度	重构中尺度涡内部的温盐剖面, 用于数据融合
SST	https://podaac.jpl.nasa.gov/dataset/MUR-JPL-L4-GLOB-v4.1	天	0.01°	温度	用于补充表层温度
SSS	https://data.remss.com/smap/SSS/V06.0/FINAL/L3/	天	0.25°	盐度	用于补充表层盐度

图 2.1 以 2025 年 5 月 7、10、11 日三天为例, 展示了融合数据的站点分布。

- Argo: 用于数据融合的Argo观测剖面
- Obs: 用于数据融合的历史观测剖面
- Eddy:用于数据融合的涡旋剖面
- Obs-ship: 用于检验融合结果的观测剖面
- Obs-eddy: 用于涡旋重构的观测剖面

剖面个数统计	5月7日	5月10日	5月11日
● Argo	18	15	16
● Obs	4	1	11
● Eddy(涡旋个数)	71	58	61
● Eddy(剖面个数)	12341	11510	11762
● Obs-ship	3	4	4
● Obs-eddy	3	3	5

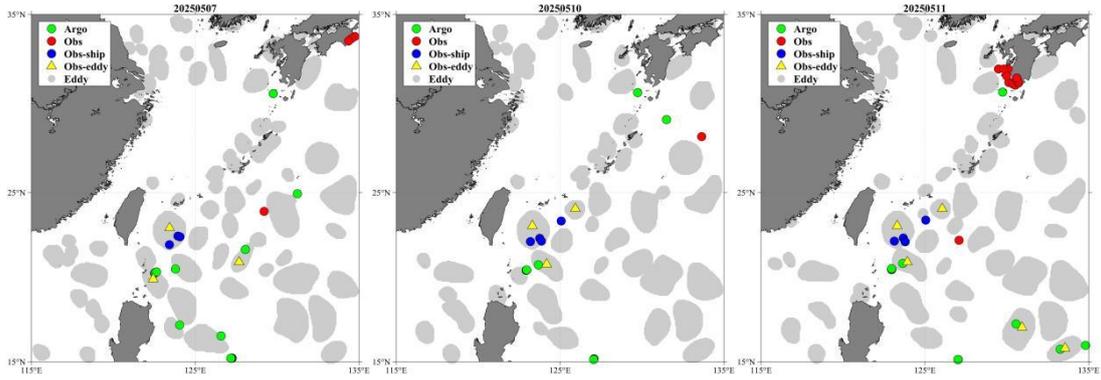


图 2.1 2025 年 5 月 7 日、10 日及 11 日观测数据分布与数量统计

三、温盐融合建模简介

EFGO.ST 融合资料集基于梯度依赖相关尺度的最优插值客观分析方法 (Zhang et al., 2013; 张春玲等 2015; Zhang et al., 2022) 构建, 融合了包括历史观测、涡旋重构数据、Argo 剖面资料、卫星海表观测等多源观测资料, 形成了 15°N~35°N, 114°E~135°E 海域, 2024 年 1 月 1 日至 2025 年 6 月 30 日期间, 空间分辨率为 1/16°, 时间分辨率为天, 垂向 44 层的高分辨率数据产品。其融合建模主要包括背景场的构建、涡旋剖面重构、观测资料标准化、多源数据融合、衍生要素计算等步骤。其构建流程如图 3.1 所示。

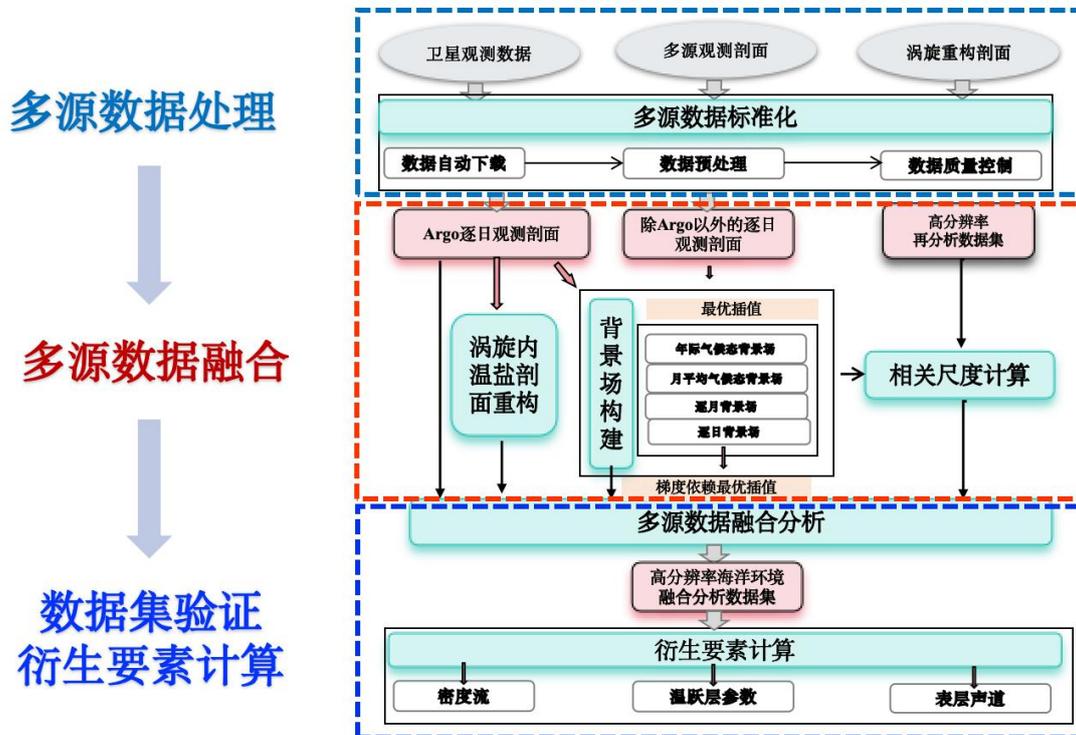


图 3.1 EFGO.ST 融合资料集构建流程图

3.1 观测资料标准化

观测资料标准化主要包括预处理和质量控制两个部分。观测数据的预处理包括各类观测数据的自动读取、变量提取、垂向匹配插值、格式转换等；而质量控制是对汇总的剖面资料统一进行水深递增检验、温盐范围检验、尖峰检验及温盐梯度检测等 (Li 等 2020)。图 3.2 以 2025 年 5 月 7 日为例，给出了观测资料质量控制前、后的剖面对比。

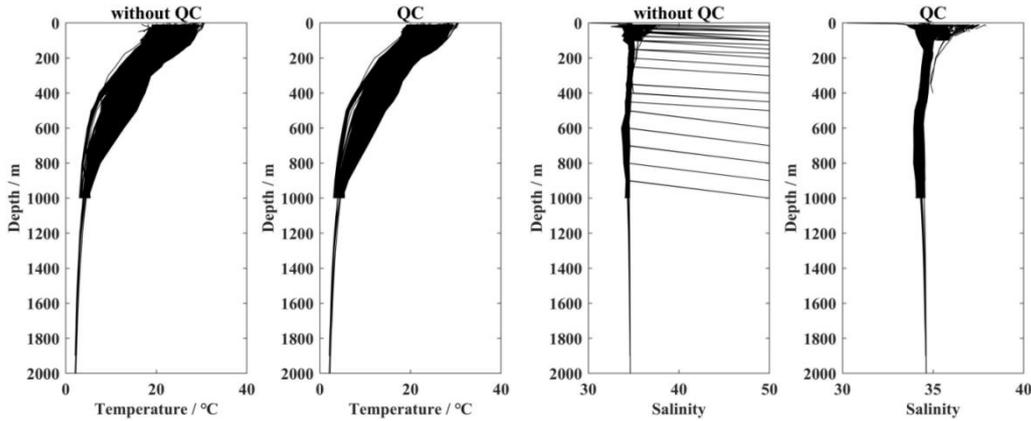


图 3.2 2025 年 5 月 7 日温、盐度观测剖面标准化处理前、后对比

3.2 温盐背景场构建

先将构建区域划分成 $(1/16)^\circ \times (1/16)^\circ$ 的网格，用零场作为多年气候态的初猜场；以此背景场，利用传统的最优插值方法，融合可利用的观测数据、气候态再分析数据，形成多年气候态背景场；然后，以气候态背景场为初猜场，利用传统最优插值，融合多年季节平均观测数据、再分析数据，构建多年季节平均的背景场；然后，以季节平均的背景场为初猜场，融合多年月平均观测数据、再分析数据，构建多年月平均的背景场；再以月平均场为初猜场，通过传统最优插值融合逐年逐月的观测资料和再分析数据，得到逐年逐月的背景场；进而，以逐月平均场为初猜场，通过传统最优插值，融合逐年逐周的观测资料和再分析数据，得到逐周的背景场；最后，以逐周的平均场为背景，融合每天的观测数据，形成 $1/16^\circ \times 1/16^\circ$ 逐日的高分辨率融合背景场，垂向分为不等间隔的 44 个水层：0m, 5m, 10m, 15m, 20m, 25m, 30m, 35m, 50m, 75m, 100m, 125m, 150m, 175m, 200m, 250m, 300m, 350m, 400m, 450m, 500m, 600m, 700m, 800m, 900m, 1000m, 1100m, 1200m, 1300m, 1400m, 1500m, 1600m, 1700m, 1750m, 1800m, 1900m, 2000m, 2500m, 3000m, 3500m, 4000m, 4500m, 5000m, 5500m。

对应的构建流程如图 3.3 所示。

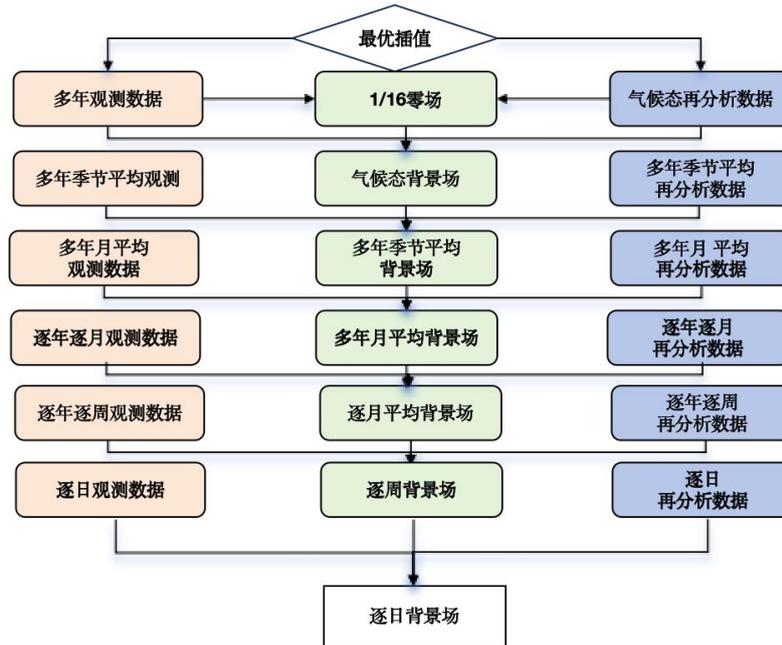


图 3.3 EFGO.ST 融合数据集的背景场构建流程

3.3 涡旋内温盐剖面重构

随着观测频次的增加与观测手段的丰富，采用 XCTD、漂流浮标 (Argo) 及潜标阵列等手段，可以获取中尺度涡内单个或多个零散的现场观测剖面，中尺度涡内的一个观测剖面包含了中尺度涡重要的垂向结构信息，对于中尺度涡三维重构有重要价值。因此，Wang 等 (2025) 提出一种基于现场观测增强的中尺度涡三维结构重构方法：将涡内现场观测与基于中尺度涡统一结构模型改进的密度重构算法相结合，增强中尺度涡垂向结构信息，实现对中尺度涡密度结构的重构，以此为基础，将中尺度涡密度、中尺度涡表观特征要素 (极性、涡心位置及涡旋半径等) 及海表要素 (温度、盐度及动力高度等) 作为输入数据，构建基于机器学习的温盐剖面重构模型，实现中尺度涡三维温盐结构的重构。重构的温盐剖面作为涡旋区的加密数据，用于多源数据融合分析。技术路线如图 3.4 所示。

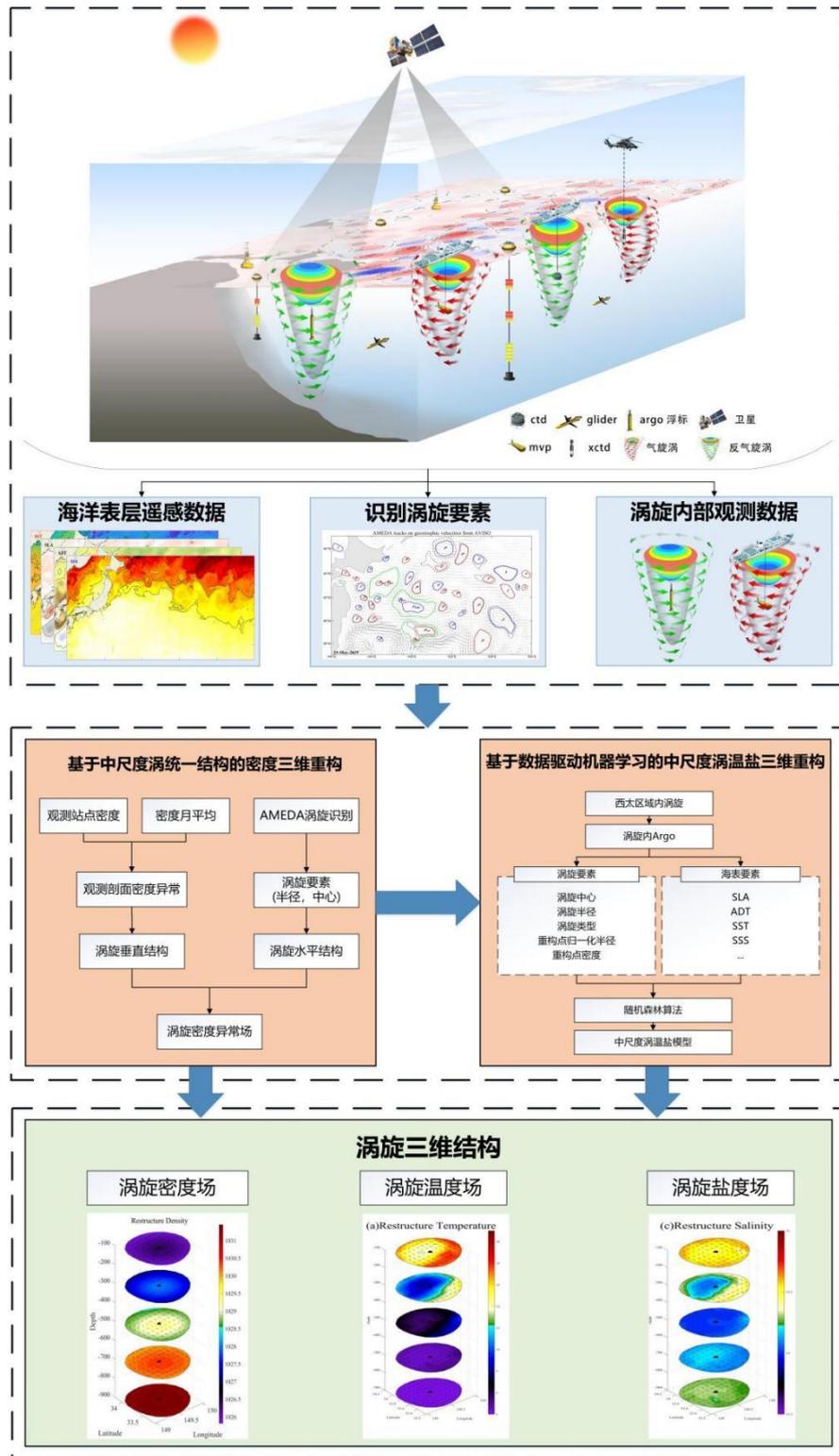


图 3.4 基于物理模型与数据驱动机器学习的中尺度涡三维结构重构技术路线图

3.4 多源数据融合

多源数据融合采用梯度依赖最优插值方法，该方法基于最小二乘理论，根据扩散原理改进了背景误差相关尺度模型。网格点上的分析值为背景值与基于最优

权重的观测增量的叠加，式(1)中给出了受 M 个观测值影响的分析格点的订正方程。该方案的一个关键部分为估计最优权重，其可根据最小二乘理论，通过求解式(2)加以确定：

$$v_i^a = v_i^b + \sum_j^M \omega_{ij} \delta y_j^0 \quad (1)$$

$$\sum_{j=1}^M \omega_{ij} \mu_{jk} + \eta_k \omega_{ik} = \mu_{ik}, \quad k=1, \dots, M \quad (2)$$

其中 v_i^a 为分析值， v 可以为任何环境变量，如温度、盐度； v_i^b 为网格点上背景场对应的初猜值，下标 i 表示网格点数， j 和 k 表示可用剖面点数；观测增量 $\delta y_j^0 = y_j^0 - H(v_j^b)$ 利用观测算子 H 将背景场转化为观测值 y_j^0 的初猜值。每个观测增量都对应一个与背景误差相关 μ_{ik} 的最优权重 ω_{ik} 。 μ_{jk} 和 μ_{ik} 分别是两个观测点 j 和 k ，以及分析格点 i 和观测点 k 之间的背景误差相关系数。参数 η_k 是基于背景误差的相对观测误差的平方，通常设置为一个常数 η ，并调整以改变观测值的权重。

背景误差相关通常遵循一个高斯指数函数，并与距离成反比，如式(3)和(4)所示：

$$\mu_{ik} \sim \exp \left[-\frac{(x_i - x_k)^2}{(L_\theta / G_x)^2} - \frac{(y_i - y_k)^2}{(L_\theta / G_y)^2} \right] \quad (3)$$

$$G_x = 1 + \frac{|\partial v / \partial x|}{E(|\partial v / \partial x|)}, G_y = 1 + \frac{|\partial v / \partial y|}{E(|\partial v / \partial y|)} \quad (4)$$

其中 x 和 y 分别表示经度和纬度， L_θ / G 取决于罗斯贝变形半径和分析要素的水平梯度。相关尺度 L_θ 为相关尺度常数与分析格点所在纬度 θ 的余弦值的乘积。梯度依赖相关尺度参数 G ，与位置 i 的水平梯度相关，包含一个纬向分量 G_x 和一个经向分量 G_y 。

梯度依赖最优插值方法涉及的重点参数包括相对梯度阶数 n （确保相关尺度收敛且缓变）、相对梯度极值 G （在提高计算效率基础上确保观测信息最大化）、相关尺度常数 L （决定相邻网格点间误差的关联强度）、搜索半径 R （定义参与

插值的有效邻域范围)、观测与背景误差比(赋予观测信息的相对权重)。EFGO.ST融合数据集采用的具体相关参数信息如表 4.1 所示。

表 3.1 多源数据融合关键参数信息

相对梯度阶数	梯度极值	相关尺度常数	搜索半径	观测与背景误差比
n=1	G=3	L=3°	R=1°	$\eta=0.25$

3.5 衍生要素计算

利用温、盐度融合数据，基于海水状态方程计算密度流，并通过计算声速场确定声层深度(钱潮等, 2021)，通过阈值法和最大角度法(王学华等, 2023)确定混合层深度、温跃层上界、温跃层下界、温跃层强度等衍生要素。

四、EFGO.ST 数据集效果检验

通过理论检验(包括相对误差分析、均方根误差统计、谱分析等)和实测数据检验(包括 Argo 观测剖面对比、历史观测断面对比)来验证融合数据集 EFGO.ST 的可靠性，同时与其他涡分辨率的再分析数据产品进行中尺度涡温盐结构类比，进一步检验 EFGO.ST 再现中尺度过程动力特征的能力。本节中，融合的温盐数据统一标注为 EFGO。

4.1 相对分析误差

相对分析误差 $p_i = \sum_{k=1}^M w_{ik} u_{ik}$ 是最优插值的理论误差，该参数是由背景误差标准化的无量纲数，代表了分析结果误差占背景误差的百分比，表征着融合结果相对于背景场的矫正程度和精度。

图 4.1-4.2 给出了 2024 年每个月 15 日 1000m 深度内温度和盐度的平均相对分析误差分布。由图可见，无论温度还是盐度，融合分析场的订正效果均主要体现在中尺度过程区域，分析误差在观测密度较大的区域明显较小 (<0.1 ，不超过背景误差的 1/10)，观测不足的边缘区域分析误差相对较大，但平均相对分析误差最大不超过 0.5 (背景误差的 1/2)。

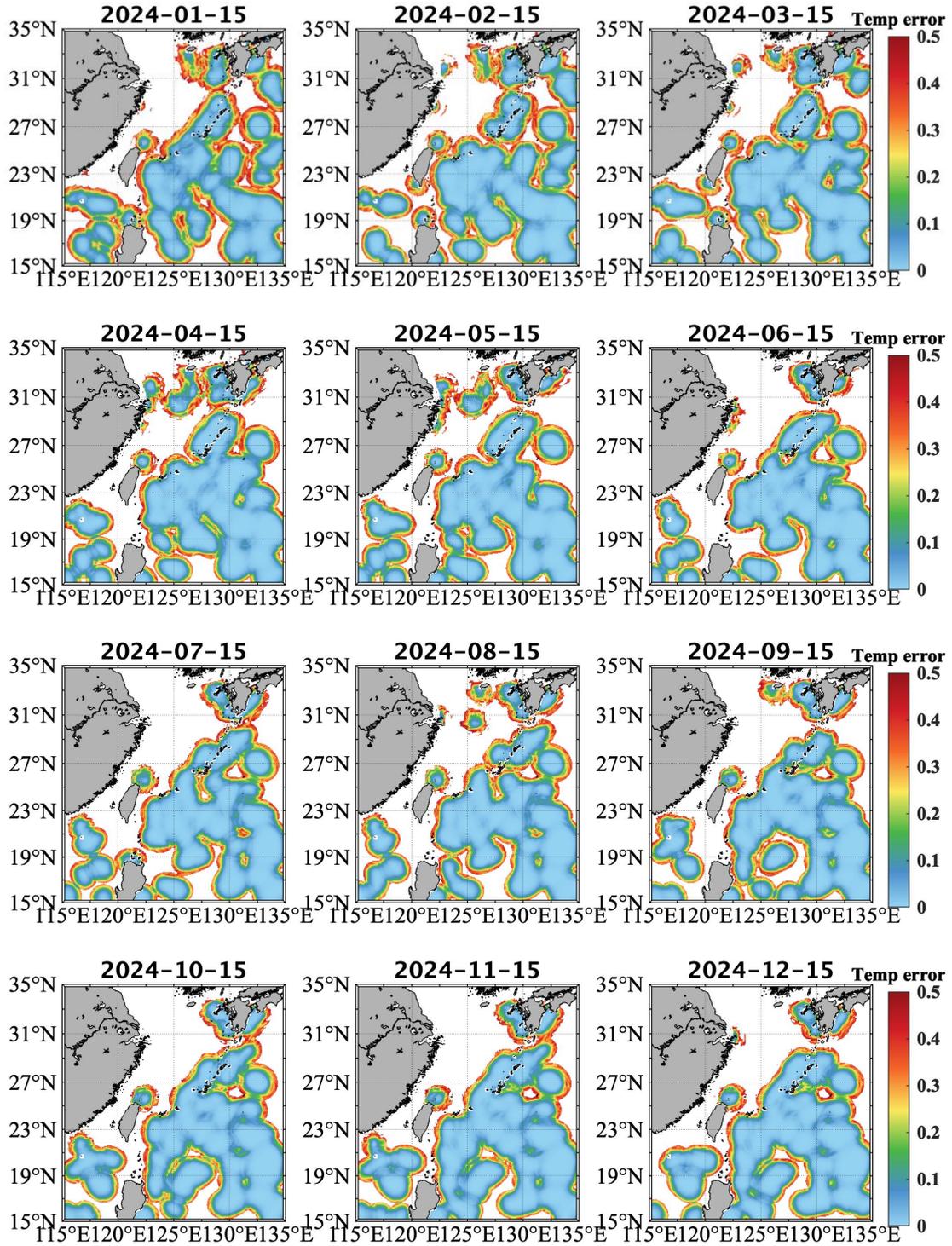


图 4.1 2024 年 1000m 深度内垂向平均的温度相对分析误差空间分布。空白表示无观测订正区域。

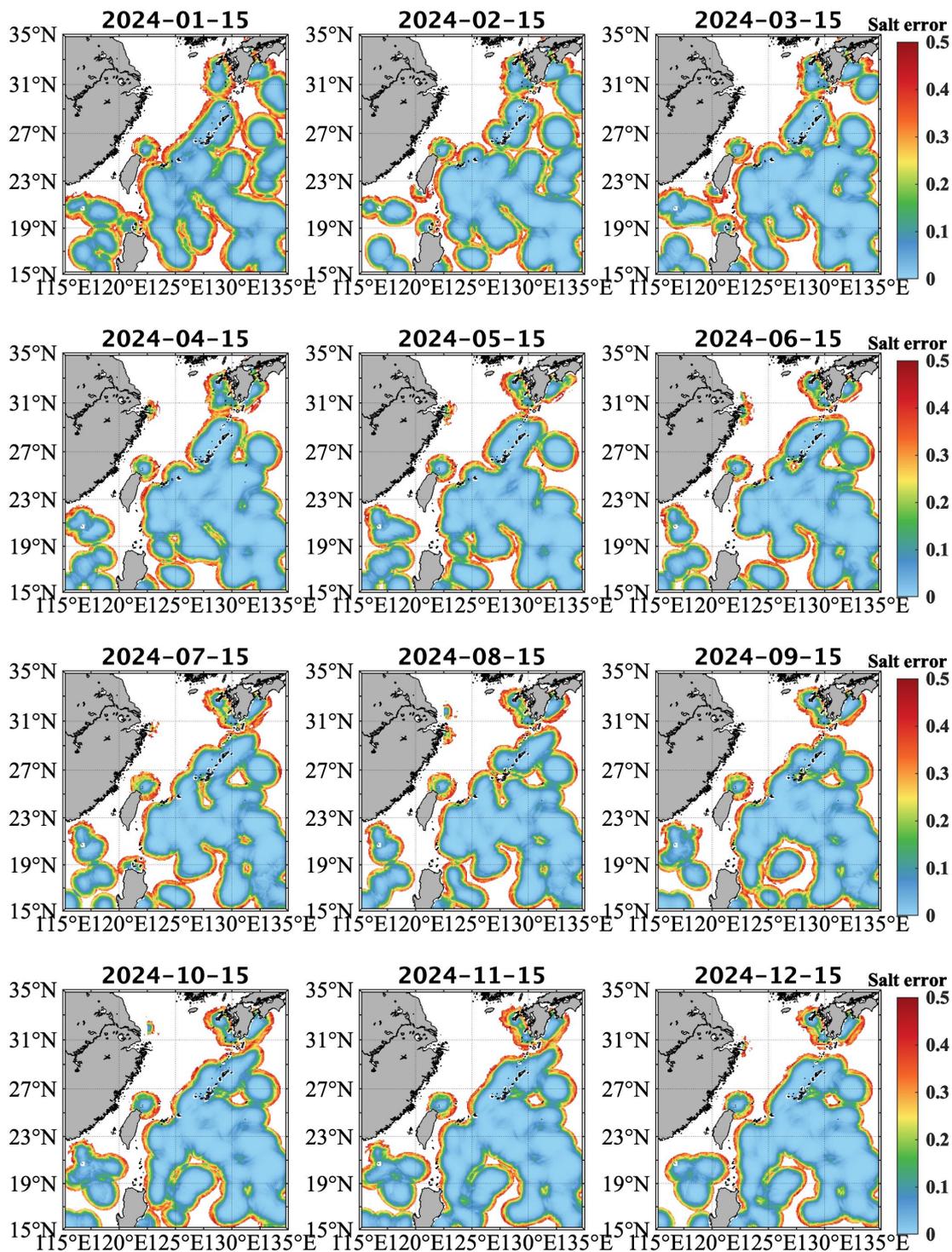


图 4.2 2024 年 1000m 深度内垂向平均的盐度相对分析误差空间分布。空白为无观测订正区域。

将相对分析误差分为 <0.1 , $0.1-0.2$ 、 $0.2-0.3$, $0.3-0.4$, >0.4 等五个量级, 图 4.3 统计了每个水层, 2024 年 1 月 1 日至 2025 年 6 月 30 日期间不同量级平均分析误差所占的比例。总体上, 无论温度还是盐度, 每个水层有着类似的占比分布,

温、盐度误差占比随着误差增大而减小，相对分析误差大于 0.5 的格点数仅占 5% 左右。各深度上温盐分析误差<0.1 的格点数占比高达约 70%，超过 0.1 的总体占 30%左右。其中 0.1-0.2 量级占比在 12%左右，分析误差介于 0.2-0.3 之间的约占 8%，0.3-0.4 之间的约为 5%，而分析误差大于 0.4 的格点数仅占约 5%。随着深度增加，不同量级误差的分析格点数所占的比例变化不大，变化范围在 1%-2% 之间浮动。

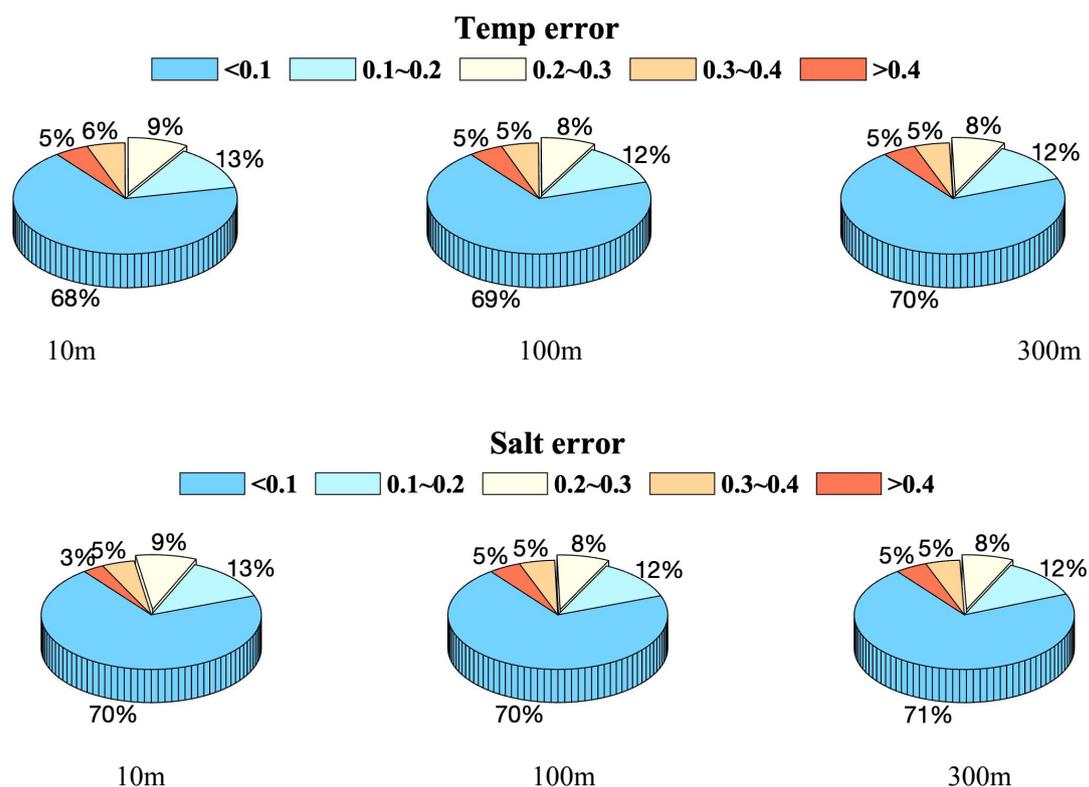


图 4.3 2024.01.01-2025.06.30 期间平均的温、盐度相对分析误差统计。

4.2 均方根误差

均方根误差 (Root Mean Square Error, 简称 RMSE 或 RMS) 常用来衡量观测值同真值之间的偏差。这里选取未参与融合的实测数据特例 (如 Argo 观测) 作为真值, 统计了融合分析产品 EFGO 的温度、盐度均方根误差, 并与 GLORYS12、和 HYCOM 两种数据进行均方根误差对比。

由均方根误差不同深度的时间序列可见, 三种数据与 Argo 观测的温度均方根误差大多小于 1.2°C (图 4.4)。GLORYS12 和 EFGO 两种数据的温度 RMSE 高值($>1^{\circ}\text{C}$)主要集中在 200m 以浅, 且两者的温度均方根误差随着时间的变化规律十分相似: 2024 年 5 月下旬至 6 月中旬, 1000m 以内的温度均方根误差整体

较大，除此之外，多数时间的温度 RMSE 仅在 150m 左右的跃层区出现，200m 以下温度均方根误差多小于 1°C ，且随着深度的增加逐渐减小，500m 以下均小于 0.5°C 。同时间同水层相比，EFGO 的温度 RMSE 比 GLORYS12 小 $0.1\text{-}0.5^{\circ}\text{C}$ 。而 HYCOM 的结果中，400m 以内均有温度均方根误差大于 $>1^{\circ}\text{C}$ 的值，600m 以浅温度 RMSE 约为 $0.5\text{-}2^{\circ}\text{C}$ ，且其在 2024 年 5 月下旬至 6 月中旬期间的整体温度误差也较大，这应该与该时段的观测数据相关。

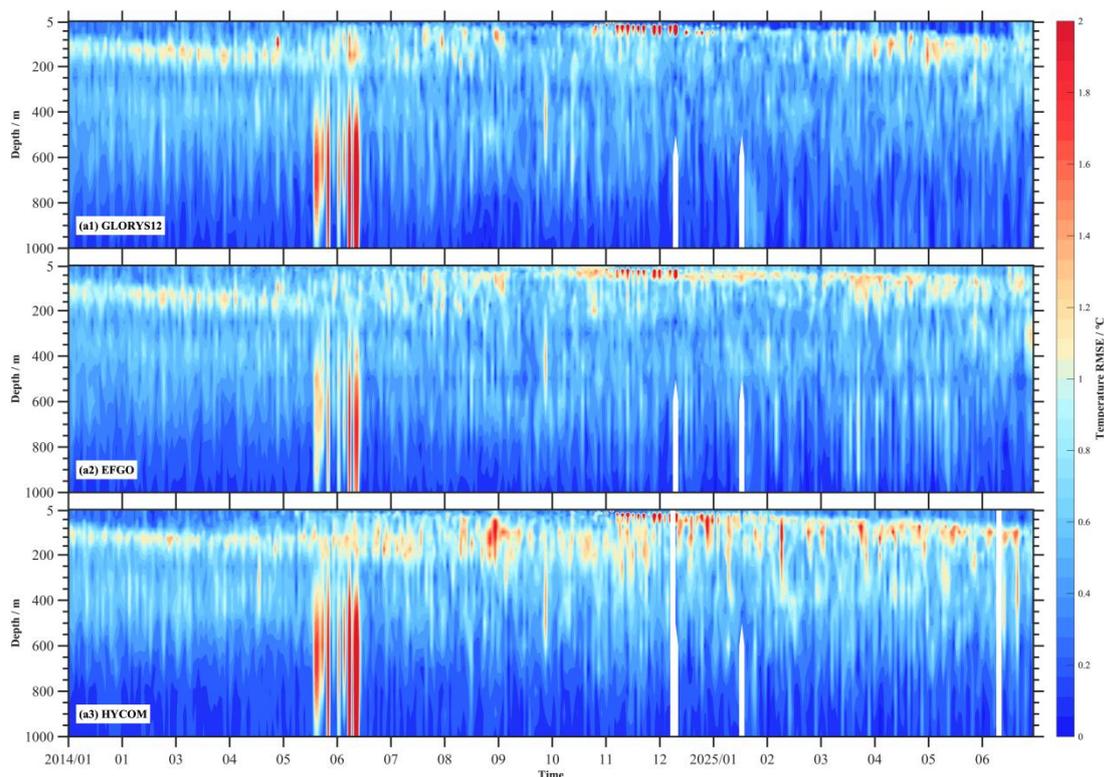


图 4.4 不同数据产品相对于 Argo 观测资料的温度均方根误差随时间和深度的分布。

三种数据的盐度均方根误差随着深度和时间的分布特征较为类似 (图 5.5)，盐度均方根误差高值 (>0.15) 主要分布在 100m 以浅，且随着时间推移，表层盐度 RMSE 逐渐增大，而后又有所减小。2024 年 11 月之后，盐度均方根误差高值影响深度有所加深，200m 深度处的有部分盐度 RMSE 超过 0.15。200m 以下，盐度均方根误差多小于 0.05。其中，EFGO 误差范围最小、GLORYS12 其次、HYCOM 误差范围最大；GLORYS12 近表层自 2024 年 9 月开始，盐度误差明显增大，一直到 2025 年 3 月，其盐度误差均大于 0.15，而 EFGO.ST 的结果则在 2024 年 11 月开始增大，2025 年 3 月以后显著降低。HYCOM 则从 2024 年 7 月开始，近表

层盐度误差就大于 0.15，一直持续到 2025 年 5 月。100m 以深，各个数据的盐度 RMSE 多小于 0.1，EFGO 的盐度误差相对其他数据更低，EFGO 在 200m 以下盐度误差多在 0.05 以下，HYCOM 则仍有部分盐度误差超过 0.1。

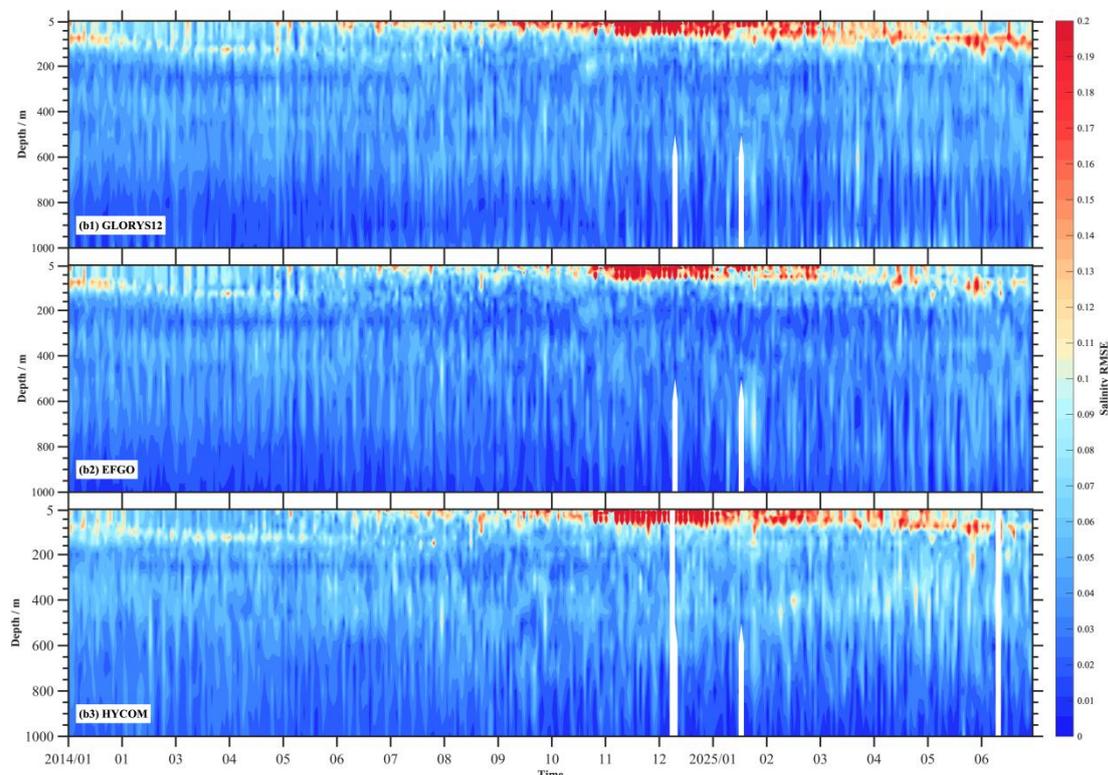


图 4.5 不同数据产品相对于 Argo 观测资料的盐度均方根误差随时间和深度的分布。

图 4.6 给出了温、盐度均方根误差的空间分布差异。对于温度均方根，日本岛周边海域的观测点上，三种数据结果均存在密集的高值误差($>1.2^{\circ}\text{C}$)分布，其他海域 EFGO 的温度 RMSE 明显低于 GLORYS12 和 HYCOM 两种数据，EFGO 温度均方根误差大多在 0.75°C 以下，GLORYS12 略高多为 $0.15\text{-}0.9^{\circ}\text{C}$ 之间，而 HYCOM 则存在大量高值误差 ($>0.45^{\circ}\text{C}$)。对于盐度均方根误差，三种数据都在台湾岛东部以及日本岛周边海域存在误差高值(>0.1)，远海区域盐度误差相对较小。其中，HYCOM 误差最大，大部分观测点上的均方根误差超过了 0.12；GLORYS12 与 EFGO 的盐度误差相对小很多，基本都在 0.1 以下，而 EFGO 比 GLORYS12 更小，其大多数盐度均方根误差均小于 0.04。

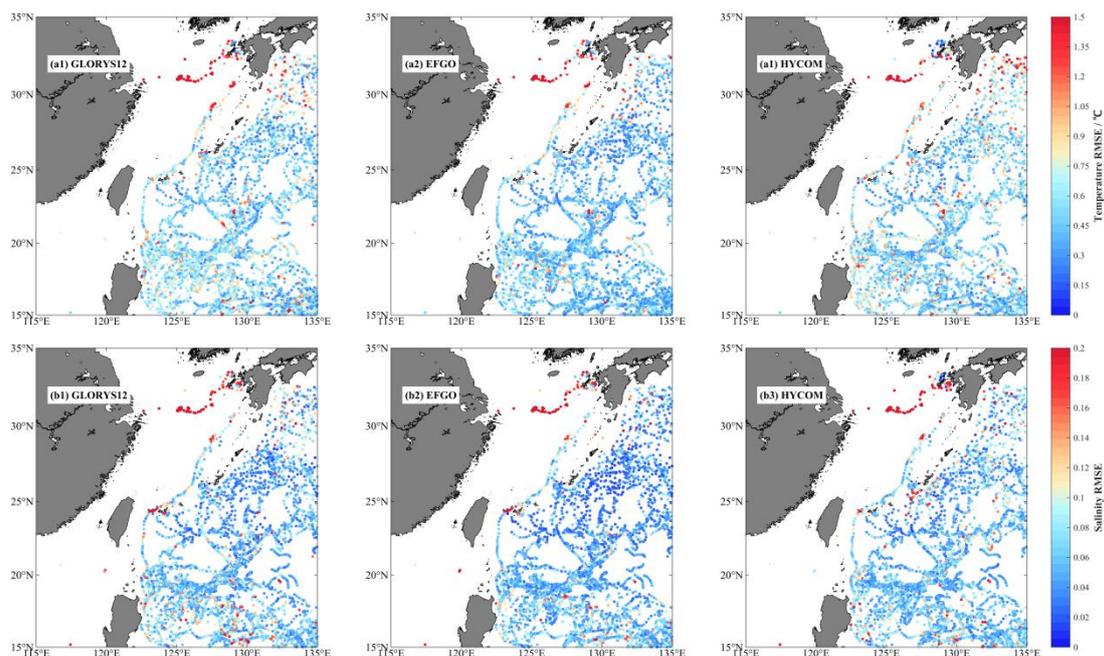


图 4.6 不同数据产品 1000m 深度内相对于 Argo 观测资料的平均均方根误差散点分布。

4.3 谱分析

为检验融合分析结果提取海洋高频信号的能力，基于快速傅里叶变换、理想高通滤波，对 GLORYS12、EFGO、HYCOM 三种温、盐度距平数据中保留的中小尺度信号进行了提取对比。图 4.7-4.8 以 2025 年 5 月 7、12、17、22 日，10m 层的温、盐度距平场为例，将其经过傅利叶变换提取频域信息后，再以滤波器去除大于 40 个步长的低频率信号，进而通过反傅利叶变化得到温、盐度距平场。距平等值线的空间连续性与温盐变化频率成反比。

如图 4.7 所示，各个数据产品温度谱分析结果随着时间变化，其温度距平的空间分布特征类似：南海、东海、以及 130°E 以东 20-30°N 之间的海域，以负温度距平为主，而 120-130°E 之间的开阔海域正温度距平为主导。但三种数据产品的温度距平反映的温度变化频率存在较大差异。GLORYS12 等值线比较平滑，温度的正负异常极值也相对最大，更多地体现了大尺度环流模式下的温度变化，EFGO 在中尺度过程的特征提取上更精细，HYCOM 则在边界处（包括海陆边界和动力过程剪切边界）呈现出较多的动力学变化特征。具体而言，由于 HYCOM 时间分辨率为 3 小时，其温度距平在空间上体现了较多的小尺度信息的变化，但其正、负温度距平的空间分布相对集中，而 GLORYS12 与 EFGO 则多为正、负温度距平交替分布的空间特征。并且，GLORYS12 与 EFGO 在 130°E 以东的负距平海域，以及 20°N 以南的正距平结构变化存在明显的空间分布差异：5 月 7

日 GLORYS12 负距平海域呈单核分布、EFGO 则有着上下相邻的双核特征，正距平海域两者均呈现自西向东的三处高距平中心，EFGO 高距平中心范围相对较小，精细结构更充分；5 月 12 日-5 月 22 日，GLORYS12 正距平向北延伸，而 EFGO 正距平被东侧负距平拦截于 20°N 以南，在此期间，EFGO 展现了更多的中小尺度正负距平分布，其中，5 月 22 日更为明显，130°E、23°N 附近 GLORYS12 存在大范围的正负距平相邻结构，而 EFGO 则是北部多个负距平结构围绕南部正距平中心。

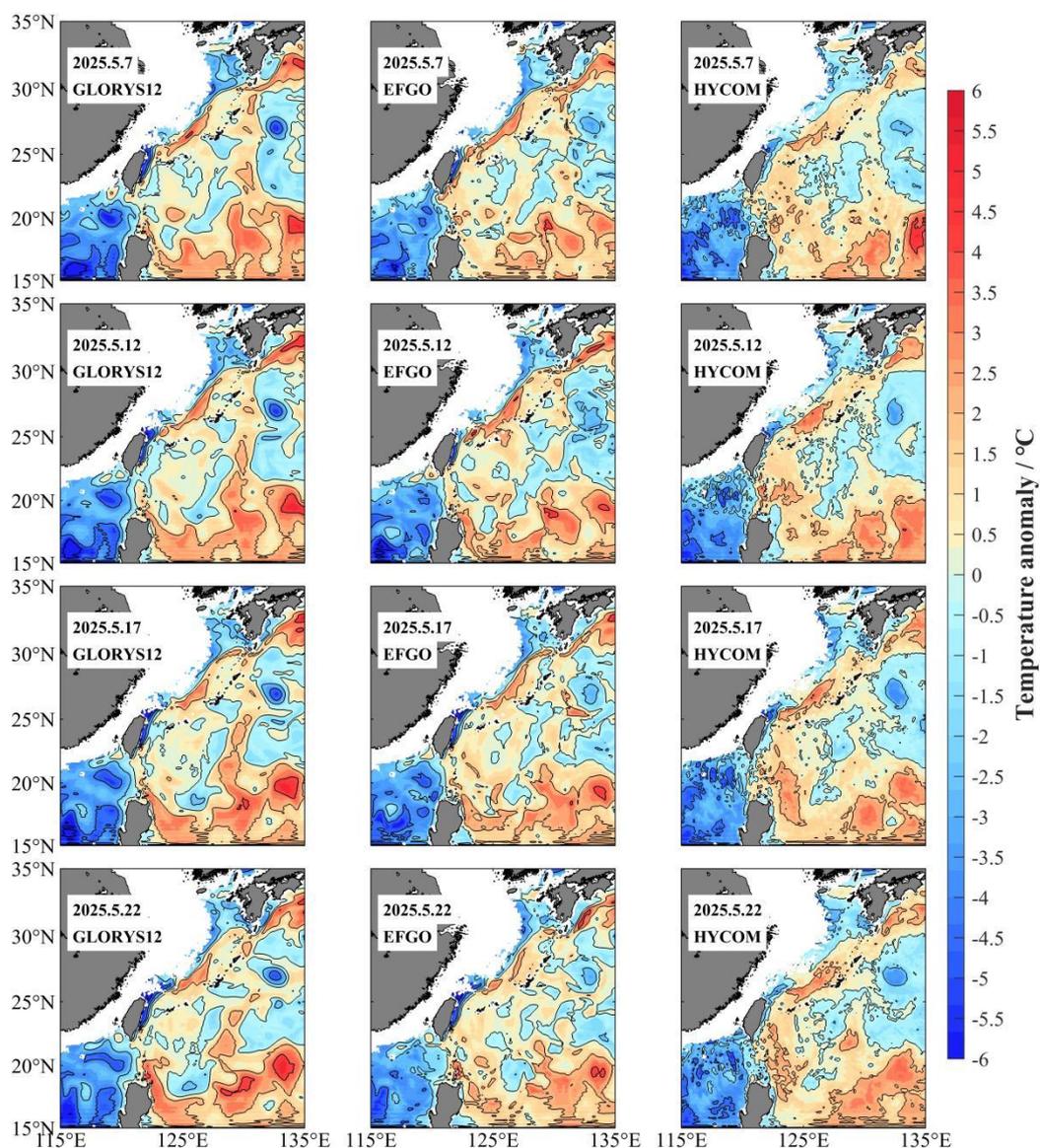


图 4.7 不同数据产品的温度二维谱分析结果对比。

盐度整体呈现西部负距平、东部正距平的分布趋势。与温度类似，GLORYS12 反映的大、中尺度盐度变化较为明显，EFGO 反映了较多地中尺度变化特征，而

HYCOM 盐度距平中存在较多地小尺度变化信息,且这些小尺度信息在边界出比较明显,开阔水域的盐度距平相对均匀。相较之下,EFGO 比 GLORYS12 存在更多的盐度精细变化,如 5 月 7 日在 133°E、27°N 的负距平中心结构附近,EFGO 结构有着明显的正距平结构,5 月 12 日台湾岛东岸 EFGO 结果存在负距平包含正距平中心结构,5 月 22 日在 125°E、17°N 处 EFGO 展现了细致的负距平中心结构。

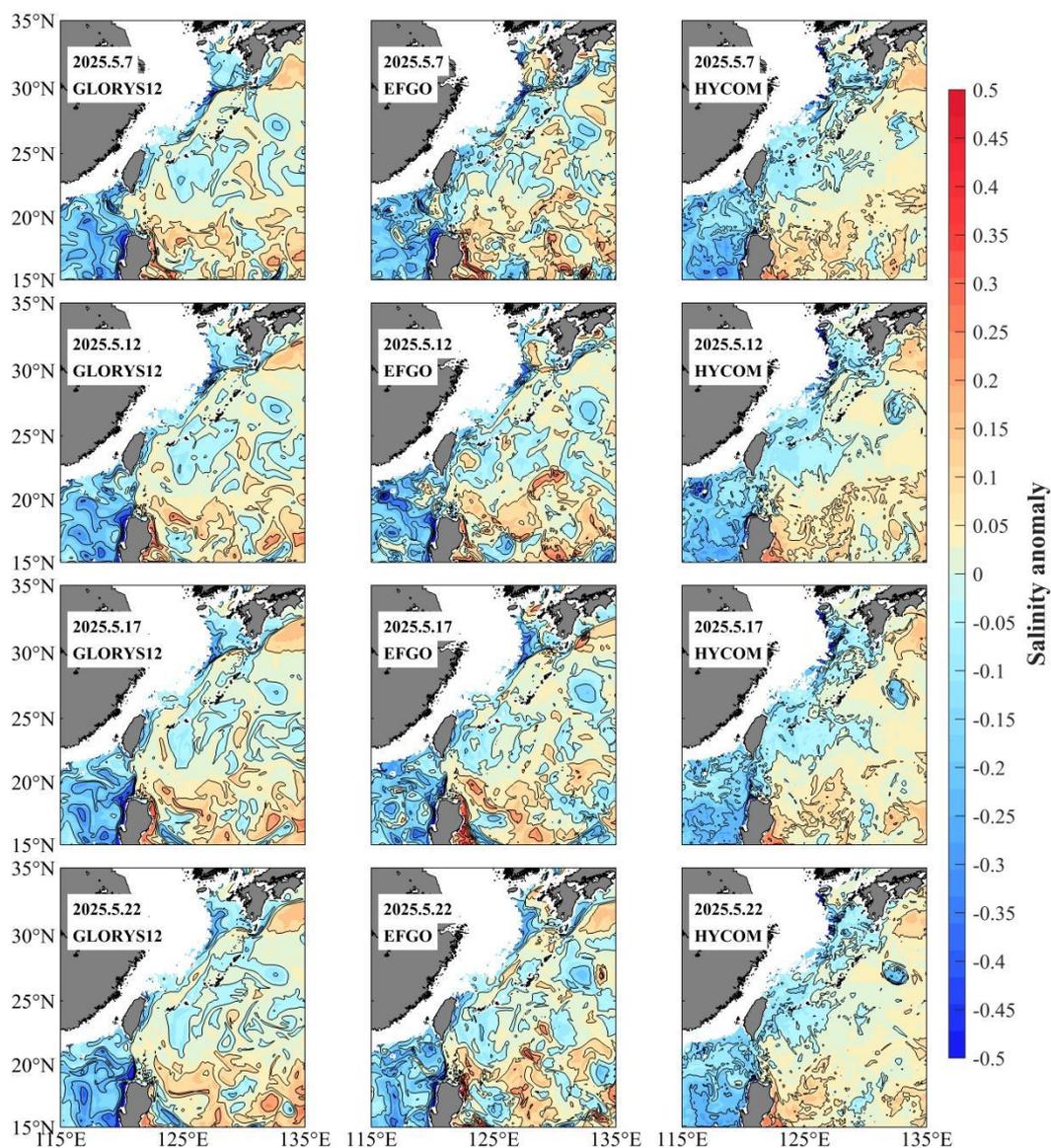


图 4.8 不同数据产品的盐度二维谱分析结果对比。

4.4 实测剖面对比

综合考虑观测站位的区域分布、观测时间、观测深度,以及温度和盐度观测剖面的完整性,在 2024 年 01 月 01 日至 2025 年 06 月 30 日期间,每间隔一个月,

选取 15 日前、后三天内（根据站位位置调整时间）观测深度大于 500m 且温度和盐度剖面均相对较完整的 Argo 观测资料，对比 EFGO、GLORYS12、HYCOM 等三种数据的温、盐度剖面。

表 4.1 用于对比验证的 Argo 剖面观测信息表

站位	P01	P02	P03	P04	P05	P06	P07	P08	P09	P10
经度 /°E	123.798	124.053	131.246	129.210	126.583	128.065	130.568	127.091	126.416	128.763
纬度 /°N	20.527	17.201	24.928	23.903	16.550	21.677	17.266	22.218	17.850	19.717
观测 时间	2024/ 01/18	2024/ 03/15	2024/ 05/15	2024/ 07/15	2024/ 09/14	2024/ 11/15	2025/ 01/16	2025/ 03/15	2025/ 05/15	2025/ 06/13
实测 深度	575m	1980m	2015m	2009m	1999m	1998m	1992m	1948m	1999m	2027m

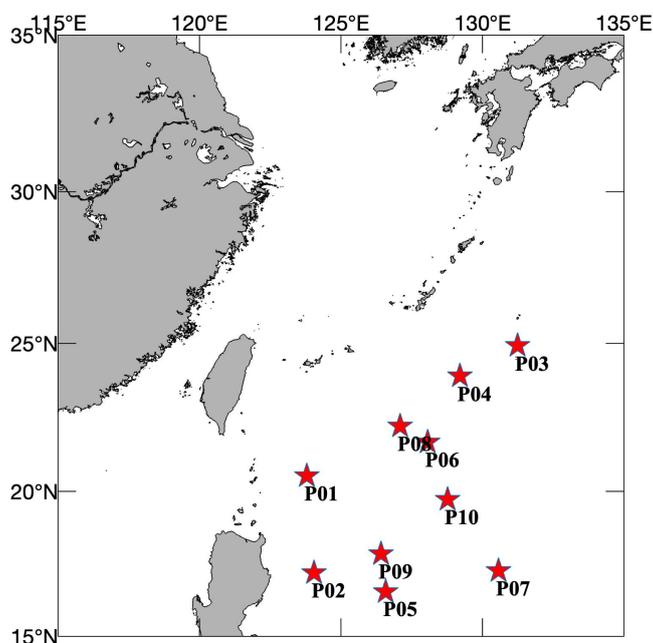


图 4.9 用于实测对比的 Argo 剖面观测站位。

对于温度剖面，EFGO、GLORYS12、HYCOM 三种数据产品与观测剖面均具有较好的一致性，各数据在 400m 以深与观测数据的温度偏差多小于 0.5°C，400m 以上尤其是温度变化剧烈的 200m 左右，各数据的温度偏差相对较大，最大温度偏差不超过±2°C。相比之下，除 P09 站位之外，其他 9 个剖面上 EFGO 与观测剖面温度偏差都最小，GLORYS12 其次，HYCOM 最大。P09 剖面上，三种数据与 Argo 观测剖面均十分接近（最大温度偏差小于 1°C），相较而言，HYCOM 的温度偏差较小，但 EFGO 与 GLORYS12 的偏差垂向变化更一致。

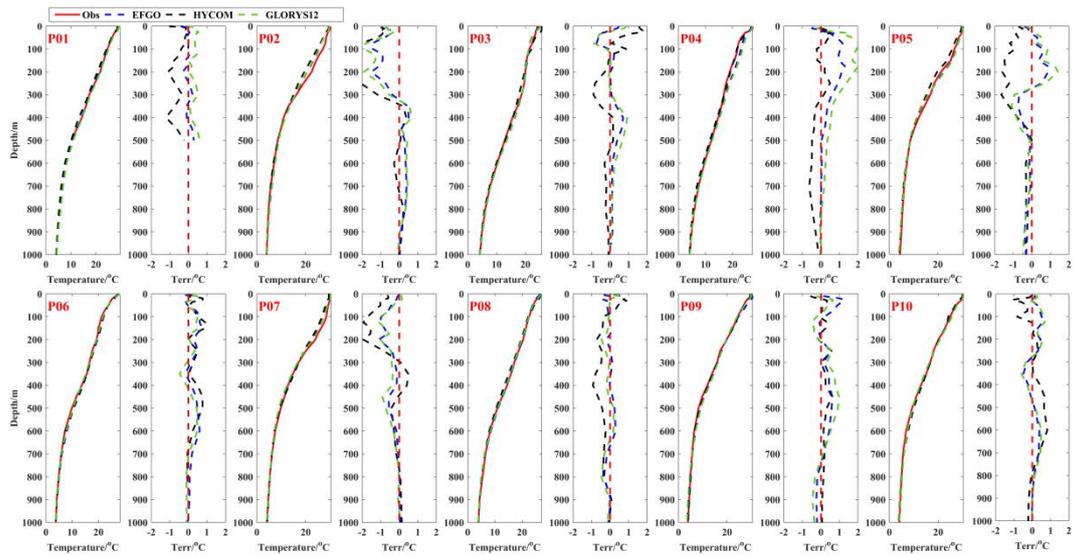


图 4.10 各站位不同数据产品与 Argo 剖面观测的温度剖面曲线。

盐度剖面的对比结果相对较复杂，随着深度的增加，盐度呈现先增加后减小又增大的变化趋势，各数据均与观测剖面呈现同样趋势，但不同深度各数据与观测剖面偏差有所不同。P01、P04、P05、P08、P10 几个剖面上，EFGO 与 Argo 观测剖面在 1000m 以内均有非常好的一致性，对应的盐度偏差小于 ± 0.05 ，这几个剖面上，GLORYS12 与观测盐度偏差约为 $-0.05-0.1$ ，而 HYCOM 的盐度偏差则约为 ± 0.1 ；P02、P03、P06 剖面上，100-600m 深度范围内，EFGO 与观测值比较接近，近表层和 800 以深的盐度偏差稍大 ($>\pm 0.1$)，GLORYS12 的结果与 EFGO 类似，HYCOM 在这几个剖面上同水层的盐度偏差均大于 EFGO 和 GLORYS12；P07 站位上，几种产品与观测的盐度差异主要在 100m 以浅的近表层，EFGO 与 GLORYS12 最大盐度偏差均约为 ± 0.2 ，HYCOM 的最大盐度偏差大于 0.2；P09 站位上，几种数据与观测的最大盐度偏差集中在 200-500m 之间，相比之下，EFGO 最小 ($<\pm 0.1$)，GLORYS12 其次，HYCOM 最大。

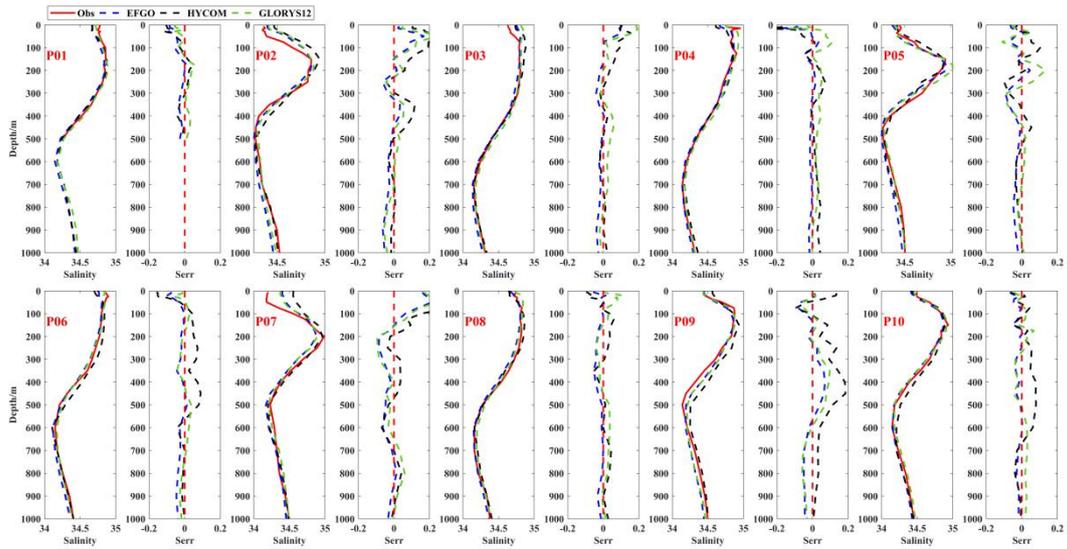


图 4.11 各站位不同数据产品与 Argo 剖面观测的盐度剖面曲线。

图 4.12 选取一个 WOD 经向断面 (S01) 和一个 Argo 纬向断面 (S02)，对比了融合重构涡旋和观测数据 (剔除对比观测断面) (EFGO+Obs)、只融合重构涡旋 (EFGO)、GLORYS12、HYCOM 几种结果的异同点。

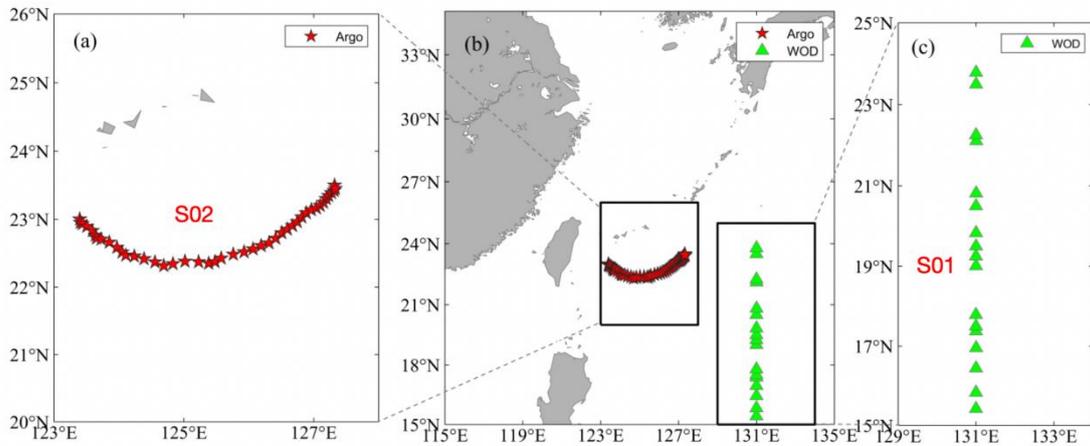


图 4.12 S01、S02 观测断面站位分布。

由 WOD 温度断面分布可见，几种数据结果的温度随着深度和纬度的特征变化规律均与观测断面很相似。但融合多源数据的结果 EFGO+Obs 与观测断面更接近，特别是 200-500m 之间温度等值线随着纬度的沉降规律十分类似，只是融合结果相对平滑；只融合重构涡旋 EFGO 的温度断面与 GLORYS12、HYCOM 更相似，如 16°N 站点上温度等值线的上涌趋势。几种结果与 WOD 断面的温度偏差中，EFGO+Obs 更接近观测数据，整个断面的温度偏差均在 $\pm 0.5^{\circ}\text{C}$ 之内；其次是 EFGO，其大部分温度偏差在 $\pm 0.75^{\circ}\text{C}$ 之内，仅 100-200m 之间，20-21°N

两个站点上的温度偏差达 -2°C 。GLORYS12 与 EFGO 的温度偏差特征相似，但有不少温度偏差超过 $\pm 1.0^{\circ}\text{C}$ 。HYCOM 的温度在 400m 以浅明显低于观测值，大多温度偏差达 -1.5°C 以上。

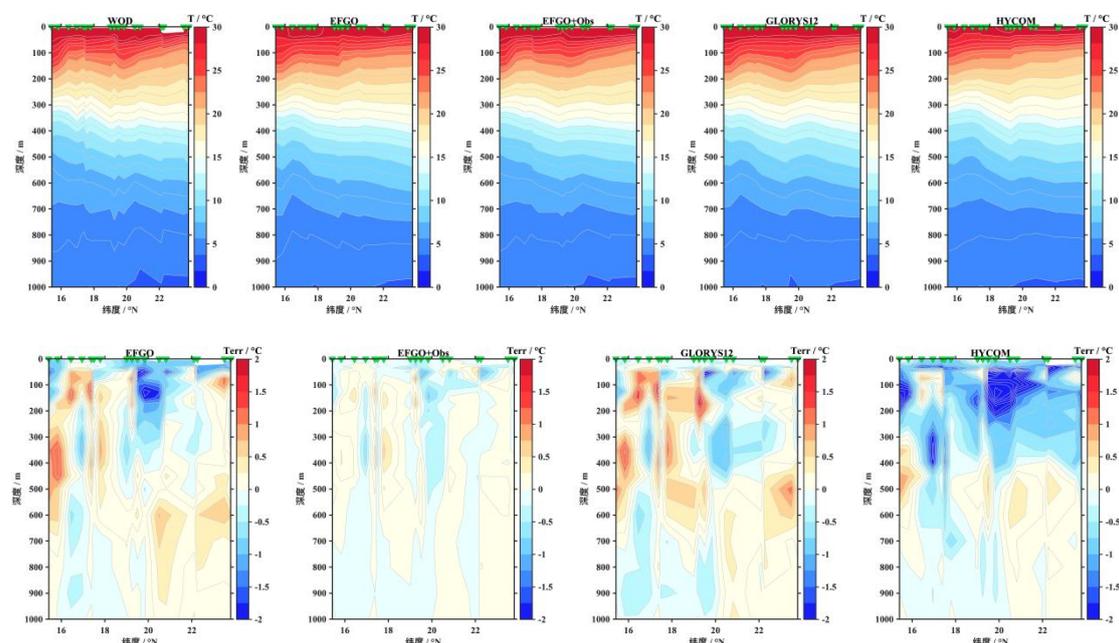


图 4.13 WOD 观测断面 (S01) 上不同数据产品的经向温度断面特征及其与观测断面的温度偏差分布。绿色三角为观测位置。

与温度类似, 几种数据结果在 S01 观测断面上的盐度特征随着深度和纬度的变化规律均与观测断面很相似。100-200m 之前存在一个高盐中心, 400-800m 之间自北向南存在显著的低盐水入侵现象。相比之下, EFGO+Obs 与观测断面最接近, EFGO 的盐度断面与 GLORYS12 较为相似, HYCOM 近表层的高盐中心的南-北向分布特征与观测断面有明显差异。从盐度偏差也可以看出, 多源融合结果 EFGO+Obs 的盐度偏差最小 (约为 ± 0.05), 其次是融合涡旋的结果 (EFGO), 其大多盐度偏差 $-0.05-0.1$ 之间, 仅 100m 以内, 18-20°N 四个站点上盐度偏差小于 -0.15 。而 GLORYS12 与观测断面的盐度偏差大多在 ± 0.1 之间, HYCOM 的盐度偏差则多超过 ± 0.15 。

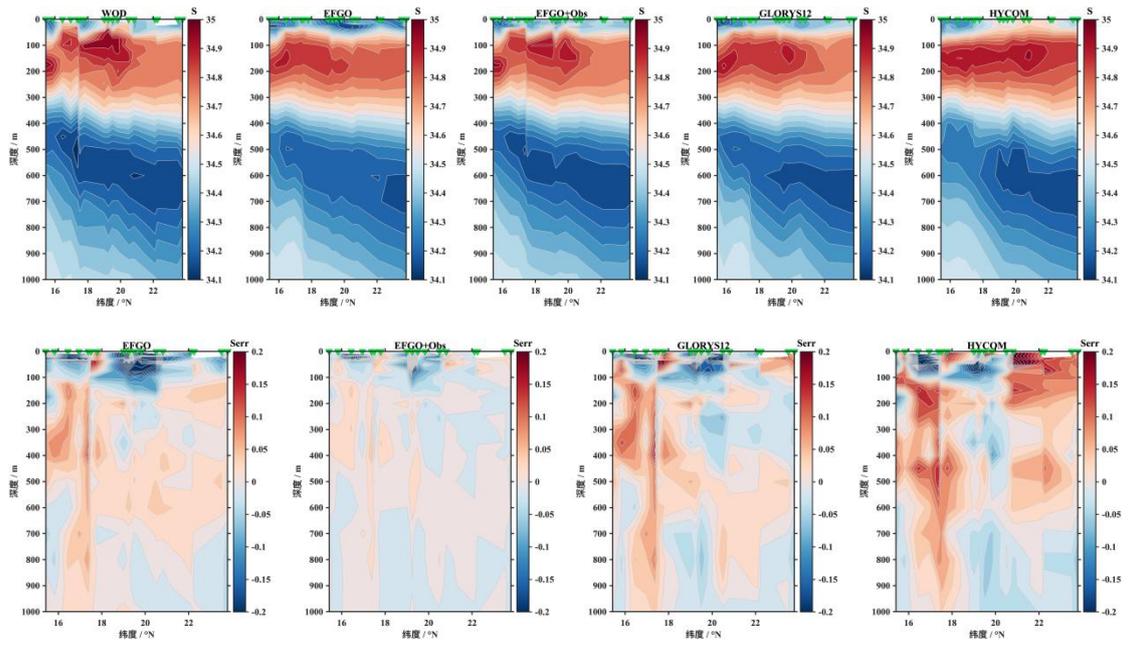
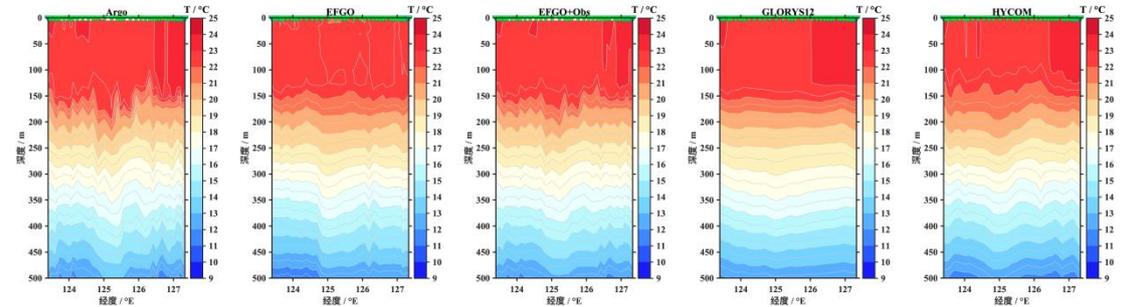


图 4.14 WOD 观测断面 (S01) 上不同数据产品的经向盐度断面特征及其与观测断面的温度偏差分布。绿色三角为观测位置。

Argo 观测断面上，多源融合结果 EFGO+Obs 的温度等值线随经度的沉降波动特征与观测断面极为相似，只融合涡旋的结果 EFGO 相对稍显平滑，HYCOM 与 EFGO 类似，而 GLORYS12 的温度等值线最为平滑，温度东-西向差异不大。对应的，多源融合结果与 Argo 观测断面的温度偏差也最小，整个断面的温度偏差均在 $\pm 0.5^{\circ}\text{C}$ 之内，其次是 EFGO 结果，其最大温度偏差约为 $\pm 0.75^{\circ}\text{C}$ ，GLORYS12 与 Argo 观测的温度偏差则在 $\pm 0.75^{\circ}\text{C}$ 之间，且两种融合结果与 GLORYS12 的温度偏差分布特征较为一致。而 HYCOM 多低于 Argo 观测的温度值，大多温度偏差在 $-1.5-0.25^{\circ}\text{C}$ 之间。



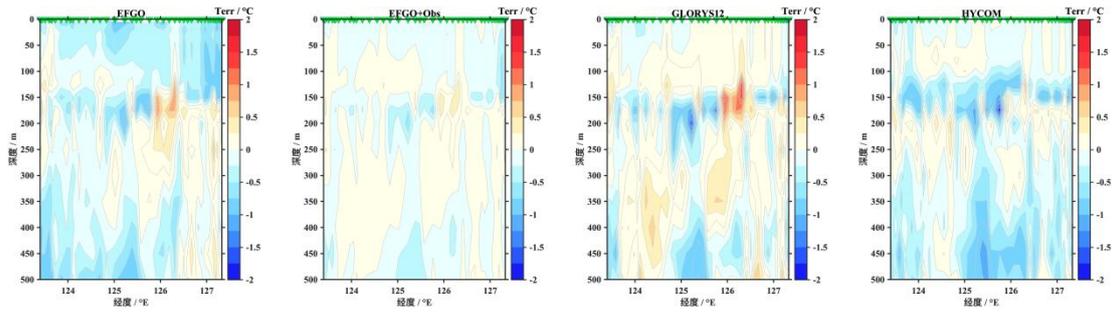


图 4.15 Argo 观测断面 (S02) 上不同数据产品的经向温度断面特征及其与观测断面的温度偏差分布。绿色三角为观测位置。

对于 Argo 观测断面上的盐度，EFGO+Obs 的盐度变化规律与观测断面最为相似：200m 以内盐度相对均匀（约为 34.7-35.0），500m 以内盐度随深度增加逐渐减小，自西向东，盐度呈现不同程度的波动 125.5°E 处的站点上的盐度等值线下沉显著。EFGO 的结果 250m 以下与观测断面也较为相似，但 200m 以内的盐度相对偏低，相比之下，GLORYS12 在该深度内的盐度比观测偏高，整体等值线较平滑，东西盐度差异不大，而 HYCOM 的盐度等值线波动与观测也较为一致，但对应的盐度偏差在-0.15-0.1 之间，GLORYS12 的盐度偏差则多处于 ± 0.1 之间，EFGO 的盐度偏差也集中在 ± 0.1 之间，多为负值，多源融合结果 EFGO+Obs 的盐度偏差均在 ± 0.05 之间。

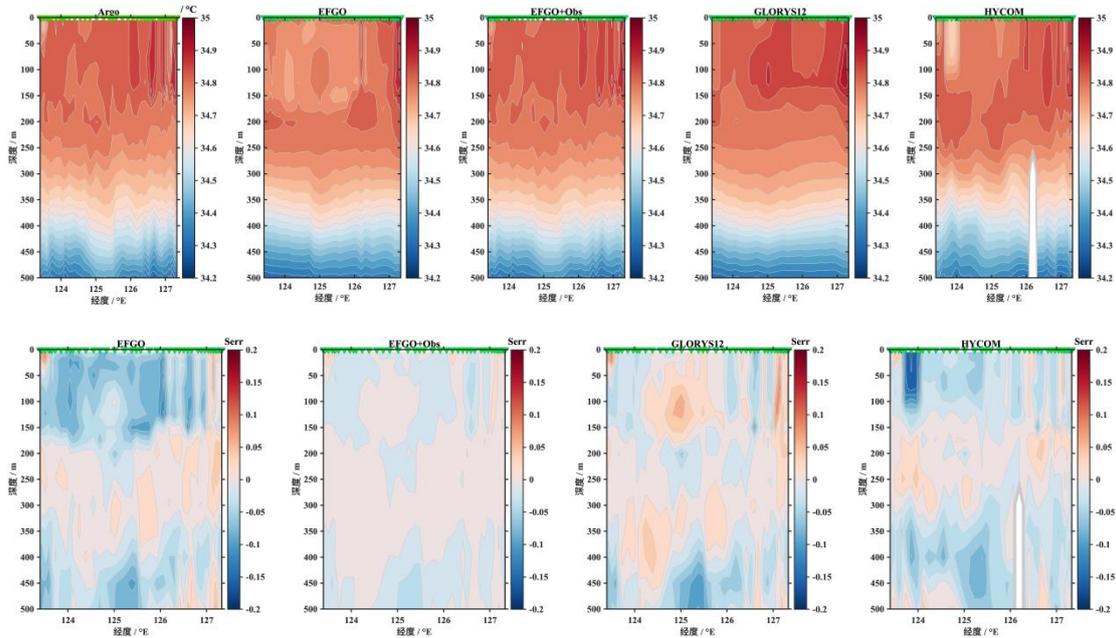


图 4.16 Argo 观测断面 (S02) 上不同数据产品的经向盐度断面特征及其与观测断面的盐度偏差分布。绿色三角为观测位置。

4.5 涡旋结构对比

以台湾东部的暖涡 (122°E-125°E, 22°N-24°N)、131°E-134.5°E, 25.5°N-28.5°N 之间的冷涡为例 (图 4.17), 选取了 2025 年 5 月 9 日两种不同类型的典型中尺度涡的温、盐度三维结构, 给出了融合数据 (EFGO)、GLORYS12、HYCOM 三种数据呈现的中尺度涡内部结构特征的异同。

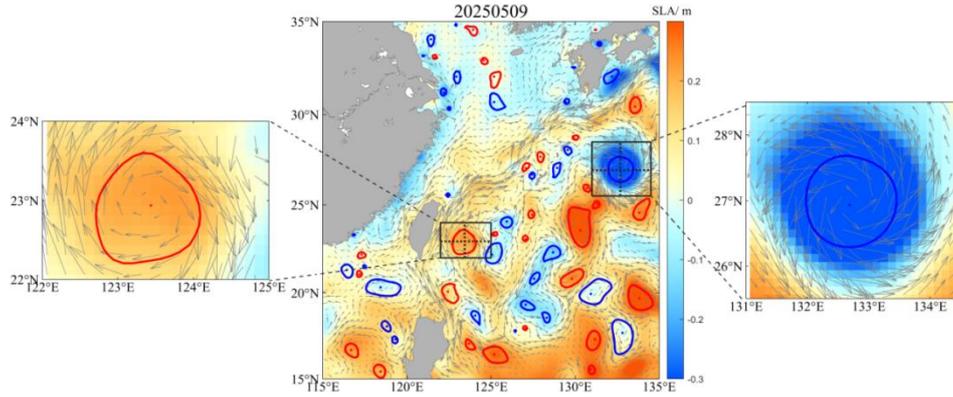
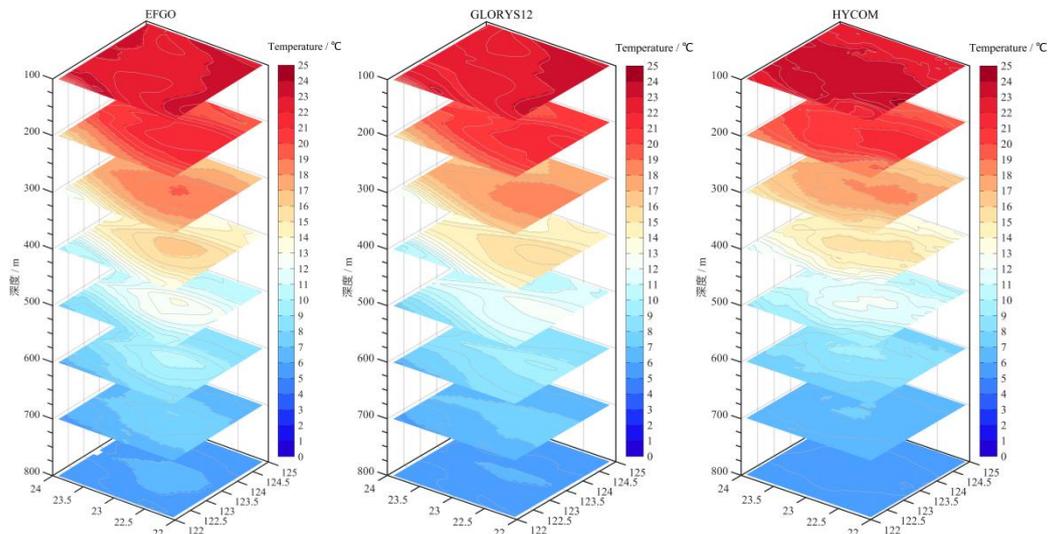


图 4.17 2025 年 5 月 9 日卫星观测的海表面高度、海表流场及对应中尺度涡分布。

如图 4.18 所示, 典型暖涡区域, 无论是温度还是盐度, 三种数据的结果均体现了中尺度涡内部的水团均一性特征。但其展示的水团强度具有明显差异。相比之下, 融合数据 (EFGO) 的温度和盐度等值线更为闭合, 水团边界更为显著, 且其展示的涡心处的高温中心和低盐中心与表层卫星观测结果高度相符; GLORYS12 的温盐度结果在涡心处具有较强的均一性, 但等值线大体与经线平行, 且涡旋内、外的温盐梯度较小, 水团特性较弱; HYCOM 的温盐等值线也呈较为明显的闭合特征, 但其内部存在明显的等值线小尺度波动, 且 600m 以下, 整个区域的温度梯度显著减小, 涡旋特征不明显。



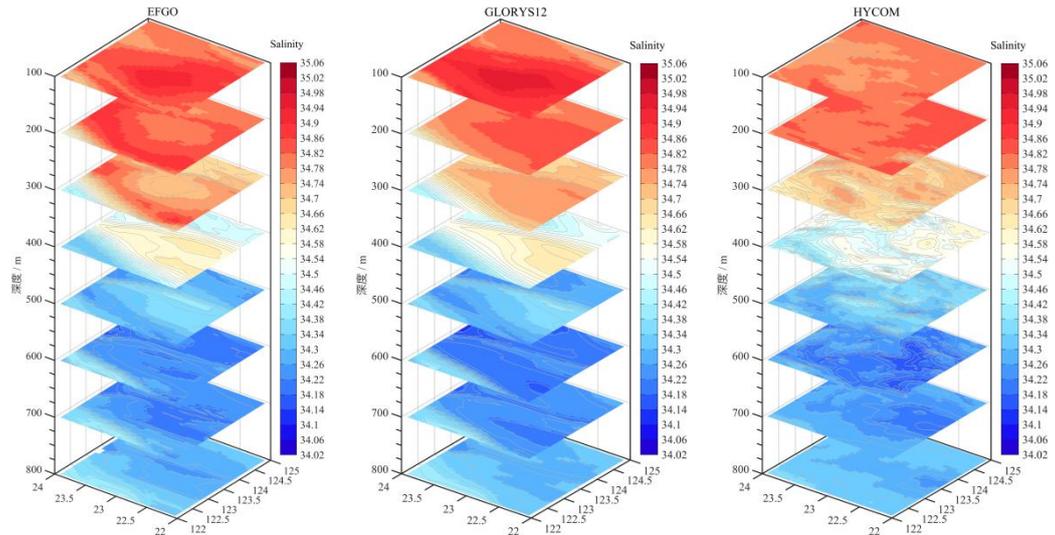
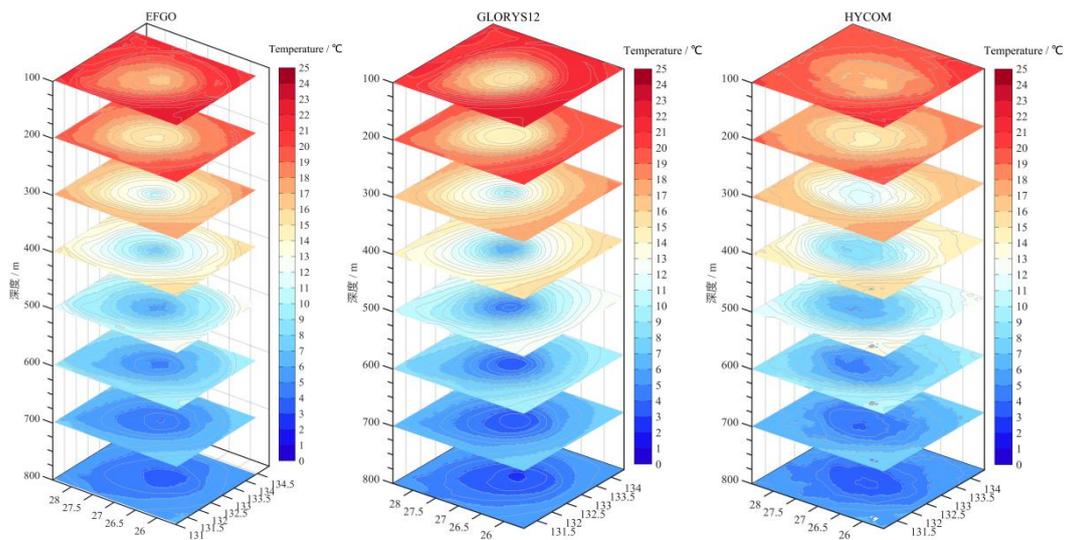


图 4.18 2025 年 5 月 9 日典型暖涡（反气旋涡）区域，EFGO、GLORYS12、HYCOM 数据三维温、盐结构。

典型冷涡区域（图 4.19），无论是温度还是盐度，三种数据的结果也都较好地展示了冷涡内部低温、低盐的水团特性，且三种数据的温盐度等值线都显著闭合，涡旋边界明显。相比之下，GLORYS12 的温盐度结果在涡心处具有较强的均一性，等值线更为规则和平滑；则与暖涡区类似，HYCOM 的冷涡内部温盐等值线存在明显的小尺度波动；无论是温度还是盐度，EFGO 的结果综合呈现了 GLORYS12 和 HYCOM 的分布特征，温度和盐度等值线具有明显的规则闭合特征，且其展示的涡旋内部温度和盐度变化比 GLORYS12 更细致，如，200-500m 深度涡心处的温盐度梯度更明显，即，其能够更充分地呈现中小尺度特征。



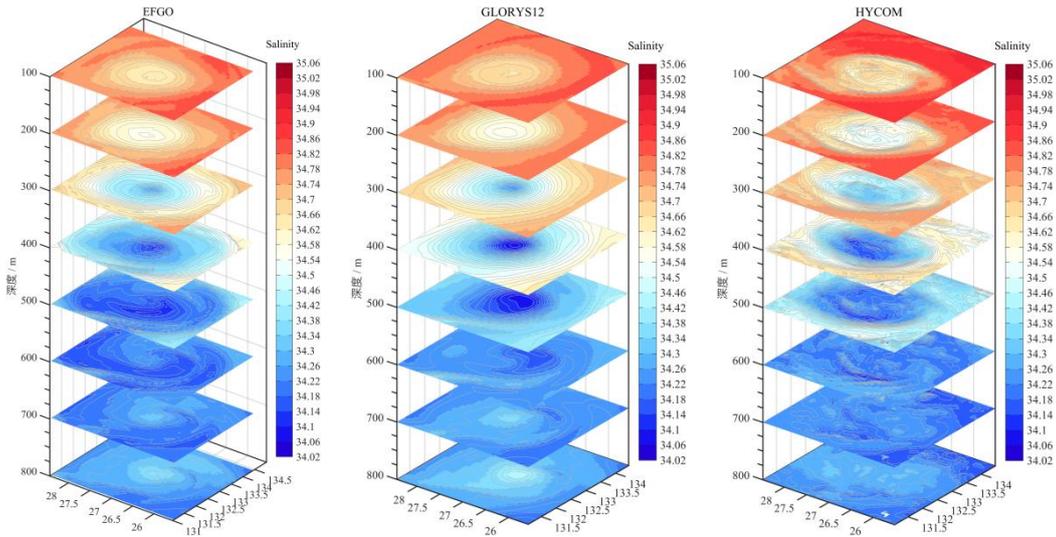


图 4.19 2025 年 05 月 09 日典型冷涡（气旋涡）区域，EFGO、GLORYS12、HYCOM 数据的三维温、盐结构。

五、EFGO.ST 数据产品总结

高分辨率涡旋建模融合数据集 (EFGO.ST) 是基于观测增强的中尺度涡三维结构重构模型以及梯度依赖最优插值客观分析方法，融合多源观测数据构建的区域高分辨率温盐数据集。

第一版于2025年12月发布：采用各项异性的梯度依赖最优插值方法，融合了涡旋重构剖面、以及卫星、Argo、历史观测等多源数据，实现了不同时空尺度海洋信息的有机融合。进而形成2024年1月1日至2025年6月30日期间，15°N~35°N，115°E~135°E海域，天平均、1/16°×1/16°、多参数变量的融合资料集。每天的数据文件除了包含温度、盐度、密度流、跃层参数、声层深度等主要水文要素外，还涵括该海域每天的涡旋信息，包括涡心位置、涡旋边界、涡旋极性、涡旋内部重构的温盐剖面等，共计28个水文变量，以NetCDF格式存储。该数据集每天重构的涡旋个数多大于40个，融合的涡旋区重构剖面超过10,000条/天，涡旋内部融合分析误差均小于0.05，边缘区分析误差约为0.2-0.5；相对于实测资料，涡旋区的温、盐度均方根误差分别为0.45°C和0.05，非涡区则分别约为0.3-1.2°C、0.04-0.15。

第二部分 EFGO.SS 数据分册

目录

一、EFGO.SS 数据集基本信息	28
二、EFGO.SS 融合数据说明	29
三、声速融合建模简介	31
3.1 声速背景场重构	31
3.2 涡旋内声速场重构	33
3.3 多尺度声速场变分融合	34
四、EFGO.SS 数据集效果检验	35
4.1 涡旋声速场重构效果检验	35
4.2 融合数据集效果检验	38
五、EFGO.SS 数据产品总结	39

一、EFGO.SS 数据集基本信息

研究团队基于多源卫星遥感资料和 Argo 实测数据，结合机器学习算法和中尺度涡统一结构动力模型，提出了面向中尺度涡的水下三维声速场重构模型，通过输入海表面温度、高度异常等海表环境信息和涡旋特征参数，快速重构中尺度涡内三维声速场；随后将中尺度涡声速与背景场声速进行多尺度变分融合，构建了一套高分辨率涡旋建模融合声速剖面数据集 (EFGO.SS)，突破了传统方式只能基于温盐场数据通过声速经验公式计算得到声速剖面诊断产品、计算误差较大的局限。

表 1.1 EFGO.SS 数据集基本信息

数据集名称	EFGO.SS	
水平空间范围	15°N~35°N, 115°E~135°	
垂直深度范围	0 ~ 2000m	
时间范围	2023 年 1 月 1 日至 2023 年 12 月 31 日	
空间分辨率	1/8°	
垂直分辨率	20m	
时间分辨率	逐日	
数据格式	EFGO.SS_YYYYMMDD.nc	
环境要素融合分析变量	sound speed (海洋声速)	
nc 文件存储变量	变量名 (全称)	要素 (维数)
	longitude (Longitude)	经度 (160×1)
	latitude (Latitude)	纬度 (160×1)
	time (Time)	时间: yyyyymmdd (1×1)
	depth (Sea water depth)	垂向分层 (101×1)
	sound speed (Sea water sound speed)	声速场 (1×101×160× 160)

二、EFGO.SS 融合数据说明

EFGO.SS 数据集的构建综合采用了海表卫星遥感数据、气候态数据和多源剖面观测数据。核心数据包括来自 Argo、GTSP、WOD 等数据库的多源剖面观测数据, 这些温盐剖面作为训练数据服务于大尺度声速背景场和精细化涡旋声速场的重构, 其中靠近涡心的剖面被直接用于中尺度涡旋结构的刻画。同时, 海表卫星遥感数据 (包括 SST、SLA、ADT 和 SSS 等) 为声速重构提供了关键的海表边界信息, 其 SST 与 SSS 数据也有效弥补了表层观测的缺失。此外, ARMOR3D 再分析数据作为气候态背景场, 为中尺度涡旋的识别与重构提供了重要的环境参考。上述多源数据的协同应用, 系统支撑了从背景场构建、涡旋精细重构到最终数据融合的全过程。详细数据信息请参见表 2.1。

表 2.1 融合数据信息

名称	数据源	时间分辨率	空间分辨率	使用变量	数据使用情况
Argo	http://www.argo.org.cn/	天	散点	温度、盐度	(1) 构建模型的训练集; (2) 靠近涡心的剖面用于中尺度涡旋重构
GTSP	https://www.ncei.noaa.gov/products/global-temperature-and-salinity-profile-programme	天	散点	温度、盐度	补充重构海域内实测剖面的数量 (剔除了重复剖面)
WOD	https://www.ncei.noaa.gov/products/world-ocean-database	—	散点	温度、盐度	主要选取近海 CTD、XBT、XCTD、锚系浮标(MRB)、漂流浮标观测数据
DUACS	https://marine.copernicus.eu/	天	1/8°	SLA、ADT 和地转流速	(1) 用于声速背景场和中尺度涡旋声速场重构; (2) 用于中尺度涡旋识别和涡旋特征参数的提取。
GHRSSST	https://marine.copernicus.eu/	天	1/20°	SST	(1) 用于声速背

	opernicus.eu/				景场和中尺度涡旋声速场重构; (2) 用于补充表层温度
CNR	https://marine.copernicus.eu/	天	1/8°	SSS	(1) 用于声速背景场重构; (2) 用于补充表层盐度
ARMOR 3D	https://marine.copernicus.eu/	周	1/4°	温度、盐度	用于提供中尺度涡旋密度背景场

三、声速融合建模简介

EFGO.SS 数据集基于 n 维数据变分插值分析 (Data Interpolating Variational Analysis in n Dimensions, DIVAnd) 方法, 融合了包括历史观测、重构涡旋、卫星海表观测等多源数据资料, 形成了 15° – 35° N, 115° – 135° E 海域, 2023 年 1 月 1 日至 2023 年 12 月 31 日期间, 空间分辨率为 $1/8^{\circ}$, 时间分辨率为天, 垂向 101 层的高分辨率数据产品。该数据集通过三个逻辑严密的步骤, 系统构建高精度三维声速场: 首先, 构建分区差异化的背景声速场 (3.1 节), 以刻画大尺度声速结构与区域分异特征; 进而, 采用专门算法实现中尺度涡旋的精细化声速重构 (3.2 节), 准确捕捉其高梯度变化的局部异常特征; 最终, 基于变分融合技术将背景场与涡旋场进行协同整合 (3.3 节), 生成空间连续、物理一致的全海域三维声速分布。整体技术路线如图 3.1 所示。

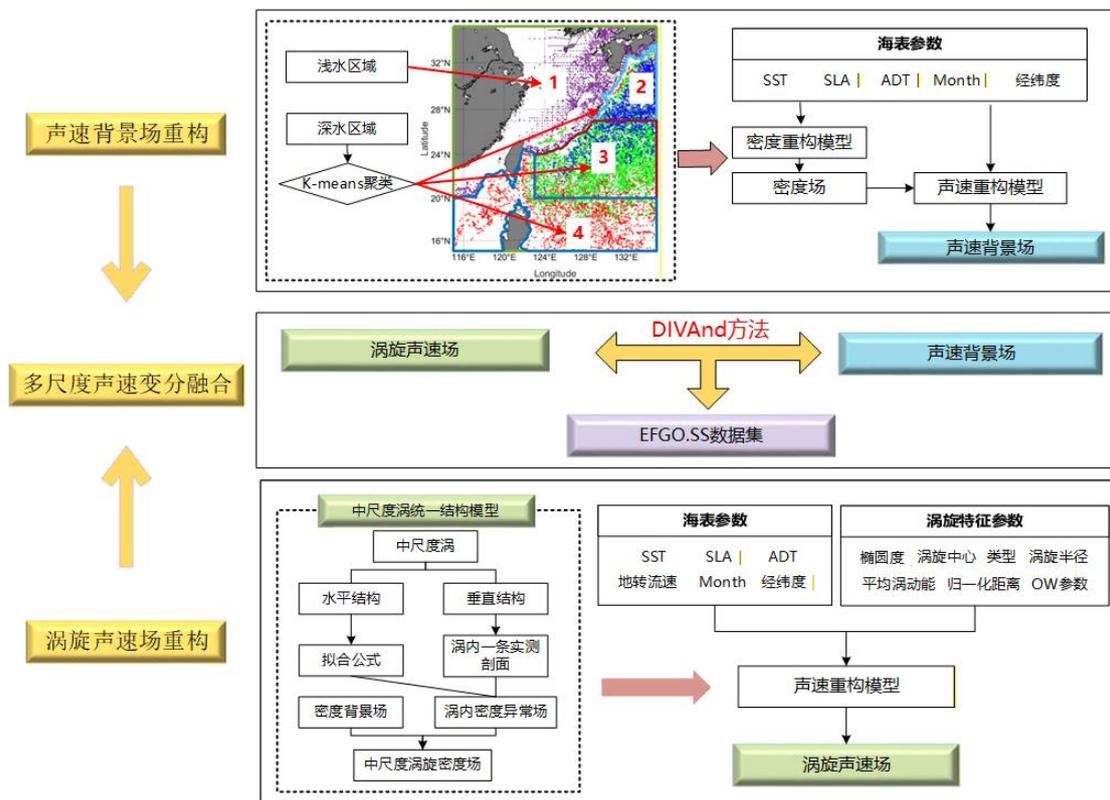


图 3.1 声速重构总体技术路线图。

3.1 声速背景场构建

为实现背景场的差异化重构, 首先依据海底地形特征, 以 1000 米等深线为界将研究区划分为浅水区与深水区。浅水区受底摩擦与地形效应影响显著, 声速结构主要受上层混合层及局地热盐变化控制, 因此采用统一建模方式。深水区水

团结构复杂、动力过程多样，声速剖面空间分异特征明显。为此，基于历史 Argo 声速剖面数据，采用 K-Means 聚类算法，依据剖面形态对深水区进行子区域划分。通过轮廓系数对聚类质量进行量化评估，确定最优聚类数目为 3 类，分别对应三种典型的深海声速剖面结构。结合浅水区声速剖面，最终将整个研究区域的声速剖面划分为 4 个类别，分类结果如图 3.2 所示。

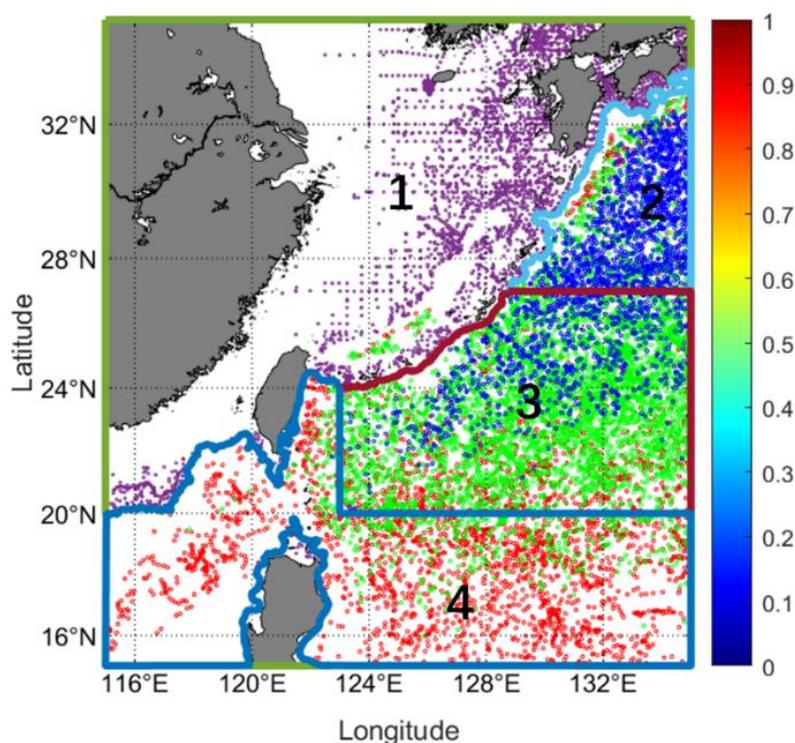


图 3.2 声速剖面分类结果及子区域划分。

依据上述分类结果，针对各子区域分别建立声速重构模型，以准确刻画不同水团与动力过程对声速结构的区域性影响。为此，本研究基于多源卫星遥感数据和 Argo 剖面数据，采用随机森林算法，首先建立海表遥感信息（SST、SLA、ADT 和 SSS）、地理时间信息（Longitude、Latitude 和 Month）与密度之间的统计映射关系，重构得到三维密度场；进而在海表环境和地理时间信息基础上，创新性地增加水下密度作为声速计算的输入，提出一种新的三维声速场重构模型（RF-SDR）。具体的技术路线如图 3.3 所示。

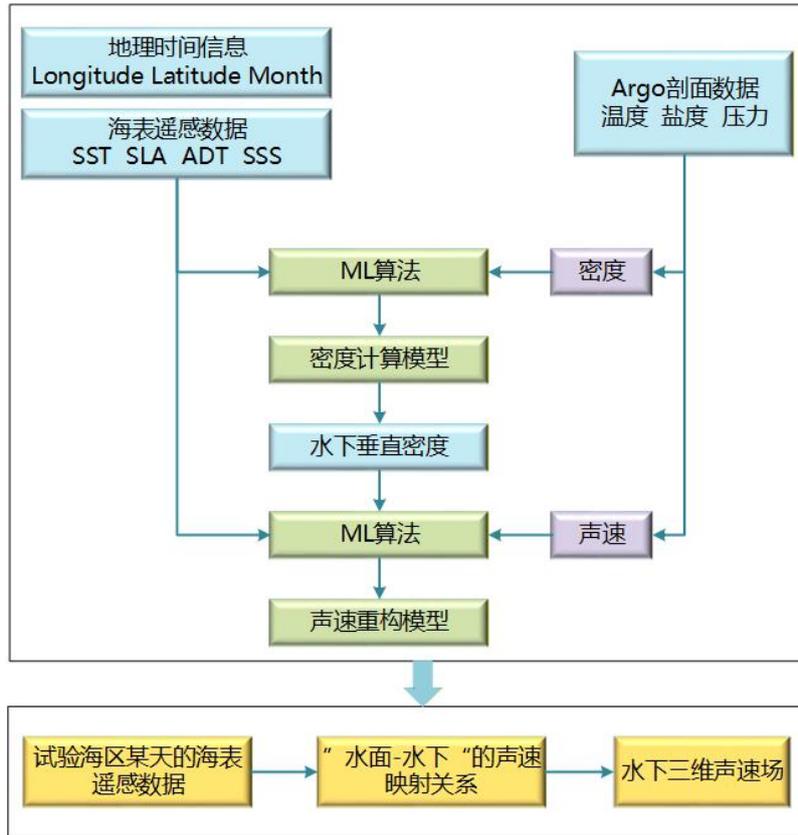


图 3.3 RF-SDR 模型技术流程。

3.2 涡旋内声速场重构

众所周知，中尺度涡旋内部与外部海域在温盐特性上存在显著差异，导致其声场垂直结构异常复杂。现有的大尺度三维背景声速场重构模型对中尺度涡旋区域的特征学习往往不够充分，致使该区域的声速重构误差明显高于其他海域，因此需要对中尺度涡这类复杂动力系统进行针对性建模。

本研究将涡旋内的 Argo 密度数据、海表环境参数（SST、SLA、ADT 和地转流速）和涡旋要素（涡旋类型、椭圆度、涡旋半径和涡旋动能等）作为 SSP 重构的输入，采用 RF 建立涡旋内由海表参数至水下声速的映射关系。依靠涡旋内一条观测剖面，采用中尺度涡统一结构模型重构涡旋内密度场，将海表环境参数、涡旋特征参数和涡旋重构密度输入映射关系，输出涡旋内 SSP，建立一种融合机器学习和物理模型的中尺度涡 SSP 重构模型，即 PIRF-DEN 模型。具体的技术路线如图 3.4 所示。

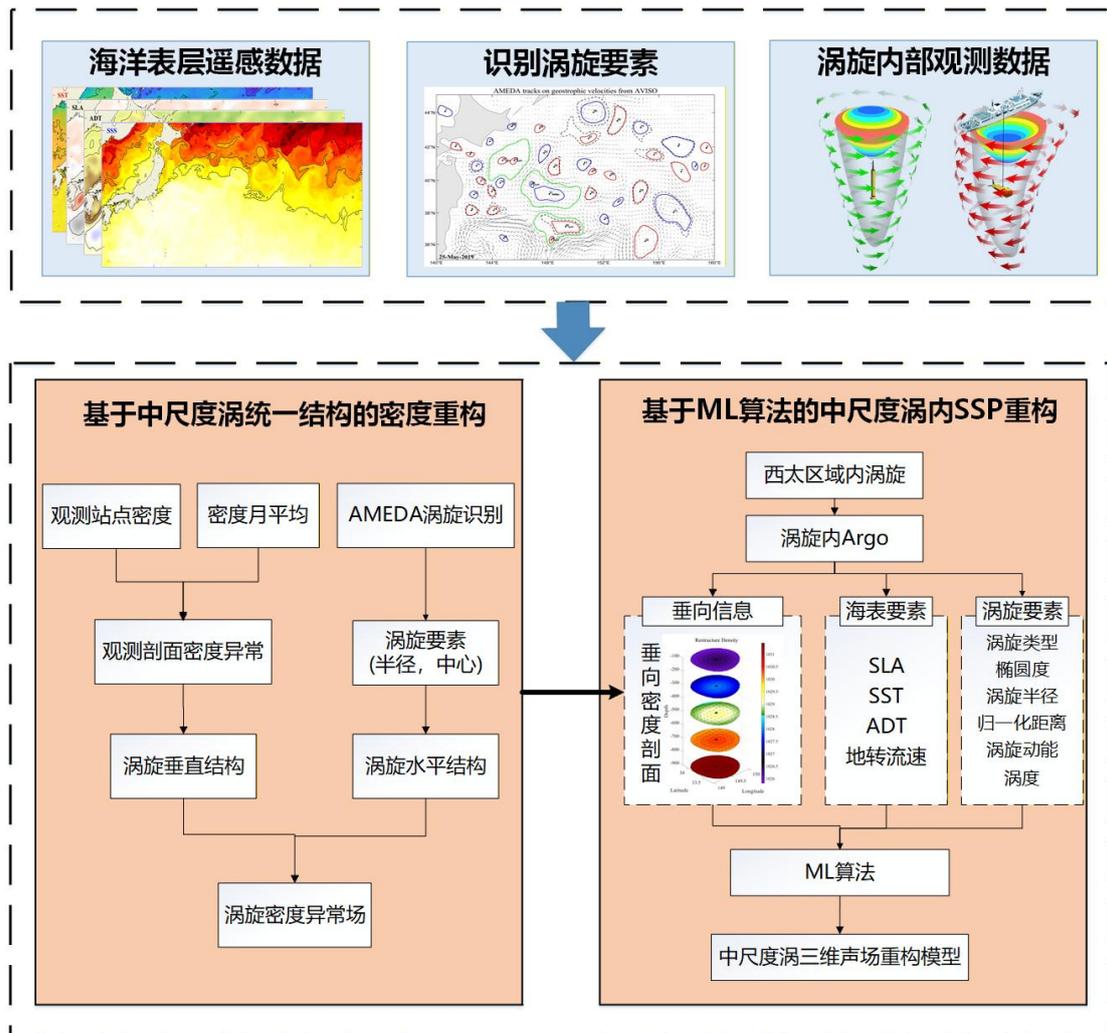


图 3.4 PIRF-DEN 模型技术流程。

3.3 多尺度声速场变分融合

在完成分区背景场与涡内异常场重构的基础上，本方案采用“背景场—局部异常场”相结合的融合思路。分区重构结果作为背景场，刻画大尺度声速结构与区域差异；涡内重构结果作为局部异常场，用于捕捉高梯度变化特征。为避免直接叠加导致的边界不连续问题，引入 DIVAnd 方法，将局部异常信息作为高精度伪观测数据，通过变分优化过程将其有机融入背景场中，最终生成连续、协调且物理一致的三维声速场，实现多尺度信息的有效整合与精度提升。其中权重参数 λ 设置为 1.5，相关尺度 len 设置为 50km。

Divand 算法旨在利用海洋观测数据实现 n 维变化场的分析与重构。该算法特别适用于海洋观测数据的插值与空间分析，因为不同水团往往具有显著差异的物理属性。该方法通过构建一个包含多项约束条件的代价函数，对分析场与观测

值之间的偏差进行惩罚，从而迫使最终的分析结果与原始观测数据相吻合。该代价函数主要包含以下约束项：

$$J[\Phi]=\sum_{j=1}^{N_d} \mu_j [d_j-\phi(x_j)]^2 + \|\phi-\phi_b\|^2 + J_c(\phi) \quad (1)$$

随后，可通过变分分析方法解析求解该二次代价函数的最小值。假设 \mathbf{x}^a 为该代价函数的最小值点，则对 \mathbf{x}^a 施加一个微小位移 $\delta\mathbf{x}$ 不会在 $\delta\mathbf{x}$ 的一阶意义上改变 J 的值。因此，通过求解 J 的相应变化（其值等于零），即可按如下方式求得最小值。

$$\begin{aligned} \delta J &= J(\mathbf{x}^a + \delta\mathbf{x}) - J(\mathbf{x}^a) \\ &= 2\delta\mathbf{x}^T \mathbf{B}^{-1}(\mathbf{x}^a - \mathbf{x}^b) + 2\delta\mathbf{x}^T \mathbf{H}^T \mathbf{R}^{-1}(\mathbf{H}\mathbf{x}^a - \mathbf{y}^o) = 0 \end{aligned} \quad (2)$$

其中， \mathbf{y}^o 为所有观测数据的重组结果， \mathbf{x}^b 为离散化的背景场。 \mathbf{H} 是一个局部插值算子，该算子同样经过离散化处理，用于在原始观测位置与格点化场之间建立可比性。 \mathbf{R} 和 \mathbf{B} 分别为观测场与背景估计场的误差协方差矩阵。此时，该表达式乘以 $\delta\mathbf{x}^T$ 应等于零，因此极值状态 \mathbf{x}^a 可表示为

$$\mathbf{x}^a = \mathbf{x}^b + (\mathbf{B}^{-1} + \mathbf{H}^T \mathbf{R}^{-1} \mathbf{H})^{-1} \mathbf{H}^T \mathbf{R}^{-1} (\mathbf{y}^o - \mathbf{H}\mathbf{x}^b) \quad (3)$$

四、EFGO.SS 数据集效果检验

为了评估融合声速场的精度，使用以下两个指标来衡量声速估计值与测试样本之间的误差。

$$\text{RMSE} = \left[\frac{1}{M} \sum_{i=1}^M (\widehat{C}_{Z_i} - C_{Z_i})^2 \right]^{\frac{1}{2}} \quad (1)$$

$$\text{MAE} = \frac{1}{M} \sum_{i=1}^M |\widehat{C}_{Z_i} - C_{Z_i}| \quad (2)$$

其中，RMSE 为均方根误差，MAE 为平均绝对误差， \widehat{C}_{Z_i} 和 C_{Z_i} 分别为融合和测量声速剖面在深度 Z_i 处的声速值。

4.1 涡旋声速场重构效果检验

基于中尺度涡内 Argo 温盐声数据集和遥感数据，以海表环境参数（SST、SLA、SSS 和 SSD）、涡旋特征参数和 Argo 密度作为模型输入，将数据集的 90%

作为训练集, 10%作为测试集, 开展模型重构性能测试, 并与 sEOF-r 法、sEOF-RF 法和 RF 法进行对比分析。图 4.1 给出了四种方法重构的声速在不同深度的 MAE 和 RMSE。整体而言, PIRF-DEN 模型和 RF 法要优于基于 EOF 的 SSP 重构模型 (sEOF-RF 和 sEOF-r), 可能原因在于基于 EOF 法提取主要模态进行声速建模会造成信息丢失, 从而降低了重构精度。

在 0~1000 m 范围内, PIRF-DEN 模型显著提升了 SSP 的重构精度, 各深度内都优于其余三种模型, 在 840 m 以浅改进效果最明显。在 0~840 m 范围内, PIRF-DEN 的声速误差为 0.89 m/s, RF 的误差为 4.20 m/s, sEOF-RF 的误差为 6.67 m/s, sEOF-r 的误差为 7.50 m/s, PIRF-DEN 较 sEOF-r 重构精度提高近 88.1%, 较 RF 重构精度提高 78.8%, 说明引入海水垂向密度信息, PIRF-DEN 的改进效果显著。在 840~1000 m 范围内, 四种模型的重构声速误差均降至 2 m/s, PIRF-DEN 的重构声速误差小于 0.7 m/s。

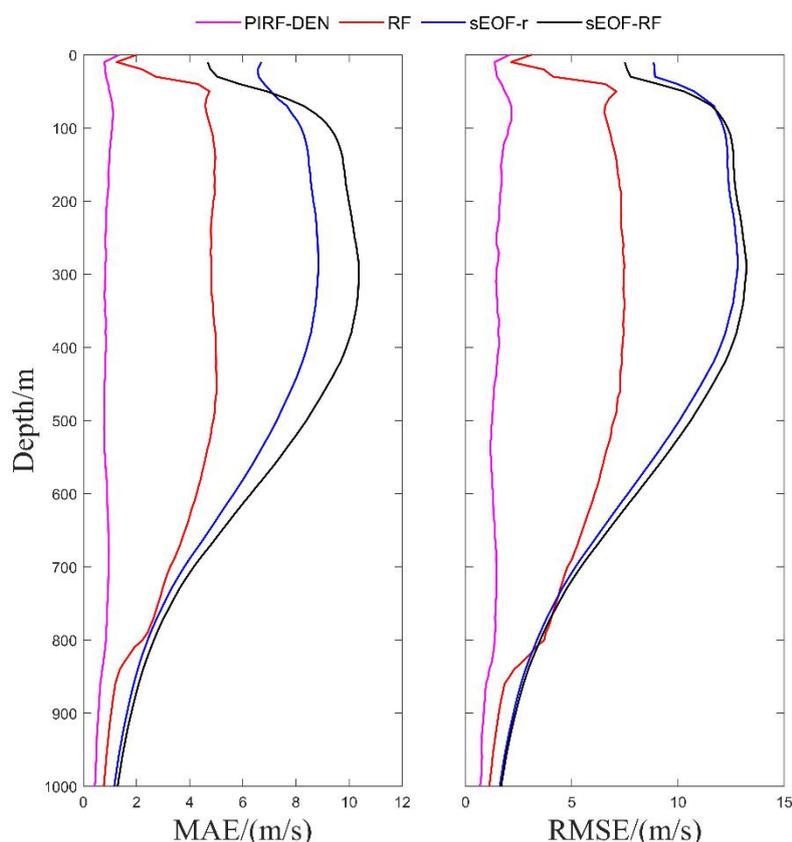


图 4.1 四种 SSP 重构方法在各深度层的声速 MAE 和 RMSE。

表 4.1 列出了四种模型针对涡旋内声速的平均重构精度: PIRF-DEN > RF > sEOF-RF > sEOF-r, PIRF-DEN 的重构精度相比于经典的 sEOF-r 法提高了 87.3%;

针对两种基于 EOF 法的 SSP 重构模型，利用机器学习模型构建的非线性框架 (sEOF-RF) 明显优于传统的线性框架 (sEOF-r)，模型的重构精度提高了 11%；对比 PIRF-DEN 模型和 RF 法，引入海水的垂向密度信息，模型的重构能力显著提高，重构精度提高了 77.28%。在锋面和中尺度涡等动力结构复杂的海域，声速重构误差显著增大。本研究提出的 PIRF-DEN 模型通过引入涡旋重构密度作为输入，有效地降低了中尺度涡旋区域的声速重构误差，将其控制在 1~2 m/s 之内，并且该模型的表现优于目前大多数的声速重构结果。

表 4.1 四种 SSP 重构方法的评估系数

评估系数	SSP 重构方法			
	PIRF-DEN	RF	sEOF-RF	sEOF-r
MAE	0.83	3.66	5.84	6.56
RMSE	1.39	5.42	8.28	8.51

为了进一步对比四种 SSP 重构方法的重构效果，图 4.2 给出了四种方法在各 Argo 站位处的 SSP 重构误差。由图 4.2 可知，四种模型的重构精度： $PIRF-DEN > RF > sEOF-RF > sEOF-r$ ，PIRF-DEN 展现出了远超其余三种模型的重构精度，绝大部分站位的声速误差低于 1 m/s，只有少数站位的误差大于 2 m/s，零散地分布在亚北极锋海域。综上，PIRF-DEN 模型通过中尺度涡统一结构模型引入密度作为输入，准确描述了中尺度涡垂向结构，显著降低了中尺度涡经常发生海域的声速重构误差。

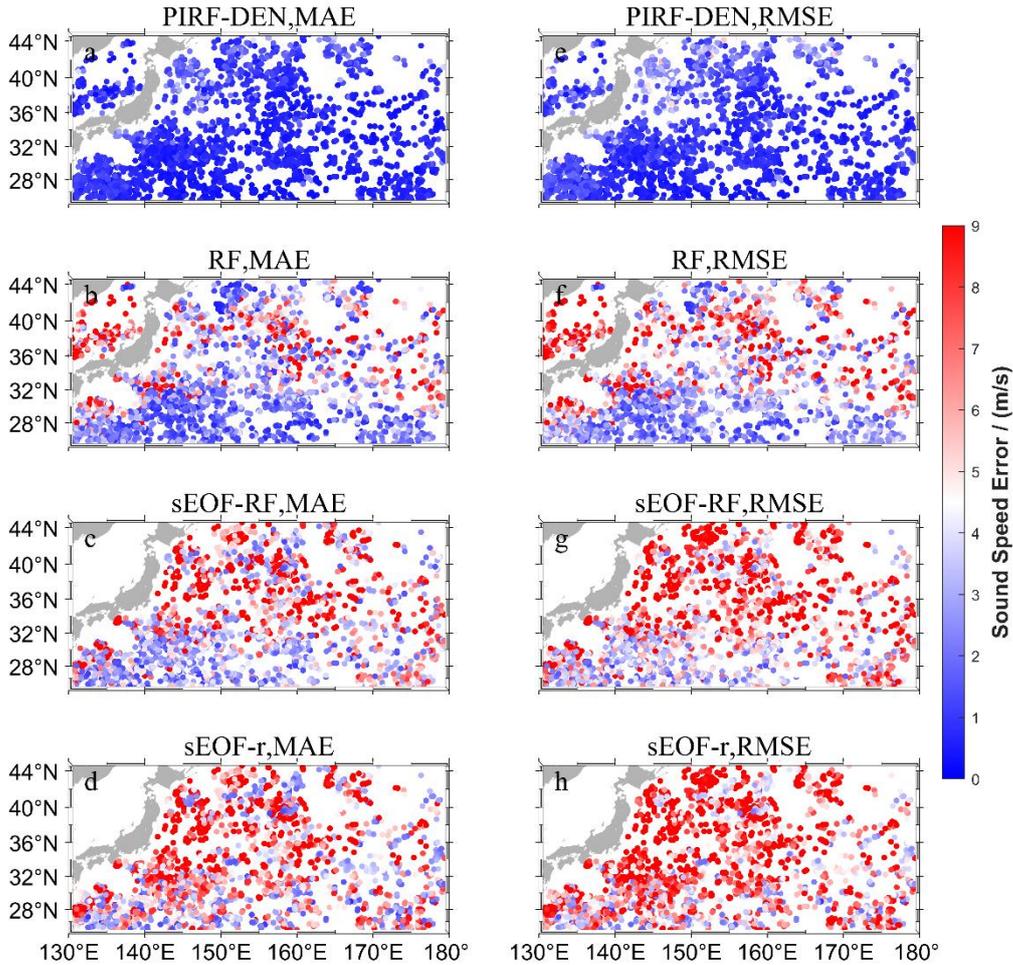


图 4.2 四种 SSP 重构方法在各 Argo 站位处的声速误差，其中 1-4 行分别为 PIRF-DEN、RF、sEOF-RF 和 sEOF-r 的声速平均误差，1-2 列分别为四种方法的声速平均 MAE 和平均 RMSE。

4.2 融合数据集效果检验

为评估本数据集产品 (EFGO.SS) 的声速重构性能，基于 2023 年 5964 条 Argo 剖面数据，将其与 CMEMS 再分析数据计算的声速剖面进行对比。结果显示，CMEMS SSP 的均方根误差与平均绝对误差分别为 2.58 m/s 和 1.47 m/s。相比之下，EFGO.SS 的 RMSE 和 MAE 分别显著降低了 45.3% 和 40.8%，充分证明了 EFGO.SS 产品在声速重构方面的优越性。

为进一步检验其在涡旋内部区域的表现，基于 1356 个涡内 Argo 站位的独立验证表明，融合后声速场的 RMSE 从背景场的 1.63 m/s 降至 0.91 m/s，MAE 从 1.0 m/s 降至 0.52 m/s，改善幅度分别达 44% 和 48%。这表明，EFGO.SS 产品成功弥补了单一背景场在涡旋区域的不足，有效抑制了由温盐扰动引起的重构误差。图 1 所示为 1356 个涡内 Argo 站位融合前后的误差对比图。

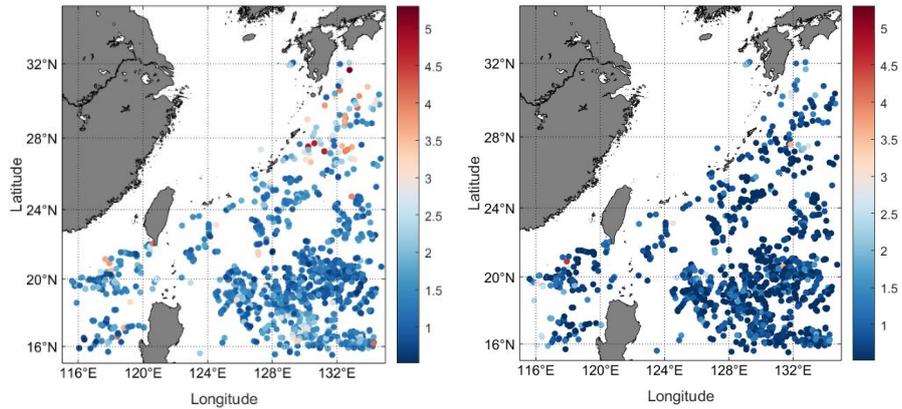


图 4.3 2023 年涡内声速剖面融合前（左）与融合后（右）误差

五、EFGO.SS 数据产品总结

本数据集基于研究团队提出的首个专门针对中尺度涡内三维声速场重构模型，建立了动力-统计融合建模框架，利用中尺度涡统一结构动力模型构造涡旋密度垂直模态，作为声速重构垂向约束，同时引入涡特征参数输入机器学习算法，有效提升了涡旋内声速重构精度。创新将海水密度垂直结构作为声速重构约束嵌入网络，有效提升了声速重构精度；同时引入空间特征提取机制，通过移位窗口对声速场高效编码，更好地刻画了声速空间分布特征，本数据集产品有效弥补了高精度水下三维声速数据的不足。