JGR Oceans

RESEARCH ARTICLE
10.1029/2023JC019941

Key Points:

- The deep operator learning algorithm based on physical constraints is proposed for density field inversion
- The combination of deep learning and dynamics-based methods provides interpretability while enabling fast density reconstruction
- The proposed reconstruction framework outperforms proposed purely data-driven algorithm and traditional dynamics algorithm

Correspondence to:
Z. Gao,
zhengao@ouc.edu.cn

Citation:
Chen, Y., Liu, L., Yuan, C., Sun, X., Chen, X., Wei, Z., & Gao, Z. (2024). Physics-informed deep operator learning based on reduced-order modeling for retrieving the ocean interior density from the surface. *Journal of Geophysical Research: Oceans*, *129*, e2023JC019941. https://doi.org/10.1029/2023JC019941

Received 20 APR 2023
Accepted 31 JAN 2024

Author Contributions:
Data curation: Li Liu
Formal analysis: Xiang Sun, Xueen Chen, Zhen Gao
Funding acquisition: Zhiqiang Wei, Zhen Gao
Investigation: Yuanhong Chen
Methodology: Yuanhong Chen
Project administration: Zhiqiang Wei, Zhen Gao
Validation: Yuanhong Chen
Writing – original draft: Yuanhong Chen
Writing – review & editing: Li Liu, Chunxin Yuan, Xiang Sun, Xueen Chen, Zhiqiang Wei, Zhen Gao

Physics-Informed Deep Operator Learning Based on Reduced-Order Modeling for Retrieving the Ocean Interior Density From the Surface

Yuanhong Chen[1], Li Liu[2], Chunxin Yuan[1], Xiang Sun[1], Xueen Chen[2], Zhiqiang Wei[3], and Zhen Gao[1]

[1]School of Mathematical Sciences, Ocean University of China, Qingdao, China, [2]Frontier Science Center for Deep Ocean Multispheres and Earth System (FDOMES) and Physical Oceanography Laboratory, Ocean University of China, Qingdao, China, [3]College of Information Science and Engineering, Ocean University of China, Qingdao, China

Abstract Exploring methods to reconstruct the ocean interior from surface data is a crucial focus in the study of ocean processes and phenomena due to the shortage of subsurface and deep-sea data. Nonetheless, the existing methods predominantly concentrate on either data-driven or dynamical methodologies, with limited exploration of integrating the strengths of both approaches. To combine the advantages of these two methods for reconstructing the subsurface density field from surface data, a novel dynamics-constrained deep operator learning network based on reduced-order model is proposed. Encoding the mean-squared residuals of the reduced-order equation along with the mean-squared error between the network outputs and targets into the loss function effectively merges the dynamical and data constraints during the training process. This integration makes the network outputs and inputs approximately satisfy a specific form of the equation, allowing for interpretability, and once the network is well-trained, rapid reconstruction evaluation can be performed. The reduced-order equation is established by the Galerkin projection of quasi-geostrophic equation onto the low-dimensional subspace identified via reduced-basis, which explains the vertical variation of ocean density. The developed model can tackle the challenge of directly measuring subsurface potential vorticity and predicting subsurface density. Evaluation is conducted using simulation data from the Max-Planck-Institute ocean model, indicating that it can offer precise estimations, outperforms the purely data-driven algorithm presented in the paper and the interior plus surface quasi-geostrophic method, and enables model sharing across different regions.

Plain Language Summary Reconstructing the interior density of the ocean from surface data is a valuable way to make up for the lack of ocean observations. Most of the previous works are purely data-driven or dynamics-based methods. In order to employ the advantages of dynamic and data-driven models at the same time, a physics-informed deep operator learning algorithm is proposed to reconstruct the density field. The proposed method shows superior predictive performance and greater robustness when compared with the pure data-driven and dynamics-based methods presented in the paper. These findings underscore the potential of deep learning algorithms incorporating physical constraints for marine scientific research.

1. Introduction

As a key component of the Earth's climate system, the ocean plays a vital role in regulating the planet's temperature, and in driving global weather patterns (Cheng et al., 2020; Levitus et al., 2012). A primary feature of the ocean is the presence of stratification, in which the ocean is distributed unevenly in both horizontal and vertical directions. Stratification changes strongly affect ocean dynamics and marine biogeochemical processes, and also have an important impact on submarine acoustic detection and safe navigation, due to the large amplitude internal waves generated in the stratified ocean. Density distribution provides important information about the physical attributes and dynamic mechanisms of the ocean, which significantly influence stratification and mixing processes. Even slight changes in density throughout the vast ocean can lead to extraordinary consequences. Therefore, comprehending ocean density distribution is vital for studying maritime safety, ocean heat balance, and climate change. This necessitates the acquisition of high-quality observational data with extensive spatial and temporal coverage (Cunningham et al., 2007; Klemas & Yan, 2014).

Advances in satellite and radar technology (Roarty et al., 2019; Torres et al., 2012; Xue et al., 2021) have greatly improved the understanding of the ocean by providing high-resolution, long-term ocean surface data on a global scale. However, the electromagnetic waves used by these sensors are limited in their ability to penetrate into the deep ocean, preventing direct observations of subsurface and deep ocean. In situ measurements can provide direct ocean interior observations, however, limited by the cost-effectiveness and challenges of collecting data at sea, these observations are sparse and inhomogeneous, rendering them inadequate to support investigations of oceanic dynamic processes and mechanisms at various scales (L. Meng & Yan, 2022; Roemmich et al., 2009). In this case, using sea surface data to derive subsurface information emerges as an effective approach to attain continuous, wide-area, three-dimensional information of the ocean, as the sea surface and subsurface are closely linked, which results in many subsurface processes manifesting themselves at the surface (Hurlburt, 1984; Khedouri et al., 1983). The widely known invertibility principle of potential vorticity (PV) (Hoskins et al., 1985) enables the diagnosis of three-dimensional dynamics of a balanced flow through PV knowledge at various levels and boundaries, and Isern-Fontanet et al. (2008) demonstrates the spatial correlation between surface buoyancy and PV in the ocean interior under shared assumptions. This also provides a theoretical basis for deriving subsurface information from surface data. Despite progress in numerical modeling (Blumberg & Mellor, 1987; Mellor et al., 2002), these simulations still face challenges, including high computational costs and the difficulty of specifying parameters and initial/boundary conditions (Fox-Kemper et al., 2019). Consequently, the pursuit of diverse methodologies to derive subsurface information has become a pivotal concern in the field of oceanography, captivating the attention of a growing cohort of researchers (Khedouri et al., 1983; Klemas & Yan, 2014).

To comprehend the subsurface and deep ocean fields employing surface data necessitates the formulation of models that describe the relationship between them. In pursuit of this objective, models employed in such endeavors can be generally classified into two categories: data-driven models and dynamics-based models. Data-driven models include those statistical methods, such as one-dimensional linear regression (Fiedler, 1988; Fischer et al., 1997), multivariable regression (Carnes et al., 1994; Guinehut et al., 2012), and more advanced techniques like empirical orthogonal function (Buongiorno Nardelli et al., 2012; Buongiorno Nardelli & Santoleri, 2005; Gallaudet & Simpson, 1994; Maes et al., 2000). These models rely on the assumption that the relationship between surface data and subsurface data can be represented by known function forms, with unknown coefficients determined through regression methods. While these models are straightforward and flexible, their simplicity may not be adequate in handling complex, nonlinear problems (L. Meng et al., 2021; Tian et al., 2022). Methods based on machine learning have recently become the most widely used class of data-driven models, such as artificial neural network (Ali et al., 2004; Bao et al., 2019; Su et al., 2020), self-organizing maps (Wu et al., 2012; C. Chen et al., 2018), support vector machine (Su et al., 2015, 2018), clustering neural network (W. Lu et al., 2019), and random forest (Su et al., 2018). These methods have demonstrated a good performance in addressing complex problems, and with the advancement of powerful computers and artificial intelligence algorithms, deep learning (DL) (Reichstein et al., 2019) is emerging as a highly promising technique. Different from traditional machine learning methods, DL is based on multi-layer artificial neural networks, with its complex model architecture and overfitting prevention techniques. Moreover, DL has the ability to learn more information and connections from data, resulting in models that are more stable and accurate (He et al., 2016; Schmidhuber, 2015). Although DL is not yet widely used in marine science research, the promising results from existing studies (Buongiorno Nardelli, 2020; Ham et al., 2019; Reichstein et al., 2019; Su et al., 2021) suggest that it has great potential and is likely to become a major method in the field in the future.

The use of data-driven methods in oceanography offers several advantages, such as flexibility, high generalizability, and ease of modeling, and has proven to work effectively in many applications. Nevertheless, the paucity of interpretability inherent in data-driven models poses a formidable obstacle to their extensive implementation within the domain of marine science (Dong et al., 2022; S. Wang, Teng, & Perdikaris, 2021). Data-driven models typically rely solely on the data for learning the relationships between inputs and outputs (Doshi-Velez & Kim, 2017; Guo et al., 2016), without considering the dynamic rules. This can lead to a "black box" model, where the underlying physical connections between data are not transparently revealed (L. Meng & Yan, 2022). Dynamic-based models, on the other hand, are built upon established physical laws, which make them more transparent and physically interpretable. These models use sea surface data as control variables and apply dynamic models to infer subsurface information (Akbari et al., 2017). In particular, approaches based on the surface-quasi-geostrophic (SQG) framework (Held et al., 1995) have received considerable attention in oceanography and

have been widely applied. For instance, LaCasce and Mahadevan (2006) suggested that the exponential decay relationship between sea surface density (SSD) and internal PV should be introduced into governing equation to invert the flow field. In addition, Lapeyre and Klein (2006) developed the effective SQG, which simplifies the dynamical equation by decomposing them into three-dimensional PV control equation and surface density control equation, which has been successfully applied to various works (Isern-Fontanet et al., 2008; Qiu et al., 2016). Furthermore, to address the issue of the need for significant correlation between sea surface height (SSH) and SSD, interior + SQG (isQG) (J. Wang et al., 2013) was proposed to reconstruct subsurface velocity and density using SSD and SSH. The validity of this approach has been demonstrated by several studies (Z. Chen et al., 2020; L. Liu et al., 2014, 2017).

In the field of marine science research, a popular objective involves the harmonious harnessing of the merits inherent in data-driven approaches and dynamics-based methodologies (L. Meng & Yan, 2022; Reichstein et al., 2019; Yan et al., 2020; Zhu et al., 2017). This endeavors to alleviate the limitations of each individual approach while constructing DL models that are embedded with physical knowledge. Specifically, the dynamics-based model imposes physical constraints on the data-driven model, ensuring that the integrated model is physically consistent and remains interpretable. In addition, the integrated model can be trained to rapidly estimate subsurface information by simply inputting new sea surface data. Raissi et al. (2019) have made a significant contribution to this field by introducing physics-informed neural network (PINN), which has been widely applied (Jagtap et al., 2020; Penwarden et al., 2022; Yang et al., 2019; Zhang et al., 2019). PINN employs automatic differentiation (Baydin et al., 2018) to incorporate partial differential equations (PDEs) into the network's loss function, ensuring that the inputs and outputs of the model comply with known PDEs. However, it is important to note that for problems with large spatial and temporal domains, a substantial abundance of residuals becomes indispensable to impose required physical constraints, which inevitability leads to the potential escalation of training costs (Karniadakis et al., 2021; X. Meng et al., 2020).

In this paper, a novel dynamics-constrained deep operator learning reduced-order model (ROM) framework is proposed for reconstructing the subsurface density from SSD, which combines the strengths of physical modeling and DL. First, to reduce computational cost, the reduced basis method (RBM) (Hesthaven et al., 2015; Quarteroni et al., 2015) is employed to construct a low-complexity ROM that reduces the spatial dimension from three to one, where the reduced-order equation is obtained by projecting QG equation onto the space spanned by the reduced basis. Then, to further estimate the subsurface density, a physics-informed operator learning network (PIO-Net) framework, is designed to estimate the unknown terms in the reduced-order equation and the density projection coefficients. PIO-Net is comprised of multiple separate sub-networks, each performing a specific function. An operator learning network that consists of three independent sub-networks is used to approximate the density projection coefficient. The structure of the operator learning network draws inspiration from the profound operator learning network (DeepONet) (L. Lu et al., 2021), which has demonstrated remarkable effectiveness in addressing parameterized partial differential equations (PDEs) encompassing diverse initial and boundary conditions (Oommen et al., 2022; S. Wang, Wang, & Perdikaris, 2021). The unidentified terms within the reduced-order equation are estimated from data employing two feedforward neural networks, which enables the ROM to be flexible in adapting to the data. PIO-Net is trained by minimizing a loss function composed of two parts, one being the data constraint, which is the mean square error of the outputs of the network with respect to the target outputs, and the other being the physical constraint, which is the mean-square residuals of the reduced-order equation at the sampling points. By incorporating both types of constraints, the resulting model combines the strengths of both dynamics-based modeling and machine learning modeling, allowing rapid and accurate simulation of subsurface data in the online stage while maintaining interpretability. This study is a first step in the theoretical exploration of a new method to project sea surface information into the subsurface. The validity of the method is verified in several regions based on Max Planck Institute Ocean Model (MPI-OM) simulations (Marsland et al., 2003; Song & Chen, 2020), which exclude the most sources of error compared to observations. The research findings demonstrate that PIO-Net is capable of accurately predicting the deep density with superior precision and robustness compared to the proposed purely data-driven and isQG approaches. To demonstrate the potential of the proposed method in three-dimensional field reconstruction, a robustness verification with in situ data is given in Section 5. The three-dimensional density is estimated using the satellite-derived SSD data as the input of the trained PIO-Net model, and the RMSEs are computed and compared between the reconstructed data and observed data, as well as between the monthly climatology and observed profiles. The results show that the proposed method can improves density estimation compared to a properly derived three-dimensional monthly

climatology, particularly in the upper layers. And in the future, it is expected that the proposed model is trained and tested with real data.

The rest of the paper is organized as follows. Section 2 presents the reconstruction algorithms in detail. The data used and the processing of the data are described in Section 3. Section 4 presents the results, and Section 5 brings together the relevant conclusions.

2. Methods

2.1. isQG for Density Reconstruction

In the context of QG-balanced motions, QG theory shows that the geostrophic stream function of a flow field is determined by the knowledge of PV of ocean interior and appropriate boundary condition. This relationship can be described as follows:

$$\begin{cases} \mathcal{L}\Psi + \dfrac{\partial}{\partial z}\left(\dfrac{f_0^2}{N^2}\dfrac{\partial \Psi}{\partial z}\right) = q, \quad -H < z < 0, \quad (x,y) \in \Omega, \\ \left.\dfrac{\partial \Psi}{\partial z}\right|_{z=0} = \dfrac{b_s}{f_0}, \quad (x,y) \in \Omega, \\ \left.\dfrac{\partial \Psi}{\partial z}\right|_{z=-H} = 0, \quad (x,y) \in \Omega, \end{cases} \tag{1}$$

where $\mathcal{L} = \left(\frac{\partial^2}{\partial x^2} + \frac{\partial^2}{\partial y^2}\right)$, $\Psi = p(\rho_0 f_0)^{-1}$ denotes the stream function with p implying the pressure, f_0 is the mean Coriolis parameter in the model domain Ω, N^2 is the Brunt − Väsälä (BV) frequency and q represents the PV anomaly, and b_s denotes the surface buoyancy anomaly, which is determined from the sea surface density anomaly (SSDA) ρ_s^a and the volume mean density ρ_0 as $b_s = -g\rho_s^a\rho_0^{-1}$. Once q is known, the stream function Ψ can be calculated by solving Equation 1 and the corresponding density anomaly ρ^a is reconstructed using the formula: $\rho^a = -\frac{\rho_0 f_0}{g}\frac{\partial \Psi}{\partial z}$.

Direct observation of the PV is typically not feasible, which poses a significant challenge in the computation of stream function. To address this issue, Ψ is generally split into two parts, one with SQG solutions Ψ_s, which is computed with sea surface buoyancy anomaly

$$\begin{cases} \mathcal{L}\Psi_s + \dfrac{\partial}{\partial z}\left(\dfrac{f_0^2}{N^2}\dfrac{\partial \Psi_s}{\partial z}\right) = 0, \\ \left.\dfrac{\partial \Psi_s}{\partial z}\right|_{z=0} = \dfrac{b_s}{f_0}, \left.\dfrac{\partial \Psi_s}{\partial z}\right|_{z=-H} = 0, \\ \rho^a_{\mathrm{SQG}} = -\dfrac{\rho_0 f_0}{g}\dfrac{\partial \Psi_s}{\partial z}, \end{cases} \tag{2}$$

and the other with interior ones Ψ_i, which is associated with interior PV anomaly

$$\begin{cases} \mathcal{L}\Psi_i + \dfrac{\partial}{\partial z}\left(\dfrac{f_0^2}{N^2}\dfrac{\partial \Psi_i}{\partial z}\right) = q, \\ \left.\dfrac{\partial \Psi_i}{\partial z}\right|_{z=0} = 0, \left.\dfrac{\partial \Psi_i}{\partial z}\right|_{z=-H} = 0, \\ \rho_i^a = -\dfrac{\rho_0 f_0}{g}\dfrac{\partial \Psi_i}{\partial z}. \end{cases}$$

The SQG solution can naturally be calculated in the Fourier space using SSDA and sea surface height anomaly (SSHA) because there are no unknown terms in Equation 2. To compute the interior solution, the isQG method

relies on the assumption that the barotropic mode and the first baroclinic mode have a significant impact, and the Fourier transform of the corresponding interior solution can be expressed as

$$\hat{\Psi}_i(k,l,z) = A_0(k,l)F_0 + A_1(k,l)F_1.$$

Here k and l denote the zonal and meridional wavenumbers, respectively. A_0 and A_1 are the unknown amplitudes corresponding to the eigenfunctions F_0 and F_1. Therefore, we can calculate the amplitude A_0 and A_1 from the SSHA η and the zero bottom buoyancy anomaly hypothesis

$$\begin{cases} A_0(k,l)F_0(0) + A_1(k,l)F_1(0) + \hat{\Psi}_s(k,l,0) = \dfrac{g}{f_0}\hat{\eta}(k,l), \\ A_0(k,l)F_0(-H) + A_1(k,l)F_1(-H) + \hat{\Psi}_s(k,l,-H) = 0, \end{cases}$$

where the hat of variables denotes the Fourier transformation, then the stream function Ψ can be refactored as the sum of Ψ_s and Ψ_i. The application of isQG method is based on the following assumptions, assuming that the velocity is QG and the pressure is hydrostatic. Additionally, the buoyancy frequency and Coriolis force parameters are assumed to be regional averages. And the interior solution is represented by two characteristic modes, based on the assumption that the barotropic mode and the first baroclinic mode have significant effects, while the higher order modes contribute less. Therefore, it is applicable to the description of eddy activity on scales between 10 and 500 km, where the QG approximation is applicable, and its application is confined to small regions. However, more details about isQG are provided by J. Wang et al. (2013).

2.2. Dynamics-Constrained Deep Operator Learning of Reduced-Order Model

2.2.1. Projection-Based Reduced Basis Method

To solve QGPV Equation 1, a spatial discretization scheme is employed to discretize the equation in Ω to obtain the high-fidelity full-order model (FOM)

$$\begin{cases} \mathcal{L}_h \boldsymbol{\Psi}_h + \dfrac{d}{dz}\left(\dfrac{f_0^2}{N^2}\dfrac{d\boldsymbol{\Psi}_h}{dz}\right) = \boldsymbol{q}, \quad -H < z < 0, \\ \left.\dfrac{d\boldsymbol{\Psi}_h}{dz}\right|_{z=0} = f_0^{-1}\boldsymbol{b}_s, \\ \left.\dfrac{d\boldsymbol{\Psi}_h}{dz}\right|_{z=-H} = 0, \end{cases} \tag{3}$$

where $\boldsymbol{\Psi}_h$, $\boldsymbol{q}$ and $\boldsymbol{b}_s$ are the discretized forms of the stream function Ψ, PV anomaly q, and the surface buoyancy anomaly b_s, respectively, on a staggered space grid with dimensions N_h. Here, N_h represents the number of grid points in the FOM, which is often quite large. $\mathcal{L}_h$ denotes the linear operator derived from $\mathcal{L}$. In this paper, the finite difference method is used to solve the problem, with the second derivative approximated by the central difference scheme. Given the challenge of performing repeated model calculations with varying boundary conditions $\boldsymbol{b}_s$, replacing the FOM with ROM presents a reliable choice for reducing computation and conserving resources.

One of the most widely used ROM techniques is RBM, where the reduced basis $\{\boldsymbol{\psi}_i\}_{i=1}^n$ in the offline stage are generated with a much smaller size n than the original space size N_h. In the online stage, the solution $\boldsymbol{\Psi}_h$ is represented as a linear combination of the reduced basis. Once the reduced coefficients for new surface buoyancy anomaly $\boldsymbol{b}_s$ is obtained, the full-order solution can be recovered efficiently in the online stage by the following form

$$\boldsymbol{\Psi}_h(z, \boldsymbol{b}_s) \approx \sum_{i=1}^{n} \alpha_i(z, \boldsymbol{b}_s)\boldsymbol{\psi}_i = \mathbf{V}\boldsymbol{\alpha}(z, \boldsymbol{b}_s), \tag{4}$$

where $\mathbf{V}$ is a matrix with the reduced basis as columns, and

$$\boldsymbol{\alpha}(z, \boldsymbol{b}_s) = [\alpha_1(z, \boldsymbol{b}_s), \alpha_2(z, \boldsymbol{b}_s), \ldots, \alpha_n(z, \boldsymbol{b}_s)]^{\mathrm{T}}$$

denotes the reduced coefficient vector, and the superscript T represents vector transpose.

For an accurate approximation of the solution of the FOM, there are several strategies for generating the reduced basis, and in this paper, we employ empirical orthogonal decomposition (EOF). To perform EOF, we collect the density anomalies at N_s different times, each at $m+1$ different depths, and assemble them into a data matrix $\mathbf{X}$

$$\mathbf{X} = [\mathbf{X}_{h,1}\ \mathbf{X}_{h,2}\ \ldots\ \mathbf{X}_{h,N_s}],$$

where $\mathbf{X}_{h,i} = \left[\boldsymbol{\rho}^a_{z_0,i}\ \boldsymbol{\rho}^a_{z_1,i}\ \ldots\ \boldsymbol{\rho}^a_{z_m,i}\right]$ represents a state matrix composed of density anomaly vectors at $m+1$ depths at the i-th time. Substituting Equation 4 into Equation 3, and projecting the resulting equation onto the reduced space spanned by the reduced basis, the following reduced-order equation can be obtained:

$$\begin{cases} \mathbf{V}^{\mathrm{T}}\mathcal{L}_h(\mathbf{V}\boldsymbol{\alpha}) + \mathbf{V}^{\mathrm{T}}\left(\dfrac{d}{dz}\dfrac{f_0^2}{N^2}\dfrac{\mathbf{V}d\boldsymbol{\alpha}}{dz}\right) = \mathbf{V}^{\mathrm{T}}\boldsymbol{q}, \quad -H < z < 0, \\ \left.\dfrac{d\boldsymbol{\alpha}}{dz}\right|_{z=0} = \mathbf{V}^{\mathrm{T}} f_0^{-1}\boldsymbol{b}_s, \left.\dfrac{d\boldsymbol{\alpha}}{dz}\right|_{z=-H} = 0, \\ \boldsymbol{\rho}^a = -\dfrac{\rho_0 f_0}{g}\mathbf{V}\dfrac{d\boldsymbol{\alpha}}{dz}. \end{cases} \tag{5}$$

To mitigate the influence of spatial coordinate dimensionality and simplify the calculation, we first define a new variable $\boldsymbol{\gamma}(z) = \rho_0\boldsymbol{\alpha}(z)(500\ \mathrm{g})^{-1} \times 10^{-5}$ and then perform a transformation on the depth variable by defining $h = z/500$, thus Equation 5 becomes

$$\begin{cases} c\mathbf{V}^{\mathrm{T}}\mathcal{L}_h(\mathbf{V}\boldsymbol{\gamma}) + \mathbf{V}^{\mathrm{T}}\left(\dfrac{d}{dh}\dfrac{f^2}{n_z}\dfrac{\mathbf{V}d\boldsymbol{\gamma}}{dh}\right) = \mathbf{V}^{\mathrm{T}}\tilde{\boldsymbol{q}}, \quad -H_f < h < 0, \\ \left.\dfrac{d\boldsymbol{\gamma}}{dh}\right|_{h=0} = -f^{-1}\mathbf{V}^{\mathrm{T}}\boldsymbol{\rho}^a_s, \left.\dfrac{d\boldsymbol{\gamma}}{dh}\right|_{h=-H_f} = 0, \\ \boldsymbol{\rho}^a = -f \cdot \left(\mathbf{V}\dfrac{d\boldsymbol{\gamma}}{dz}\right), \end{cases}$$

where $c = 500^2 \times 10^5$, $f = f_0 \times 10^5$, $n_z = N^2 \times 10^5$, $\tilde{\boldsymbol{q}} = c\boldsymbol{q}^{-1}$ and $H_f = H/500$. In QG theory, BV frequency and Coriolis force parameters are generally assumed to be constant values of regional averages. Therefore, different regional models need to be built for the global application of the methods based on QG theory. Besides, QG is an inherently simplified dynamic model that lacks the ability to fully capture the intricate physical processes occurring within the ocean interior. To address these limitations to some extent, and to avoid the calculation of area-averaged BV frequency, the unknown functions g_1, g_2, and g_3 are introduced into the equation, and the following model established

$$c\mathbf{V}^{\mathrm{T}}\mathcal{L}_h(\mathbf{V}\boldsymbol{\gamma}) + g_1\left(h, \overline{\frac{d\boldsymbol{\gamma}}{dh}}\right)\frac{d\boldsymbol{\gamma}}{dh} + g_2\left(h, \overline{\frac{d\boldsymbol{\gamma}}{dh}}\right)\frac{d^2\boldsymbol{\gamma}}{d^2h} = g_3(h)\mathbf{V}^{\mathrm{T}}\boldsymbol{\rho}^a_s,$$

the bar of the vector denotes the mean of the vector, and the functions g_1 and g_2 depend on the variable h and the mean of the derivative of vector $\boldsymbol{\gamma}$ with respect to h, and g_3 is the function of h. This modification allows the established model to incorporate both prior information from QG dynamics and adaptability to the data, thus enhancing its universality. The term $g_3(h)\mathbf{V}^{\mathrm{T}}\boldsymbol{\rho}^a_s$ is derived from Isern-Fontanet et al. (2008), where it was demonstrated that the surface-trapped solution Ψ_s can effectively represent the total solution Ψ for baroclinically unstable flows within the first 500 m, driven by a large-scale density gradient. In such cases, the mesoscale anomalies of potential vorticity (PV) in the interior are correlated with the buoyancy anomalies at the surface, denoted as $q = g_3(h)b_s$. And the model ultimately simplifies to

$$\begin{cases} c\mathbf{V}^{\mathrm{T}}\mathcal{L}_h(\mathbf{V}\boldsymbol{\gamma}) + g_1\left(h, \overline{\frac{d\boldsymbol{\gamma}}{dh}}\right)\frac{d\boldsymbol{\gamma}}{dh} + g_2\left(h, \overline{\frac{d\boldsymbol{\gamma}}{dh}}\right)\frac{d^2\boldsymbol{\gamma}}{d^2h} = g_3(h)\mathbf{V}^{\mathrm{T}}\boldsymbol{\rho}_s^a, \quad -H_f < h < 0, \\ \left.\frac{d\boldsymbol{\gamma}}{dh}\right|_{h=0} = -f^{-1}\mathbf{V}^{\mathrm{T}}\boldsymbol{\rho}_s^a, \left.\frac{d\boldsymbol{\gamma}}{dh}\right|_{h=-H_f} = 0, \\ \mathbf{V}^{\mathrm{T}}\boldsymbol{\rho}^a = -f\frac{d\boldsymbol{\gamma}}{dz}, \end{cases} \tag{6}$$

Equation 6 can be viewed as a differential equation that is parameterized by $f^{-1}\mathbf{V}^{\mathrm{T}}\boldsymbol{\rho}_s^a$. Given g_1, g_2, and g_3, the subsurface density anomaly can be calculated accordingly. Inspired by DeepONet, a machine learning algorithm that can be used to learn solution operators of parametric PDEs, an operator learning network with three separate subnetworks is designed to estimate the subsurface density anomalies, the details of which are presented in the next section.

2.2.2. Learning Operators With DeepONets

DeepONet is a powerful tool for solving parametric PDEs. The parameters here can be, among others, the shape of the computational domain, initial conditions, boundary conditions, source terms, and variable coefficients in the equation. In solving such equations, traditional solvers such as finite difference methods and finite element methods require repeated simulations for different parameters. In contrast, DeepONet uses the neural network to estimate the PDE solution for different parameters in almost real-time as long as the neural network is trained offline since it involves only the forward pass of the network.

The parametric PDE can always be expressed in the following form

$$\mathcal{N}(\boldsymbol{u}, \boldsymbol{s}) = 0,$$

where $\boldsymbol{u}$ refers to the parameters (input function) and $\boldsymbol{s}$ denotes the PDE solution function. This article discusses a case in which a unique solution function $\boldsymbol{s}$ exists for each $\boldsymbol{u}$, and the resulting function $\boldsymbol{s}(\boldsymbol{y}) = G(\boldsymbol{u})(\boldsymbol{y})$ is real for any coordinate $\boldsymbol{y}$. The purpose of DeepONet is to approximate $G(\boldsymbol{u})(\boldsymbol{y})$ with $\boldsymbol{y}$ and $\boldsymbol{u}$ as inputs, for a more detailed description of the DeepONet method, see Appendix B.

In this paper, for a high-dimensional problem, that is, where $G(\boldsymbol{u})(\boldsymbol{y})$ is a vector of dimension greater than 1, we design the network with the structure shown in Figure 1, which consists of three separate sub-networks. Net 1 and Net 2 are the same as the original DeepONet, and Net 3 is a "merge net" with $\boldsymbol{f} \otimes \boldsymbol{\phi}$ as input to predict $G(\boldsymbol{u})(\boldsymbol{y})$, where $\boldsymbol{f}$ and $\boldsymbol{\phi}$ are the outputs of Net 1 and Net 2, respectively, and $\otimes$ denotes the element-wise multiplication operator.

According to the construction process of the reduced basis, we can know that the projection coefficient corresponding to the density anomaly profile $\boldsymbol{\rho}_{z_j}^a$ at depth z_j is

$$\boldsymbol{\nu}_{z_j} = \mathbf{V}^{\mathrm{T}}\boldsymbol{\rho}_{z_j}^a, \tag{7}$$

in which the projection coefficient is the optimal choice to the density anomaly profile in the reduced-order space. And the density anomaly can be estimated as $\boldsymbol{\rho}_{z_j}^a \approx \mathbf{V}\mathbf{V}^{\mathrm{T}}\boldsymbol{\rho}_{z_j}^a = \mathbf{V}\boldsymbol{\nu}_{z_j}$. Therefore, precise estimation of subsurface temperature anomalies can be achieved through accurate approximation of the projection coefficients.

Based on the collected density anomaly profiles, the DeepONet can be trained to predict the projection coefficients $\boldsymbol{\nu}$. Below, we provide specific details of DeepONet for predicting $\boldsymbol{\nu}$, which is called the projection-driven operator learning network (PDO-Net). In this case, the parameters of the PDO-Net can be thought of as projection coefficient of SSDA, that is, $\boldsymbol{u} = \boldsymbol{\nu}_{z_0}$, and $\boldsymbol{y} = h = z/500$. The output of PDO-Net is denoted as $G_{\mathrm{PDO-Net}}$, which can be learned by minimizing the mean squared error loss

$$loss_{\mathrm{PDO\text{-}Net}} = \frac{1}{N_s(m+1)}\sum_{i=1}^{N_s}\sum_{j=0}^{m}\left|G_{\mathrm{PDO\text{-}Net}}\left(\boldsymbol{\nu}_{z_0}^i\right)(h_j) - \boldsymbol{\nu}_{z_j}^i\right|^2.$$

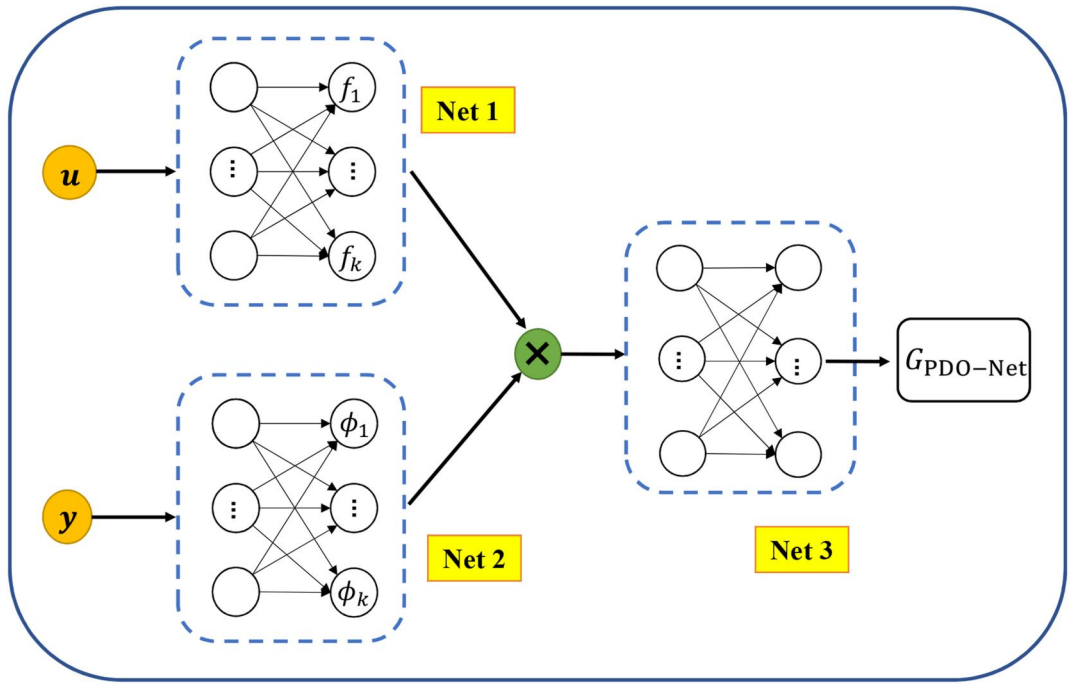

Figure 1. Structure of PDO-Net for estimate subsurface density projection coefficient.

PDO-Net is a data-driven model that infers the relationship between sea surface density projection coefficient and depth to subsurface density projection coefficient from training data, and its prediction may be physically inconsistent. To incorporate the physical laws behind the data into the training process, a physics-constrained machine learning model, PIO-Net, is devised and will be comprehensively elucidated in the forthcoming section.

2.2.3. Physical-Informed Operator Learning Network

Building a credible model that provides accurate and fast predictions while remaining interpretable presents a significant challenge. PINN has emerged as a promising approach to address this challenge, which encode the PDEs into the loss function of the neural network using automatic differentiation to ensure that the physical laws are included in the model training. Compared with traditional numerical techniques for solving PDEs, PINN represents a novel approach that does not require discretization of space-time or complex program implementation. Rather, it aims to employ a neural network to learn a regression function to approximate the solution Ψ of the PDEs. Moreover, there is no limit to the number of residuals, so it overcomes the shortcoming of insufficient training data in training the neural network. In the context of inferring solutions of PDEs using PINNs and DeepONets, a new deep learning framework for ROM is proposed, namely PIO-Net, with input $\boldsymbol{u} = f^{-1}\boldsymbol{\nu}_{z_0}$ and $\boldsymbol{y} = h$, which is designed to predict the variable $\boldsymbol{\gamma}$. The architecture of PIO-Net is shown in Figure 2, and the output of PIO-Net refers to $G_{\text{PIO-Net}}$ (G_{PI}), compared with PDO-Net, PIO-Net adds two feedforward neural networks, which are used to predict the unknown functions g_1, g_2, and g_3. In this case, the loss function of the network can be defined as

$$\mathcal{L}(\theta) = \mathcal{L}_{\text{operator}}(\theta) + \mathcal{L}_{\text{physics}}(\theta), \tag{8}$$

where $\mathcal{L}_{\text{operator}}(\theta)$ is the mean squared error between the network output and the projection coefficient of the density anomaly profile, given by

$$\mathcal{L}_{\text{operator}}(\theta) = \frac{1}{N_s m}\sum_{i=1}^{N_s}\sum_{j=1}^{m}\left| -f\frac{dG_{\text{PI}}\left(f^{-1}\boldsymbol{\nu}^i_{z_0}\right)(h_j)}{dh} - \boldsymbol{\nu}^i_{z_j}\right|^2, \tag{9}$$

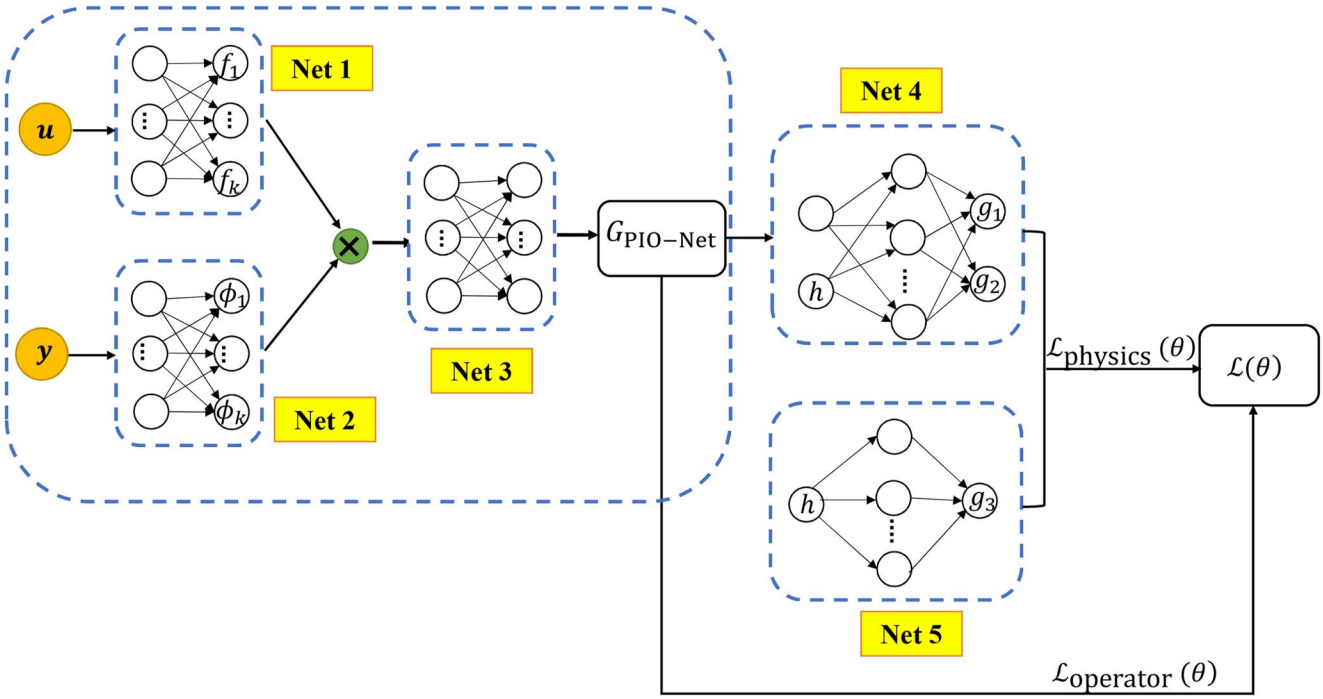

Figure 2. Structure of PIO-Net.

which is used to evaluate the network output against the coefficients of density anomaly profiles. By minimizing Equation 9 during model training enforces known data constraints, allowing model output to fit to the function $\boldsymbol{\nu} \approx -f\frac{dG_{\text{PI}}}{dh}$. This encoding of the relationships within the data results in a more accurate fit of PIO-Net output to the data, and producing a reduced-order model adapted to the data. And $\mathcal{L}_{\text{physics}}(\theta)$ is the mean squared residual error of the reduced-order equation Equation 2 on a set of residual points to encode physical laws and prior physical knowledge into the PIO-Net with the specific expressions are shown in Appendix C. By minimizing the loss $\mathcal{L}_{\text{physics}}(\theta)$, the network's input and output approximately satisfy the governing equation Equation 2 while maintaining physical consistency, thus being able to provide strong theoretical constraints on the basis of the data.

The loss function of the PIO-Net contains not only the difference between the output and the target, but also the residuals of the governing equation, therefore, PIO-Net is reliant on both the provided data and the established physical equation. The setting of the loss function Equation 8 transfers the difficulty of solving the equation to the optimization of the PIO-Net. With the neural network's powerful ability to approximate nonlinear functions, PIO-Net can reliably estimate the physical model while achieving accurate predictions. For huge data sets, a computationally efficient mini-batch setting can be readily employed in optimization algorithms (Goodfellow et al., 2016).

The proposed reduced-order model in this work not only reduces the amount of computation, but also avoids the problem of the large gap between the dimensions of spatial coordinates.

2.3. Experimental Setup

The flowchart for subsurface density anomalies estimation via PIO-Net is shown in Figure 3. First, according to the collected density anomaly profiles, we calculate the reduced basis matrix $\mathbf{V}$ and the corresponding projection coefficients by Equation 7. By combining the matrix $\mathbf{V}$ and the QG equation, the reduced-order equation Equation 2 can be constructed. Then, PIO-Net is trained using a defined loss function Equation 8 based on the collected projection coefficients and the reduced-order equation Equation 2. The inputs to PIO-Net consist of $\boldsymbol{u} = f^{-1}\boldsymbol{\nu}_{z_0}$ and $\boldsymbol{y} = h$, while the outputs are represented by G_{PI}, which is used to predict $\boldsymbol{\gamma}$ and can be utilized to estimate the projection coefficient as $\boldsymbol{\nu} \approx -f\frac{dG_{\text{PI}}}{dh}$. Once PIO-Net has been trained, estimating the density anomaly

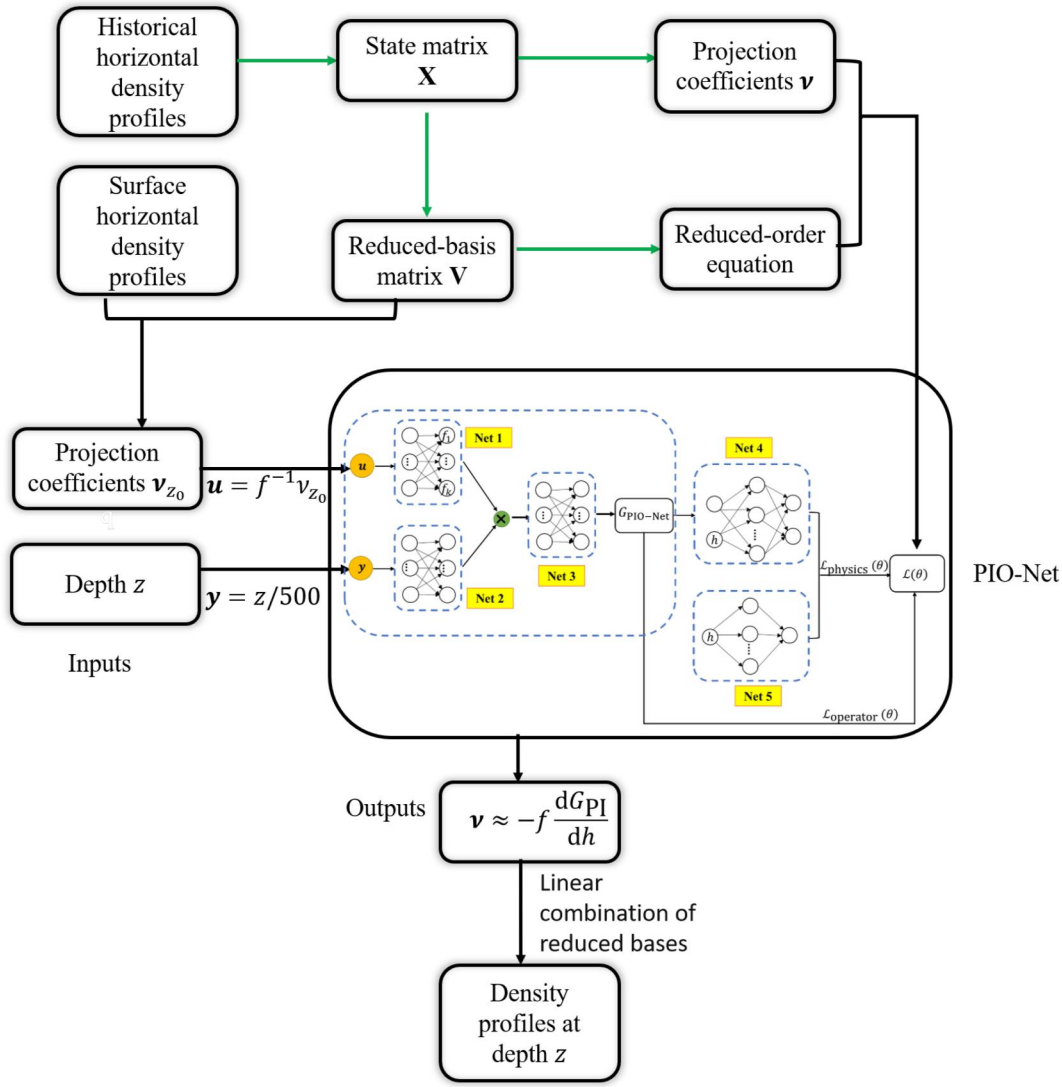

Figure 3. The flowchart for subsurface density anomalies estimation via PIO-Net.

profile at a new depth z only requires the projection coefficient of sea surface density anomaly and the depth z. This estimation is achieved through a linear combination of the reduced-basis.

The structure of each sub-network is shown in Table 2, both PIO-Net and PDO-Net are trained by minimizing the loss function for 10, 000 iterations of gradient descent using the Adam optimizer with a batch size of 100, and the number of reduced-order basis (modes) of $n = 45$. The numbers of training and testing data are presented in Table 1. In Section 4.1, based on the MPI-OM's simulations of temperature-salt data in 2011, we uniformly select half of the calculated density data as the training set and the remaining half as the test set, resulting in $N_s = 189$ parameters inputs for Net 1. Additionally, for each parameter $\boldsymbol{u}^i = f^{-1}\boldsymbol{\nu}^i_{z_0}$ of PIO-Net, the depth residuals are the same, 100 points $\left\{z_j^{(i)}\right\}_{j=1}^{100}$ uniformly distributed within $[-200, -5]$ and 100 points $\left\{z_j^{(i)}\right\}_{j=101}^{200}$ uniformly distributed within $[-1{,}250, -200]$ are selected, thus $N_t = 200$ residual points $\left\{h_j^{(i)}\right\}_{j=1}^{N_t} = \left\{z_j^{(i)}/500\right\}_{j=1}^{N_t}$ are chosen as inputs to Net 2. In this case, the training time for PIO-Net is 21,273.52 s using a laptop with four Intel(R) Core(TM) i7-6770 CPU @ 3.40 GHz and 8 GB of memory. The prediction time to achieve a 3D field is 0.98 s. Once the model is well trained, refactoring can be accomplished quickly as the time required is much less than the training time. Additionally, we find that a majority of the training time is consumed by second-order derivation of

Table 1
The Numbers of Training and Testing Data

Section	Number of training parameters	Number of residual points	Number of test parameters
Section 4.1	189	200	176
Section 4.2	316	200	1,144

variables related to depth. If the automatic differentiation for the second derivative is replaced by numerical differentiation

$$\left.\frac{d^2G_{\mathrm{PI}}}{dh^2}\right|_{h=h_j} = \left(\left.\frac{dG_{\mathrm{PI}}}{dh}\right|_{h=h_j} - \left.\frac{dG_{\mathrm{PI}}}{dh}\right|_{h=h_j-10^{-5}}\right)/10^{-5},$$

the training time reduces significantly without compromising accuracy, only 1828.13 s are necessary. In addition, the impact of network size on accuracy of PIO-Net is provided in Section 4.1, some visualizations and the impact of batch size are provided in Appendix D.

In Section 4.2, one-fifth of the data from each of the four regions in 2011 are uniformly selected as the training set, with the remainder used as the test set. Furthermore, 100 points $\left\{z_j^{(i)}\right\}_{j=1}^{100}$ uniformly distributed in range of [−200, −5] and another 100 points $\left\{z_j^{(i)}\right\}_{j=101}^{200}$ uniformly distributed in range of [−1,625, −200] are selected, resulting in a total of $N_t = 200$ residuals $\left\{h_j^{(i)}\right\}_{j=1}^{N_t} = \left\{z_j^{(i)}/500\right\}_{j=1}^{N_t}$ as inputs for the Net 2.

3. Data Preprocess

The data for training the PIO-Net model includes Hybrid Coordinate Ocean Model (HYCOM)/Navy Coupled Ocean Data Assimilation (NCODA) reanalysis data (Chassignet et al., 2007) and simulation data. The MPI-OM (Marsland et al., 2003) is selected to simulation ocean state as the test bed for our study, which is the successor to the Hamburg Ocean Primitive Equation Model (Wolff et al., 1997). In this study, the model configuration and settings that align with those employed by Song and Chen (2020), except for tidal force. The configuration of the model is briefly introduced. An improved PP parameterization scheme (Pacanowski & Philander, 1981) is incorporated into the vertical mixing scheme, on which the wind-induced mixing parameter is added, as the PP parameterization scheme underestimates wind-induced mixing at the sea surface. The parametrization scheme of GM95 (Gent et al., 1995) is adopted for vortex mixing at subgrid scale. Furthermore, to more accurately characterize the transport phenomena within the bottom boundary layer, the model integrates the parameterization scheme for the bottom boundary layer, as proposed by Legutke and Maier-Reimer (2002), allows for a more precise representation of density-driven flow processes. In the horizontal dimension, the curved orthogonal Arakawa-C grid is employed with grid that range from 3 to 100 km. The two extreme points of grid resolution are on the continents of China and Australia, resulting in a denser grid configuration in regions such as the South China Sea and Northwest Pacific. In the vertical dimension, a Z-coordinate system with 40 layers is employed, while the time step is set to 300 s. The model's topography is derived from ETOPO2 (Center, 2006) through interpolation. To establish a stable ocean circulation state, the model is initially run for a period of 10 years. The initial temperature and salinity fields are obtained by interpolating the PHC data set (Steele et al., 2001), and surface forcing is interpolated from the German Ocean Model Intercomparison Project (OMIP) database (Röske, 2006). Throughout the model computation, temperature and salinity fields at the sea surface are relaxed toward the climate state data, ensuring a more accurate representation of the climatological ocean circulation. The model is initialized using the final results of a 10-year climatological simulation, with the forcing fields derived from the authentic 2011 JRA55 data set (Agency, 2013). The upper surface forcing incorporates momentum, heat, and freshwater fluxes, with variables such as wind stress, 10-m wind speed, cloud cover, short-wave radiation, precipitation, sea surface pressure, sea surface air temperature, and 2-m dew point temperature. It should be noted that both upward long-wave radiation and seawater evaporation are calculated by the model, and the influence of runoff is excluded and set to zero. Subsequently, the model is integrated for an additional 10-year period, and the output from the final year is utilized as the model data for this study. And we utilize temperature and salinity data from MPI-OM to calculate the density (Fofonoff & Millard, 1983). Furthermore, SSHA, SSDA, and vertical stratification profile $N^2(z)$ are taken as input data in the application of the different reconstruction algorithms. To obtain

Table 2
Neural Network Architectures

Network number	Net 1	Net 2	Net 3	Net 4	Net 5
Number of hidden layers	4	3	3	2	2
Number of neurons in output layer	40	40	45	2	1
Number of neurons in hidden layer	40	40	40	40	40
Activation function	tanh	tanh	tanh	tanh	tanh
Inputs	$f^{-1}\boldsymbol{\nu}_{z_0}$	h	$\boldsymbol{f}\otimes\boldsymbol{\phi}$	$\frac{\overline{dy}}{dh}$, h	h
Outputs	$\boldsymbol{f}$	$\boldsymbol{\phi}$	G_{PI}	g_1, g_2	g_3

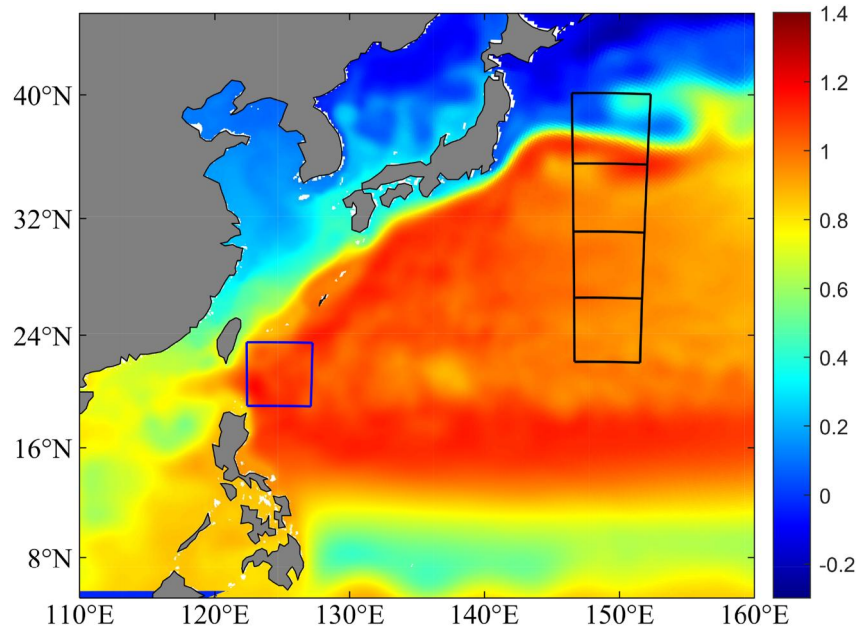

Figure 4. Annual mean surface height (m) in the MPI-OM simulation, with the five tested regions identified by the corresponding boxes.

anomaly fields, we remove the large-scale background information from the SSD and SSH fields. The respective mean monthly data are considered as background fields in our analysis. The area-averaged BV frequency $N^2(z)$ is calculated as:

$$N^2(z) = -\frac{g}{\rho_0}\frac{\partial \overline{\rho}}{\partial z},$$

where $\overline{\rho}$ is the horizontal mean of density in target regions. Several subdomains are employed to validate the proposed method. The first subdomain (marked by the blue boxes in Figure 4) encompasses the eastern side of the Luzon Strait, a region characterized by vigorous eddy activity. Furthermore, to evaluate the performance of the proposed method in different regions, four distinct areas (marked by the black boxes in Figure 4) with varying latitudes are selected and a model is trained for prediction.

This study uses the Level 4 sea surface salinity product produced by the International Pacific Research Center (Melnichenko et al., 2016), the daily sea surface temperature produced by the Operational Sea Surface Temperature and Ice Analysis system (Donlon et al., 2012), the in situ data set from the global ocean Array for Real-time Geostrophic Oceanography (Argo) temperature and salinity profile scatter data set (Z. Liu et al., 2021), and the monthly World Ocean Atlas 2018 (WOA18) (Boyer et al., 2018) in the subdomain presented in Section 4.1 to evaluate the effectiveness of PIO-Net for robust validation with in situ data.

4. Results

4.1. Density Reconstruction in a Region

To demonstrate PIO-Net's capability of reconstructing density, several metrics have been defined:

- The approximation error at fixed depth z of the prediction $\hat{\boldsymbol{\rho}}^a(z)$, measured by the daily RMSE with respect to the numerical solution $\boldsymbol{\rho}^a(z)$:

$$\mathrm{RMSE} = \sqrt{\frac{1}{N_h}\left\|\boldsymbol{\rho}_i^a(z) - \hat{\boldsymbol{\rho}}_i^a(z)\right\|_2^2},$$

with N_h is the number of horizontal grid points in the reconstructed region.

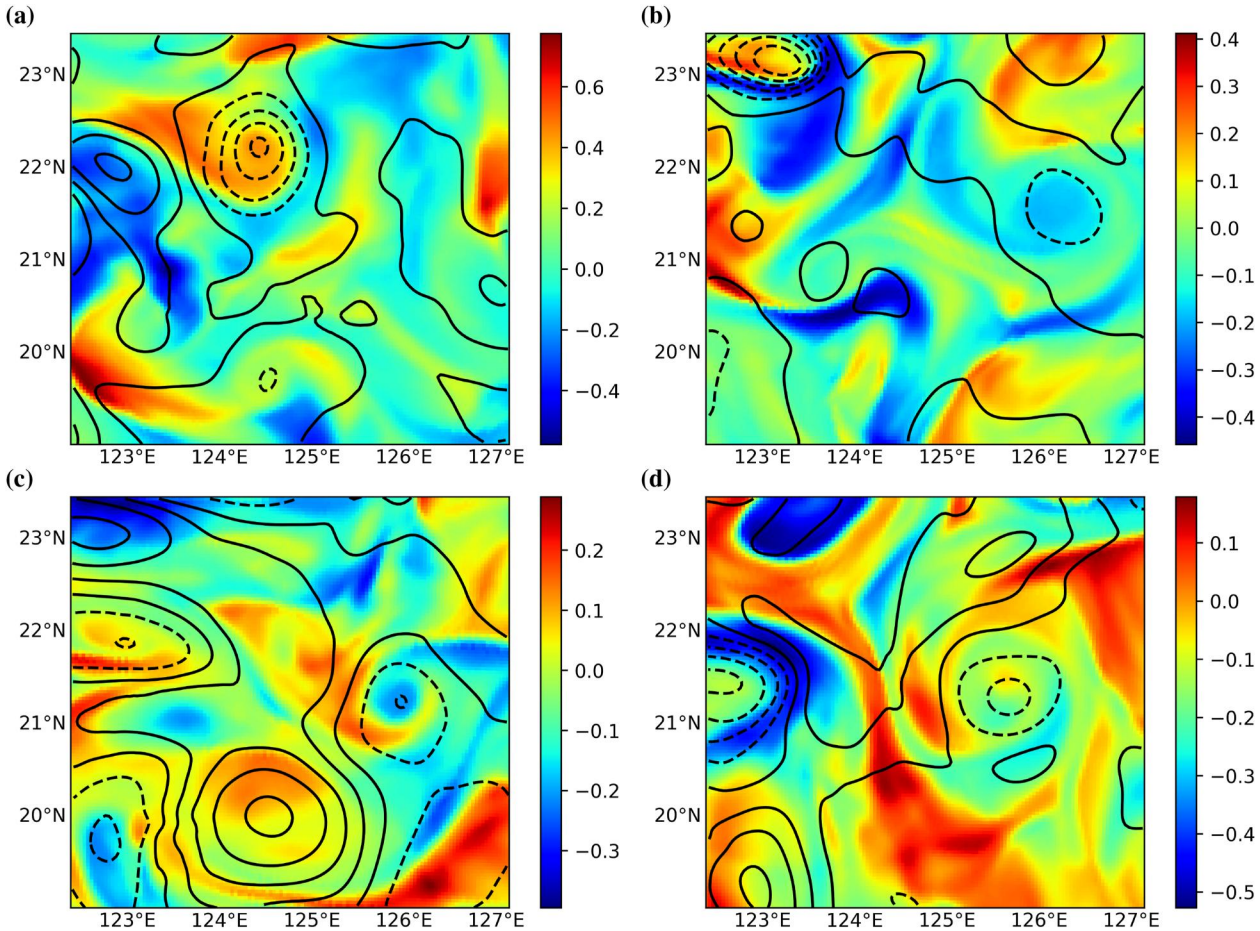

Figure 5. (a) SSDA (shading) and SSHA (contour) from MPI-OM in 16th February 2011. Panels (b)–(d) are the same as panel (a), respectively, except in 15th May 2011, in 15th August 2011 and in 15th November 2011.

- The daily RMS at fixed depth z of the density anomaly $\boldsymbol{\rho}^a(z)$:

$$\text{RMS} = \sqrt{\frac{1}{N_h}\left\|\boldsymbol{\rho}_i^a(z)\right\|_2^2}.$$

and the averaged RMSE and averaged RMS are computed as the mean of their respective values across all testing instances.

The objective of PIO-Net and PDO-Net is to learn the relationship between surface density and the interior density of the ocean. We first consider the subfield presented in Figure 5, which illustrate that the SSDA (shading) and SSHA (contour) at four different times in this region show a weak spatial correlation, indicating that SSD is decoupled from the internal dynamics. Thus, the inversion results of isQG in this region are likely to be unsatisfactory. This is further confirmed by Figure 6, which shows the vertical distribution of the averaged root mean square (RMS) of the density anomalies for different methods. It can be seen that the seasonally averaged RMS of the density anomaly of isQG solution (magenta dash-dot line) is similar to that of the MPI-OM in spring and winter, but there is a significant difference in summer and autumn, especially between −400 and −50 m in the upper layers, which significantly underestimates the true state. This may be attributed to the fact that the interior solution cannot be fully represented by the barotropic mode and the first baroclinic mode. Additionally, the vertical distribution of the averaged RMS of the density anomaly reveals that the vertical variation is stronger in summer, which is likely to be the cause of the poor performance of isQG method.

As shown in Figure 6, PIO-Net (red dashed line) has significant performance in achieving satisfactory reconstruction for all time, with the averaged RMS of PIO-Net not only similar to that of MPI-OM, but their amplitudes

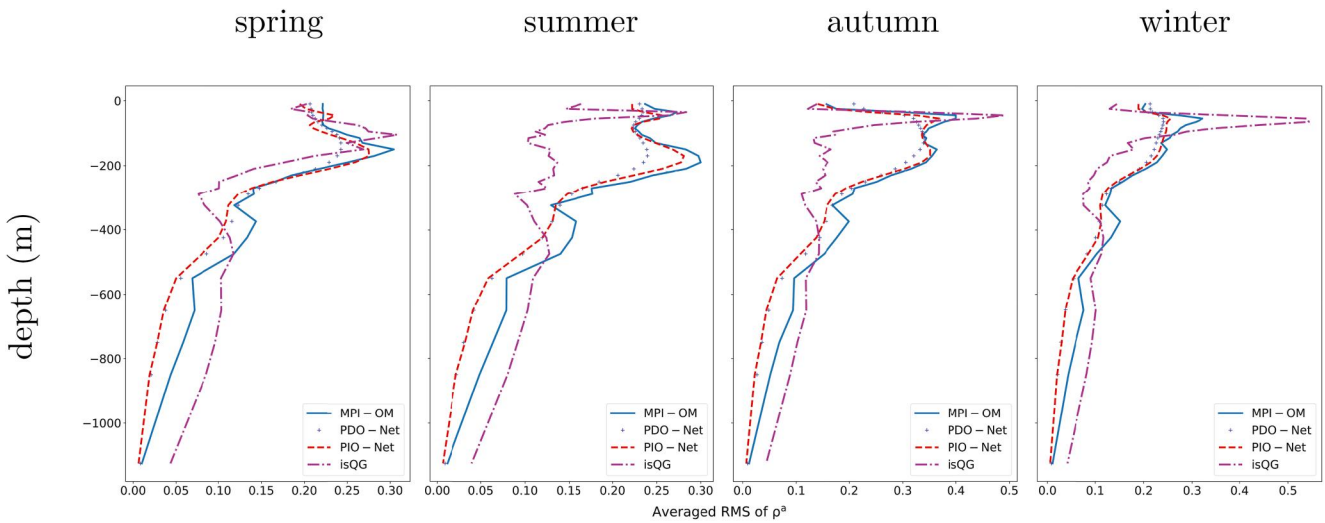

Figure 6. The seasonally averaged RMS of the density anomaly derived from MPI-OM data (blue solid line), from PDO-Net solution (blue dotted line), from PIO-Net solution (red dashed line), and from isQG solution (magenta dash-dot line) in different times.

are pretty comparable. In particular, PIO-Net effectively captures the subsurface maxima, although with a slightly lower amplitude than MPI-OM. The incorporation of EOF allows the selection of multiple functions to represent the density anomaly profile, overcoming the shortcomings of isQG modes in reconstructing the fine structure of the ocean. Furthermore, the imposition of physical constraints allows PIO-Net not to rely solely on sea surface information to reconstruct the internal structure of the ocean. This results in reliable predictions even when the correlation between SSHA and SSDA is weak. By analyzing the averaged RMS of the density anomaly of PDO-Net (Figure 6, blue dotted line), it can be inferred that PDO-Net is able to capture the vertical structure of the anomaly. However, its amplitude is significantly smaller than that of MPI-OM. The performance of PDO-Net is generally better than that of isQG, but not as excellent as that of PIO-Net, especially at depths above −230 m. This implies that surface fitting alone is not sufficient to capture the complicated dynamical processes in the ocean, whereas machine learning based on physical laws has considerable advantages.

The density anomaly distributions at −210 m from three reconstruction methods are shown in Figure 7. On the horizontal plane, the structure of PIO-Net's solution is in a good agreement with MPI-OM. Although the inversion of PIO-Net is generally weaker than MPI-OM, it accurately captures the structure and position of the vortex. It can be concluded that PIO-Net successfully inverts the density field of the region. In contrast, the results of PDO-Net in spring and autumn are obviously less accurate than those of PIO-Net, and the structure of the density field inverted by isQG is similar to the actual state, but the amplitudes are weaker than MPI-OM, particularly in summer and autumn, which further confirms the result of Figure 6.

To further analyze the prediction performance of PIO-Net, we average the density anomaly with respect to depth on 16th February and present the logarithmic spectrum diagram corresponding in Figure 8. The similar amplitude distributions can be observed between PIO-Net and MPI-OM, reinforcing the reliability of PIO-Net's predictions. In addition, Figure 9 shows the distribution of 3D density anomaly fields from different models, and similar patter can be observed in both models. Although the details of PIO-Net are lacking due to the low dimensional approximation, the overall approximate structure is satisfactory. These results further demonstrate the reliability and accuracy of PIO-Net.

The averaged root mean square error (RMSE) is employed as a performance metric to assess the reconstruction performance of the different methods. The averaged RMSEs of the three methods shown in the left of Figure 10, exhibit a similar variation with depth. From those figures, one can easily observe that the RMSEs increase rapidly from the sea surface and reach a maximum around −100 m, before gradually decreasing and becoming stable below −300 m. Furthermore, at almost depth levels, the averaged RMSE of PIO-Net is the lowest. In particular, the averaged RMSE of PIO-Net has a maximum value of 0.17, PDO-Net has a maximum averaged RMSE of 0.23, and isQG has a maximum averaged RMSE of 0.31. In total, these results indicate that PIO-Net shows superior performance compared with the other models. Additionally, the vertical distribution of the standard deviation

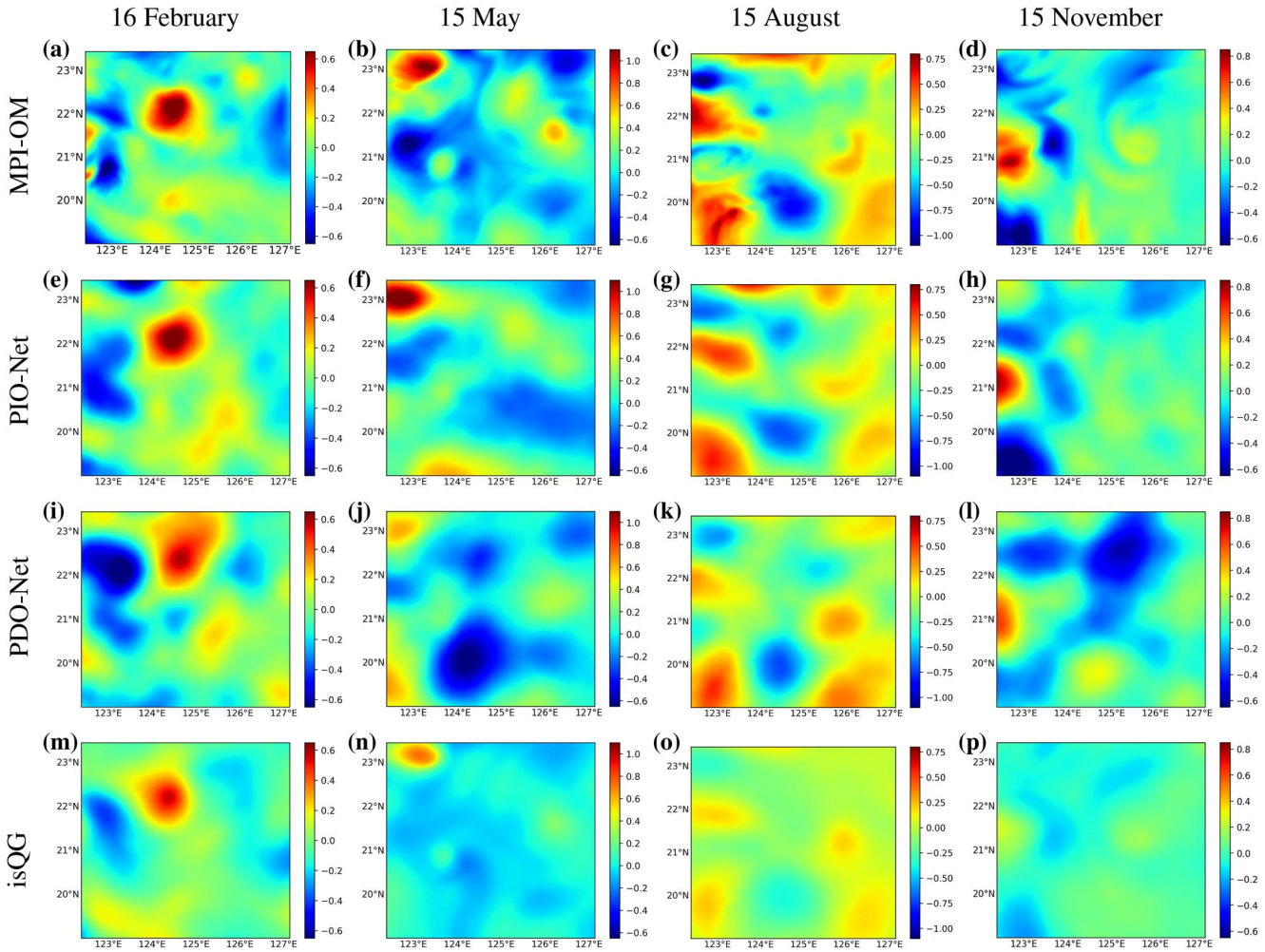

Figure 7. Horizontal distributions of density anomaly at −210 m from MPI-OM (a)–(d), PIO-Net (e)–(h), PDO-Net (i)–(l), and isQG (m)–(p) in four days.

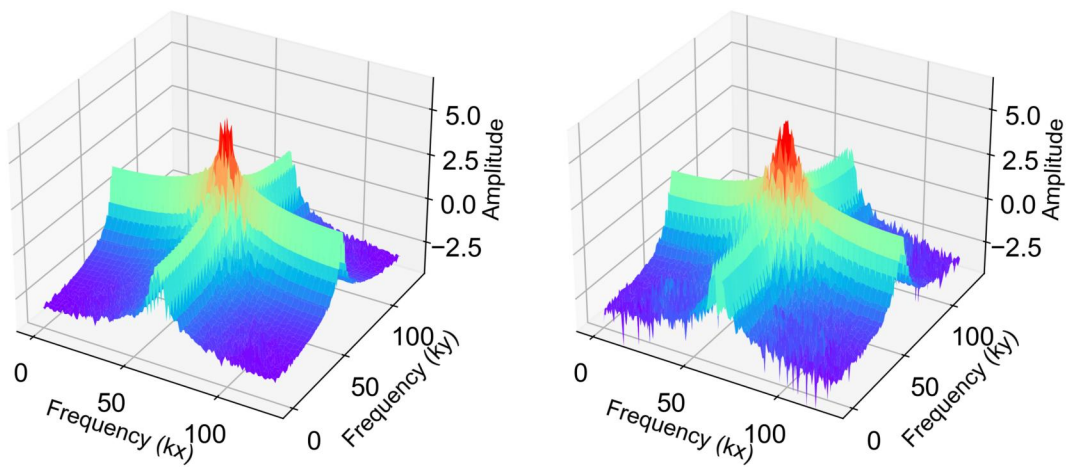

Figure 8. Statistical distribution of the spectral amplitudes from PIO-Net (left) and from MPI-OM (right) calculated in 16th February 2011.

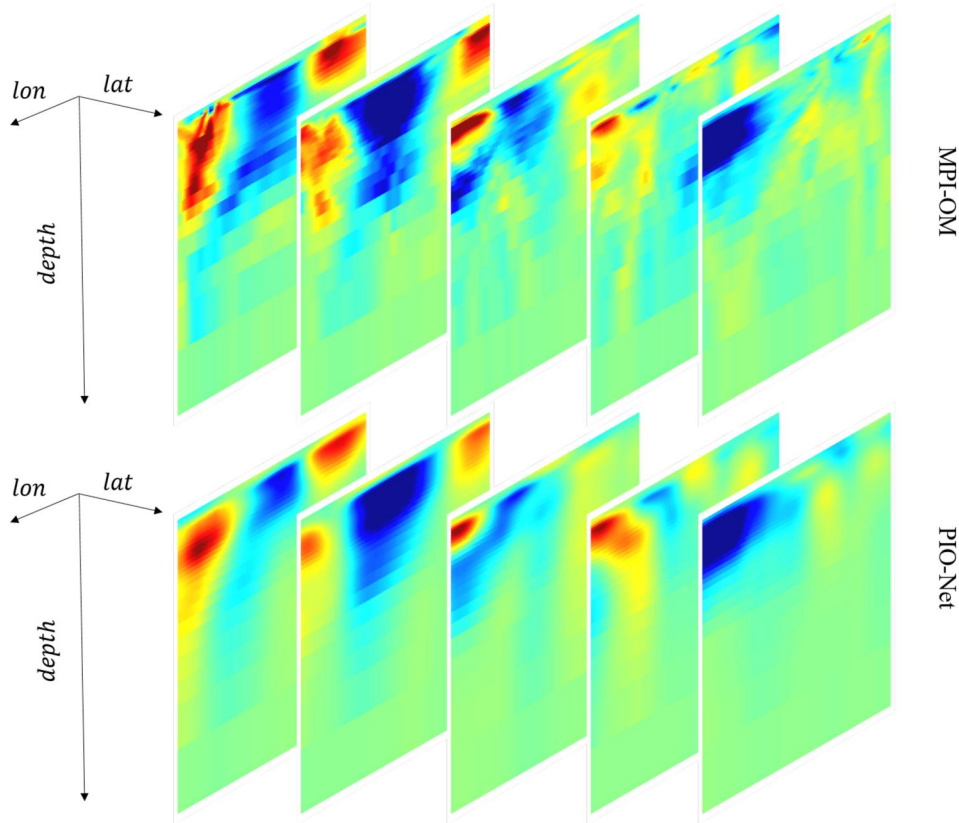

Figure 9. 3D density field distribution from MPI-OM (top) and from PIO-Net (bottom) calculated in 16th February 2011.

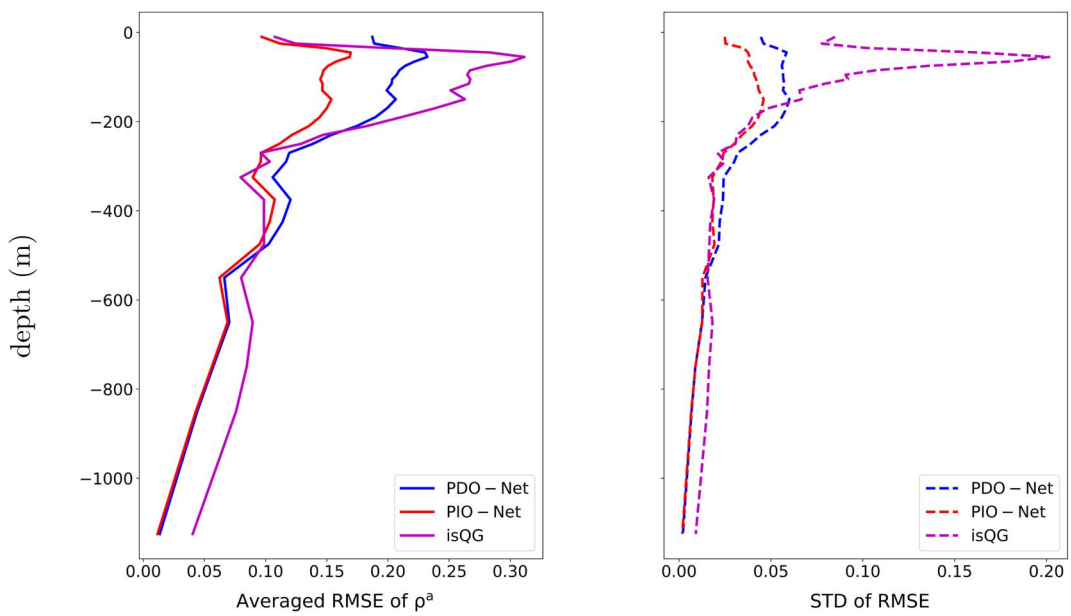

Figure 10. The averaged RMSE of the density anomaly derived from PDO-Net solution (blue solid line), from PIO-Net solution (red solid line), and from isQG solution (magenta solid line). And the STD of the RMSE derived from PDO-Net solution (blue dashed line), from PIO-Net solution (red dashed line), and from isQG solution (magenta dashed line).

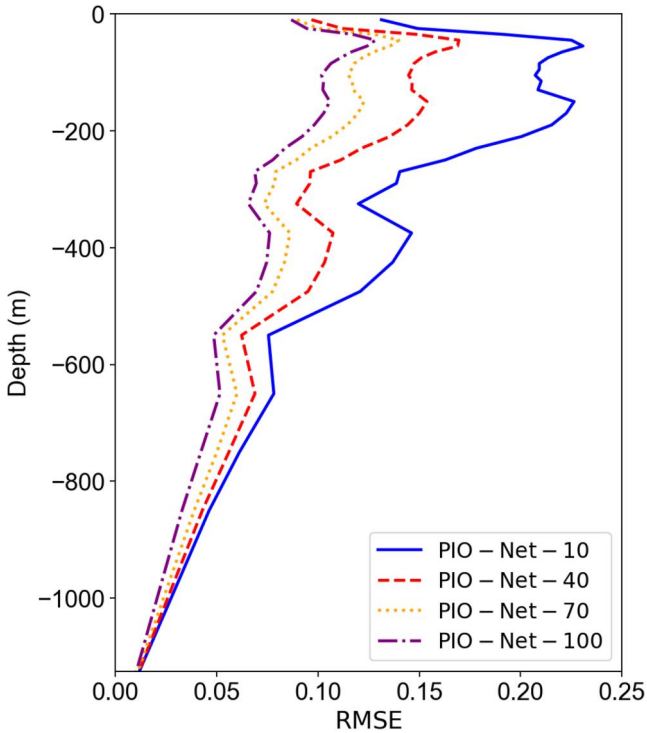

Figure 11. The averaged RMSE of the density anomaly derived from PIO-Net with different neurons.

(STD) of RMSEs of different methods plotted in the right of Figure 10, which shows that PIO-Net has stronger robustness, as its STD is the smallest at all depths. In contrast, isQG, especially above −200 m, has the largest variance, reaching 0.2 around −55 m, which again confirms the results of Figure 6 that isQG has a considerable difference in the accuracy of reconstruction at different times.

To investigate the impact of network size on accuracy of PIO-Net, an experiment are conducted, which is based on networks with varying hidden layer widths. The number of hidden layers is fixed, while the number of neurons in each hidden layer varied from 10 to 100 with a step increment of 10. The averaged RMSE of four different models is depicted in Figure 11. The results show that although the RMSE curves exhibit similar trends, increasing the number of neurons leads to improved accuracy. This is because larger networks have a stronger ability to learn and capture more comprehensive relationships. However, it is important to note that the performance becomes less significant once the complexity reaches a certain threshold. To enhance the clarity of the model's accuracy, we further average the calculated RMSE with respect to the depth. The results of the prediction error with respect to the number of neurons are depicted in Figure 12, which unequivocally demonstrates that a more complex network achieves higher accuracy, particularly when comparing models with 100 neurons to those with 10 neurons. It reveals a significant 47% decrease in error, indicating that careful exploration and testing of our model structure can lead to more accurate results. An additional experiments is performed to investigate the influence of network size on PDO-Net, and the prediction errors are shown in Figure 12. Similar conclusions to those of PIO-Net can be drawn. Notably, we observe that PIO-Net consistently achieves superior accuracy compared with PDO-Net when employing the same network structure settings. Furthermore, the PIO-Net's utilization of a larger training data set to impose physical constraints, enabling the imposition of constraints and facilitating enhanced learning of the intricate relationships among data points. The results obtained from investigating the network structure clearly demonstrate that the careful selection of appropriate network settings can significantly enhance the performance of PIO-Net in forecasting. Moreover, it is evident that the deep learning-based approach holds greater potential compared with the dynamic-based approach.

4.2. Density Reconstruction in Different Regions

Four regions of different latitudes are selected to investigate the ability of PIO-Net to reconstruct the density of different regions. The Coriolis parameter of each region is taken as the region averaged, which is entered into PIO-Net when calculating the loss function.

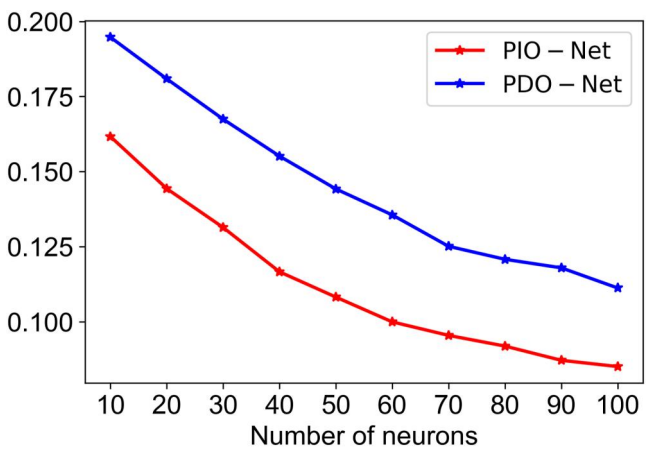

Figure 12. Prediction error with different neurons.

The SSHA (contour) and SSDA (shading) of the four regions on 1st September are shown in Figure 13, where the correlation between SSHA and SSDA is not very strong. As shown in Figure 14, the averaged RMSs of the density anomalies from different methods are calculated. It can be observed that the predictions of PIO-Net match well with those of MPI-OM. And PIO-Net is successful in reconstructing the vertical structure of MPI-OM data across all regions. Specifically, the upper layer structure (above −200 m) of region 1 is accurately captured by PIO-Net, while other methods do not. However, the extreme values at around −400 m of all regions are not accurately reconstructed. This discrepancy is more obvious in region 1 and region 3, where the solutions of PIO-Net and MPI-OM are significantly different between −400 m and about −800 m, which may be attributed to the relatively weak fluctuation of SSDA in these two regions.

PIO-Net's performance is acceptable in region 1 and region 4, where changes in the vertical structure of the density anomalies in the MPI-OM data are also

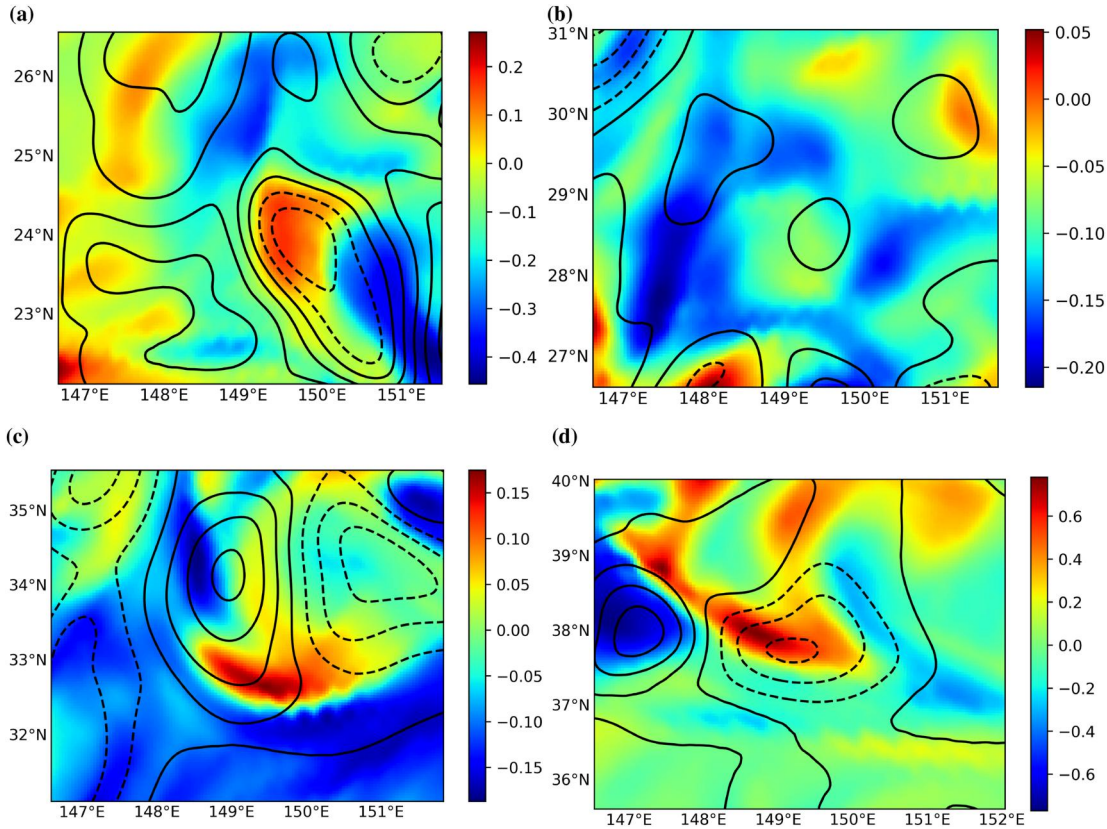

Figure 13. SSDA (shading) and SSHA (contour) from MPI-OM in 1st September 2011 for (a) region 1, (b) region 2, (c) region 3 and (d) region 4.

be observed in the PIO-Net solution, with all maxima being successfully retrieved, albeit with a slightly reduced amplitude. The variations in the vertical structure of the averaged RMS of PDO-Net and PIO-Net are quite similar, but the two diverge above −75 m. In region 1 and region 4, PDO-Net fails to capture the extreme values of the averaged RMS of density anomaly, and the averaged RMS of density anomaly reconstructed by PDO-Net decreases with increasing depth. This further demonstrates that fitting based on surface information alone is insufficient, and the incorporation of dynamic laws into the reconstruction algorithm is essential to generate more

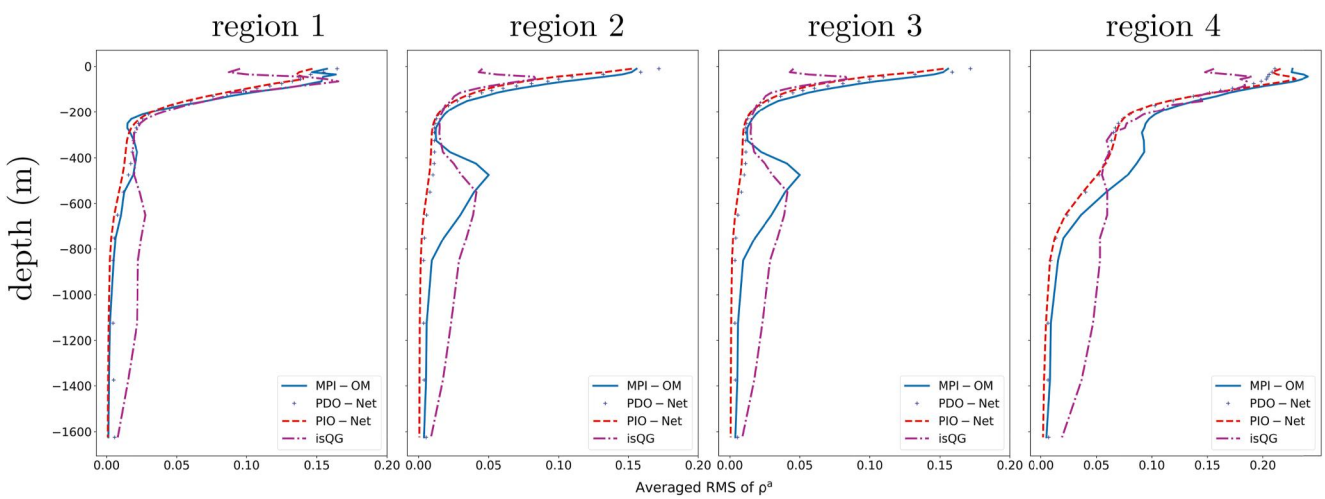

Figure 14. The averaged RMS of the density anomaly derived from MPI-OM data (blue solid line), from PDO-Net solution (blue dotted line), from PIO-Net solution (red dashed line), and from isQG solution (magenta dash-dot line) for different regions.

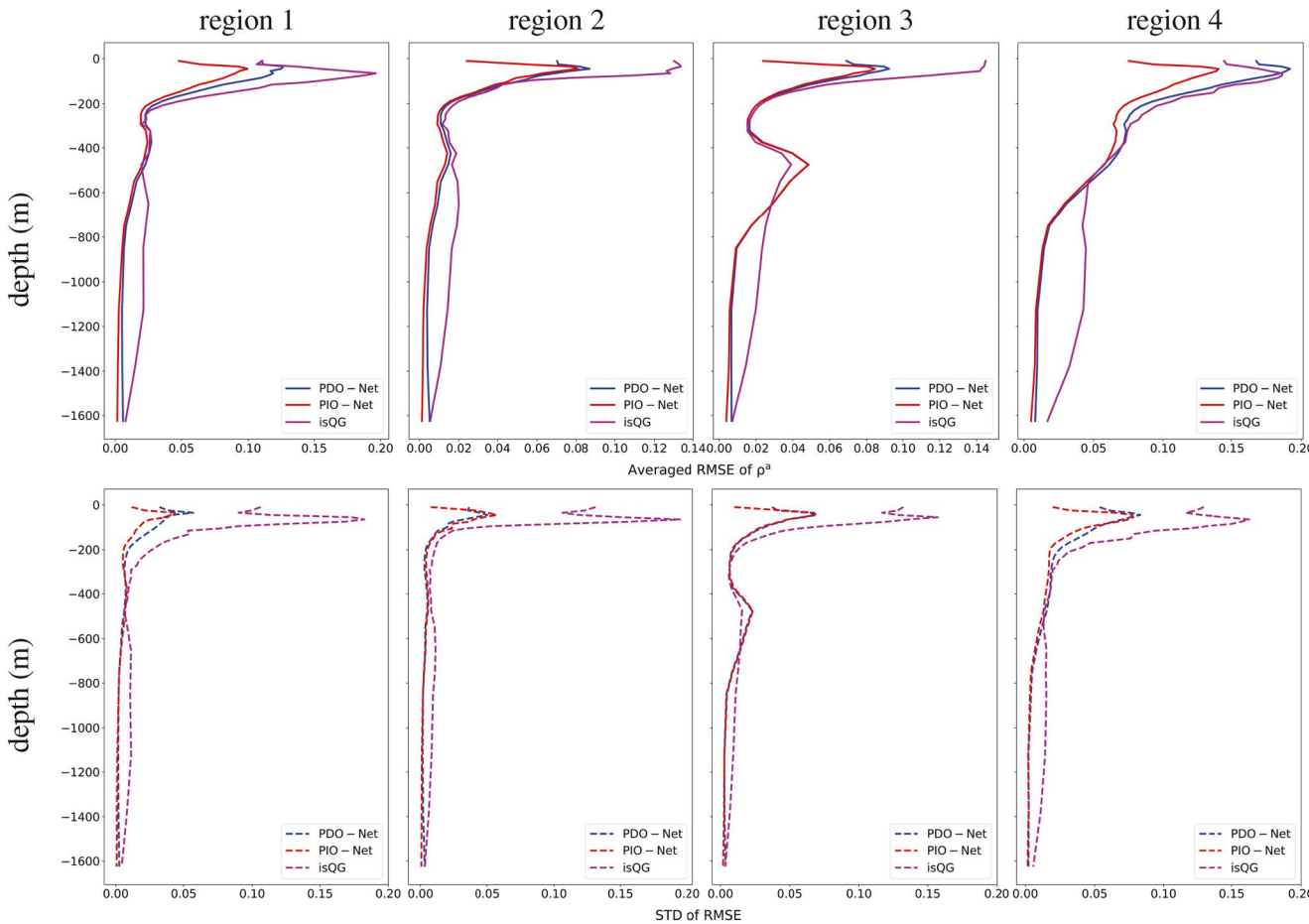

Figure 15. The averaged RMSE of the density anomaly derived from PDO-Net (blue solid line), from PIO-Net (red solid line), and from isQG (magenta solid line) for different regions (top row). STD of the RMSE derived from PDO-Net, from PIO-Net (red dashed line), and from isQG (magenta dashed line) for different regions.

precise predictions. Different from PIO-Net, isQG does not perform well in the upper ocean, which may be attributed to the lack of significant signal fluctuation of SSHA. At this point, the data from MPI-OM are not applicable to isQG theory. Nevertheless, in regions 1–3, the trend of the curve of averaged RMS near −400 m is captured by isQG, although with inconsistent locations of extreme points and underestimated or overestimated amplitudes.

From the perspective of the averaged RMSE estimated by different methods (Figure 15), PIO-Net has the smallest averaged RMSE in almost all regions, except for depth from about −325 to −650 m of region 3, where its averaged RMSE is larger than that of isQG. In regions 1–3, PIO-Net and PDO-Net present averaged RMSE curves with similar changes. Above −200 m, the averaged RMSE of both are significantly better than that of isQG, which is consistent with the averaged RMS result in Figure 14. This is because that isQG do not successfully estimate the density anomaly in the upper ocean. At the same time, PDO-Net has been performing no better than PIO-Net, with its averaged RMSE larger than PIO-Net's at all depths, indicating that the introduction of dynamic constraints is successful and can significantly improve the reconstruction accuracy of pure data-driven machine learning algorithm. In region 4, the averaged RMSE curve varies similarly for the three reconstruction methods. However, above −500 m, the PIO-Net reconstruction is significantly better than the other algorithms, with a maximum averaged RMSE of 0.14 (−45 m), whereas the value in PDO-Net reconstruction is 0.19 and is 0.17 in isQG reconstruction. Furthermore, it is evident that the performance of PIO-Net is highly consistent, with similar averaged RMSE curves across different regions. Figure 15 also shows the STD of the different models, where PIO-Net presents the lowest STD, indicating that it has the most stable performance across regions and times. DeepONet based on physical information not only reduces the

averaged RMSE, but also decreases the standard deviation of the reconstructed density profile. Consequently, PIO-Net, a machine learning algorithm based on QG, demonstrates its superiority in both performance and robustness.

5. Discussion and Conclusion

The combination of dynamics-based models and deep learning algorithms is the main goal of current marine science research, but it is currently in its infancy due to the challenge of combining them. In this study, we primarily explore the integration method in deep ocean remote sensing and develop a novel dynamics-constrained deep operator learning reduced-order model framework that incorporates both dynamics-based and machine learning techniques using the reduced-order technique as an attempt to integrate the method. This framework can effectively learn the unknown parameters in the equation, as well as estimate the subsurface density field. The integrated approach benefits from the strengths of both dynamics-based and data-driven methods, with the former providing interpretable background constraints and the latter enabling fast prediction and flexible model building. The ROM is the Galerkin projection of the QG onto reduced-basis extracted from historical data via EOF. An operator learning network is used to predict the subsurface density projections, and feedforward neural networks are used to estimate the unknown parameters in the reduced-order equation, which makes the ROM more adaptable to the data. The loss function is set to the combination of mean square residuals from the reduced-order equation and mean square errors between the model outputs and target outputs, ensuring that both physical and data constraints are accounted for during network training. By only using depth as the independent variable in the reduced-order equation, the training cost of the network is greatly reduced and it avoids untrained models in addressing high-dimensional spatial problems. Using MPI-OM simulation data, we evaluate the results of the proposed method, the purely data-driven method described in the paper, and isQG. The proposed method exhibits the smallest RMSE, indicating higher prediction accuracy than other methods. In addition, the proposed method shows good performance in different regions and at different times, highlighting its stability and robustness.

This paper aims to explore an efficient deep learning model with physical constraints to reconstruct internal data from the surface. However, the proposed method ultimately aims to achieve three-dimensional reconstruction of ocean conditions from remote sensing data, there are still concerns about its practical performance. Therefore, a robust experiment is conducted to evaluate PIO-Net's accuracy compared to climatology. The HYCOM reanalysis data is applied to train the PIO-Net, which is considered to be a more accurate representation of the ocean state than pure simulations. The collected satellite-derived sea surface density data is then input into the well-trained PIO-Net for three-dimensional density reconstruction. To test the model's performance and compare the reconstructed density with the Argo profiles, the reconstructed density and the monthly WOA18 climatology are interpolated to match the observed locations (marked with asterisks in Figure 16), and more detailed model configuration and data descriptions are given in Appendix E. Figure 16 shows the density distribution of various models above −450 m at one location (indicated by red asterisk), illustrating the effectiveness of the proposed method in capturing the trend of density variation with depth. The RMSEs estimated from different models, as shown in Figure 16, reveal that the RMSE of PIO-Net is smaller than that of the climatology at nearly all depths, particularly for densities above −700 m. The PIO-Net model's performance is heavily influenced by the training data, and discrepancies between HYCOM data and observed data are a significant source of reconstruction error. However, the results of this study demonstrate the potential of incorporating dynamic and data-driven models in density reconstruction, thereby enhancing confidence in the subsequent development of the integration methods.

The proposed model has the potential to be applied in two aspects. Firstly, it can assist in detecting and studying the ocean interior's structure and spatio-temporal changes comprehensively by reconstructing density distribution from surface data. This can provide a valuable reference and foundation for the analysis of mesoscale ocean processes and climate change. Secondly, the reconstructed data could improve the accuracy of the numerical model through model initialization and data assimilation, but it is essential to validate its reliability before implementation. The proposed method still has some limitations. Firstly, it is derived under the background of QG framework, which is a valuable tool for understanding large-scale seawater motion, but inadequate when considering finer details phenomena. Although unknown functions learned from the data are introduced into the equation to make the established model closer to the data, it considers only spatially dependent equation and fails to account for spatio-temporal dependencies in ocean system. Furthermore, it does not fully consider the complex interdependence among multiple variables. In addition, in order to reduce the number of unknowns thus reduce the burden of network training, reduced-order modeling technique is used to construct low-dimensional physical

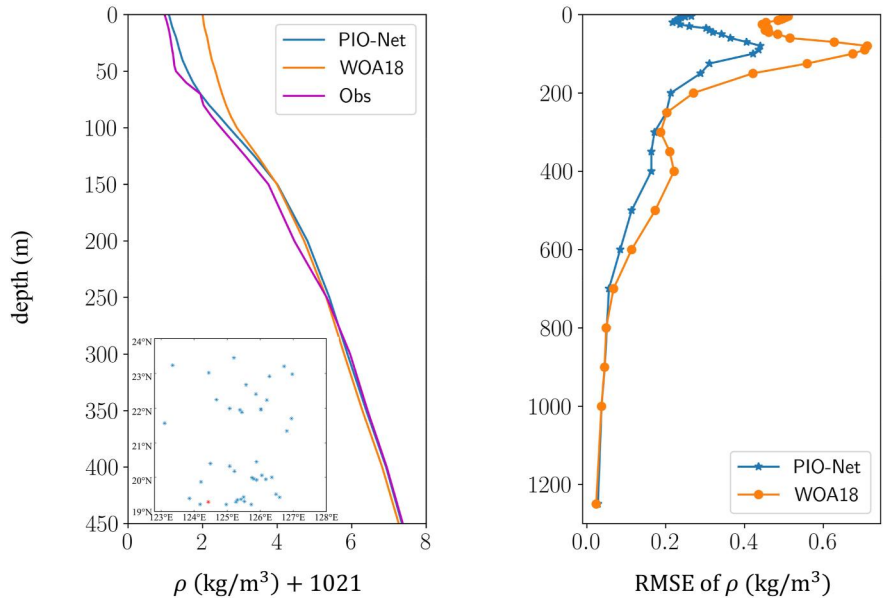

Figure 16. Density distribution derived from different models (left) and RMSEs derived from PIO-Net and WOA18 climatology (right).

equation, which is another source of error. Consequently, the model may fail to accurately capture the intricate dynamics of the ocean system. Secondly, the method does not consider additional processing for bottom topography, which may limit the reconstruction of the deeper ocean in the global region. It may be necessary to divide the reconstruction areas into smaller subdomains and group them differently based on water depth, with separate models trained for each group. In addition, in order to apply uniform bottom boundary conditions, the bottom boundary conditions should be set as far away from the surface during PIO-Net training (i.e., H should be large).

However, the present study focused on the integration of a data-driven method with a dynamics-based model for 3D reconstruction, a more detailed verification of the proposed method is also beyond the scope of this paper, there may be some potential limitations of this study. First, the verification of PIO-Net is based on simulated data from the numerical model, which is insufficient to accurately describe the real state of the ocean, and does not take into account all sources of error, including sampling, coverage, noise, and interpolation errors that affect existing observational products. Thus, Applying the proposed framework in real ocean is significantly more intricate than utilizing model simulations alone, and it is necessary to construct and verify the proposed framework with real data. Second, there is insufficient coverage to evaluate the applicability of the model to different regions. Although the model is developed within the framework of geostrophic equations, the final model form is derived from the data, so the validity in different regions needs to be determined. Third, optimal configuration may not be identified, the individual networks in the combinatorial network are equipped with the flexibility to select optimal parameters, including network depth and width, activation function, and other relevant training parameters. However, this paper solely examines the effects of some network structures and parameters, and observes that networks constructed in different structures perform differently. Increasing the number of neurons in the hidden layer leads to enhanced model accuracy, albeit this increment weakens after a certain number of neurons. Overall, an appropriate network architecture plays an important role in the prediction of PIO-Net, and the impact of network parameters and training process needs further investigation.

To address the above issues, we present some future research directions and methods. Initial priority must lie in verifying the efficacy and regional applicability utilizing more realistic data. However, using observational data presents multiple challenges as it is not as comprehensive as the model-generated data, with lower temporal and spatial resolution, and inconsistencies between satellite data and Argo. Although reanalysis data is not of the same quality as observational data, it still provides a more accurate and realistic estimate of the ocean state than pure numerical model because reanalysis data utilizes ocean observations to modify the numerical model output. And

importantly, reanalysis data can provide high spatiotemporal resolution data that cannot be achieved with current observational data. Therefore, the proposed model can be first further evaluated with reanalysis data. On this basis, transfer learning can be used to alleviate the challenge of insufficient observable data, and the PIO-Net model can be retrained and tested using Argo data. The second direction is to find the optimal network configuration to improve the estimation of PIO-Net. The third point is that only density data is used in the training of the proposed model to achieve the estimation of density fields at different depths. Future model enhancements can involve incorporating different variables into the network's training process, with the relationship between variables encoded in the loss function to predict varying variables. Finally, the ultimate objective is to create a deep learning model that can both reconstruct and forecast. One possible means is to integrate outputs as inputs to construct an iterative forecasting system. Another possibility is to construct a forecasting system that incorporates temporal and spatial variables by combining physical constraints that are more physically consistent with reality, taking full account of nonlinear spatiotemporal correlations.

Appendix A: A Detailed Introduction of Reduced Basis Method

The FOM Equation 3 is derived by discretizing Equation 1 in the domain Ω through the application of the finite difference method on a finely meshed grid. Specifically, to solve Equation 1 numerically, the spatial domain is divided into a grid with uniformly spaced intervals in both the x and y directions. The grid intervals in the x and y directions are denoted by Δx and Δy, while the grid points are represented as (x_i, y_j), where i and j correspond to the indices of the x and y directions. The central difference scheme is utilized in this study to approximate the second-order partial derivatives of the unknown function Ψ at a grid point (x_i, y_j) as follows

$$\frac{\partial^2 \Psi(x_i,y_j)}{\partial x^2} \approx \frac{\Psi_{i+1,j} - 2\Psi_{i,j} + \Psi_{i-1,j}}{\Delta x^2},$$

$$\frac{\partial^2 \Psi(x_i,y_j)}{\partial y^2} \approx \frac{\Psi_{i,j+1} - 2\Psi_{i,j} + \Psi_{i,j-1}}{\Delta y^2},$$

where $\Psi_{i,j}$ denotes the numerical solution Ψ at grid point (x_i, y_j). Then Equation 1 can be discretized as

$$\frac{\Psi_{i,j+1} - 2\Psi_{i,j} + \Psi_{i,j-1}}{\Delta y^2} + \frac{\Psi_{i+1,j} - 2\Psi_{i,j} + \Psi_{i-1,j}}{\Delta x^2} + \frac{\partial}{\partial z}\left(\frac{f_0^2}{N^2}\frac{\partial \Psi_{i,j+1}}{\partial z}\right) = q_{i,j}.$$

All equations discretized at each grid point are combined into a single system to obtain the FOM. For a more detailed explanation of finite difference, refer to Zhou (1993). Computing the FOM solution can be expensive due to the large grid size. As a result, our objective is to replace the FOM solution with an approximation in the lower dimensional space in following form

$$\boldsymbol{\Psi}_h(z, \boldsymbol{b}_s) \approx \sum_{i=1}^{n} \alpha_i(z, \boldsymbol{b}_s)\boldsymbol{\psi}_i = \mathbf{V}\boldsymbol{\alpha}(z, \boldsymbol{b}_s),$$

and the reduced basis is generated using proper orthogonal decomposition (POD) (also know as EOF) in this paper. POD computes the reduced basis matrix $\mathbf{V} \in \mathbf{R}^{N_h \times n}$ by solving the following minimum problem

$$\min_{\mathbf{W} \in \mathbf{Y}_n} \left\| \mathbf{X} - \mathbf{W}\mathbf{W}^{\mathrm{T}}\mathbf{X} \right\|_F^2, \tag{A1}$$

where $\mathbf{Y}_n = \left\{ \mathbf{W} \in \mathbf{R}^{N_h \times n} : \mathbf{W}^{\mathrm{T}}\mathbf{W} = \mathbf{I}_n \right\}$, and $\|\cdot\|_F$ is the Frobenius norm. To solve this minimization problem, one approach is to compute the singular value decomposition of $\mathbf{X}$

$$\mathbf{X} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{Z}^{\mathrm{T}}.$$

The matrices $\mathbf{U} \in \mathbf{R}^{N_h \times N_h}$ and $\mathbf{Z} \in \mathbf{R}^{N_s(m+1) \times N_s(m+1)}$ are orthogonal matrix, which made up of the left and right singular vectors of $\mathbf{X}$, respectively, and $\mathbf{\Sigma}$ is a diagonal matrix whose diagonal elements are the singular values of $\mathbf{X}$ and are ordered from largest value to smallest one. It can be shown that the first n columns of $\mathbf{U}$ is the solution of Equation A1, and satisfies

$$\left\| \mathbf{X} - \mathbf{V}\mathbf{V}^{\mathrm{T}}\mathbf{X} \right\|_F^2 = \sum_{i=n+1}^{\min(N_h, N_s(m+1))} \sigma_i^2, \tag{A2}$$

where σ_i is the i-th largest singular value of $\mathbf{X}$. According to the given accuracy, the optimal reconstruction of $\mathbf{X}$ in the reduced-order space can be obtained by the appropriate value of n computed through Equation A2. By replacing Equation 4 in Equation 3 and projecting the outcome onto the reduced space spanned by the reduced basis, the reduced-order model Equation 5 can be derived.

Appendix B: Brief Introduction of DeepONet

DeepONet provides a simple and intuitive structure to approximate G, the structure of unstacked DeepONet shown in Figure B1, which consists of two separate networks, one called "brunch net," which takes parameter $\boldsymbol{u} = [u(\boldsymbol{x}_1), u(\boldsymbol{x}_2), \ldots, u(\boldsymbol{x}_p)]$ as input and corresponding output characteristic $\boldsymbol{f} = \left[f_1, f_2, \ldots f_k\right]^{\top}$, where $\{\boldsymbol{x}_i\}_{i=1}^p$ denotes a collection of fixed locations for input function. The other network, called "trunk net," takes continuous coordinates $\boldsymbol{y}$ as input and produces an output vector $\boldsymbol{\phi} = \left[\phi_1, \phi_2, \ldots \phi_k\right]^{\top}$. DeepONet's final output $G_{\boldsymbol{\theta}}$ is the inner product of $\boldsymbol{f}$ and $\boldsymbol{\phi}$:

$$G_{\boldsymbol{\theta}}(\boldsymbol{u})(\boldsymbol{y}) = \sum_{l=1}^{k} \underbrace{f_l(\boldsymbol{u}(\boldsymbol{x}_1), \boldsymbol{u}(\boldsymbol{x}_2), \ldots, \boldsymbol{u}(\boldsymbol{x}_p))}_{\text{branch}} \underbrace{\phi_l(\boldsymbol{y})}_{\text{trunk}}.$$

The loss function of DeepONet is defined as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\theta}) &= \frac{1}{NP} \sum_{i=1}^{N} \sum_{j=1}^{P} \left| G_{\boldsymbol{\theta}}\left(\boldsymbol{u}^{(i)}\right)\left(\boldsymbol{y}_j^{(i)}\right) - G\left(\boldsymbol{u}^{(i)}\right)\left(\boldsymbol{y}_j^{(i)}\right) \right|^2 \\ &= \frac{1}{NP} \sum_{i=1}^{N} \sum_{j=1}^{P} \left| \sum_{k=1}^{q} f_k\left(\boldsymbol{u}^{(i)}(\boldsymbol{x}_1), \ldots, \boldsymbol{u}^{(i)}(\boldsymbol{x}_m)\right) \phi_k\left(\boldsymbol{y}_j^{(i)}\right) - G\left(\boldsymbol{u}^{(i)}\right)\left(\boldsymbol{y}_j^{(i)}\right) \right|^2, \end{aligned} \tag{B1}$$

where $\left\{\boldsymbol{u}^{(i)}\right\}_{i=1}^N$ denotes N separate input functions sampled from parameter space. For each $\boldsymbol{u}^{(i)}$,$\left\{\boldsymbol{y}^{(i)}\right\}_{j=1}^P$ are P points sampled in the definitional domain of $G(\boldsymbol{u})$, and the solution function $G\left(\boldsymbol{u}^{(i)}\right)$ evaluated at each point $\boldsymbol{y}_j^{(i)}$ is denoted by $G\left(\boldsymbol{u}^{(i)}\right)\left(\boldsymbol{y}_j^{(i)}\right)$.

For example, L. Lu et al. (2021) apply DeepONet to solve ordinary differential equations of the following form

$$\frac{\mathrm{d}G(x)}{\mathrm{d}x} = u(x),\ x \in [0,1],$$

with an initial condition $G(0) = 0$. The purpose of DeepONet is to predict the function $G(x)$ of arbitrary x for any function $u(x)$. To generate the training data, $p = 100$ fixed observation points $\{x_i\}_{i=1}^p$ from a uniform distribution in the interval [0,1] are selected to generate the "brunch net" input $\boldsymbol{u} = [u(\boldsymbol{x}_1), u(\boldsymbol{x}_2), \ldots, u(\boldsymbol{x}_p)]$, representing the $u(x)$ which can be considered a parameter, and the function $G(x)$ is evaluated at N different points $\left\{\boldsymbol{y}^{(i)}\right\}_{j=1}^N = \left\{x^{(i)}\right\}_{j=1}^N$ for the fixed $u(x)$. As a result, there are $p \times N$ data pairs in the training data set, each with the following format

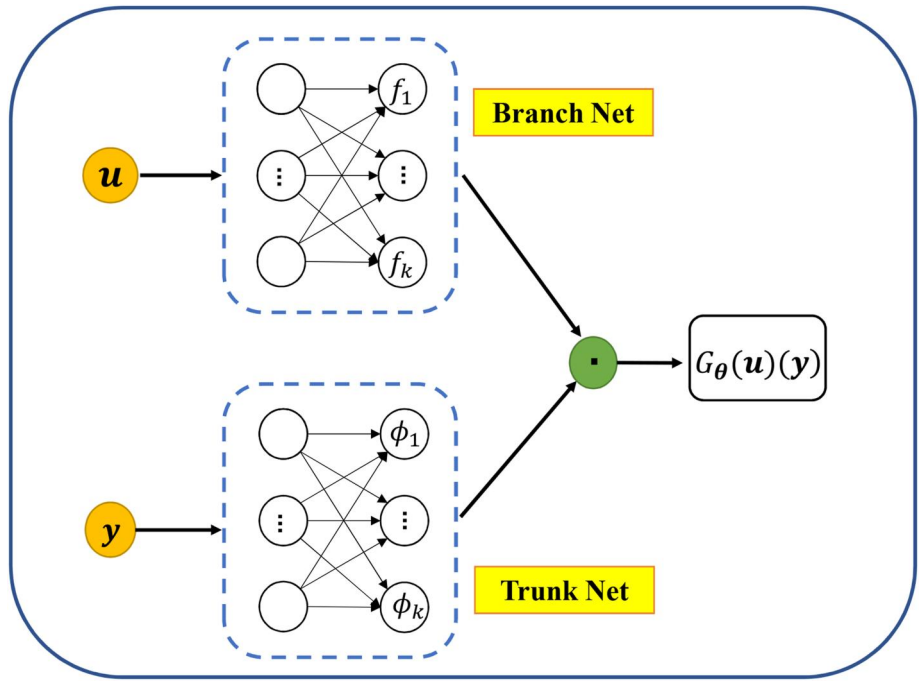

Figure B1. Structure of DeepONet.

$$[\boldsymbol{u}, \boldsymbol{y}, G(\boldsymbol{u})(\boldsymbol{y})] = [[u(x_1), u(x_2), \ldots, u(x_m)], x, G(u)(x)],$$

where $\boldsymbol{u}$ and $\boldsymbol{y}$ are the inputs of "brunch net" and "trunk net," respectively, $G(\boldsymbol{u})(\boldsymbol{y})$ is the output to be approximated by the DeepONet. The network is trained by minimizing the loss function Equation B1. After the network has been trained, the corresponding $\boldsymbol{u} = [u(x_1), u(x_2), \ldots, u(x_m)]$ can be computed for various functions $u(x)$, and solution operator $G(x)$ is then derived by feeding $\boldsymbol{u}$ into the network. Further implementation information can be found in the reference (L. Lu et al., 2021).

In this article, we employ a combination of the reduced-order model and the DeepONet to estimate the SDA projection coefficient of varying depths, which is a high-dimensional vector. Therefore, we add a third network called "Net 3," which computes the final output of PIO-Net using inputs computed from both the "brunch net" and the "trunk net," based on the original structure shown in Figure B1. Additionally, PIO-Net's loss function includes the reduced-order equation to enforce physical constraints. To predict the unknown functions g_1 and g_2 that depend on PIO-Net output and depth, two simple feedforward neural networks "Net 4" and "Net 5" are also integrated into PIO-Net. Therefore, PIO-Net is composed of five subnets, and the parameters of all networks are obtained by the minimization loss function Equation 8.

Appendix C: A Concrete Representation of PIO-Net's Loss Function

In the training of the PIO-Net to predict $\boldsymbol{\gamma}$, we can sample a large number of training points and define the loss function of the following form to encode Equation 2 into the neural network

$$\begin{aligned}\mathcal{L}_{\text{physics}}(\theta) = &\frac{1}{N_s N_t}\sum_{i=1}^{N_s}\sum_{j=1}^{N_t}\left|g_{nn}\left(f^{-1}\boldsymbol{\nu}^i_{z_0}\right)\left(h_j^{(i)}\right)\right|^2 \\ &+\frac{1}{N_s}\sum_{i=1}^{N_s}\left|\frac{dG_{\text{PI}}\left(f^{-1}\boldsymbol{\nu}^i_{z_0}\right)(h_0)}{dh}+f^{-1}\mathbf{V}^{\text{T}}\boldsymbol{\rho}^i_{z_0}\right|^2 \\ &+\frac{1}{N_s}\sum_{i=1}^{N_s}\left|\frac{dG_{\text{PI}}\left(f^{-1}\boldsymbol{\nu}^i_{z_0}\right)(-H_f)}{dh}\right|^2,\end{aligned}$$

with

$$
\begin{aligned}
g_{nn}\left(f^{-1}\boldsymbol{\nu}^i_{z_0}\right)\left(h_j^{(i)}\right) = {} & c\mathbf{V}^{\mathrm{T}}\mathcal{L}_h\left(\mathbf{V}G_{\mathrm{PI}}\left(f^{-1}\boldsymbol{\nu}^i_{z_0}\right)\left(h_j^{(i)}\right)\right) \\
& + g_1\left(h, g_j^i\right)\frac{dG_{\mathrm{PI}}\left(f^{-1}\boldsymbol{\nu}^i_{z_0}\right)\left(h_j^{(i)}\right)}{dh} \\
& + g_2\left(h, g_j^i\right)\frac{d^2G_{\mathrm{PI}}\left(f^{-1}\boldsymbol{\nu}^i_{z_0}\right)\left(h_j^{(i)}\right)}{dh^2} \\
& - g_3(h)\mathbf{V}^{\mathrm{T}}\boldsymbol{\rho}_s,
\end{aligned}
$$

where $\left\{h_j^{(i)}\right\}_{i=1}^{N_t}$ are N_t collocation points uniformly selected in $[-H_f, 0]$, and

$$
g_j^i = \overline{\frac{dG_{\mathrm{PI}}\left(f^{-1}\boldsymbol{\nu}^i_{z_0}\right)\left(h_j^{(i)}\right)}{dh}}.
$$

Appendix D: Visualization of PIO-Net

Figure D1 shows the loss history over number of iterations, where all the losses are converging. The data loss $\mathcal{L}_{\mathrm{operator}}$ diminishes beyond 30,000 iterations, a larger value leads to greater losses due to lack of preprocessing before neural network training. The residual of the equation $\mathcal{L}_{\mathrm{physics}}$ decreases quickly at first, then increases before stabilizing.

To observe the impact of the batch size on the network performance, we plotted the losses $\mathcal{L}_{\mathrm{operator}}$ with different batch size in Figure D2, where it can be clearly seen that with the increase of the batch size, the loss function has a decrease. However, the accuracy of the estimation also decreases with the increase of batch size, as illustrated in Figure D3.

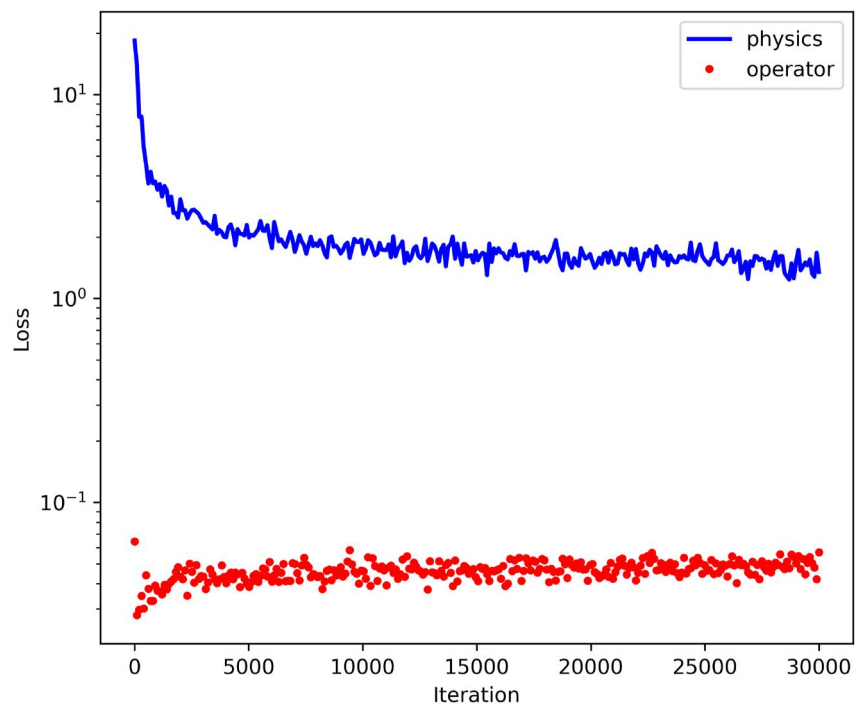

Figure D1. Loss history over number of iterations.

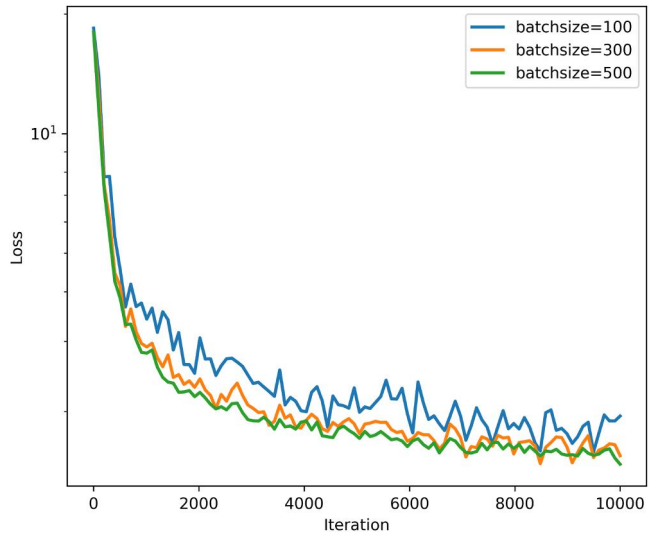

Figure D2. Loss history of PIO-Net trained using different batch size.

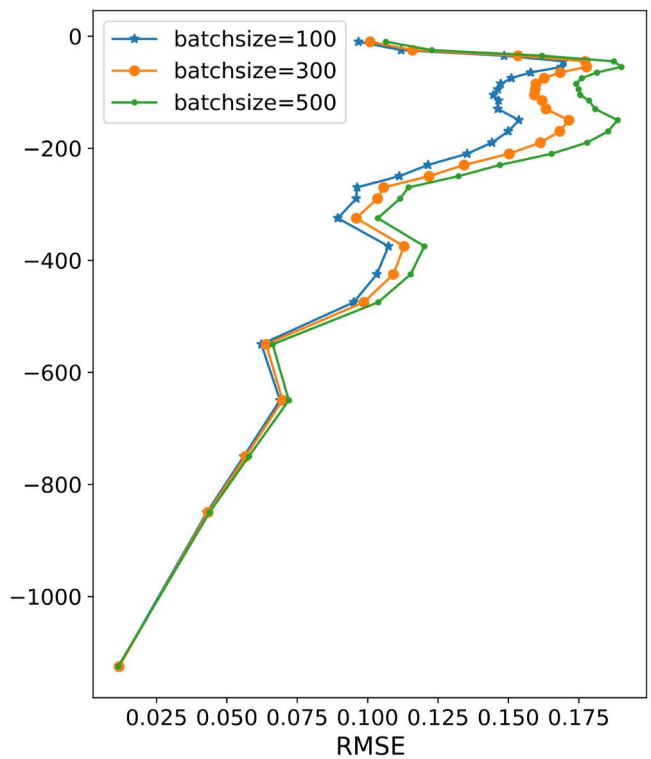

Figure D3. Prediction error of PIO-Net trained using different batch size.

Appendix E: An Introduction to Robust Validation With Argo Data

The PIO-Net is trained using HYCOM data from year 2012, following the same data processing and network training procedures described in the paper. Based on the results presented in Section 4.1, the neurons of middle layers of the PIO-Net is set to 100. Figure E1 displays the RMSE of PIO-Net in the test set to verify the

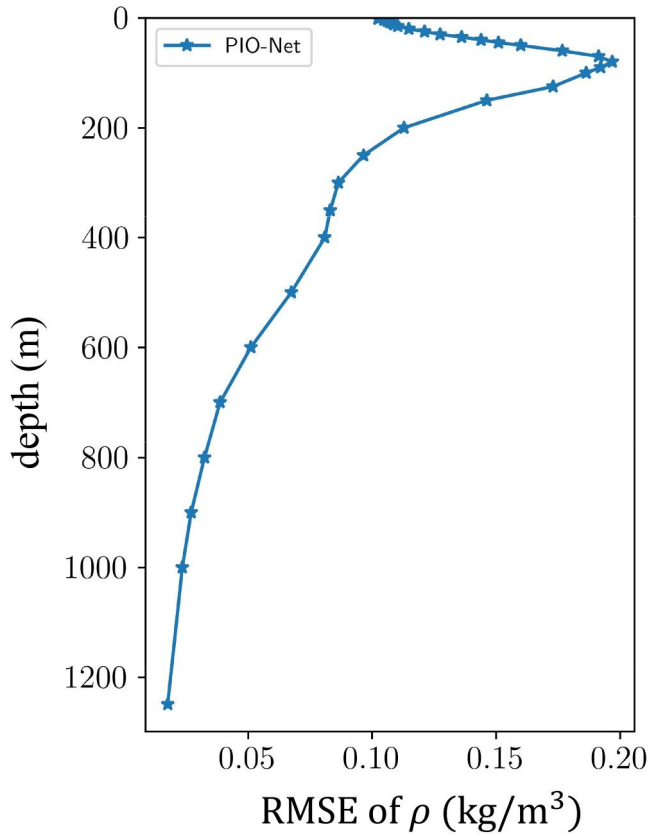

Figure E1. RMSE of density anomalies between the original reanalysis data and reconstructed profiles.

applicability of the proposed method in reanalysis data. The results demonstrate that PIO-Net provides a accurate estimate.

In order to calculate the input of PIO-Net and achieve three-dimensional spatial density reconstruction, the satellite-derived sea surface temperature and salinity are interpolated onto a resolution 4 km grids using bicubic spline interpolation and then used to calculate SSD. The total of 41 validated Argo profiles with records down to a depth of 1,200 m in the training region are selected as the test set to illustrate the refactoring performance of PIO-Net. To compute the RMSE between the observed and reconstructed profiles, bilinear interpolation is applied to interpolate the reconstructed data and the WOA18 data to the same matching point as the Argo profiles. Furthermore, linear interpolation is used vertically to interpolate the data to a uniform depth.

Data Availability Statement

The data are from the MPI-OM without tides, and the information required to replicate the model simulations are listed in Section 3. Additionally, the data used for training and testing the proposed method can be found in Y. Chen (2023b). The computer codes used here are available in Y. Chen (2023a).

Acknowledgments
The authors would like to thank the reviewers for their valuable suggestions and professional guidance, which plays an important role in improving the quality of the manuscript. This study is partially supported by the National Key Research and Development Program of China (2021YFF0704000), Taishan Scholars Program (tsqn202211059), and the Fundamental Research Funds for the Central Universities (202265005, 202264007, 202261100).

References

Agency, J. M. (2013). *Jra-55: Japanese 55-year reanalysis, daily 3-hourly and 6-hourly data*. Research Data Archive, National Center for Atmospheric Research, Computational and Information Systems Laboratory.

Akbari, E., Alavipanah, S. K., Jeihouni, M., Hajeb, M., Haase, D., & Alavipanah, S. (2017). A review of ocean/sea subsurface water temperature studies from remote sensing and non-remote sensing methods. *Water*, *9*(12), 936. https://doi.org/10.3390/w9120936

Ali, M. M., Swain, D., & Weller, R. A. (2004). Estimation of ocean subsurface thermal structure from surface parameters: A neural network approach. *Geophysical Research Letters*, *31*(20), L20308. https://doi.org/10.1029/2004GL021192

Bao, S., Zhang, R., Wang, H., Yan, H., Yu, Y., & Chen, J. (2019). Salinity profile estimation in the Pacific Ocean from satellite surface salinity observations. *Journal of Atmospheric and Oceanic Technology*, *36*(1), 53–68. https://doi.org/10.1175/JTECH-D-17-0226.1

Baydin, A. G., Pearlmutter, B. A., Radul, A. A., & Siskind, J. M. (2018). Automatic differentiation in machine learning: A survey. *Journal of Marchine Learning Research*, *18*, 1–43.
Blumberg, A. F., & Mellor, G. L. (1987). A description of a three-dimensional coastal ocean circulation model. In *Three-dimensional coastal ocean models* (pp. 1–16). American Geophysical Union (AGU). https://doi.org/10.1029/CO004p0001
Boyer, T. P., Garcia, H. E., Locarnini, R. A., Zweng, M. M., Mishonov, A. V., Reagan, J. R., et al. (2018). World ocean atlas 2018. Retrieved from https://www.ncei.noaa.gov/archive/accession/NCEI-WOA18/
Buongiorno Nardelli, B. (2020). A deep learning network to retrieve ocean hydrographic profiles from combined satellite and in situ measurements. *Remote Sensing*, *12*(19), 3151. https://doi.org/10.3390/rs12193151
Buongiorno Nardelli, B., Guinehut, S., Pascual, A., Drillet, Y., Ruiz, S., & Mulet, S. (2012). Towards high resolution mapping of 3-d mesoscale dynamics from observations. *Ocean Science*, *8*(5), 885–901. https://doi.org/10.5194/os-8-885-2012
Buongiorno Nardelli, B., & Santoleri, R. (2005). Methods for the reconstruction of vertical profiles from surface data: Multivariate analyses, residual gem, and variable temporal signals in the North Pacific Ocean. *Journal of Atmospheric and Oceanic Technology*, *22*(11), 1762–1781. https://doi.org/10.1175/JTECH1792.1
Carnes, M. R., Teague, W. J., & Mitchell, J. L. (1994). Inference of subsurface thermohaline structure from fields measurable by satellite. *Journal of Atmospheric and Oceanic Technology*, *11*(2), 551–566. https://doi.org/10.1175/1520-0426(1994)011⟨0551:IOSTSF⟩2.0.CO;2
Center, N. G. D. (2006). *2-minute gridded global relief data (ETOPO2) v2*. National Geophysical Data Center.
Chassignet, E. P., Hurlburt, H. E., Smedstad, O. M., Halliwell, G. R., Hogan, P. J., Wallcraft, A. J., et al. (2007). The hycom (hybrid coordinate ocean model) data assimilative system. *Journal of Marine Systems*, *65*(1), 60–83. (Marine Environmental Monitoring and Prediction). https://doi.org/10.1016/j.jmarsys.2005.09.016
Chen, C., Yang, K., Ma, Y., & Wang, Y. (2018). Reconstructing the subsurface temperature field by using sea surface data through self-organizing map method. *IEEE Geoscience and Remote Sensing Letters*, *15*(12), 1812–1816. https://doi.org/10.1109/LGRS.2018.2866237
Chen, Y. (2023a). Code for physics-informed deep operator learning based on reduced-order modelling for retrieving the ocean interior density from the surface [Software]. Zenodo. https://doi.org/10.5281/zenodo.8175390
Chen, Y. (2023b). Data for training and testing the PIO-Net [Dataset]. Zenodo. https://doi.org/10.5281/zenodo.8191286
Chen, Z., Wang, X., & Liu, L. (2020). Reconstruction of three-dimensional ocean structure from sea surface data: An application of ISQG method in the southwest Indian Ocean. *Journal of Geophysical Research: Oceans*, *125*(6), e2020JC016351. https://doi.org/10.1029/2020JC016351
Cheng, L., Abraham, J. P., Zhu, J., Trenberth, K. E., Fasullo, J. T., Boyer, T. P., et al. (2020). Record-setting ocean warmth continued in 2019. *Advances in Atmospheric Sciences*, *37*(2), 137–142. https://doi.org/10.1007/s00376-020-9283-7
Cunningham, S. A., Kanzow, T., Rayner, D., Baringer, M. O., Johns, W. E., Marotzke, J., et al. (2007). Temporal variability of the Atlantic meridional overturning circulation at 26.5°N. *Science*, *317*(5840), 935–938. https://doi.org/10.1126/science.1141304
Dong, C., Xu, G., Han, G., Bethel, B. J., Xie, W., & Zhou, S. (2022). Recent developments in artificial intelligence in oceanography. *Ocean-Land-Atmosphere Research*, *2022*. https://doi.org/10.34133/2022/9870950
Donlon, C. J., Martin, M., Stark, J., Roberts-Jones, J., Fiedler, E., & Wimmer, W. (2012). The operational sea surface temperature and sea ice analysis (OSTIA) system. *Remote Sensing of Environment*, *116*, 140–158. (Advanced Along Track Scanning Radiometer(AATSR) Special Issue). https://doi.org/10.1016/j.rse.2010.10.017
Doshi-Velez, F., & Kim, B. (2017). *Towards a rigorous science of interpretable machine learning*. arXiv: Machine Learning.
Fiedler, P. C. (1988). Surface manifestations of subsurface thermal structure in the California current. *Journal of Geophysical Research*, *93*(C5), 4975–4983. https://doi.org/10.1029/JC093iC05p04975
Fischer, M., Latif, M., Flügel, M., & Ji, M. (1997). The impact of data assimilation on ENSO simulations and predictions. *Monthly Weather Review*, *125*(5), 819–830. https://doi.org/10.1175/1520-0493(1997)125⟨0819:TIODAO⟩2.0.CO;2
Fofonoff, N., & Millard, R. (1983). Algorithms for computation of fundamental properties of seawater. In *UNESCO technical papers in marine science* (Vol. 44).
Fox-Kemper, B., Adcroft, A., Böning, C. W., Chassignet, E. P., Curchitser, E., Danabasoglu, G., et al. (2019). Challenges and prospects in ocean circulation models. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00065
Gallaudet, T. C., & Simpson, J. J. (1994). An empirical orthogonal function analysis of remotely sensed sea surface temperature variability and its relation to interior oceanic processes off Baja California. *Remote Sensing of Environment*, *47*(3), 375–389. https://doi.org/10.1016/0034-4257(94)90105-8
Gent, P. R., Willebrand, J., McDougall, T. J., & McWilliams, J. C. (1995). Parameterizing eddy-induced tracer transports in ocean circulation models. *Journal of Physical Oceanography*, *25*(4), 463–474. https://doi.org/10.1175/1520-0485(1995)025⟨0463:PEITTI⟩2.0.CO;2
Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. Retrieved from http://www.deeplearningbook.org
Guinehut, S., Dhomps, A.-L., Larnicol, G., & Le Traon, P.-Y. (2012). High resolution 3-d temperature and salinity fields derived from in situ and satellite observations. *Ocean Science*, *8*(5), 845–857. https://doi.org/10.5194/os-8-845-2012
Guo, Y., Liu, Y., Oerlemans, A., Lao, S., Wu, S., & Lew, M. S. (2016). Deep learning for visual understanding: A review. *Neurocomputing*, *187*, 27–48. (Recent Developments on Deep Big Vision). https://doi.org/10.1016/j.neucom.2015.09.116
Ham, Y.-G., Kim, J.-H., & Luo, J.-J. (2019). Deep learning for multi-year enso forecasts. *Nature*, *573*(7775), 1568–1572. https://doi.org/10.1038/s41586-019-1559-7
He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* (pp. 770–778). https://doi.org/10.1109/CVPR.2016.90
Held, I. M., Pierrehumbert, R. T., Garner, S. T., & Swanson, K. L. (1995). Surface quasi-geostrophic dynamics. *Journal of Fluid Mechanics*, *282*, 1–20. https://doi.org/10.1017/S0022112095000012
Hesthaven, J. S., Rozza, G., & Stamm, B. (2015). Certified reduced basis methods for parametrized partial differential equations.
Hoskins, B. J., McIntyre, M. E., & Robertson, A. W. (1985). On the use and significance of isentropic potential vorticity maps. *Quarterly Journal of the Royal Meteorological Society*, *111*(470), 877–946. https://doi.org/10.1002/qj.49711147002
Hurlburt, H. E. (1984). The potential for ocean prediction and the role of altimeter data. *Marine Geodesy*, *8*(1–4), 17–66. https://doi.org/10.1080/15210608409379497
Isern-Fontanet, J., Lapeyre, G., Klein, P., Chapron, B., & Hecht, M. W. (2008). Three-dimensional reconstruction of oceanic mesoscale currents from surface information. *Journal of Geophysical Research*, *113*(C9), C09005. https://doi.org/10.1029/2007JC004692
Jagtap, A. D., Kharazmi, E., & Karniadakis, G. E. (2020). Conservative physics-informed neural networks on discrete domains for conservation laws: Applications to forward and inverse problems. *Computer Methods in Applied Mechanics and Engineering*, *365*, 113028. https://doi.org/10.1016/j.cma.2020.113028
Karniadakis, G. E., Kevrekidis, I. G., Lu, L., Perdikaris, P., Wang, S., & Yang, L. (2021). Physics-informed machine learning. *Nature Reviews Physics*, *3*(6), 422–440. https://doi.org/10.1038/s42254-021-00314-5

Khedouri, E., Szczechowski, C., & Cheney, R. (1983). Potential oceanographic applications of satellite altimetry for inferring subsurface thermal structure. In *Proceedings oceans '83* (pp. 274–280). https://doi.org/10.1109/OCEANS.1983.1152138

Klemas, V., & Yan, X.-H. (2014). Subsurface and deeper ocean remote sensing from satellites: An overview and new results. *Progress in Oceanography*, *122*, 1–9. https://doi.org/10.1016/j.pocean.2013.11.010

LaCasce, J. H., & Mahadevan, A. (2006). Estimating subsurface horizontal and vertical velocities from sea-surface temperature. *Journal of Marine Research*, *64*(5), 695–721. https://doi.org/10.1357/002224006779367267

Lapeyre, G., & Klein, P. (2006). Dynamics of the upper oceanic layers in terms of surface quasigeostrophy theory. *Journal of Physical Oceanography*, *36*(2), 165–176. https://doi.org/10.1175/JPO2840.1

Legutke, S., & Maier-Reimer, E. (2002). The impact of a downslope water-transport parametrization in a global ocean general circulation model. *Climate Dynamics*, *18*(7), 611–623. https://doi.org/10.1007/s00382-001-0202-z

Levitus, S., Antonov, J. I., Boyer, T. P., Baranova, O. K., Garcia, H. E., Locarnini, R. A., et al. (2012). World Ocean heat content and thermosteric sea level change (0–2000 m), 1955–2010. *Geophysical Research Letters*, *39*(10), L10603. https://doi.org/10.1029/2012GL051106

Liu, L., Peng, S., & Huang, R. X. (2017). Reconstruction of ocean's interior from observed sea surface information. *Journal of Geophysical Research: Oceans*, *122*(2), 1042–1056. https://doi.org/10.1002/2016JC011927

Liu, L., Peng, S., Wang, J., & Huang, R. X. (2014). Retrieving density and velocity fields of the ocean's interior from surface data. *Journal of Geophysical Research: Oceans*, *119*(12), 8512–8529. https://doi.org/10.1002/2014JC010221

Liu, Z., Li, Z., Lu, S., Wu, X., Sun, C., & Xu, J. (2021). Scattered dataset of global ocean temperature and salinity profiles from the international ARGO program. *Journal of Global Change Data & Discovery*, *5*(3), 312–321. https://doi.org/10.3974/geodp.2021.03.09

Lu, L., Jin, P., Pang, G., Zhang, Z., & Karniadakis, G. E. (2021). Learning nonlinear operators via DeepONet based on the universal approximation theorem of operators. *Nature Machine Intelligence*, *3*(3), 218–229. https://doi.org/10.1038/s42256-021-00302-5

Lu, W., Su, H., Yang, X., & Yan, X.-H. (2019). Subsurface temperature estimation from remote sensing data using a clustering-neural network method. *Remote Sensing of Environment*, *229*, 213–222. https://doi.org/10.1016/j.rse.2019.04.009

Maes, C., Behringer, D., Reynolds, R. W., & Ji, M. (2000). Retrospective analysis of the salinity variability in the western tropical Pacific Ocean using an indirect minimization approach. *Journal of Atmospheric and Oceanic Technology*, *17*(4), 512–524. https://doi.org/10.1175/1520-0426(2000)017⟨0512:RAOTSV⟩2.0.CO;2

Marsland, S., Haak, H., Jungclaus, J., Latif, M., & Röske, F. (2003). The max-planck-institute global ocean/sea ice model with orthogonal curvilinear coordinates. *Ocean Modelling*, *5*(2), 91–127. https://doi.org/10.1016/S1463-5003(02)00015-X

Mellor, G., Hakkinen, S., Ezer, T., & Patchen, R. (2002). In *Ocean forecasting: Conceptual basis and applications* (pp. 55–72).

Melnichenko, O., Hacker, P., Maximenko, N., Lagerloef, G., & Potemra, J. (2016). Optimum interpolation analysis of Aquarius sea surface salinity. *Journal of Geophysical Research: Oceans*, *121*(1), 602–616. https://doi.org/10.1002/2015JC011343

Meng, L., Yan, C., Zhuang, W., Zhang, W., & Yan, X.-H. (2021). Reconstruction of three-dimensional temperature and salinity fields from satellite observations. *Journal of Geophysical Research: Oceans*, *126*(11), e2021JC017605. https://doi.org/10.1029/2021JC017605

Meng, L., & Yan, X.-H. (2022). Remote sensing for subsurface and deeper oceans: An overview and a future outlook. *IEEE Geoscience and Remote Sensing Magazine*, *10*(3), 72–92. https://doi.org/10.1109/MGRS.2022.3184951

Meng, X., Li, Z., Zhang, D., & Karniadakis, G. E. (2020). PPINN: Parareal physics-informed neural network for time-dependent PDEs. *Computer Methods in Applied Mechanics and Engineering*, *370*, 113250. https://doi.org/10.1016/j.cma.2020.113250

Oommen, V., Shukla, K., Goswami, S., Dingreville, R., & Karniadakis, G. E. (2022). Learning two-phase microstructure evolution using neural operators and autoencoder architectures. *Npj Computational Materials*, *8*(1), 190. https://doi.org/10.1038/s41524-022-00876-7

Pacanowski, R. C., & Philander, S. G. H. (1981). Parameterization of vertical mixing in numerical models of tropical oceans. *Journal of Physical Oceanography*, *11*(11), 1443–1451. https://doi.org/10.1175/1520-0485(1981)011⟨1443:POVMIN⟩2.0.CO;2

Penwarden, M., Zhe, S., Narayan, A., & Kirby, R. M. (2022). Multifidelity modeling for physics-informed neural networks (PINNS). *Journal of Computational Physics*, *451*, 110844. https://doi.org/10.1016/j.jcp.2021.110844

Qiu, B., Chen, S., Klein, P., Ubelmann, C., Fu, L.-L., & Sasaki, H. (2016). Reconstructability of three-dimensional upper-ocean circulation from SWOT sea surface height measurements. *Journal of Physical Oceanography*, *46*(3), 947–963. https://doi.org/10.1175/JPO-D-15-0188.1

Quarteroni, A., Manzoni, A., & Negri, F. (2015). Reduced basis methods for partial differential equations: An introduction. https://doi.org/10.1007/978-3-319-15431-2

Raissi, M., Perdikaris, P., & Karniadakis, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

Reichstein, M., Camps-Valls, G., Stevens, B., Jung, M., Denzler, J., Carvalhais, N., & Prabhat (2019). Deep learning and process understanding for data-driven Earth system science. *Nature*, *566*(7743), 195–204. https://doi.org/10.1038/s41586-019-0912-1

Roarty, H., Cook, T., Hazard, L., George, D., Harlan, J., Cosoli, S., et al. (2019). The global high frequency radar network. *Frontiers in Marine Science*, *6*. https://doi.org/10.3389/fmars.2019.00164

Roemmich, D., Johnson, G., Riser, S., Davis, R., Gilson, J., Owens, W., et al. (2009). The ARGO program: Observing the global ocean with profiling floats. *Oceanography*, *22*(2), 34–43. https://doi.org/10.5670/oceanog.2009.36

Röske, F. (2006). A global heat and freshwater forcing dataset for ocean models. *Ocean Modelling*, *11*(3), 235–297. https://doi.org/10.1016/j.ocemod.2004.12.005

Schmidhuber, J. (2015). Deep learning in neural networks: An overview. *Neural Networks*, *61*, 85–117. https://doi.org/10.1016/j.neunet.2014.09.003

Song, P., & Chen, X. (2020). Investigation of the internal tides in the northwest Pacific Ocean considering the background circulation and stratification. *Journal of Physical Oceanography*, *50*(11), 3165–3188. https://doi.org/10.1175/JPO-D-19-0177.1

Steele, M., Morley, R., & Ermold, W. (2001). PHC: A global ocean hydrography with a high-quality arctic ocean. *Journal of Climate*, *14*(9), 2079–2087. https://doi.org/10.1175/1520-0442(2001)014<2079:pagohw>2.0.co;2

Su, H., Li, W., & Yan, X.-H. (2018). Retrieving temperature anomaly in the global subsurface and deeper ocean from satellite observations. *Journal of Geophysical Research: Oceans*, *123*(1), 399–410. https://doi.org/10.1002/2017JC013631

Su, H., Wang, A., Zhang, T., Qin, T., Du, X., & Yan, X.-H. (2021). Super-resolution of subsurface temperature field from remote sensing observations based on machine learning. *International Journal of Applied Earth Observation and Geoinformation*, *102*, 102440. https://doi.org/10.1016/j.jag.2021.102440

Su, H., Wu, X., Yan, X.-H., & Kidwell, A. (2015). Estimation of subsurface temperature anomaly in the Indian Ocean during recent global surface warming hiatus from satellite measurements: A support vector machine approach. *Remote Sensing of Environment*, *160*, 63–71. https://doi.org/10.1016/j.rse.2015.01.001

Su, H., Zhang, H., Geng, X., Qin, T., Lu, W., & Yan, X.-H. (2020). OPEN: A new estimation of global ocean heat content for upper 2000 meters from remote sensing data. *Remote Sensing*, *12*(14), 2294. https://doi.org/10.3390/rs12142294

Tian, T., Cheng, L., Wang, G., Abraham, J., Wei, W., Ren, S., et al. (2022). Reconstructing ocean subsurface salinity at high resolution using a machine learning approach. *Earth System Science Data*, *14*(11), 5037–5060. https://doi.org/10.5194/essd-14-5037-2022

Torres, R., Snoeij, P., Geudtner, D., Bibby, D., Davidson, M., Attema, E., et al. (2012). Gmes sentinel-1 mission. *Remote Sensing of Environment*, *120*, 9–24. (The Sentinel Missions—New Opportunities for Science). https://doi.org/10.1016/j.rse.2011.05.028

Wang, J., Flierl, G. R., LaCasce, J. H., McClean, J. L., & Mahadevan, A. (2013). Reconstructing the ocean's interior from surface data. *Journal of Physical Oceanography*, *43*(8), 1611–1626. https://doi.org/10.1175/JPO-D-12-0204.1

Wang, S., Teng, Y., & Perdikaris, P. (2021). Understanding and mitigating gradient flow pathologies in physics-informed neural networks. *SIAM Journal on Scientific Computing*, *43*(5), A3055–A3081. https://doi.org/10.1137/20M1318043

Wang, S., Wang, H., & Perdikaris, P. (2021). Learning the solution operator of parametric partial differential equations with physics-informed deeponets. *Science Advances*, *7*(40), eabi8605. https://doi.org/10.1126/sciadv.abi8605

Wolff, J.-O., Maier-Reimer, E., & Legutke, S. (1997). The hamburg ocean primitive equation model hope Technical Report/Deutsches Klimarechenzentrum.

Wu, X., Yan, X.-H., Jo, Y.-H., & Liu, W. T. (2012). Estimation of subsurface temperature anomaly in the North Atlantic using a self-organizing map neural network. *Journal of Atmospheric and Oceanic Technology*, *29*(11), 1675–1688. https://doi.org/10.1175/JTECH-D-12-00013.1

Xue, S., Geng, X., Meng, L., Xie, T., Huang, L., & Yan, X.-H. (2021). HISEA-1: The first C-band SAR miniaturized satellite for ocean and coastal observation. *Remote Sensing*, *13*(11), 2076. https://doi.org/10.3390/rs13112076

Yan, H., Wang, H., Zhang, R., Chen, J., Bao, S., & Wang, G. (2020). A dynamical-statistical approach to retrieve the ocean interior structure from surface data: SQG-mEOF-R. *Journal of Geophysical Research: Oceans*, *125*(2), e2019JC015840. https://doi.org/10.1029/2019JC015840

Yang, X. I. A., Zafar, S., Wang, J.-X., & Xiao, H. (2019). Predictive large-eddy-simulation wall modeling via physics-informed neural networks. *Physical Review Fluids*, *4*(3), 034602. https://doi.org/10.1103/PhysRevFluids.4.034602

Zhang, D., Lu, L., Guo, L., & Karniadakis, G. E. (2019). Quantifying total uncertainty in physics-informed neural networks for solving forward and inverse stochastic problems. *Journal of Computational Physics*, *397*, 108850. https://doi.org/10.1016/j.jcp.2019.07.048

Zhou, P.-b. (1993). Finite difference method. In *Numerical analysis of electromagnetic fields* (pp. 63–94). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-50319-1_3

Zhu, X. X., Tuia, D., Mou, L., Xia, G.-S., Zhang, L., Xu, F., & Fraundorfer, F. (2017). Deep learning in remote sensing: A comprehensive review and list of resources. *IEEE Geoscience and Remote Sensing Magazine*, *5*(4), 8–36. https://doi.org/10.1109/MGRS.2017.2762307