

Physics-Informed Neural Networks with Two Weighted Loss Function Methods for Interactions of Two-Dimensional Oceanic Internal Solitary Waves*

SUN Junchao · CHEN Yong · TANG Xiaoyan

DOI: 10.1007/s11424-024-3500-x

Received: 16 November 2023 / Revised: 2 December 2023

©The Editorial Office of JSSC & Springer-Verlag GmbH Germany 2024

Abstract The multiple patterns of internal solitary wave interactions (ISWI) are a complex oceanic phenomenon. Satellite remote sensing techniques indirectly detect these ISWI, but do not provide information on their detailed structure and dynamics. Recently, the authors considered a three-layer fluid with shear flow and developed a (2+1) Kadomtsev-Petviashvili (KP) model that is capable of describing five types of oceanic ISWI, including O-type, P-type, TO-type, TP-type, and Y-shaped. Deep learning models, particularly physics-informed neural networks (PINN), are widely used in the field of fluids and internal solitary waves. However, the authors find that the amplitude of internal solitary waves is much smaller than the wavelength and the ISWI occur at relatively large spatial scales, and these characteristics lead to an imbalance in the loss function of the PINN model. To solve this problem, the authors introduce two weighted loss function methods, the fixed weighing and the adaptive weighting methods, to improve the PINN model. This successfully simulated the detailed structure and dynamics of ISWI, with simulation results corresponding to the satellite images. In particular, the adaptive weighting method can automatically update the weights of different terms in the loss function and outperforms the fixed weighting method in terms of generalization ability.

Keywords Internal solitary wave interactions, KP equation, PINN method, weighted loss function method.

SUN Junchao

School of Mathematical Sciences, Key Laboratory of Mathematics and Engineering Applications (Ministry of Education) & Shanghai Key Laboratory of PMMP, East China Normal University, Shanghai 200241, China.

Email: 52215500032@stu.ecnu.edu.cn.

CHEN Yong

School of Mathematical Sciences, Key Laboratory of Mathematics and Engineering Applications (Ministry of Education) & Shanghai Key Laboratory of PMMP, East China Normal University, Shanghai 200241, China; College of Mathematics and Systems Science, Shandong University of Science and Technology, Qingdao 266590, China. Email: ychen@sei.ecnu.edu.cn.

TANG Xiaoyan (Corresponding author)

School of Mathematical Sciences, Key Laboratory of Mathematics and Engineering Applications (Ministry of Education) & Shanghai Key Laboratory of PMMP, East China Normal University, Shanghai 200241, China.

Email: xytang@sist.ecnu.edu.cn.

*This work was supported by the National Natural Science Foundation of China under Grant Nos. 12275085, 12235007, and 12175069, and Science and Technology Commission of Shanghai Municipality under Grant Nos. 21JC1402500 and 22DZ2229014.

◇ This paper was recommended for publication by Guest Editor YAN Zhenya.

1 Introduction

Internal solitary waves (ISW) are a special fluctuation phenomenon that occurs in the interior of the ocean where the density is stably stratified^[1–3]. The wavelengths of oceanic ISW can reach hundreds or even thousands of meters, and their amplitudes usually range from a few meters to tens of meters. Because of the important role that ISW play in marine ecosystems, marine geology and coastal engineering, particular interest has been shown in the observation of ISW in the oceans over the past few decades^[4–6]. After ISW radiate outward from their place of generation, a variety of factors, such as changes in the ocean background, cause two or more ISW to interact with each other, and there are many types of such interactions^[7–10]. In fact, although internal solitary wave interactions (ISWI) occur within the ocean, they exhibit surface features in the form of small modulations in surface roughness. Furthermore, these interactions have a large horizontal scale and can propagate over long distances within the ocean while maintaining their wave shapes unchanged. These characteristics allow these interactions to be captured and recorded through satellite imagery. For instance, the RADARSAT-1 satellite captured the Miles resonance along the coast of Washington State^[11]; Astronaut photograph (STS036-082-76) depicted the ordinary interaction in the South African Sea^[12]; and the ERS-2 satellite equipped with SAR observed the asymmetric interaction in the Andaman Sea^[13].

However, satellite images usually cannot provide the detailed structure and dynamical behavior of ISWI, so scholars have mostly relied on building mathematical models to more deeply investigate ISW, such as KdV equation, BO equation, and NLS equation^[14–16]. Recently, we have considered the three-layer fluid with shear flow and continuous density distribution that approximates ocean stratification, and developed the KP model, which is capable of describing five different types of two-dimensional ISWI, including O-type, P-type, TO-type, TP-type, and Y-shaped^[17]. With the assistance of satellite imagery, we found the real-world correspondences to these interactions. Although we have obtained precise solutions for these interactions, the complexity of the solving methods and the typically limited data and information available in real-world scenarios have prompted us to develop a method for simulating ISWI. This method aims to accelerate the simulation process, reduce computational costs, and enable broader applications, particularly in addressing issues related to scarce data.

The physics-informed neural network (PINN) method is an important application of deep learning in solving physics problems, which combines the powerful approximation ability of neural networks with mathematical models in physics problems^[18–21]. The PINN method requires only a small amount of initial and boundary value data to simulate the solution of physics equations with high accuracy^[22–24]. In recent years, PINN methods have been widely used in the field of fluids. Raissi, et al.^[25] encoded the Navier-Stokes equations into the neural networks to learn velocity and pressure fields from flow visualizations, Mao, et al.^[26] investigated the forward and inverse problems of the Euler equations describing high-speed aerodynamic flows, and Cai, et al.^[27] demonstrated the effectiveness of the PINN in dealing with three-dimensional wake flows, supersonic flows, and biomedical flows.

Despite the remarkable progress of the PINN method in the field of fluids, the application

of PINN in the field of ISW, especially their interaction, is relatively few. The current research mainly focuses on the introduction of deep learning methods for the identification or simulation of ISW by using satellite images^[28–31]. However, to obtain the detailed structures and dynamical behaviors of ISWI, relying solely on satellite remote sensing techniques is not sufficient to meet the demand. Therefore, the application of PINN methods incorporating physical modeling is necessary. It is worth noting that the classical PINN method is not applicable to all fields due to its own structure as well as stability issues^[32, 33]. In order to make the PINN algorithm more widely applicable, the classical PINN algorithm needs to be improved. For example, fractional PINN for solving fractional equations^[34], Bayesian PINN for positive and inverse nonlinear problems described by noisy data^[35], and Navier-Stokes flow nets, which were developed by considering the velocity-pressure and vorticity-velocity formulations of the Navier-Stokes equations^[36]. In addition, Chen's team proposed an integrable deep learning framework that successfully incorporates special integrable structures and properties into PINN, such as conservation laws, symmetries, and Miura transformations^[37–39]. However, among the numerous PINN methods, there is a lack of methods specially designed for oceanic ISWI. Therefore, we aim to consider the properties of ISWI and develop a PINN method for simulating these interactions based on their mathematical model.

The rest of the paper is organized as follows. In Section 2, we introduce mathematical models describing five types of ISWI. In Section 3, combining the properties of the ISWI and the experimental results, we analyze the reasons for the failure of the classical PINN method and propose a scheme to solve the imbalance of the loss function. In Section 4, we design an improved PINN algorithm applied to simulate two-dimensional oceanic ISWI by introducing the fixed weighing and adaptive weighting methods into the loss function, and successfully simulate the detailed structures and dynamical behaviors of five types of ISWI. The last section is devoted to the conclusion and discussions.

2 Physics Background and Model of Internal Waves

Considering an inviscid, incompressible, stratified fluid. The fluid is bounded by a rigid boundary $z = -h$ as the lower boundary, a free surface $z = \psi(x, y, t)$ as the upper boundary, and an equilibrium position $z = 0$ at the upper boundary. Thus the basic governing equations and boundary conditions describing the internal waves are

$$\begin{aligned}
 \rho \frac{\partial U}{\partial T} + U \frac{\partial U}{\partial X} + V \frac{\partial U}{\partial Y} + W \frac{\partial U}{\partial Z} + \frac{\partial P}{\partial X} &= 0, \\
 \rho \frac{\partial V}{\partial T} + U \frac{\partial V}{\partial X} + V \frac{\partial V}{\partial Y} + W \frac{\partial V}{\partial Z} + \frac{\partial P}{\partial Y} &= 0, \\
 \rho \frac{\partial W}{\partial T} + U \frac{\partial W}{\partial X} + V \frac{\partial W}{\partial Y} + W \frac{\partial W}{\partial Z} + \frac{\partial P}{\partial Z} + \rho g &= 0, \\
 \frac{\partial \rho}{\partial T} + U \frac{\partial \rho}{\partial X} + V \frac{\partial \rho}{\partial Y} + W \frac{\partial \rho}{\partial Z} &= 0, \\
 \frac{\partial U}{\partial X} + \frac{\partial V}{\partial Y} + \frac{\partial W}{\partial Z} &= 0,
 \end{aligned} \tag{1}$$

$$W = 0 \mid_{Z=-h}, \quad P = 0 \mid_{Z=\psi(X,T,T)},$$

$$W = \frac{\partial \psi}{\partial T} + U \frac{\partial \psi}{\partial X} + V \frac{\partial \psi}{\partial Y} \mid_{Z=\psi(X,Y,T)},$$

where X , Y , and Z represent the spatial coordinates, while U , V , and W describe the fluid velocities. ρ is the fluid density, P is the pressure, g is the gravitational acceleration, and $\psi(x, y, t)$ is the vertical displacement of the free surface. Figure 1(a) can be used as a schematic illustration of the above description.

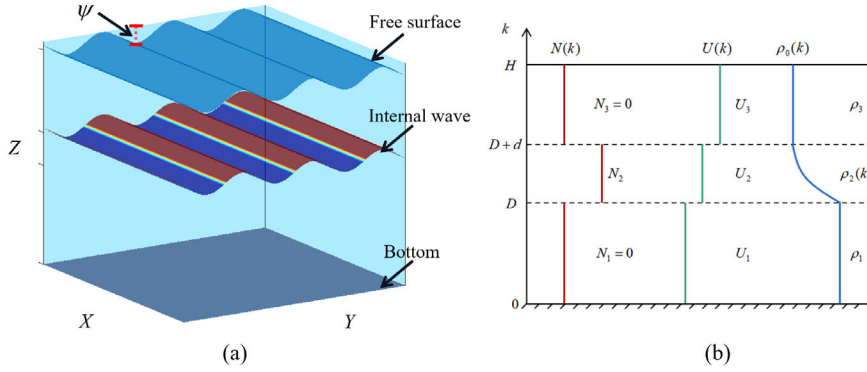


Figure 1 (Color online) Schematic diagrams: (a) The flow coordinate system; (b) The three-layer fluid

Next, Equation (1) is simplified by the multiscale method. The slow variables $x = \varepsilon(X - cT)$, $y = \varepsilon^2 Y$, $t = \varepsilon^3 T$ are introduced, and the velocity field in the X -direction is decomposed into the fundamental component $U(k)$ and the perturbation component, while the lower-order terms of the vertical displacement are separated into the variables $u(x, y, t) \Phi(k)$. The KP equation describing the internal waves is finally obtained^[17].

$$\frac{\partial}{\partial x} \left(\frac{\partial u}{\partial t} + a_1 u \frac{\partial u}{\partial x} + a_2 \frac{\partial^3 u}{\partial x^3} \right) + a_3 \frac{\partial^2 u}{\partial y^2} = 0, \quad (2)$$

where

$$a_1 = \frac{3 \int_{-h}^0 \rho_0(k) (U(k) - c)^2 \left(\frac{d\Phi}{dk} \right)^3 dk}{2 \int_{-h}^0 \rho_0(k) (c - U(k)) \left(\frac{d\Phi}{dk} \right)^2 dk}, \quad a_2 = \frac{\int_{-h}^0 \rho_0(k) (U(k) - c)^2 \Phi^2 dk}{2 \int_{-h}^0 \rho_0(k) (c - U(k)) \left(\frac{d\Phi}{dk} \right)^2 dk},$$

$$a_3 = \frac{- \int_{-h}^0 \int_0^k \rho_0(k) N(k)^2 \Phi dk' \frac{d\Phi}{dk} dk}{2 \int_{-h}^0 \rho_0(k) (c - U(k)) \left(\frac{d\Phi}{dk} \right)^2 dk}, \quad N^2(k) = - \frac{1}{\sigma \rho_0(k)} \frac{d\rho_0(k)}{dk}. \quad (3)$$

Here, Φ can be obtained by solving the following eigenvalue problem

$$\frac{d}{dk} \left[\rho_0(k) (U(k) - c)^2 \frac{d\Phi}{dk} \right] + \rho_0 N^2 \Phi = 0, \quad \Phi = \sigma (U(k) - c)^2 \frac{d\Phi}{dk} \Big|_{k=0}, \quad \Phi = 0 \mid_{k=-h}. \quad (4)$$

Consider a three-layer configuration similar to stratification in the ocean, as shown in Figure 1(b). The depth of the upper layer is $H - d - D$, the depth of the middle layer

is d , and the depth of the lower layer is D . Densities ρ_3 and ρ_1 of the upper and lower layers are constant, while, the density of the middle layer is a depth-dependent function $\rho_2(k) = \rho_3 e^{\frac{1}{d} \ln \frac{\rho_1}{\rho_3} (D+d)} e^{-\frac{1}{d} \ln \frac{\rho_1}{\rho_3} k}$, and $\rho_1 > \rho_2(k) > \rho_3$. From the density distribution, the buoyancy frequency of both the upper and lower layers is zero, i.e., $N_3 = N_1 = 0$, and the buoyancy frequency of the middle layer is N_2 . Furthermore, assume that each fluid layer has a constant flow rate U_i (where $i = 1, 2, 3$) and satisfies $U_1 < U_2 < U_3$. Using these information from this configuration one can solve Equation (4) and further determine the expressions for the coefficients of the KP equation.

The detailed derivation of Equation (2) is given in [17]. It is noteworthy that we also succeeded in obtaining five different types of ISWI (O-type, P-type, Y-shaped, TO-type, and TP-type) described by Equation (2), and relied on high-resolution satellite imagery to provide evidence of the real existence of these interactions in the oceans. The exact solutions of these interactions are employed to evaluate the performance of PINN method simulations.

3 The Classical PINN Method in Two-Dimensional Space

PINN is a method that combines deep learning techniques with the laws of physics, and is mainly used to solve the forward and inverse solution problems of partial differential equations (PDE). It is unique in that it embeds physical information into the loss function, making it possible to learn the physical field from a small number of data points. In this section, in order to simulate the five types of two-dimensional ISWI described by Equation (2), we introduce the basic framework of the classical PINN method in two-dimensional space. Then, we analyze the limitation of the classical PINN method in simulating two-dimensional ISWI by incorporating the properties of ISWI. It is worth mentioning that we find this limitation is caused by the imbalanced loss function.

3.1 The Basic Framework

We construct a deep neural network with the number of layers L , and the number of nodes in the i th layer is M_i . Layer 0 is the input layer with three nodes ($M_0 = 3$), which is responsible for inputting temporal and spatial data into the network. The output layer L has one node ($M_L = 1$) and is designed to output the learned ISWI. The other intermediate layers are called hidden layers tasked with feature extraction and better linear partitioning of different types of data. The input of each node is equal to the output of the previous node after the nonlinear transformation

$$\mathbf{V}^{(i)} = \mathcal{L} \left(\mathbf{W}^{(i)} \mathbf{V}^{(i-1)} + \mathbf{b}^{(i)} \right), \quad i = 1, 2, \dots, L, \quad (5)$$

where $\mathbf{V}^{(i)}$, $\mathbf{W}^{(i)}$ and $\mathbf{b}^{(i)}$ are the state vector, the weight matrix and the bias vector of the network at the layer i , respectively. \mathcal{L} is the nonlinear activation function, which is chosen as the hyperbolic function \tanh . Initialization plays an important role in model training for deep learning, and an unreasonable initialization can easily cause the model to fail to converge. We initialize the bias term as a zero tensor, and apply the Xavier initialization to initialize the

weight term

$$\mathbf{W}^{(i)} \sim N(0, \sigma_i^2), \quad \sigma_i^2 = \frac{2}{M_i + M_{i+1}}, \quad i = 1, 2, \dots, L, \quad (6)$$

where $N(0, \sigma_i^2)$ denotes a Gaussian distribution, such that the initialization of the weights keeps the variance of the inputs and outputs unchanged.

In order to provide a small amount of sample data for the neural network, we consider the Dirichlet boundary conditions of Equation (2):

$$\begin{cases} u(x, y, -t_1) = u_0(x, y, -t_1), & \forall (x, y) \in G, \\ u(x, y, t) = u_0(x, y, t), & \forall (x, y) \in \partial G, \quad t \in [-t_1, t_1], \end{cases} \quad (7)$$

where $G = [x_1, x_2] \times [y_1, y_2]$ is a closed rectangle.

In difference to the case of one-dimensional space, the small amount of data required for the neural network in the two-dimensional space comes from the five planes (four boundary planes and one initial value plane) of the rectangular cube $G \times [-t_1, t_1]$, as shown in Figure 2. We construct three types of discrete data points, which play a crucial role in the training and optimization process of the network. By dividing the spatial domain $[x_1, x_2]$ and $[y_1, y_2]$ into n_x and n_y equal parts, and the temporal domain $[-t_1, t_1]$ into n_t equal parts, we obtain a three-dimensional grid. The e -type points consist of $n_e = (n_x + 1)(n_y + 1)(n_t + 1)$ points. The b -type points are the intersections of the e -type points with four boundary surfaces and one initial value surface, and the number of b -type points is defined as n_b . The Latin hypercube sampling (LHS) method is used to randomly select n_f points in $G \times [-\tau_1, \tau_1]$ to form f -type points. The b -type, e -type, and f -type points are defined as $\{(x_b^i, y_b^i, t_b^i), u_0(x_b^i, y_b^i, t_b^i)\}_{i=1}^{n_b}$, $\{(x_e^i, y_e^i, t_e^i), u_0(x_e^i, y_e^i, t_e^i)\}_{i=1}^{n_e}$, and $\{(x_f^i, y_f^i, t_f^i)\}_{i=1}^{n_f}$, respectively.

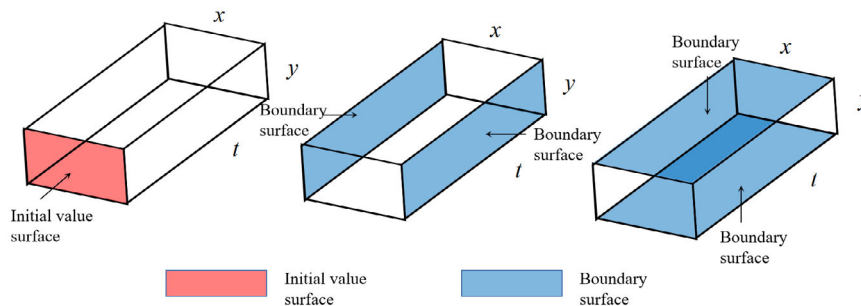


Figure 2 (Color online) A schematic representation of the boundary surfaces and initial surface

The loss function consists of constraints from the boundary and physical equation (2),

$$\text{Loss}(\theta) = \text{Loss}_{KP}(\theta) + \text{Loss}_b(\theta), \quad (8)$$

where

$$\begin{aligned}\text{Loss}_{KP}(\theta) &= \frac{1}{n_f} \sum_{i=1}^{n_f} |f(x_f^i, y_f^i, t_f^i; \tilde{u}(x, y, t, \theta))|^2, \\ \text{Loss}_b(\theta) &= \frac{1}{n_b} \sum_{i=1}^{n_b} |\tilde{u}(x_b^i, y_b^i, t_b^i, \theta) - u_0(x_b^i, y_b^i, t_b^i)|^2.\end{aligned}\quad (9)$$

Here, $\theta = \{\mathbf{W}^{R(i)}, \mathbf{b}^{(i)}\}_{i=1}^L$ denotes the parameter space. \tilde{A} is the output of the neural network and f is derived from the left hand side of Equation (2). Finally, the e -type points are used to evaluate the generalization ability of PINN models,

$$\text{Error}_g = \frac{\sqrt{\sum_{i=1}^{n_e} |\tilde{u}(x_e^i, y_e^i, t_e^i, \theta^*) - u_0(x_e^i, y_e^i, t_e^i)|^2}}{\sqrt{\sum_{i=1}^{n_e} |u_0(x_e^i, y_e^i, t_e^i)|^2}}, \quad (10)$$

where θ^* represents the optimal parameter.

3.2 Analysis of Simulation Failures

Within the framework mentioned above, we conduct simulations of five distinct oceanic ISWI, unfortunately, all such simulations encounter challenges. In this subsection, we take O-type interactions as a case to investigate why the classical PINN method struggles in capturing the structure and dynamics of ISW. Furthermore, we propose strategies to enhance the performance of the classical PINN method.

The results obtained using the classical PINN method for simulating ISWI are presented in Figure 3. It is evident that the learned results at different time points converge to a zero plane, suggesting that the amplitude of the interactions is neglected during the learning process. Figure 4 provides a comparison between the exact density map and the learned density map, revealing a significant disparity between them. This further emphasizes the difficulty of the classical PINN approach in simulating ISWI.

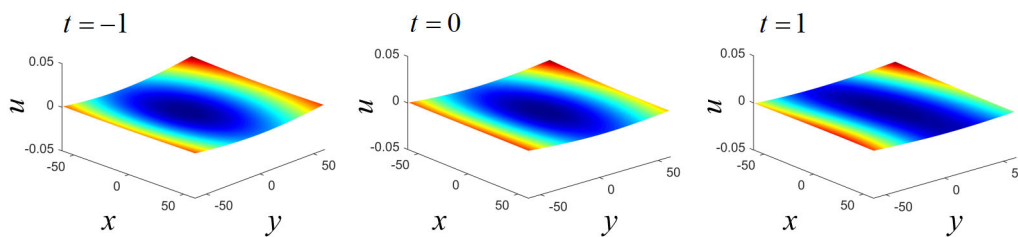


Figure 3 (Color online) The simulation results of the classical PINN

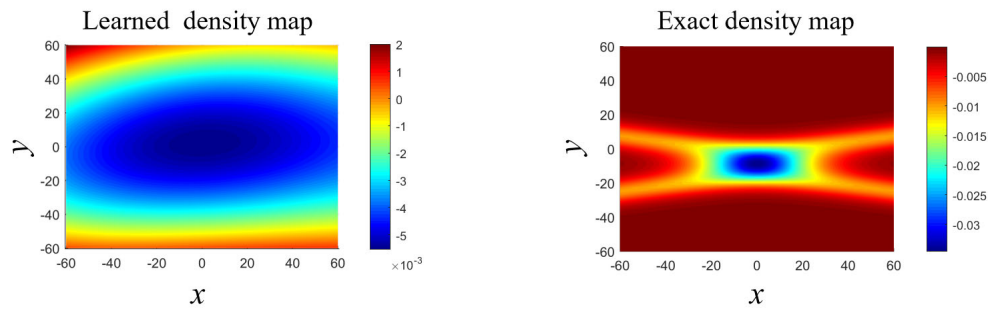


Figure 4 (Color online) The learned density map of classical PINN and the exact density map

What causes the result of learning to approximate a zero plane? In fact, it stems from the properties of ISW themselves. First, Oceanic ISW typically exhibit wavelengths ranging from hundreds to thousands of meters, while their amplitudes usually fall within the range of a few meters to tens of meters. Consequently, the amplitude of these waves is significantly smaller than their wavelength. As depicted in Figure 5(a), the xy -plane possesses dimensions of 100×60 , while the maximum amplitude of the interaction is approximately 0.03. This illustrates that the range of amplitude variations is quite limited. Consequently, in numerical simulations, the amplitude may be neglected. This leads to the convergence of the classical PINN method towards a zero plane. Secondly, ISWI occur on relatively large spatial scales. As shown in Figure 5(b), in order to obtain a complete view of these interactions in the real ocean environment, the horizontal dimensions probed by satellites are $80 \text{ km} \times 84 \text{ km}$. This implies that most of the entire spatial region is free of fluctuations (can be considered as the zero plane), with only a small portion exhibiting interactions of ISW. Therefore, in simulations, wave amplitudes may be averaged or canceled out, resulting in a lack of discernible amplitude variations in the simulation results.

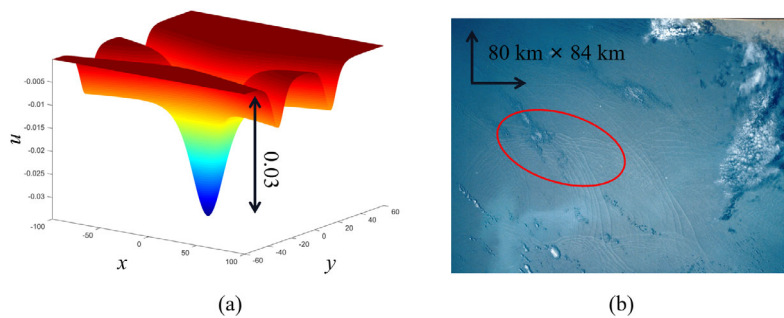


Figure 5 (Color online) The internal solitary wave interaction at small amplitude and large spatial scale: (a) The exact solution; (b) The satellite imagery (Astronaut photograph (STS036-082-76) acquired on 1 March 1990 at 1254 UTC), from [40]

Further analysis reveals that these characteristics of ISW lead to an imbalance in the loss function of the PINN model. As shown in Figure 6, the change in the loss function during the training process, it can be observed that the value of Loss_{KP} remains around 10^{-9} to 10^{-10} ,

while the value of Loss_b ranges from 10^{-4} to 10^{-5} , indicating a significant discrepancy between them. This suggests that during training, the PINN model tends to optimize the residuals of the partial differential equations while neglecting the constraints of boundary and initial conditions. This also explains why the model's output results approximate a zero solution since the zero solution satisfies the physical equations.

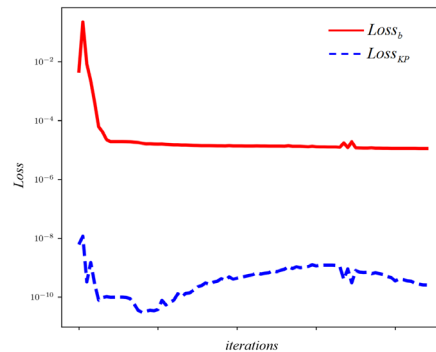


Figure 6 (Color online) The imbalance phenomenon in the loss function

There are various ways to deal with the imbalance of the loss function, such as hyperparameter tuning, adding data and data resampling^[41]. In view of the aforementioned problem and analysis, our proposed strategy is to introduce weighted loss function methods in the PINN model to enhance the constraint tendency of the model to optimize the boundary conditions and initial value conditions.

4 Data-Driven Internal Solitary Wave Interactions

In this section, based on the PINN method in two-dimensional space, we introduce the fixed weighing and the adaptive weighting methods to simulate the structural and dynamical behavior of oceanic ISWI.

4.1 The Weighted Loss Function Methods

The fixed weighing method is a simple but effective strategy that directly multiplies constants on the different terms of the loss function and adjusts the size of the different constants by analyzing the results. The form of the method is as follows

$$\text{Loss}(\theta) = c_1 \text{Loss}_{KP}(\theta) + c_2 \text{Loss}_b(\theta), \quad (11)$$

where c_1 and c_2 are constants. According to the analysis in Subsection 3.2, we need to make sure that c_2 is larger than c_1 . Therefore, in our experiments, we adjust the ratio of c_1 to c_2 by setting it to 1:500.

Despite the coding simplicity and effectiveness of the fixed weighing method, the weights need to be constantly adjusted manually in order to obtain excellent simulation results, especially when the terms in the loss function are increased, which makes this adjustment even

more cumbersome. Therefore, we introduce an adaptive weighting method based on Gaussian probabilistic models^[41], which enables the PINN model to automatically update the learning weights during the training process

$$\text{Loss}(\theta) = \frac{1}{2\beta_1^2} \text{Loss}_{KP}(\theta) + \frac{1}{2\beta_2^2} \text{Loss}_b(\theta) + \log \beta_1 \beta_2, \quad (12)$$

where β_1 and β_2 describe the adaptive weights of loss terms. $\frac{1}{2\beta_1^2}$ and $\frac{1}{2\beta_2^2}$ describe the total weights of loss terms. We utilize the L-BFGS-B algorithm to optimize loss functions, and the goal is to find the best weights.

The form of Equation (12) is obtained by the following derivation. Construct a Gaussian probability model with output u

$$p(u | \hat{u}(x, y, t; \theta), \beta^2) = \frac{1}{\sqrt{2\pi\beta^2}} e^{-\frac{(u - \hat{u})^2}{2\beta^2}}, \quad (13)$$

where β represents the weights. From the minimization objective, we have

$$-\log p = \frac{1}{2} \log(2\pi\beta^2) + \frac{(u - \hat{u})^2}{2\beta^2} \propto \frac{1}{2\beta^2} \|u - \hat{u}(x, t; \theta)\|^2 + \log \beta. \quad (14)$$

The framework diagram for the PINN with weighted loss function methods is provided in Figure 7. The overall environment for our code is Python 3.7, tensorflow 1.15, and a DELL Precision 7920 vertical computer equipped with a 2.10 GHz 8-core Xeon Silver 4110 processor.

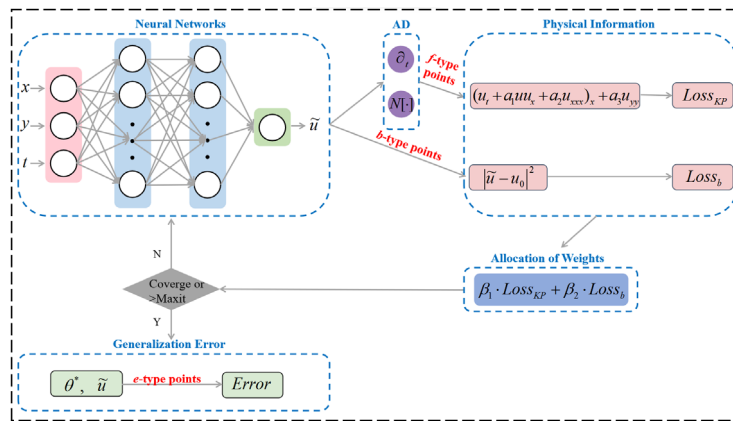


Figure 7 (Color online) A schematic diagram of the basic framework of the PINN with the weighted loss function method

4.2 Data-Driven O-Type Interactions

In simulating all types of ISWI, we let $M_i = 40$ ($i = 1, 2, \dots, L - 1$), $\xi_1 = \theta_1 = -60$, $\xi_2 = \theta_2 = 60$, $\tau_1 = 2$, $n_x = n_y = 65$, and $n_t = 33$. In fact, since not all b -type points are needed in the training process, only 6000 b -type points ($n_b = 6000$) are randomly selected. Using the LHS method, 50000 f -type points are selected ($n_f = 50000$).

Table 1 Generalization errors under different weighted loss function methods

	O-type	P-type	TO-type	TP-type	Y-shaped
No weight	7.83×10^{-1}	1.58×10^{-1}	8.18×10^{-1}	8.30×10^{-1}	8.38×10^{-1}
Fixed weighting	4.99×10^{-2}	1.73×10^{-2}	6.53×10^{-2}	6.27×10^{-2}	7.81×10^{-2}
Adaptive weighting	3.73×10^{-2}	1.56×10^{-2}	3.65×10^{-2}	3.30×10^{-2}	3.43×10^{-2}

The generalization error is a crucial factor in ensuring the applicability, reliability, and generalization capability of the PINN model. Table 1 presents a comparison of the generalization errors between two weighting methods and the classical PINN method. It is evident that the classical PINN performs poorly in learning ISWI. However, the introduction of weighted loss functions significantly reduces the generalization error, especially the adaptive weighting method, which exhibits relatively smaller generalization error and demonstrates superior generalization capability. Next, we focus on presenting the simulation results of the adaptive weighting method.

We construct a 12-layer feedforward neural network to simulate O-type interactions. After training, the PINN model completes 7582 iterations in approximately 4636.36 seconds. The loss functions Loss_{KP} and Loss_b are reduced to 5.696×10^{-11} and 1.980×10^{-9} , respectively. Our experimental findings are visually represented in Figure 8. Two ISW of the same amplitude produce wave peaks in the interaction region that are higher than their respective original amplitudes. The predicted solution curve closely aligns with the exact solution curve through the two-dimensional section diagram of the O-type interaction at different time points. Additionally, the density plots of the error illustrate that the regions of interaction produce large errors. Figure 9 presents the exact density map, the learned density map with adaptive weights, the learned density map with fixed weights, and corresponding satellite images. It is evident that the learned density map closely match the exact density map and correspond well with the satellite images. Moreover, O-type interactions induce phase shifts.

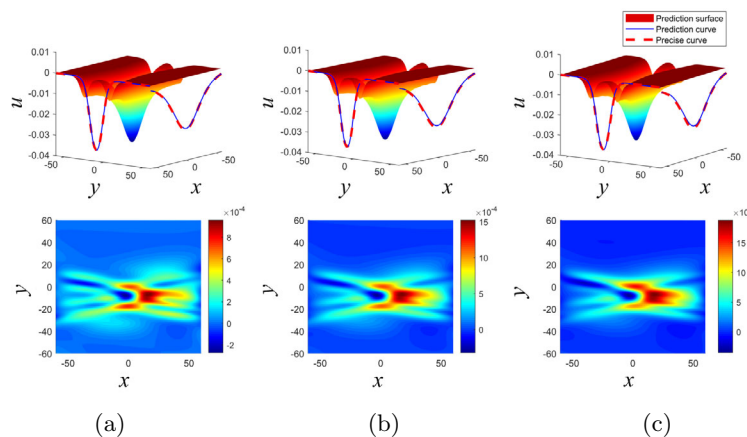


Figure 8 (Color online) The data-driven O-type interactions and the corresponding density map of the error at $\tau = -1$ (a), $\tau = 0$ (b), and $\tau = 1$ (c)

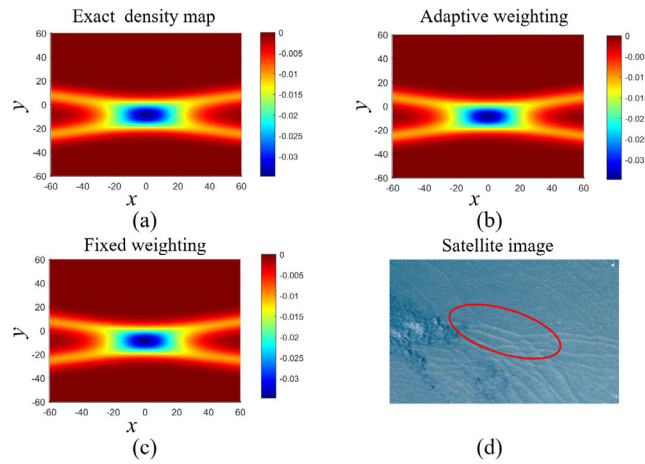


Figure 9 (Color online) The density map of the O-type interactions: (a) The exact density map; (b) The learned density map of the adaptive weighting; (c) The learned density map of the fixed weighting; (d) The satellite imagery (Astronaut photograph (STS036-082-76) acquired on 1 March 1990 at 1254 UTC), from [40]

The convergence curves of adaptive weights β_1 and β_2 are depicted in Figure 10. Over the course of iterations, β_1 consistently decreases, while β_2 exhibits notable fluctuations during the initial iterations. In the final 1000 iterations, β_1 remains stable, whereas β_2 continues to iterate. The ultimate values for β_1 and β_2 are 49.69 and -4.455×10^{-5} , respectively.

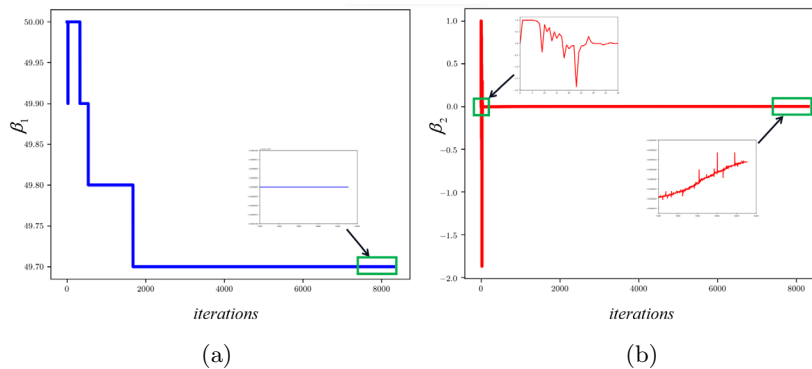


Figure 10 (Color online) The numerical convergence curve of adaptive weights for the data-driven O-type interactions

4.3 Data-Driven P-Type Interactions

With L set to 12, we achieve convergence of the loss functions, Loss_{KP} and Loss_b , to 4.224×10^{-9} and 1.131×10^{-8} , respectively, after 6000 iterations, completing in approximately 3766.71 seconds. This successful convergence allowed us to obtain data-driven P-type interactions. Figure 11 presents three-dimensional plots of the P-type interaction at various time points, along with corresponding density plots of the errors. The P-type interactions emerge from the interaction of two ISW with different amplitudes, and the amplitude within the interaction

region falls between those of the two ISW. The curves of the predicted and exact solutions fit relatively well. The main source of errors comes from the region with larger gradients.

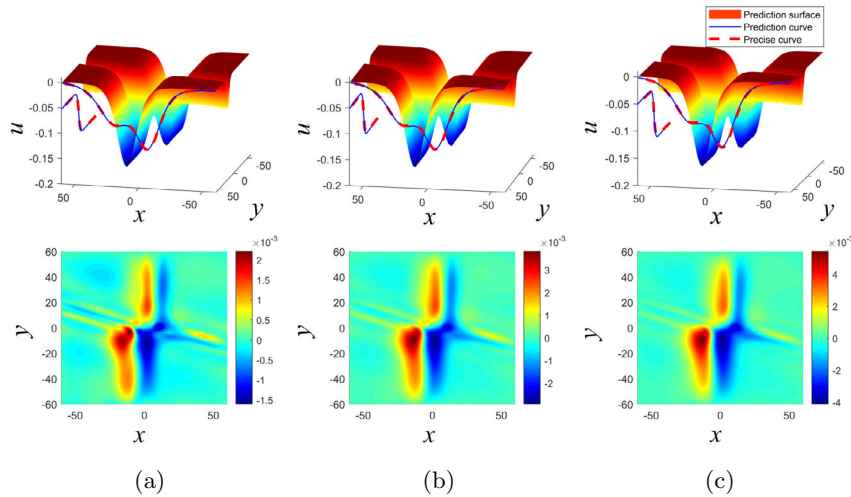


Figure 11 (Color online) The data-driven P-type interactions and the corresponding density map of the error at $\tau = -1$ (a), $\tau = 0$ (b), and $\tau = 1$ (c)

Figure 12 showcases the exact density map of P-type interactions, the density map with adaptive weight learning, the density map with fixed weight learning, and the corresponding satellite images. It is clear that the simulated results match the precise results very well and can be correlated with the ISWI in the Amanda Sea.

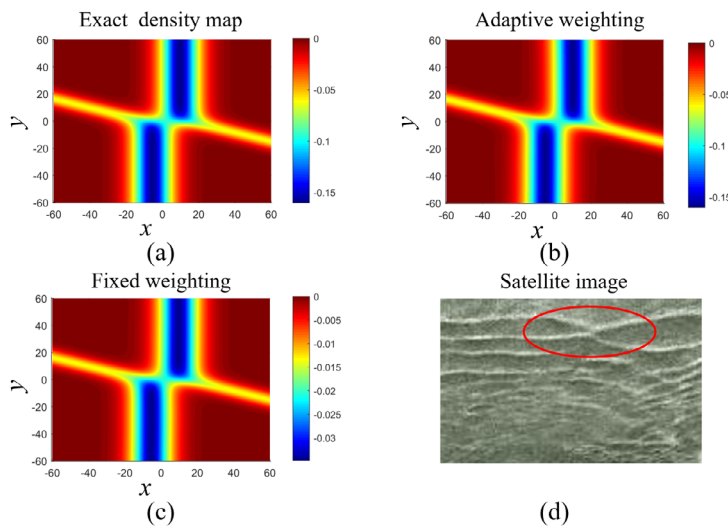


Figure 12 (Color online) The density map of the P-type interactions: (a) The exact density map; (b) The learned density map of the adaptive weighting; (c) The learned density map of the fixed weighting; (d) The satellite imagery (ERS-2 SAR image of the Andaman acquired on 11 February 1997 at 0359 UTC), from [40]

The convergence curves of adaptive weights β_1 and β_2 during the simulation process of the P-type interaction are illustrated in Figure 13. β_1 steadily decreases, while β_2 exhibits noticeable fluctuations during the initial iterations. In the last 1000 iterations, β_1 remains stable, while β_2 continues to iterate. The final iterative values for β_1 and β_2 are 48.11 and -1.065×10^{-4} , respectively.

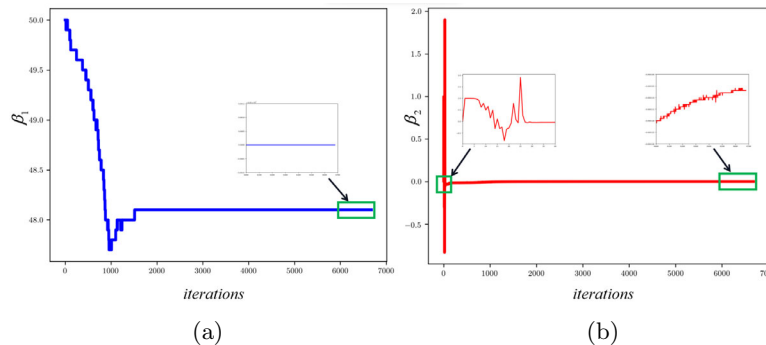


Figure 13 (Color online) The numerical convergence curve of adaptive weights for the data-driven P-type interactions

4.4 Data-Driven TO-Type Interactions

For capturing the structural and dynamical behavior of the TO-type interactions, we configure $L = 11$. Following 6753 iterations completing in approximately 3771.70 seconds, the Loss_{KP} and Loss_b impressively converge to 2.998×10^{-11} and 5.655×10^{-9} , respectively. Figure 14 presents three-dimensional plots of the TO-type interaction at various time instances, along with density plots illustrating the associated errors. Different from the O-type interactions, the amplitude of the interaction region of the TO-type interactions is generated by two different amplitudes of the ISW.

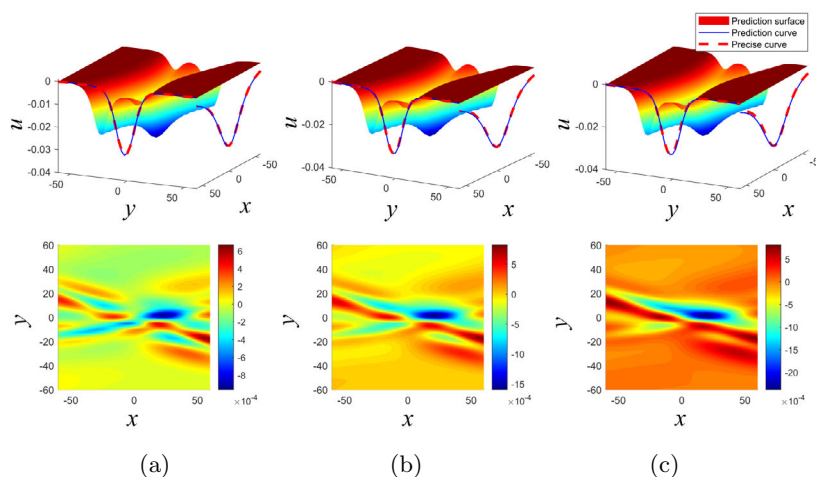


Figure 14 (Color online) The data-driven TO-type interactions and the corresponding density map of the error at $\tau = -1$ (a), $\tau = 0$ (b), and $\tau = 1$ (c)

The exact density map of the TO-type interaction, the density map with adaptive weight learning, the density map with fixed weight learning, and the corresponding satellite images are illustrated in Figure 15. It is worth noting that the distinctive feature of this interaction lies in the absence of any noticeable phase shift.

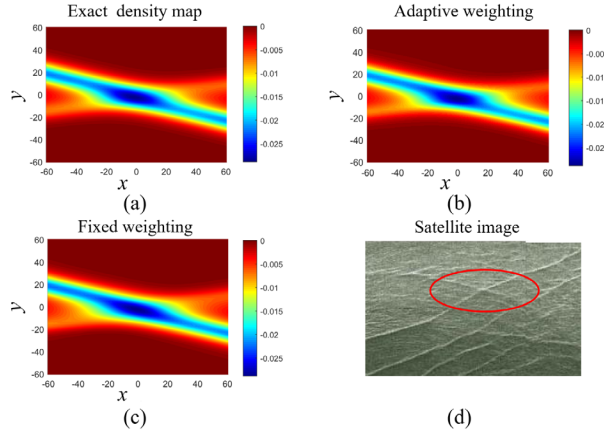


Figure 15 (Color online) The density map of the TO-type interactions: (a) The exact density map; (b) The learned density map of the adaptive weighting; (c) The learned density map of the fixed weighting; (d) The satellite imagery (ERS-2 SAR image of the Andaman acquired on 11 February 1997 at 0359 UTC), from [40]

The convergence curves of the adaptive weights β_1 and β_2 during the simulation process of TO-type interactions are shown in Figure 16. β_1 gradually decreases, while β_2 exhibits noticeable fluctuations in the initial iterations. In the last 1000 iterations, β_1 remains unchanged, while β_2 continues to iterate. The final iterative values for β_1 and β_2 are 49.23 and 7.516×10^{-5} , respectively.

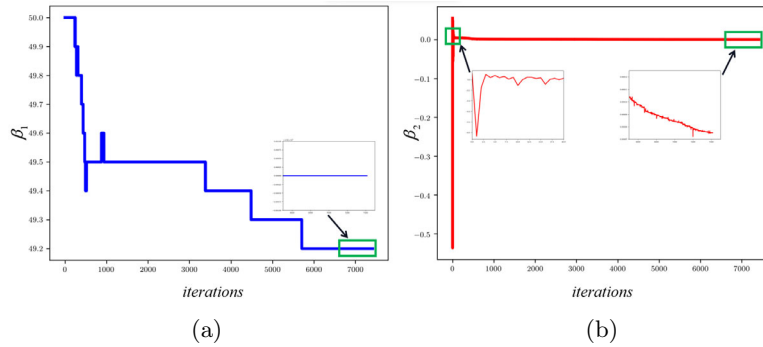


Figure 16 (Color online) The numerical convergence curve of adaptive weights for the data-driven TO-type interactions

4.5 Data-Driven TP-Type Interactions

To model the TP-type interactions, we employ a 12-layer feedforward neural network. The PINN model converge after 10063 iterations, which takes approximately 5549.18 seconds. The loss functions Loss_{KP} and Loss_b impressively reached values of 7.195×10^{-11} and 1.401×10^{-9} ,

respectively. Figure 17 offers three-dimensional representations of the TP-type interaction, complemented by density plots illustrating the associated errors. In contrast to P-type interactions, the amplitude within the TP-type interaction region is lower than that of any of the ISW. This visual aspect makes it appear as if two ISW are not interacting. Figure 18 presents the exact density map of the TP-type interaction, the density map with adaptive weight learning, the density map with fixed weight learning, and corresponding satellite images. This type of interaction has been rarely reported, and corresponds to an ISW pattern captured along the coast of Washington State by the RADARSAT-1 satellite.

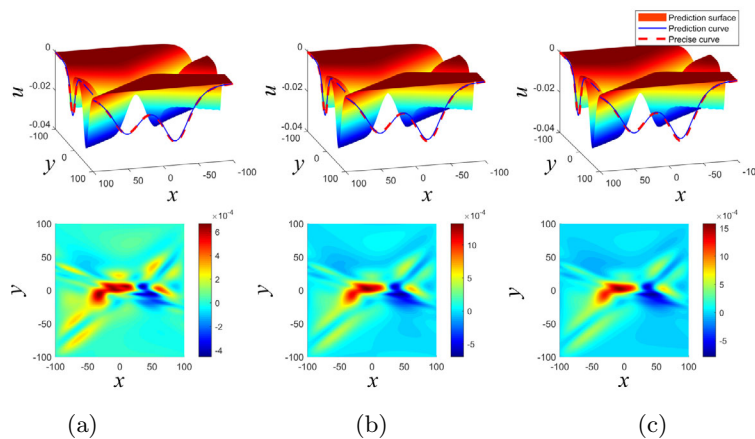


Figure 17 (Color online) The data-driven TP-type interactions and the corresponding density map of the error at $\tau = -1$ (a), $\tau = 0$ (b), and $\tau = 1$ (c)

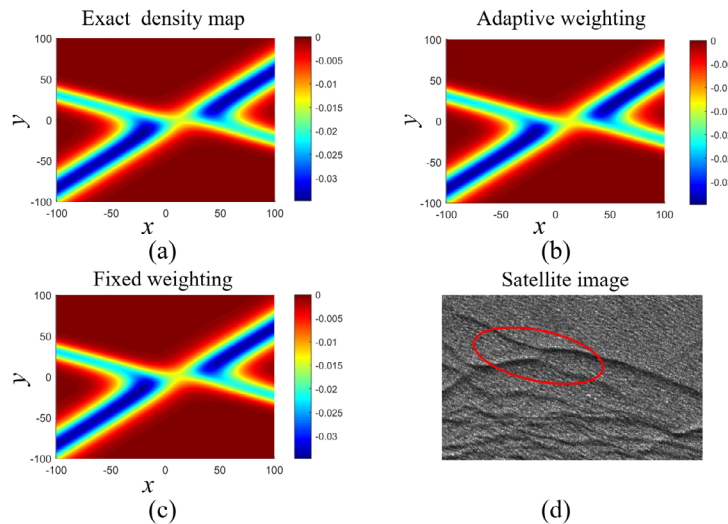


Figure 18 (Color online) The density map of the TP-type interactions: (a) The exact density map; (b) The learned density map of the adaptive weighting; (c) The learned density map of the fixed weighting; (d) The satellite imagery (RADARSAT-1 image showing internal waves off the coast of Washington State, acquired 9 August 1999 at 0155 UTC), from [40]

Convergence curves of the adaptive weights β_1 and β_2 during the simulation of TP-type interactions are depicted in Figure 19. The process of iteration is similar to the previous interactions, and the final iterative values for β_1 and β_2 are 48.98 and 3.742×10^{-5} , respectively.

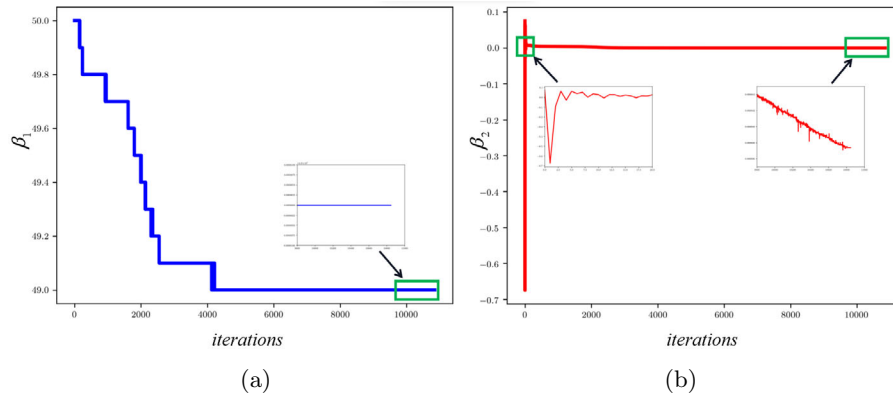


Figure 19 (Color online) The numerical convergence curve of adaptive weights for the data-driven TP-type interactions

4.6 Data-Driven Y-Shaped Interactions

Finally, we design an 8-layer neural network that successfully learns the Y-shaped interactions. The model undergoes a total of 8079 iterations, which are accomplished in a mere 3149.89 seconds. The loss functions Loss_{KP} and Loss_b reach 1.089×10^{-10} and 2.420×10^{-9} , respectively. The experimental results are shown in Figure 20. The Y-shaped interactions are formed by the resonant interaction of three ISW at a vertex, which demonstrates that the interaction of ISW can produce a Miles resonance. Figure 21 presents the exact density map of the Y-shaped interaction, the density map with adaptive weight learning, the density map with fixed weight learning, and the corresponding satellite images.

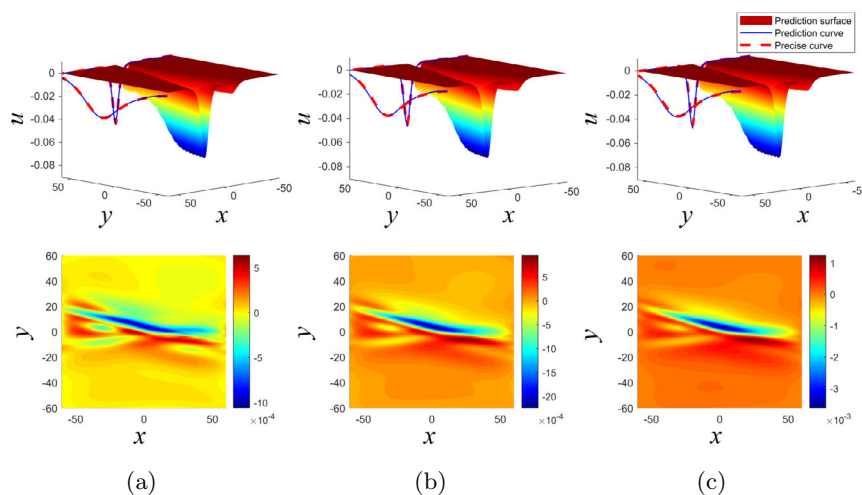


Figure 20 (Color online) The data-driven Y-shaped interactions and the corresponding density map of the error at $\tau = -1$ (a), $\tau = 0$ (b), and $\tau = 1$ (c)

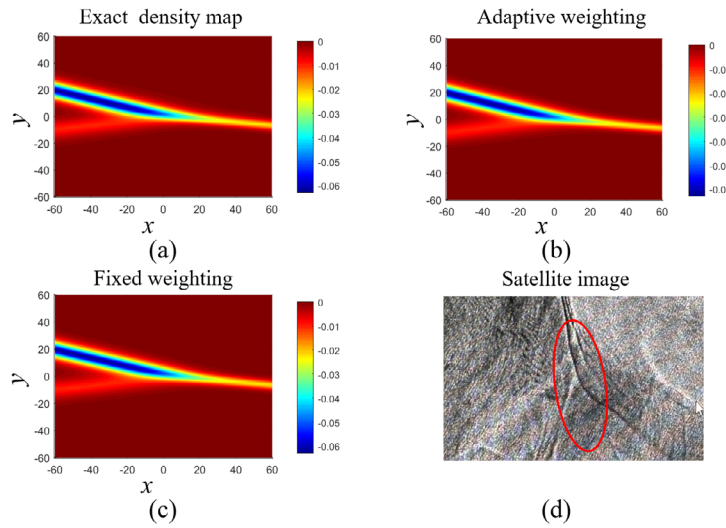


Figure 21 The density map of the Y-shaped interactions: (a) The exact density map; (b) The learned density map of the adaptive weighting; (c) The learned density map of the fixed weighting; (d) The satellite imagery (RADARSAT-1 image showing internal waves off the coast of Washington State, acquired 9 August 1999 at 0155 UTC), from [40]

The convergence curves for the adaptive weights, β_1 and β_2 , throughout the simulation of Y-shaped interactions are illustrated in Figure 22. The concluding values for β_1 and β_2 are 48.68 and -4.925×10^{-5} , respectively.

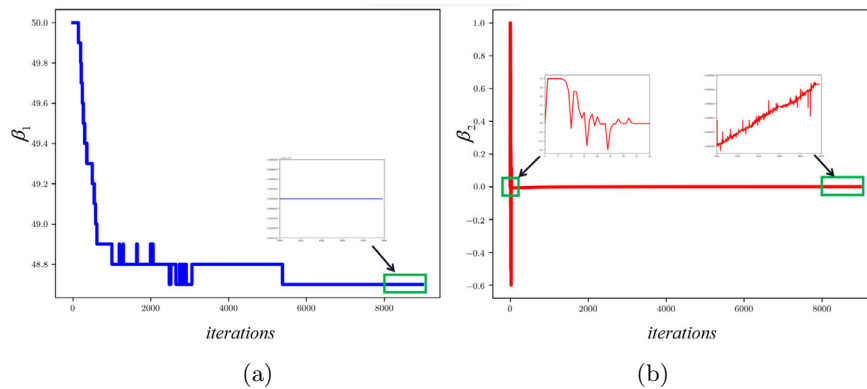


Figure 22 (Color online) The numerical convergence curve of adaptive weights for the data-driven Y-shaped interactions

We compile the final converged values of Loss_{KP} and Loss_b for all the experimental results, as detailed in Table 2. The values of each loss function term are very small, and there is no imbalance as shown in Figure 6.

Table 2 Convergence values of different terms in the loss function

	O-type	P-type	TO-type	TP-type	Y-shaped
Loss _{KP}	5.696×10^{-11}	4.224×10^{-9}	2.998×10^{-11}	7.195×10^{-11}	1.089×10^{-10}
Loss _b	1.980×10^{-9}	1.131×10^{-8}	5.655×10^{-9}	1.401×10^{-9}	2.420×10^{-9}

The PINN with weighted loss function method is specially designed for simulating oceanic ISWI. It successfully solves the problem of loss function imbalance caused by the properties of ISWI. It is able to accurately simulate the detailed structure and dynamical behavior of ISWI, and performs well in the simulation of many types of interactions.

5 Conclusions and Discussions

We introduce fixed and adaptive weighting methods into the loss function to design an improved PINN method suitable for simulating two-dimensional oceanic ISWI. At first, the classical PINN method in the two-dimensional space is used to simulate different types of ISWI with the support of three types of data points, and the simulation results show that these interactions are close to the zero plane, which is caused by two properties of the ISW. One is the significantly smaller amplitude compared to the wavelength. The other is the occurrence of these interactions on a relatively large spatial scale. Furthermore, these properties cause the loss function imbalance, i.e., the PINN model is biased towards optimizing the residuals of PDE during the training process, while ignoring the constraints of the initial boundary value conditions.

To solve this problem, we introduce the weighted loss function methods to balance the different terms in the loss function. Applying this improved PINN model we successfully simulate five different types of ISWI and obtain their detailed structural and dynamical behavior. The exact and predicted solution curves for all the results are in high agreement, with the errors mainly originating from the region of large gradients, that is, the interaction region. The learned density maps are in general agreement with the exact density maps and are able to correspond to satellite images of oceanic ISWI. In addition, both Loss_{KP} and Loss_b converge to a smaller value, indicating that the imbalance problem of the loss function is solved. The adaptive weighting method can automatically update the weights of different terms without manually adjusting the weights, in which the weights of Loss_{KP} are decreasing and converge preferentially, and the weights of Loss_b fluctuate more at the beginning and converge slowly. The PINN method with adaptive weighting exhibits a small generalization error, showing that it has a better generalization ability compared with other methods.

As is known, there are many different types of internal waves in the ocean, so it is important to establish new systems to describe these diverse types of internal waves, as well as to improve the PINN method to adapt to different types of internal waves. These will be our future research.

Acknowledgements

The authors would like to thank Dr. Zhengwu Miao for his kindly support and help.

Conflict of Interest

The authors declare no conflict of interest.

References

- [1] Boegman L and Stastna M, Sediment resuspension and transport by internal solitary waves, *Annual Review of Fluid Mechanics*, 2019, **51**: 129–154.
- [2] Alford M H, Peacock T, MacKinnon J A, et al., The formation and fate of internal waves in the South China Sea, *Nature*, 2015, **521**(7550): 65–69.
- [3] Cavaliere D, la Forgia G, Adduce C, et al., Breaking location of internal solitary waves over a sloping seabed, *Journal of Geophysical Research: Oceans*, 2021, **126**(2): e2020JC016669.
- [4] Ferrari R and Wunsch C, Ocean circulation kinetic energy: Reservoirs, sources, and sinks, *Annual Review of Fluid Mechanics*, 2009, **41**: 253–282.
- [5] Yuan C, Pan L, Gao Z, et al., Combined effect of topography and rotation on oblique internal solitary wave-wave interactions, *Journal of Geophysical Research: Oceans*, 2023, **128**(6): e2023JC019634.
- [6] Wang Z, Wang Z, and Yuan C, Oceanic internal solitary waves in three-layer fluids of great depth, *Acta Mechanica Sinica*, 2022, **38**(2): 321473.
- [7] Xue J, Graber H C, Romeiser R, et al., Understanding internal wave-wave interaction patterns observed in satellite images of the Mid-Atlantic Bight, *IEEE Transactions on Geoscience and Remote Sensing*, 2014, **52**(6): 3211–3219.
- [8] Kodama Y, Young diagrams and N-soliton solutions of the KP equation, *Journal of Physics A: Mathematical and General*, 2004, **37**(46): 11169.
- [9] Ablowitz M J and Baldwin D E, Nonlinear shallow ocean-wave soliton interactions on flat beaches, *Physical Review E*, 2012, **86**(3): 036305.
- [10] Guo L, Chen L, Mihalache D, et al., Dynamics of soliton interaction solutions of the Davey-Stewartson I equation, *Physical Review E*, 2022, **105**(1): 014218.
- [11] Chen G Y, Liu C T, Wang Y H, et al., Interaction and generation of long-crested internal solitary waves in the South China Sea, *Journal of Geophysical Research: Oceans*, **2011**, **116**(C6), DOI: 10.1029/2010JC006392.
- [12] Zheng Q, Klemas V, Yan X H, et al., Digital orthorectification of space shuttle coastal ocean photographs, *International Journal of Remote Sensing*, 1997, **18**(1): 197–211.
- [13] Alpers W, Wang-Chen H, and Hock L, Observation of internal waves in the Andaman Sea by ERS SAR, *IGARSS'97, 1997 IEEE International Geoscience and Remote Sensing Symposium Proceedings, Remote Sensing — A Scientific Vision for Sustainable Development*, 1997, **4**: 1518–1520.
- [14] Cui J, Dong S, and Wang Z, Study on applicability of internal solitary wave theories by theoretical and numerical method, *Applied Ocean Research*, 2021, **111**: 102629.

- [15] Grimshaw R H J, Smyth N F, and Stepanyants Y A, Interaction of internal solitary waves with long periodic waves within the rotation modified Benjamin-Ono equation, *Physica D: Nonlinear Phenomena*, 2021, **419**: 132867.
- [16] Bokaeeeyan M, Ankiewicz A, and Akhmediev N, Bright and dark rogue internal waves: The Gardner equation approach, *Physical Review E*, 2019, **99**(6): 062224.
- [17] Sun J, Tang X, and Chen Y, Oceanic internal solitary wave interactions via the KP equation in a three-layer fluid with shear flow, 2023, arXiv: 2311.07990.
- [18] Hornik K, Stinchcombe M, and White H, Multilayer feedforward networks are universal approximators, *Neural Networks*, 1989, **2**(5): 359–366.
- [19] Raissi M, Perdikaris P, and Karniadakis G E, Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations, *Journal of Computational Physics*, 2019, **378**: 686–707.
- [20] Wang L and Yan Z, Data-driven rogue waves and parameter discovery in the defocusing nonlinear Schrödinger equation with a potential using the PINN deep learning, *Physics Letters A*, 2021, **404**: 127408.
- [21] Sheng C, Wang L, Huang Z, et al., Transformer-based deep learning network for tooth segmentation on panoramic radiographs, *Journal of Systems Science & Complexity*, 2023, **36**(1): 257–272.
- [22] Tian S, Niu Z, and Li B, Mix-training physics-informed neural networks for high-order rogue waves of cmKdV equation, *Nonlinear Dynamics*, 2023, **111**(17): 16467–16482.
- [23] Zhu Q, Liu Z, and Yan J, Machine learning for metal additive manufacturing: Predicting temperature and melt pool fluid dynamics using physics-informed neural networks, *Computational Mechanics*, 2021, **67**: 619–635.
- [24] Zhong M and Yan Z, Data-driven forward and inverse problems for chaotic and hyperchaotic dynamic systems based on two machine learning architectures, *Physica D: Nonlinear Phenomena*, 2023, **446**: 133656.
- [25] Raissi M, Yazdani A, and Karniadakis G E, Hidden fluid mechanics: Learning velocity and pressure fields from flow visualizations, *Science*, 2020, **367**(6481): 1026–1030.
- [26] Mao Z, Jagtap A D, and Karniadakis G E, Physics-informed neural networks for high-speed flows, *Computer Methods in Applied Mechanics and Engineering*, 2020, **360**: 112789.
- [27] Cai S, Mao Z, Wang Z, et al., Physics-informed neural networks (PINNs) for fluid mechanics: A review, *Acta Mechanica Sinica*, 2021, **37**(12): 1727–1738.
- [28] Pan X, Wang J, Zhang X, et al., A deep-learning model for the amplitude inversion of internal waves based on optical remote-sensing images, *International Journal of Remote Sensing*, 2018, **39**(3): 607–618.
- [29] Zhang X and Li X, Satellite data-driven and knowledge-informed machine learning model for estimating global internal solitary wave speed, *Remote Sensing of Environment*, 2022, **283**: 113328.
- [30] Liang K, Zhang M, Li Z X, et al., A new method to estimate the speed of internal solitary waves based on a single optical remote sensing image, *International Journal of Remote Sensing*, 2022, **43**(17): 6430–6444.
- [31] Zhang M, Hu H, Du P, et al., Detection of an internal solitary wave by the underwater vehicle based on machine learning, *Physics of Fluids*, 2022, **34**(11): 115137.
- [32] Wang S, Yu X, and Perdikaris P, When and why PINNs fail to train: A neural tangent kernel perspective, *Journal of Computational Physics*, 2022, **449**: 110768.
- [33] Goswami S, Anitescu C, Chakraborty S, et al., Transfer learning enhanced physics informed

- neural network for phase-field modeling of fracture, *Theoretical and Applied Fracture Mechanics*, 2020, **106**: 102447.
- [34] Pang G, Lu L, and Karniadakis G E, fPINNs: Fractional physics-informed neural networks, *SIAM Journal on Scientific Computing*, 2019, **41**(4): A2603–A2626.
- [35] Yang L, Meng X, and Karniadakis G E, B-PINNs: Bayesian physics-informed neural networks for forward and inverse PDE problems with noisy data, *Journal of Computational Physics*, 2021, **425**: 109913.
- [36] Jin X, Cai S, Li H, et al., NSFnets (Navier-Stokes flow nets): Physics-informed neural networks for the incompressible Navier-Stokes equations, *Journal of Computational Physics*, 2021, **426**: 109951.
- [37] Lin S and Chen Y, A two-stage physics-informed neural network method based on conserved quantities and applications in localized wave solutions, *Journal of Computational Physics*, 2022, **457**: 111053.
- [38] Lin S and Chen Y, Physics-informed neural network methods based on Miura transformations and discovery of new localized wave solutions, *Physica D: Nonlinear Phenomena*, 2023, **445**: 133629.
- [39] Miao Z and Chen Y, VC-PINN: Variable coefficient physics-informed neural network for forward and inverse problems of PDEs with variable coefficient, *Physica D: Nonlinear Phenomena*, 2023, 133945.
- [40] An atlas of internal solitary-like waves and their properties, <http://www.internalwaveatlas.com/>.
- [41] Xiang Z, Peng W, Liu X, et al., Self-adaptive loss balanced physics-informed neural networks, *Neurocomputing*, 2022, **496**: 11–34.