



# Predicting particle catchment areas of deep-ocean sediment traps using machine learning

Théo Picard<sup>1</sup>, Jonathan Gula<sup>2,3,5</sup>, Ronan Fablet<sup>4,5</sup>, Jeremy Collin<sup>1</sup>, and Laurent Mémery<sup>1</sup>

<sup>1</sup>Laboratoire des Sciences de l’Environnement Marin (LEMAR), Univ Brest,  
CNRS, IRD, Ifremer, IUEM, Plouzané, France

<sup>2</sup>Laboratoire d’Océanographie Physique et Spatiale (LOPS), Univ Brest,  
CNRS, IRD, Ifremer, IUEM, Plouzané, France

<sup>3</sup>Institut Universitaire de France (IUF), Paris, France

<sup>4</sup>IMT Atlantique, Lab-STICC, Plouzané, France

<sup>5</sup>ODYSSEY, Inria, Brest, France

**Correspondence:** Théo Picard (theo.picard@univ-brest.fr)

Received: 23 November 2023 – Discussion started: 5 December 2023

Revised: 31 May 2024 – Accepted: 22 July 2024 – Published: 19 September 2024

**Abstract.** The ocean’s biological carbon pump plays a major role in climate and biogeochemical cycles. Photosynthesis at the surface produces particles that are exported to the deep ocean by gravity. Sediment traps, which measure deep-carbon fluxes, help to quantify the carbon stored by this process. However, it is challenging to precisely identify the surface origin of particles trapped thousands of meters deep due to the influence of ocean circulation on the sinking path of carbon. In this study, we conducted a series of numerical Lagrangian experiments in the Porcupine Abyssal Plain region of the North Atlantic and developed a machine learning approach to predict the surface origin of particles trapped in a deep-ocean sediment trap. Our numerical experiments support the predictive performance of the machine learning approach, and surface conditions appear to provide valuable information for accurately predicting the source area, suggesting a potential application with satellite data. We also identify factors that potentially affect prediction efficiency, and we show that the best predictions are associated with low kinetic energy and the presence of mesoscale eddies above the trap. This new tool could provide a better link between satellite-derived sea surface observations and deep-ocean sediment trap measurements, ultimately improving our understanding of the biological-carbon-pump mechanism.

## 1 Introduction

The biological carbon pump (BCP) plays a major role in climate and biogeochemical cycles. The BCP reduces unperturbed atmospheric CO<sub>2</sub> by 35 % to 50 % (Williams and Follows, 2011) by exporting organic matter to the deep ocean, thereby supporting abyssal food webs (Billett et al., 1983; Rembauville et al., 2018). The BCP is driven by photosynthesis that occurs within the euphotic layer, typically between 0 and 200 m, producing gravitationally sinking particulate organic carbon (POC). Sinking particles have a wide range of vertical velocities, from neutral buoyancy to more than 600 m d<sup>-1</sup> (Villa-Alfageme et al., 2016), and are usually considered the main contributors to the BCP (Armstrong et al., 2001; Alonso-González et al., 2010; Siegel et al., 2014; Le Moigne, 2019). While most POC is remineralized in the euphotic and mesopelagic zones (200–1000 m), a small but significant fraction of POC reaches the deep ocean below 1000 m, where it is sequestered for hundreds or thousands of years (Lampitt et al., 2008; Burd et al., 2016). Despite its critical importance, the annual estimate of global carbon export remains poorly constrained, ranging from 5 to over 12 Pg C yr<sup>-1</sup> (Turner, 2015). Therefore, quantifying the biological carbon pump is key to understanding the global carbon cycle and how the BCP will respond to climate change (Passow and Carlson, 2012; Henson et al., 2022).

Export production has historically been observed using sediment traps (STs) from the upper thermocline down to several thousand meters (Le Moigne et al., 2013). However, linking local primary production to observed exported carbon fluxes is a challenging task. To effectively interpret ST time series, the catchment area, defined as the surface domain containing all likely positions from which particles entering the trap could originate (Deuser et al., 1988), must be clearly identified. Traditionally, the catchment area is considered to be the zone directly above the traps (Deuser and Ross, 1980; Alldredge and Gotschalk, 1988; Asper et al., 1992; Armstrong et al., 2001; Lampitt et al., 2023). While this approach demonstrates a first-order coupling, it is based on a strong hypothesis: POC export via gravitational sinking is considered from a quasi-one-dimensional perspective. However, ocean circulation significantly affects the sinking paths of particles. Thus, the origins of particles reaching a deep ST can be distributed over a large domain that mainly depends on ocean dynamics at all scales (Siegel et al., 1990; Deuser et al., 1990; Burd et al., 2010; Liu et al., 2018; Dever et al., 2021)

Recent studies have focused on the influence of ocean circulation and have evaluated funnel statistics for specific sediment trap locations using particle backtracking and numerical simulations (Siegel et al., 2008; Liu et al., 2018; Wekerle et al., 2018; Wang et al., 2022). These studies have shown that the catchment area is highly dependent on several factors, such as sinking velocities, trap depth, and regional and seasonal advective processes. The application of Lagrangian backtracking approaches with assimilated ocean currents has been proposed to relate ocean circulation to real ST observations (Frigstad et al., 2015; Ma et al., 2021). However, a full 3D reconstruction of ocean states from available satellite-derived and in situ observations is highly challenging and prone to significant biases in the retrieval of mesoscale and submesoscale dynamics (Cutolo et al., 2022). This leads to uncertainties in the prediction of catchment areas and in the assessment of the BCP.

Machine learning tools are increasingly being used to tackle these complex problems, and they generally offer an end-to-end formulation that is both easier to develop and computationally cheaper. Oceanography studies have already demonstrated the benefits of machine learning, particularly in predicting the properties of the ocean interior from observations of the ocean surface (Chapman and Charantonis, 2017; Bolton and Zanna, 2019; George et al., 2021; Manucharyan et al., 2021; Pauthenet et al., 2022). These tools can achieve state-of-the-art performance capabilities or even outperform standard operational interpolation data assimilation schemes when focused on specific ocean variables (Manucharyan et al., 2021; Beauchamp et al., 2022; Cutolo et al., 2022). They can also significantly reduce the computational complexity of numerical simulations, including those for Lagrangian particle trajectories (Jenkins et al., 2022). In addition, while it can be complex to investigate how the sur-

face constrains particle motion at depth using 3D reanalysis, machine learning offers new ways of assessing the main drivers of particle displacement. This can lead to a better understanding of the variables and processes involved in particle sinking and a clearer identification of the temporal/spatial resolution required for effective particle trajectory reconstruction.

In this study, we use deep learning schemes to study the catchment areas of deep-ocean particles. Our main contributions are as follows: (i) we formulate the prediction of the catchment area of particles trapped in STs as the supervised learning of a regression model from ocean circulation variables, (ii) we investigate whether ocean surface conditions constrain the paths of sinking particles, and (iii) we analyze the main factors affecting prediction performance. We report on numerical experiments conducted in a case study area in the North Atlantic using realistic high-resolution simulation data.

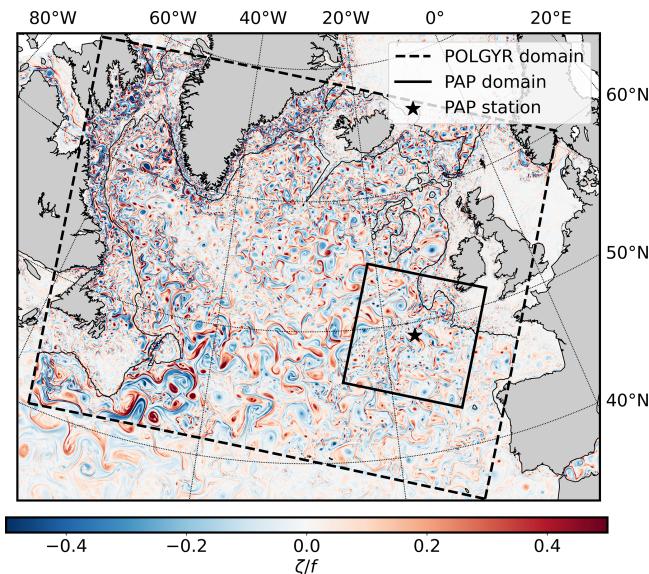
The paper is organized as follows. Section 2 introduces the methodology used to design the databases for training and performance evaluation. Section 3 presents the neural-network structure. In Sect. 4, we evaluate the accuracy of the predictions, compare different configurations, and analyze the relationship between input conditions and prediction performance. In Sect. 5, we discuss potential improvements and future work. Our conclusions are presented in Sect. 6.

## 2 Data and case study area

In a case study, we focus on the origin of particles captured by moored STs at the Porcupine Abyssal Plain (PAP) station, located in the UK ( $49^{\circ}$  N,  $16.5^{\circ}$  W), which provides fluxes with a temporal coverage of more than 30 years (Hartmann et al., 2021). This strategy is based on Wang et al. (2022), who characterized the origins of particles collected at the PAP station using a simulation-based experimental setup consisting of realistic Coastal and Regional Ocean COmmunity (CROCO) model simulations and particle-backtracking experiments. In this study, we simulate, in the same way, particles falling into a fictive PAP ST located at 1000 m of depth. We focus on particles with a vertical sinking velocity of  $50 \text{ m d}^{-1}$ , which represents the slow range of sinking particles observed at depth in the region (Villa-Alfageme et al., 2014). Slow particles are more susceptible to being influenced by ocean circulation (Wang et al., 2022), which, in turn, makes the prediction of catchment areas under these conditions more challenging.

### 2.1 Numerical simulation around the PAP station

We use the realistic North Atlantic subpolar gyre simulation (called POLGYR) designed and validated by Le Corre et al. (2020). This simulation is run using the CROCO model, based on the Regional Ocean Modeling System (ROMS)



**Figure 1.** Snapshot of the North Atlantic subpolar gyre simulation (POLGYR) taken on 2 June 2008, showing relative vorticity. The dashed square indicates the domain of the simulation. The solid square indicates the subdomain we are focusing on. The location of the PAP station is indicated by the black star. The black contour indicates the bathymetry at 1000 m of depth.

(Shchepetkin and McWilliams, 2005), which solves hydrostatic primitive equations for momentum and state variables. The domain has  $2000 \times 1600$  grid points and a horizontal grid resolution of 2 km, resolving the mesoscale level and part of the submesoscale level. The model has 80 vertical sigma levels, with a vertical spacing of  $\sim 5$  m at the surface and up to 100 m in the intermediate layer. After a 2-year spin-up period, two simulations, called POLGYR1 and POLGYR2, are run. POLGYR1 runs between 1 January 2002 and 4 December 2009 (8 years) and is used to create a training and evaluation database, while POLGYR2 runs between 24 July 2003 and 12 January 2009 (about 6.5 years) and is used for the test database. The two simulation setups are identical in all aspects except for the initial and boundary conditions, which are perturbed in POLGYR2. After the spin-up, the chaotic evolution results in uncorrelated dynamics between the two simulations (Fig. A1). We save snapshots for both simulations every 12 h. We focus on a  $1040 \times 1040$  km subdomain centered on the PAP station (Fig. 1). This region is characterized by moderate kinetic energy compared to the western and northern parts of the subpolar gyre, with a mean flow of about  $0.05 \text{ m s}^{-1}$  (Le Cann, 2005).

## 2.2 Lagrangian backtracking of catchment areas

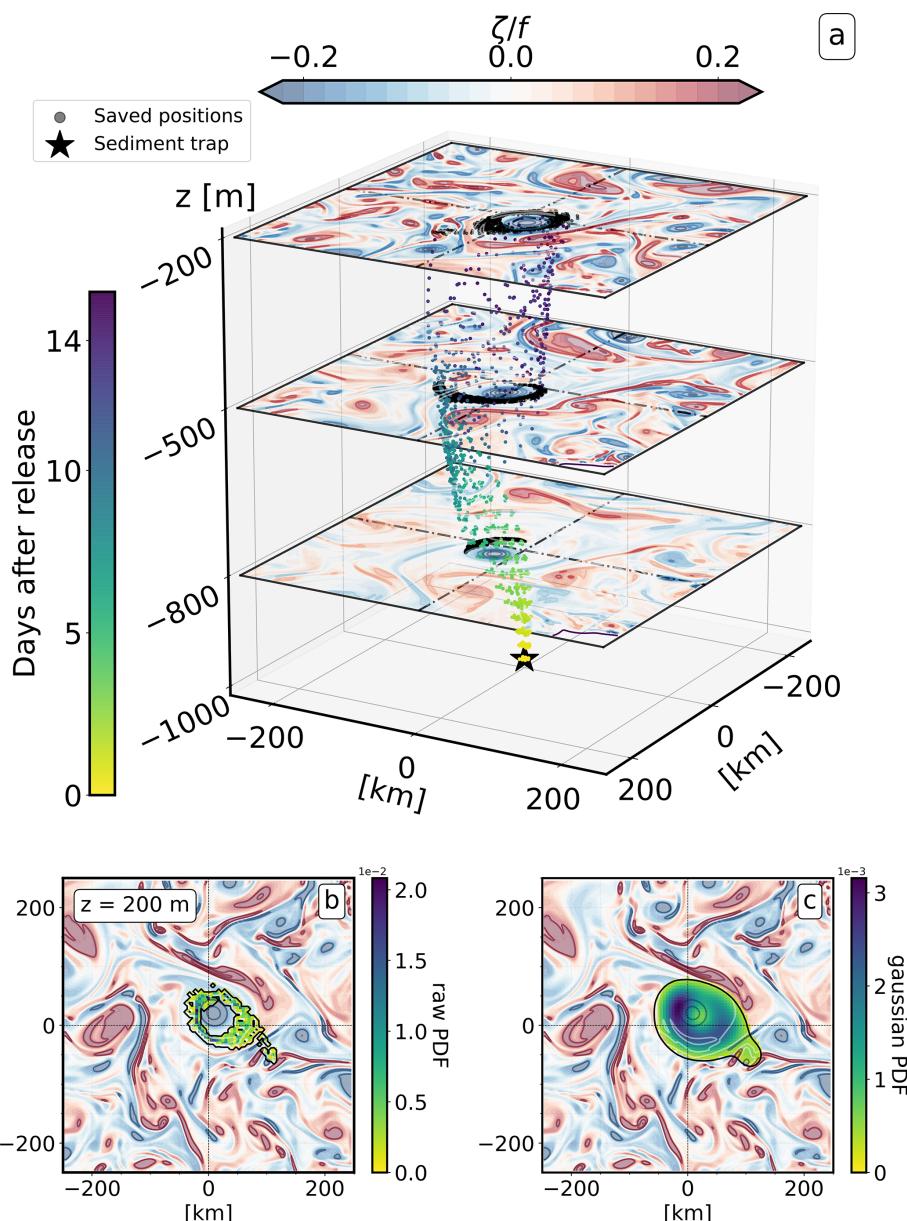
A series of Lagrangian experiments are performed offline using “Pyicles” (Gula and Collin, 2021), employing 3D velocities from the POLGYR1 and POLGYR2 simulations. Pyicles is a parallel Fortran–Python Lagrangian tool for offline

3D advection and has the ability to include particle behavior. Transport is performed along the native Arakawa C-grid and terrain-following vertical coordinates of the ocean model. The 3D velocity fields are linearly interpolated at particle positions, which are advected using a “Runge–Kutta 4” numerical scheme. In addition to passive advection, a negative vertical velocity is applied to simulate the sinking of dense particles. The numerical schemes have been shown to be robust to trajectory reversibility (Wang et al., 2022), employing 12-hourly input fields and a time step of 120 s.

For each experiment, we apply the following procedure. At 1000 m of depth, we release 36 particles in backtracking every 12 h over a period of 10 d (particle collection period) for a total of 720 particles. The particles are released uniformly over a patch measuring  $10 \text{ km} \times 10 \text{ km}$  and ascend until they reach the base of the euphotic layer at 200 m. Particles have a constant sinking velocity of  $50 \text{ m d}^{-1}$ , and their journey takes on average 15 d, but this can vary from 10 to 20 d (Wang et al., 2022). We save the particles’ positions at every 100 m step and compute the resulting probability density function (PDF). Figure 2a illustrates a typical experiment in which most of the particles are attracted to an anticyclone structure visible between 200 and 1000 m. Biological particles are mostly created between the surface and 200 m, but we initially consider that the PDF at 200 m represents the depth from which the particles are exported. The PDFs saved at vertical levels lower than 200 m provide information on the three spatial dimensions and the temporal-dimension ( $3\text{D} + T$ ) path of the particles to better constrain the training phase of the network. To improve the training efficiency, we also apply a Gaussian filter to each PDF (Fig. 2b and c). We performed sensitivity tests to ensure that increasing the number of particles and the size of the patch does not significantly affect the PDF.

For each Lagrangian experiment, we create an output tensor, expressed as  $Y_i = (n_{\text{output}} \times n_x \times n_y)$ , where  $n_x$  and  $n_y$  represent the number of points on the horizontal axis and  $n_{\text{output}}$  represents the number of vertical levels at which the PDFs are computed. In our case, we consider that  $n_{\text{output}} = 8$ , corresponding to one layer every 100 m of depth from 900 to 200 m. Based on Wang et al. (2022), who computed the source region at 200 m for particles collected by the moored sediment traps over 7 years (2002–2008), a horizontal domain of  $800 \text{ km} \times 800 \text{ km}$ , centered on the ST, was chosen to encompass all the source particle positions. Due to computational constraints and the effective resolution of the model, which is expected to be around 8 km (Soufflet et al., 2016), the horizontal resolution of all data is downsampled to 8 km, resulting in  $n_y = n_x = 100$  points.

For each output ( $Y_i$ ), we store a corresponding input tensor, expressed as  $X_i = (n_{\text{input}} \times n_x \times n_y)$ , which contains the hydrodynamic conditions of the experiment ( $i$ ). These data include data on temperature, vorticity, the horizontal velocities ( $u$  and  $v$ ), and sea surface height (SSH). All the variables are extracted on the same horizontal domain as  $Y_i$  and are



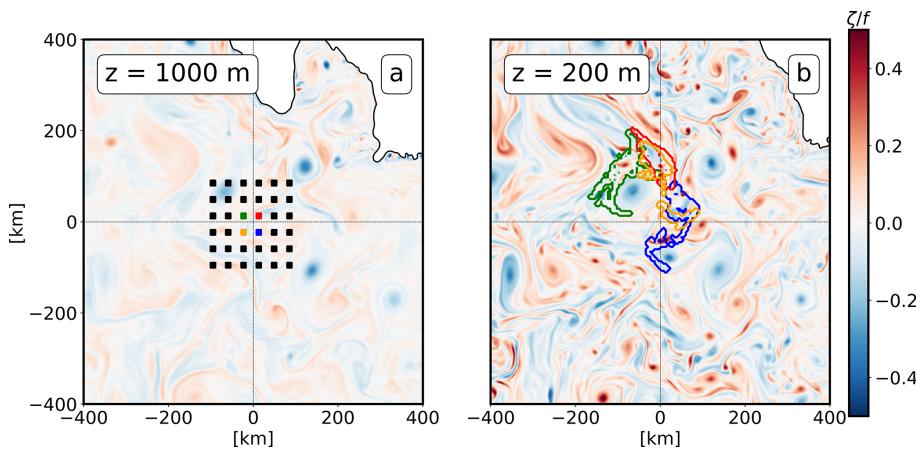
**Figure 2.** (a) 3D view of the Lagrangian experiment. The different colors of the particles represent the number of days since their release from the ST (5 % of the particles are plotted). Snapshots of relative vorticity during the particle crossing are displayed for three vertical layers ( $-800$ ,  $-500$ , and  $-200$  m). The black points in the visualization indicate all saved positions of particles at these specific vertical levels. (b) The computed raw PDF of particles at 200 m. (c) The PDF at 200 m after applying the Gaussian filter.

downscaled to a resolution of 8 km. To capture the temporal variability, we consider a 10 d sampling period over a 30 d time window (four time steps). To evaluate the importance of surface data in comparison to subsurface data (Sect. 4), we distinguish two datasets,  $D_{\text{4 layers}}$  and  $D_{\text{surf}}$ :

- $D_{\text{4 layers}}$  contains variables at four vertical levels: the surface level, 200 m, 500 m, and 1000 m. Each input tensor has the dimensions  $68 \times 100 \times 100$ , where  $n_{\text{input}} = 68$  represents the number of input fields ex-

tracted for each experiment.

- $D_{\text{surf}}$  contains only sea surface conditions. This results in input sensors with the dimensions  $20 \times 100 \times 100$ .



**Figure 3.** Snapshots of relative vorticity taken on 22 October 2003 (POLGYR1), i.e., 20 d after the particle release. **(a)** Vorticity field at 1000 m of depth, showing the 36 initial positions of the particles (black and colored squares). **(b)** Vorticity field at 200 m of depth, with the origins of the backtracked particles for the four colored patches from panel **(a)** superimposed.

### 3 Proposed deep learning scheme

This section introduces the proposed deep learning scheme, specifically the proposed neural architecture and the training scheme.

#### 3.1 Training, validation, and test data

Following the training and evaluation frameworks used in deep learning studies (Lecun et al., 2015), we consider independent training, validation, and test datasets as follows. Using the Lagrangian experiments presented above, we create a training and validation dataset using the first simulation setup, POLGYR1 (2002 to 2008). During this period, Lagrangian experiments are realized every 10 d at 36 fictive ST positions around the PAP region, resulting in a total of 10 260 samples (Fig. 3a). The STs are close enough to the PAP site to involve the same hydrodynamic conditions. We place the patch centers 36 km apart, allowing particles from two different patches to be separated by at least 26 km, which is slightly above the Rossby radius value in the region (Chelton et al., 1998). This distance is sufficient for observing significant differences in the catchment areas for two consecutive patches (Fig. 3b). We divide the samples into a training database (8604 samples from January 2002 to September 2008) and an evaluation database (1224 samples from the year 2009). Note that data from September 2008 to January 2009 are not used. The two datasets are separated by a 100 d period to ensure statistical independence. For the test dataset, 6800 samples are created with POLGYR2. These samples are independent of the training and evaluation experiments.

#### 3.2 Architecture

As shown in Fig. 4, we use a U-Net type of architecture (Ronneberger et al., 2015). U-Net is a state-of-the-art neural ar-

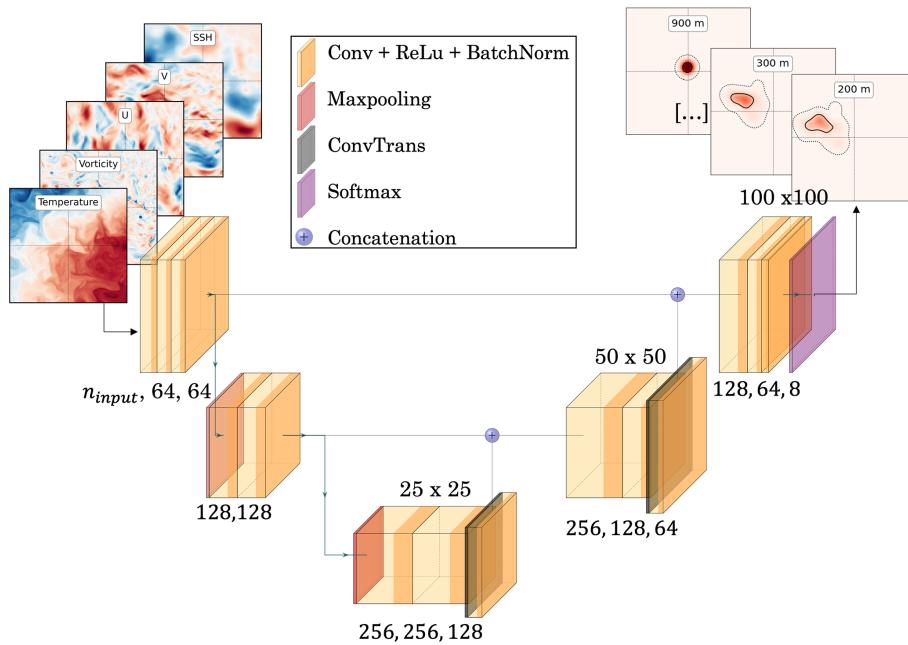
chitecture used for a wide range of image-to-image mapping tasks, including applications in ocean studies (Barth et al., 2020; Beauchamp et al., 2022). The input for the network includes hydrodynamic conditions and a tensor with the dimensions  $8 \times 100 \times 100$  that represents the initial PDF of the particles. All initial PDFs have a probability of 0.25 over the four points in the center of the map, representing the release patch (not shown). The network consists of a series of convolutional layers (with a kernel size of 5), ReLU activation, and batch normalization layers. A three-step pooling process is used to downsample the channel dimension, i.e., the dimension representing the number of dynamical images for one experiment. At each step, the number of features is doubled. This successive resolution downscaling from 8 to 32 km, combined with skip connections and concatenation (as additions over the channel dimension), may facilitate the detection of structures at different spatial scales (Ronneberger et al., 2015). The final layer of the U-Net architecture is a “softmax” layer that provides normalized PDF predictions. The output of the network is a tensor with the dimensions  $8 \times 100 \times 100$  and is composed of the predicted PDFs from 900 to 200 m (every 100 m). The last PDF at 200 m represents the predicted origin of the particles.

#### 3.3 Training scheme

We consider the Bhattacharyya coefficient (Bhattacharyya, 1943) as a training criterion to assess how well the predicted PDFs match the real ones. It is expressed as

$$\text{BC}_z = \sum_{i \in D} \sqrt{P_{i,z} Q_{i,z}}, \quad (1)$$

where  $D$  represents the PAP domain,  $P_i$  is the predicted probability, and  $Q_i$  is the true probability at point  $i$  and depth  $z$ . The Bhattacharyya coefficient provides a similarity score between two PDFs and is used to derive the Bhattacharyya



**Figure 4.** General U-Net architecture. The inputs include temperature, vorticity,  $u$ ,  $v$ , and SSH images at multiple time steps and depths, concatenated with the initial PDFs (not shown). The outputs include the PDFs at eight depths, representing the distribution of the particles for each depth. At the top of each layer, the resolution of the images is shown, and at the bottom of each layer, the number of channels is shown.

training loss (BL),

$$\text{BL}_z = 1 - \text{BC}_z. \quad (2)$$

The value of  $\text{BL}_z$  ranges from 1 to 0, with 0 representing a perfect prediction. During the training phase, we aim to minimize the loss function ( $\mathcal{L}$ ), expressed as

$$\mathcal{L} = \frac{\text{BL}_{200\text{m}} + \text{BL}_{300\text{m}} + \dots + \text{BL}_{800\text{m}} + \text{BL}_{900\text{m}}}{n_{\text{layer}}}, \quad (3)$$

where  $\mathcal{L}$  is the mean of the BL computed at each of the eight vertical layers. The aim is to force the model to consider the evolution of particles along the water column. Empirically, this method improves the performance of the trained model in comparison to experiments where training loss is based only on  $\text{BL}_{200\text{m}}$ .

We implement our deep learning scheme using PyTorch (Paszke et al., 2019). For the training phase, the Adam optimization algorithm (Kingma and Ba, 2015) is used with the following hyperparameters:  $\beta = (0.5, 0.999)$ , zero weight decay, and a learning rate of 0.001. The training process is performed using mini-batches with a size of 32. After 50 training epochs, the best model is selected based on its performance using the evaluation data. We further improve the performance and robustness of the model using a bootstrapping method with 10 replicates (Breiman, 1996). The final prediction consists of a set of PDFs computed as the median of the predictions from the 10 models, followed by a re-normalization step. We also compute the standard deviation of the 10 predictions as a confidence index (see Sect. 5).

## 4 Results

We present the results of the numerical experiments conducted in the PAP case study. First, we detail a benchmarking experiment to evaluate the performance of the proposed deep learning schemes. Second, we further analyze how hydrodynamic conditions affect prediction performance.

### 4.1 Prediction performance and comparison

We evaluate the performance of several models using the test database by employing the  $\text{BL}_{200\text{m}}$  metric (i.e., the Bhattacharyya distance between the predicted PDF and the true PDF for the catchment area at 200 m of depth). We observe an empirical polynomial relationship between the BL score and the percentage of PDFs predicted, defined as  $\sum_{i \in D} \min(P_{i,200\text{m}}, Q_{i,200\text{m}})$  (Fig. B1). We set up the following arbitrary evaluation criterion: a prediction is considered valid if the BL score is less than 0.2 and invalid if the BL score is greater than 0.3. These two thresholds correspond to percentages of predicted PDFs greater than or equal to 55 % and 45 %, respectively. For predictions with a BL score between 0.2 and 0.3, the global area is often still well predicted, but the values of the PDF are usually rough. In these cases, we create a third class called “quasi-valid”. Predictions in this class are not considered valid, but they still provide valuable information in the search for particle origins, and this information needs to be highlighted in the final score statistics.

**Table 1.** Scores and performance capabilities for different model architectures. For each model, we indicate the number of parameters, the average prediction score ( $BL_{200m}$ ), and the percentages of valid, quasi-valid, and non-valid test cases.

Models	No. of parameters	$BL_{200m}$	Valid prediction (%)	Quasi-valid prediction (%)	Non-valid prediction (%)
U-Net <sub>4 layers</sub>	1.9 M ( $\times 10$ )	0.13	81	13	6
U-Net <sub>surf</sub>	1.8 M ( $\times 10$ )	0.18	66	19	15
U-Net <sub>surf-nb</sub>	1.8 M	0.21	58	23	20
CNN <sub>surf-basic</sub>	205 K	0.61	0	1	99
Baseline	0	0.45	0	8	92

We benchmark different models in Table 1. The U-Net<sub>4 layers</sub> model is the U-Net architecture presented in Sect. 3, trained and tested with the  $D_{4 \text{ layers}}$  database, which contains both surface and subsurface information. The U-Net<sub>surf</sub> model only uses surface information from the  $D_{\text{surf}}$  database. The U-Net<sub>surf-nb</sub> model is similar to the U-Net<sub>surf</sub> model but does not use the bootstrap method. The CNN<sub>surf-basic</sub> model includes a simple convolutional-neural-network (CNN) architecture consisting of a series of convolutional layers with a ReLU activation function, which we tested with  $D_{\text{surf}}$ . We also include a baseline prediction, which corresponds to the average of all the true PDFs used in the training database. This mean PDF roughly corresponds to a 2D Gaussian distribution centered on the PAP station, as presented in Wang et al. (2022).

The baseline and CNN<sub>surf-basic</sub> models have poor performance, with zero valid cases. This highlights the complexity of the prediction task and the need for non-trivial models. It is noteworthy that the U-Net architecture seems to be well suited for this problem. The U-Net<sub>surf-nb</sub> model achieves a very good performance, with 58 % of cases being valid, increasing to 66 % with the use of bootstrapping. Taking the quasi-valid predictions into account, ~ 85 % of the predictions provide valuable information with only surface data. However, a comparison between the U-Net<sub>4 layers</sub> and U-Net<sub>surf</sub> models demonstrates the importance of subsurface information for obtaining more accurate and robust predictions. With 81 % of cases being valid and 13 % of cases being quasi-valid, the U-Net<sub>4 layers</sub> model clearly outperforms the U-Net<sub>surf</sub> model. This means that in about 10 %–15 % of cases, subsurface data are required to make valid predictions. Figure 5 shows examples of predictions from the test database using the U-Net<sub>4 layers</sub> and U-Net<sub>surface</sub> models. The PDFs are represented by two types of contour: one containing 99 % of the particles (dotted contours) and the other containing 50 % of the particles (solid contours). Both models are effective in predicting the overall catchment area. However, valid predictions are mainly characterized by a well-predicted center of the PDF, which is usually more accurate for the U-Net<sub>4 layers</sub> model. In the following, we only focus on predictions from the U-Net<sub>4 layers</sub> and U-Net<sub>surface</sub> models.

## 4.2 Statistical analysis of the performance

We quantitatively analyze the potential causes of valid or invalid predictions. The examples show that the shape of particle distributions has a significant impact on score prediction. We find that PDFs located far from the STs, with mass centers  $> 150$  km (i.e., the average distance between particles and their source), are generally not well predicted (Fig. 5; indices 16, 21, and 25 in Table 2). Similarly, complex PDF shapes characterized by multiple patches spread across the domain (see indices 2, 15, and 16 in Fig. 5) appear difficult to predict. Entropy ( $-\sum p_i \log(p_i)$ , where  $p_i$  is the PDF value at point  $i$ ) characterizes the complexity of the PDF beyond its variance. Entropy increases with the spread of particles and in multimodal distributions, and high entropy values ( $> 6.6$ ) could explain the prediction bias in some cases (indices 2, 15, and 16 in Table 1). This trend is confirmed by a statistical analysis over the entire test dataset, shown in Fig. 6. For PDFs with a mass center greater than 200 km, the probability of obtaining a valid prediction with the U-Net<sub>surf</sub> model is less than 30 %. Similarly, we observe a trend of progressive degradation in the scores for PDFs with an entropy value higher than 6.6 (Fig. 6b).

Several physical factors could be responsible for the final PDF shape. To characterize the local dynamics, we compute the mean kinetic energy (MKE) and the eddy kinetic energy (EKE) around the ST as follows:

$$\underbrace{\frac{1}{2}(\overline{u^2} + \overline{v^2})}_{\text{KE}} = \underbrace{\frac{1}{2}(\bar{u}^2 + \bar{v}^2)}_{\text{MKE}} + \underbrace{\frac{1}{2}(u'^2 + v'^2)}_{\text{EKE}}, \quad (4)$$

where  $u$  and  $v$  are the two components of the horizontal ocean velocity. Here,  $\bar{u}$  and  $\bar{v}$  are the velocities averaged over a 30 d window, corresponding to the characteristic time period of our Lagrangian experiments. Finally,  $u'$  and  $v'$  are defined as follows:

$$u' = u - \bar{u}, \quad (5)$$

$$v' = v - \bar{v}. \quad (6)$$

MKE and EKE are spatially averaged within a 400 km box around the ST and vertically averaged between 200 and 1000 m. These averages are computed every 10 d. MKE can

**Table 2.** Variables associated with the Lagrangian experiments.  $BL_{200\text{m}}^{\text{surf}}$  and  $BL_{200\text{m}}^{4\text{L}}$  indicate the Bhattacharyya loss at 200 m for the surface and four-layer models. We computed the mass centers (in kilometers) and entropy of the PDFs. MKE and EKE ( $\text{cm}^2 \text{s}^{-2}$ ) are computed with a temporal window of 30 d (the period of the Lagrangian experiment) and averaged between 200 and 1000 m in a 400 km box centered on the ST. Moreover,  $\zeta/f$  and OW/f are computed from a surface snapshot taken 20 d after the first particle release and averaged within a 100 km box centered on the ST. The values of the median, 10th percentile, and 90th percentile are given at the top. Index numbers correspond to the examples given in Fig. 5.

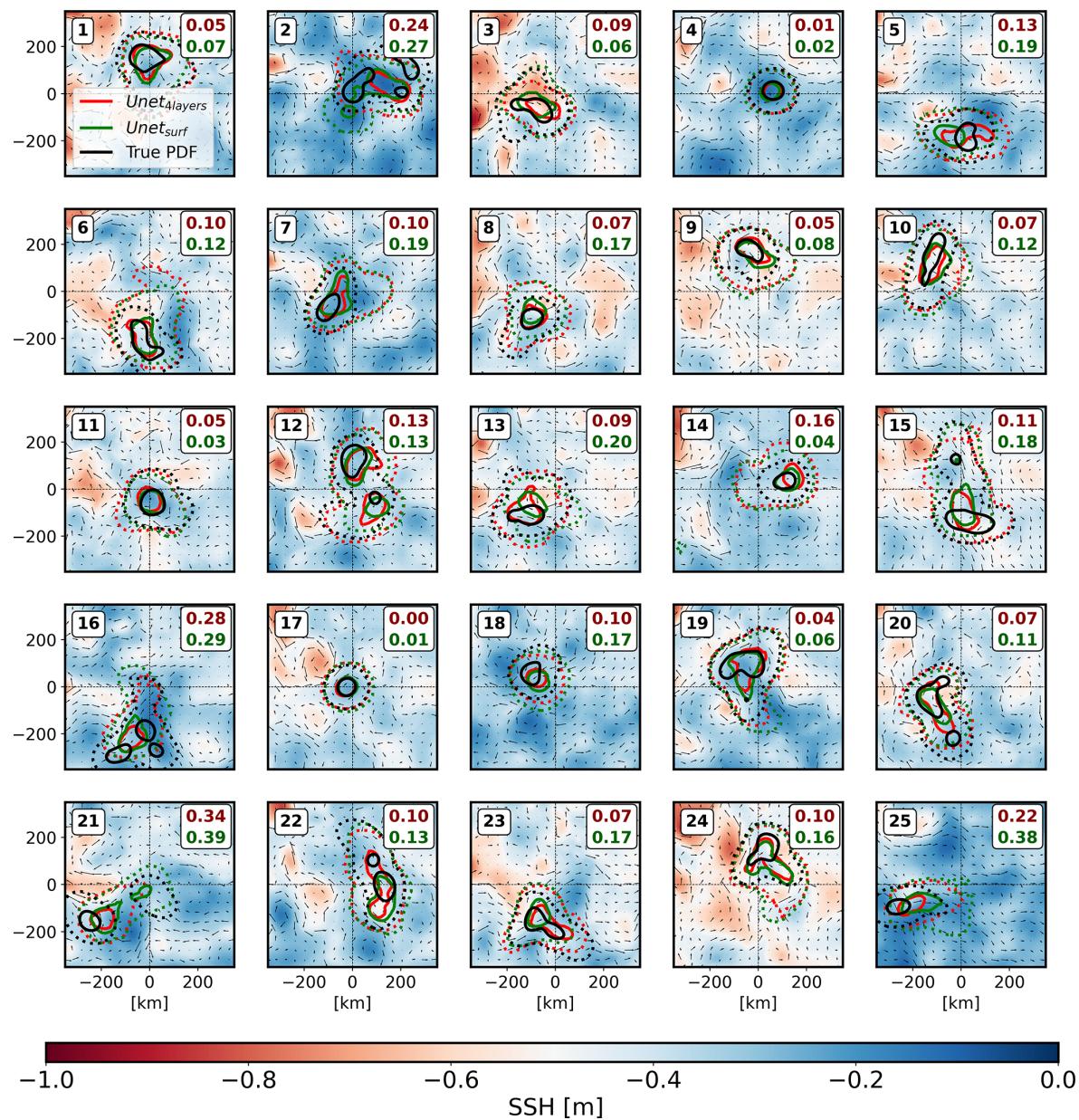
Index	$BL_{200\text{m}}^{\text{surf}}$	$BL_{200\text{m}}^{4\text{L}}$	Mass center	Entropy	MKE	EKE	$\zeta/f_{10^{-2}}$	$OW/f_{10^{-3}}$
10th percentile	0.06	0.04	36	5.7	69	32	-4.1	-1.4
50th percentile	0.15	0.11	99	6.3	95	49	-0.4	2.2
90th percentile	0.35	0.26	193	6.7	126	71	3.6	12.1
1	0.07	0.05	136	6.5	90	81	-4.6	7.4
2	0.27	0.24	136	6.9	94	65	-2.4	2.0
3	0.06	0.09	110	6.4	101	82	4.2	-1.1
4	0.02	0.01	63	5.7	124	31	-4.9	-2.6
5	0.19	0.13	183	6.1	127	46	0.9	4.0
6	0.12	0.10	216	6.4	107	50	-0.4	-0.7
7	0.19	0.10	109	6.0	96	44	-4.4	-2.0
8	0.17	0.07	164	5.9	98	57	0.8	3.3
9	0.08	0.05	172	6.0	82	39	3.0	7.9
10	0.12	0.07	178	6.3	79	47	3.2	1.3
11	0.03	0.05	62	6.2	89	58	-5.1	4.6
12	0.13	0.13	51	6.7	65	47	-2.4	1.1
13	0.20	0.09	148	6.4	74	39	2.2	0.8
14	0.04	0.16	114	5.6	84	36	1.1	1.7
15	0.18	0.11	68	6.8	114	38	1.8	3.0
16	0.29	0.28	242	6.6	91	53	-3.7	0.9
17	0.01	0.00	24	5.5	131	36	-5.0	-2.2
18	0.17	0.10	110	5.9	60	41	1.7	0.3
19	0.06	0.04	95	6.7	137	45	-2.5	2.6
20	0.11	0.07	133	6.7	85	31	2.2	3.8
21	0.39	0.34	257	6.0	89	51	-1.1	2.0
22	0.13	0.10	114	6.5	89	32	-1.0	4.4
23	0.17	0.07	181	6.4	98	35	0.2	0.3
24	0.16	0.10	140	6.3	132	40	-1.8	20.6
25	0.38	0.22	270	5.7	92	27	0.6	0.3

be associated with large-scale currents and mesoscale eddies that stay stable during the time window. High MKE implies strong velocities that are likely to transport particles far from their source, further complicating the prediction process. EKE typically indicates the presence of moving mesoscale eddies or submesoscale fronts. These smaller-scale dynamics can promote divergent flows and increase entropy, thereby affecting the prediction score.

The number of invalid predictions increases on average with both MKE and EKE (Fig. 6c and d), even though the relationship between the prediction and EKE/MKE is not always simple when considering all cases listed in Table 2. Moreover, the influence of MKE seems to explain the seasonal patterns observed during the 5-year period of temporal coverage (Fig. 7). For both U-Net models, the best performance was achieved in winter/early spring, while the highest probability of non-valid predictions was observed in summer. We found a strong correlation ( $r^2 = 0.64$ ) between this trend

and the seasonal evolution of MKE. This observation suggests that MKE could be the main driver of the score across long timescales. The EKE evolution shows a weaker correlation ( $r^2 = 0.13$ ). In particular, the EKE peak visible in April, associated with the deepest mixed layer and intensified submesoscale activity down to 500 m (Buckingham et al., 2016), is not clearly reflected in the performance score.

However, Table 2 suggests that MKE and EKE alone cannot explain the prediction score in all cases. The presence of coherent vorticity structures above the trap appears to facilitate prediction (indices 4, 11, and 17 in Fig. 5). To demonstrate this, we analyze dynamical features using three indicators: surface relative vorticity ( $\zeta = v_x - u_y$ ), the Okubo–Weiss parameter ( $OW = \sigma^2 - \zeta^2$ , where  $\sigma = (u_x - v_y) + (v_x + u_y)$ ), and the sea-level anomaly ( $SLA = SSH - <SSH>$ , where  $<SSH>$  is the average SSH within the subdomain). In order to avoid any smoothing effect, these variables are taken from a surface snapshot taken 20 d before the first par-

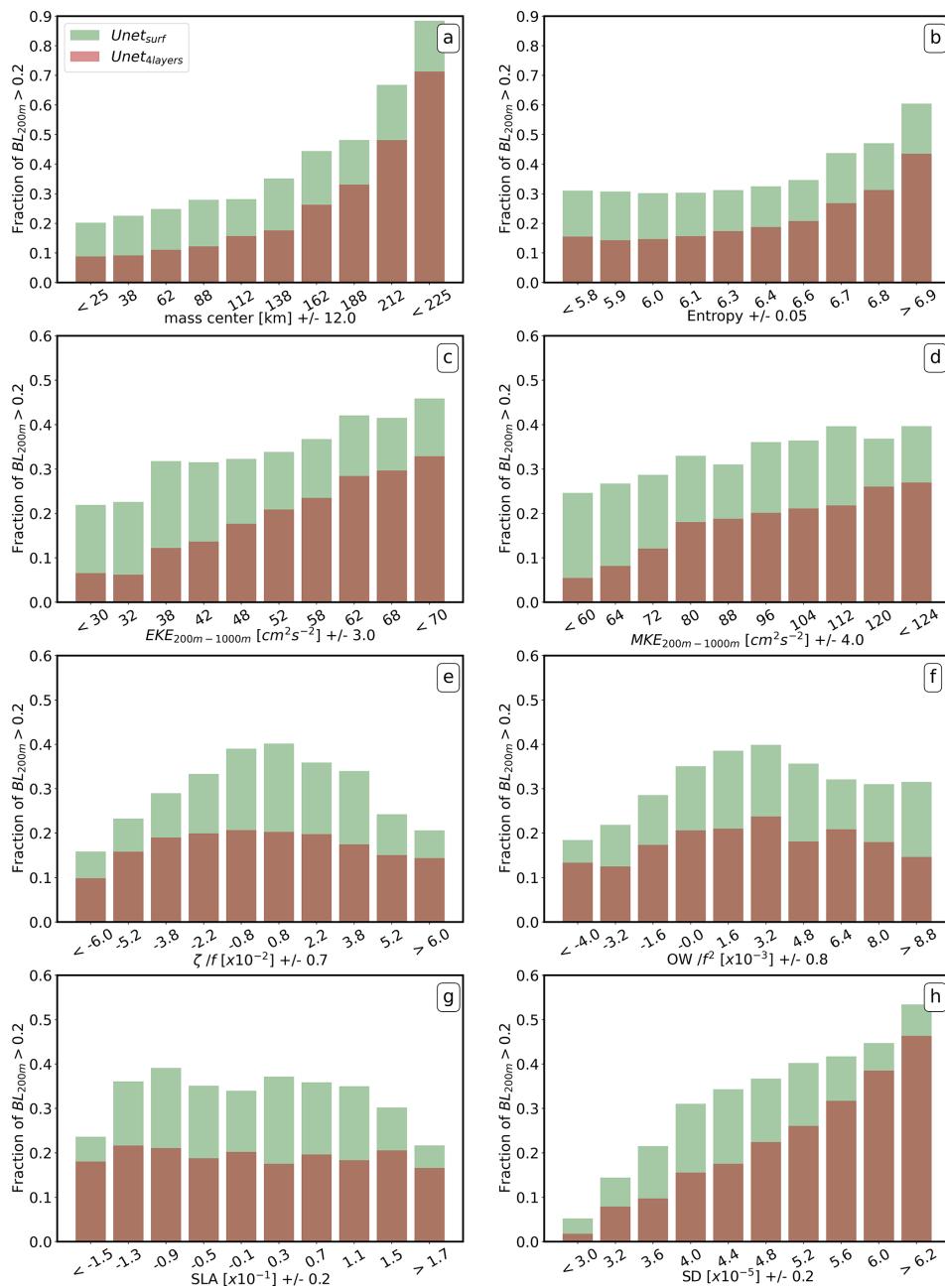


**Figure 5.** Examples of predictions. PDFs are predicted with the U-Net<sub>4layers</sub> model (red) and the U-Net<sub>surf</sub> model (green). Black contours represent the true origins of the particles. Dotted contours encompass 99 % of the particles, while solid contours encompass 50 %. The BL<sub>200m</sub> scores for both predictions are provided in the top-right corner of each plot. The background of each plot shows a snapshot of SSH taken 20 d after the first particle release, with surface velocities represented by the arrows. All the images are centered on the trap where particles are released.

ticle release (rather than from an average over the experiment period). The particles reach a depth of approximately 200 m between days −15 and −25 following the initial release. Consequently, the 20th day represents the oceanic conditions present when the majority of particles are near the surface and under the influence of surface currents. The variables are averaged within a 100 km × 100 km area centered on the ST (Fig. 6e, f, and g; Table 2). The best performances (> 80 % of valid predictions) are associated with low nega-

tive OW values, high absolute  $\zeta$  values, and significant SLA values. This situation generally corresponds to the presence of a vortex structure above the trap (Wang et al., 2022). On the other hand, chaotic situations characterized by OW values greater than 0,  $\zeta$  values of  $\sim 0$ , and SLA values of  $\sim 0$  are generally associated with the worst performance.

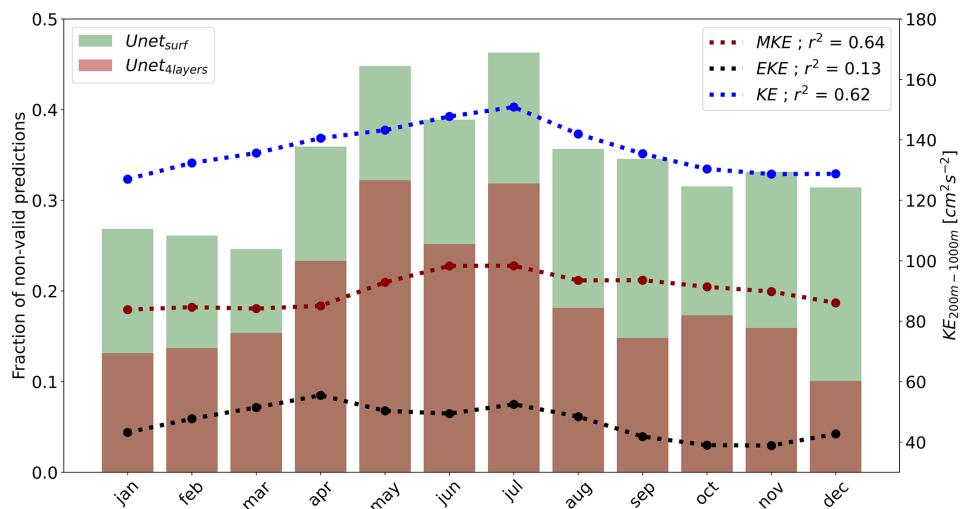
The presence of large-scale currents and eddies appears to be sufficient for explaining the prediction scores to a first-order approximation: high kinetic energy (KE) is typically



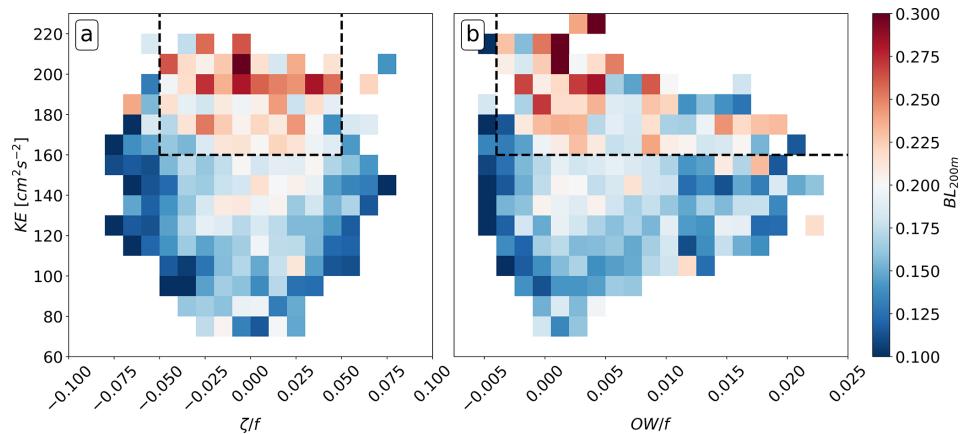
**Figure 6.** Fraction of non-valid predictions ( $BL_{200m} > 0.2$ ) with respect to the following local dynamical variables: (a) PDF mass center shift, (b) PDF Shannon entropy ( $-\sum p \log(p)$ ), (c) eddy kinetic energy (EKE), (d) mean kinetic energy (MKE), (e) relative surface vorticity ( $\zeta/f$ ), (f) the surface relative Okubo–Weiss parameter ( $OW/f^2$ ), (g) the sea-level anomaly (SLA), and (h) the standard deviation (SD) of the 10 bootstrap models. In panels (c) and (d), energy is averaged within a 400 km box centered on the ST and between 200 and 1000 m. A temporal averaging window of 30 d is used for MKE and EKE. In panels (e), (f), and (g), the local dynamical variables come from surface snapshots taken 20 d after the first particle release and are averaged within a 100 km box around the ST.

associated with lower scores, unless coherent structures are present above the trap (corresponding to high KE with negative OW values and large vorticity amplitudes). This is highlighted by the bin statistics in Fig. 8. The non-valid prediction area in Fig. 8a is defined by  $KE > 160 \text{ cm}^2 \text{ s}^{-2}$  and  $|\zeta/f| < 0.05$  and is characterized by an average Bhattacharyya

score greater than 0.2 (indicated by the red bins). This chaotic situation represents 27 % of our dataset. Similarly, the same non-valid prediction zone can be observed in Fig. 8b and is defined by  $KE > 160 \text{ cm}^2 \text{ s}^{-2}$  and  $OW/f > -0.0075$ .



**Figure 7.** Monthly performance statistics over the testing period for the U-Net<sub>surf</sub> and U-Net<sub>4layer</sub> models. The bar graph indicates the probability of non-valid predictions ( $BL_{200\text{m}} > 0.2$ ). The monthly averaged evolution of EKE (dotted) and MKE (dashed) between 200 and 1000 m is shown in black.



**Figure 8.** Averaged bin statistics of the U-Net<sub>surf</sub> prediction score with regard to (a) kinetic energy vs. relative vorticity and (b) kinetic energy vs. the surface relative Okubo–Weiss parameter. Color represents the mean Bhattacharyya score. Bins with fewer than five samples are masked. The dashed black lines delineate the non-valid prediction zone, represented by the red bins.

## 5 Discussion

In this section, we discuss possible interpretations of the results, present the limitations of the model, and explore potential improvements that will guide our future work.

### 5.1 Interpretation of the statistics and confidence index

Our results support the relevance of the proposed machine learning approach in predicting the catchment area of particles using only surface data. Although subsurface information can improve the quality of the prediction, it is not essential in most cases. Analysis of the drivers of the prediction scores reveals complex, multifactorial causes. Only a few indicators were chosen for this analysis, and we do not exclude the possibility of other links that have not been

investigated. However, the statistics appear coherent to us as they suggest that the model is more robust under weak and/or stable dynamics. Given that our model is limited in time and space by the 8 km input resolution (limited to the surface or four vertical layers, with snapshots taken every 10 d), it is clear that the predictions are sensitive to finer spatial and temporal variability. This may explain the good performance with low KE or in the presence of coherent eddies above the trap. Eddy structures can have a clear surface signature that is easy to identify and that stays coherent at depth (Fig. 2). Considering a 30 d period, these structures are generally stable, showing little temporal variability and limiting the spatial dispersion of particles trapped inside. In contrast, chaotic situations, typically characterized by randomly distributed submesoscale fronts with a horizontal size of about

10 km, are more likely to be less well predicted. Additionally, such structures can be too small to be detected. Temporal evolution can be very fast (on the order of days) and is unlikely to be adequately captured by our machine learning models. In the future, such assumptions may directly aid in developing a confidence index that depends on local dynamics. For future real-world applications, it will probably be difficult to predict the catchment area in every situation. The aim is, therefore, to understand in which situations the predictions made by the machine learning model can be trusted. In this respect, the standard deviation (SD) of the 10 model predictions from the bootstrap method (Fig. 6h) can also provide insight into quantifying the uncertainties in the predictions. The SD, computed as the mean SD of the 10 models at each prediction point, shows a linear relationship with the probability of having a valid prediction for the U-Net models. Previous studies have also explored such bootstrapping methods for uncertainty quantification (Pauthenet et al., 2022; Beauchamp et al., 2022; Haynes et al., 2023). Future work could also explore more advanced deep learning schemes with built-in uncertainty quantification properties (Haynes et al., 2023). However, it is worth noting that it would also be important to evaluate other sources of uncertainty, such as sinking-velocity distributions, simplifications of numerical simulations, and transport rates.

## 5.2 Towards real data and a more realistic representation of particles

The results strongly suggest the potential for using satellite data (SSH and sea surface temperature (SST)) to more precisely identify the source areas of particles collected in the deep ocean. A possible strategy for validating the model predictions against real observations would involve evaluating the cross-correlation between satellite chlorophyll in the prediction area and carbon fluxes measured at the PAP station (Frigstad et al., 2015). The cross-correlation coefficient can be compared with the obtained cross-correlation by considering a simplified catchment area, such as a 100 or 200 km box around the PAP location, which is a classical method still used today (Lampitt et al., 2023). If a better correlation is found, this indicates a better connection between deep fluxes at the PAP station and surface images from satellites and confirms the relevance of the model's application to real observations. As emphasized by Febvre et al. (2024) with respect to the learning-based mapping of sea surface dynamics, the representativeness of numerical simulations in reflecting real (3D + T) hydrodynamic conditions is likely to be a critical feature when applying a neural model trained on simulation data to a real-world configuration. The use of ocean reanalyses (Frigstad et al., 2015) could also be a relevant solution. In both cases, an in-depth analysis of 3D ocean circulation seems particularly important to assess the potential limitations of the proposed scheme.

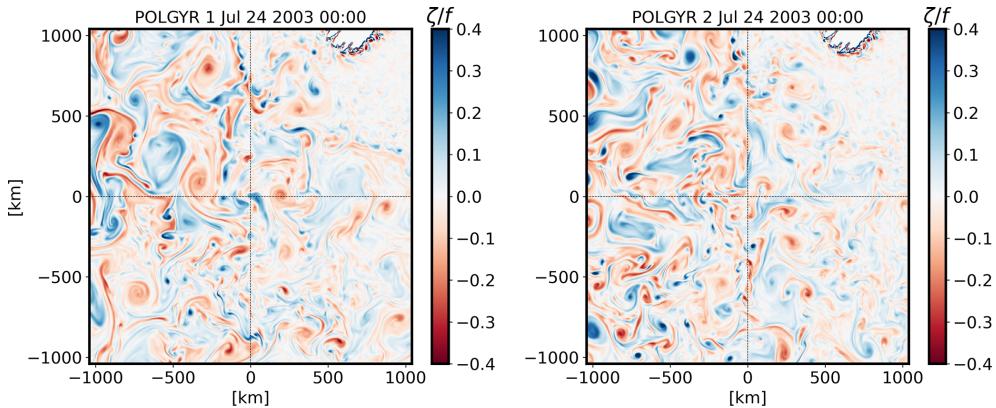
Regarding biology, we consider STs at 1000 m, whereas in reality, the STs at the PAP station are moored at a greater depth ( $\sim 3000$  m). It has been shown that below 1000 m, currents are weaker and do not significantly affect the horizontal displacement of particles (Wang et al., 2022). Therefore, although the influence of ST depth on prediction scores needs to be investigated, it can be assumed that it will not significantly affect model performance. However, a major bias in the model probably comes from the representation of carbon particles. We represent sinking particles with a settling velocity of  $50 \text{ m d}^{-1}$ , whereas, in reality, particles exhibit a wide range of velocities that can vary significantly with the seasons (Villa-Alfageme et al., 2016). The origins of particles can be very different depending on their sinking velocity, meaning it is important to consider the entire particle velocity spectrum to represent the catchment areas for particles with different velocities (Wekerle et al., 2018). Therefore, the next step of this study will focus on the effect of sinking velocities on prediction scores. We can assume that predicting catchment areas for particles with higher speeds than  $50 \text{ m d}^{-1}$  may be more efficient as the particle trajectories are less impacted by currents and turbulence (Liu et al., 2018; Wekerle et al., 2018; Ma et al., 2021; Wang et al., 2022). However, particles with lower sinking velocities are more sensitive to the flow and can easily be dispersed over a large area far away from the ST, decreasing the prediction score, as suggested in Fig. 6a and b.

The use of a coupled biogeochemical model embedded in POLGYR could provide information on surface chlorophyll distribution (i.e., sea color) and primary-production (PP) intensity (Kostadinov et al., 2009; Dunne et al., 2005). Although these relationships are not direct (Laurenceau-Cornec et al., 2020; Iversen and Lampitt, 2020; Cael et al., 2021), high PP levels are associated with large phytoplankton cells and particles, which may be associated with higher sinking rates. It is then possible to weigh the distribution of particles in terms of size and sinking velocity against the PP estimated by the model at their export locations, which should make the simulations more realistic. In addition, sea color observations, together with altimetry data, can be used as input for a machine learning system to better constrain carbon fluxes at depth, employing an approach similar to that presented here. However, the size and sinking velocity of particles are affected by numerous biological processes during their journey, which are generally not taken into account in simple Lagrangian studies. Although a Lagrangian approach could be developed (Jokulsdottir and Archer, 2016), an ad hoc parameterization, e.g., a parameterization derived from biogeochemical (BGC) models simulating particle dynamics (Aumont et al., 2015), could also be used and tested.

## 6 Conclusions

Identifying the origin of particles captured by sediment traps is important for interpreting measured fluxes and improving sampling methods. However, it remains complex to constrain particle motion at depth with surface data. In our study, we use the proposed machine learning scheme to address this issue. We demonstrate the ability of machine learning to predict, in real time, the origin of particles trapped in a PAP sediment trap at 1000 m of depth using realistic 3D numerical simulations and Lagrangian tracking. The evaluation of the machine learning models supports the use of a sea-surface-only configuration for the relevant prediction of particle catchment areas in most cases, suggesting a potential application with satellite data. The statistical analysis also shows that the prediction performance is sensitive to local dynamics. The model performs better under conditions of low KE and in the presence of coherent surface vortices above the sediment trap. The next challenges are to improve the particle modeling by considering the wide range of particle sinking velocities and to apply our model to real satellite data. Ultimately, this will improve the link between observations of surface and deep-carbon transport.

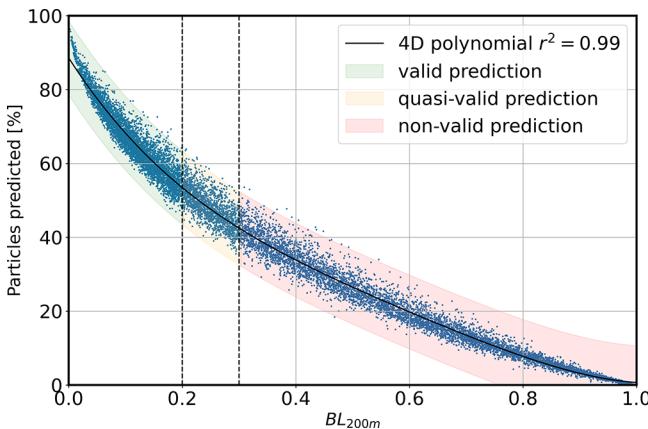
## Appendix A: Comparison between POLGYR1 and POLGYR2



**Figure A1.** Snapshots of surface relative vorticity taken on 24 July 2003 (00:00 UTC) for POLGYR1 (**a**) and POLGYR2 (**b**). Despite the fact that similar forcing is applied to the two simulations, the surface dynamics are different due to distinct initial conditions and chaotic effects.

## Appendix B: Relationship between the BL score and the percentage of PDFs predicted

*Author contributions.* TP, JG, RF, and LM: writing, analysis, and methodology. JC: writing and technical support for the simulations.



**Figure B1.** Scatterplot of the average U-Net<sub>surf</sub> Bhattacharyya score (BL<sub>200m</sub>) vs. the percentage of particles that are predicted, defined as  $\sum_{i \in D} \min(P_i, 200m Q_i, 200m)$ . The 4D polynomial relationship is shown with a black line. The vertical dashed lines represent the limits of the valid, quasi-valid, and non-valid zones.

*Code availability.* The codes used in this study are available online at <https://github.com/TheoPicrd/SPARO> (last access: 24 November 2023; Picard, 2023).

*Data availability.* The datasets used in this study are available online at <https://doi.org/10.17882/97556> (Picard et al., 2023).

*Competing interests.* The contact author has declared that none of the authors has any competing interests.

*Disclaimer.* Publisher's note: Copernicus Publications remains neutral with regard to jurisdictional claims made in the text, published maps, institutional affiliations, or any other geographical representation in this paper. While Copernicus Publications makes every effort to include appropriate place names, the final responsibility lies with the authors.

*Special issue statement.* This article is part of the special issue “Special Issue for the 54th International Liège Colloquium on Machine Learning and Data Analysis in Oceanography”. It is not associated with a conference.

*Acknowledgements.* The authors thank Mathieu Le Corre for providing the CROCO simulation outputs. They would also like to express their gratitude to the reviewers for their insightful feedback.

*Financial support.* Théo Picard received a PhD grant from École normale supérieure Paris-Saclay. This work was also supported by the ISblue project, the “interdisciplinary graduate school for the blue planet” (grant no. ANR-17-EURE-0015), and co-funded by a grant from the French government as part of the “Investissements d’Avenir” program. This paper contributes to the APERO project funded by the National Research Agency (grant no. ANR-21-CE01-

0027), AI Chair OceaniX (grant no. ANR-19-CHIA-0016), and the DEEPER project (grant no. ANR-19-CE01-0002-01). Simulations were performed using HPC resources from GENCI-TGCC (grant no. 2018-A0050107638) and Datarmor (provided by the “Pôle de Calcul Intensif pour la Mer” at Ifremer in Brest, France).

**Review statement.** This paper was edited by Matjaz Licer and reviewed by Martin Vodopivec, Gael Forget, and one anonymous referee.

## References

- Alldredge, A. L. and Gotschalk, C.: In situ settling behavior of marine snow, *Limnol. Oceanogr.*, 33, 339–351, <https://doi.org/10.4319/lo.1988.33.3.0339>, 1988.
- Alonso-González, I. J., Arístegui, J., Lee, C., and Calafat, A.: Regional and temporal variability of sinking organic matter in the subtropical northeast Atlantic Ocean: a biomarker diagnosis, *Biogeosciences*, 7, 2101–2115, <https://doi.org/10.5194/bg-7-2101-2010>, 2010.
- Armstrong, R. A., Lee, C., Hedges, J. I., Honjo, S., and Wakeham, S. G.: A new, mechanistic model for organic carbon fluxes in the ocean based on the quantitative association of POC with ballast minerals, *Deep-Sea Res. Pt. II*, 49, 219–236, [https://doi.org/10.1016/S0967-0645\(01\)00101-1](https://doi.org/10.1016/S0967-0645(01)00101-1), 2001.
- Asper, V. L., Deuser, W. G., Knauer, G. A., and Lohrenz, S. E.: Rapid coupling of sinking particle fluxes between surface and deep ocean waters, *Nature*, 357, 670–672, <https://doi.org/10.1038/357670a0>, 1992.
- Aumont, O., Ethé, C., Tagliabue, A., Bopp, L., and Gehlen, M.: PISCES-v2: An ocean biogeochemical model for carbon and ecosystem studies, *Geosci. Model Dev.*, 8, 2465–2513, <https://doi.org/10.5194/gmd-8-2465-2015>, 2015.
- Barth, A., Alvera-Azcárate, A., Licer, M., and Beckers, J.-M.: DINCAE 1.0: a convolutional neural network with error estimates to reconstruct sea surface temperature satellite observations, *Geosci. Model Dev.*, 13, 1609–1622, <https://doi.org/10.5194/gmd-13-1609-2020>, 2020.
- Beauchamp, M., Febvre, Q., Georgenthum, H., and Fablet, R.: 4DVarNet-SSH: end-to-end learning of variational interpolation schemes for nadir and wide-swath satellite altimetry, *Geosci. Model Dev.*, 16, 2119–2147, <https://doi.org/10.5194/gmd-16-2119-2023>, 2023.
- Bhattacharyya, A.: On a measure of divergence between two statistical populations defined by their probability distribution, *Bull. Calcutta Math. Soc.*, 35, 99–110, 1943.
- Billett, D. S., Lampitt, R. S., Rice, A. L., and Mantoura, R. F.: Seasonal sedimentation of phytoplankton to the deep-sea benthos, *Nature*, 302, 520–522, <https://doi.org/10.1038/302520a0>, 1983.
- Bolton, T. and Zanna, L.: Applications of Deep Learning to Ocean Data Inference and Subgrid Parameterization, *J. Adv. Model. Earth Sy.*, 11, 376–399, <https://doi.org/10.1029/2018MS001472>, 2019.
- Breiman, L.: Bagging predictors, *Mach. Learn.*, 24, 123–140, <https://doi.org/10.1007/bf00058655>, 1996.
- Buckingham, C. E., Naveira Garabato, A. C., Thompson, A. F., Brannigan, L., Lazar, A., Marshall, D. P., George Nurser, A. J., Damerell, G., Heywood, K. J., and Belcher, S. E.: Seasonality of submesoscale flows in the ocean surface boundary layer, *Geophys. Res. Lett.*, 43, 2118–2126, <https://doi.org/10.1002/2016GL068009>, 2016.
- Burd, A. B., Buchan, A., Church, M., Landry, M., McDonnell, A., Passow, U., Steinberg, D., and Benway, H.: Towards a transformative understanding of the biology of the ocean’s biological pump: Priorities for future research, Report of the NSF Biology of the Biological Pump Workshop, 19–20 February 2016, Hyatt Place New Orleans, New Orleans, LA, 67 pp., <https://doi.org/10.1575/1912/8263>, 2016.
- Burd, A. B., Hansell, D. A., Steinberg, D. K., Anderson, T. R., Arístegui, J., Baltar, F., Beaupré, S. R., Buesseler, K. O., DeHairs, F., Jackson, G. A., Kadko, D. C., Koppelman, R., Lampitt, R. S., Nagata, T., Reinhäler, T., Robinson, C., Robison, B. H., Tamburini, C., and Tanaka, T.: Assessing the apparent imbalance between geochemical and biochemical indicators of meso- and bathypelagic biological activity: What the @#\$! is wrong with present calculations of carbon budgets?, *Deep-Sea Res. Pt. II*, 57, 1557–1571, <https://doi.org/10.1016/j.dsr2.2010.02.022>, 2010.
- Cael, B. B., Cavan, E. L., and Britten, G. L.: Reconciling the Size-Dependence of Marine Particle Sinking Speed, *Geophys. Res. Lett.*, 48, e2020GL091771, <https://doi.org/10.1029/2020GL091771>, 2021.
- Chapman, C. and Charantonis, A. A.: Reconstruction of Subsurface Velocities From Satellite Observations Using Iterative Self-Organizing Maps, *IEEE Geosci. Remote Sens. Lett.*, 14, 617–620, <https://doi.org/10.1109/LGRS.2017.2665603>, 2017.
- Chelton, D. B., Deszoeke, R. A., Schlax, M. G., El Naggar, K., and Siwertz, N.: Geographical variability of the first baroclinic Rossby radius of deformation, *J. Phys. Oceanogr.*, 28, 433–460, [https://doi.org/10.1175/1520-0485\(1998\)028<0433:GVOTFB>2.0.CO;2](https://doi.org/10.1175/1520-0485(1998)028<0433:GVOTFB>2.0.CO;2), 1998.
- Cutolo, E., Pascual, A., Ruiz, S., Zarokanellos, N., and Fablet, R.: CLOINet: Ocean state reconstructions through remote-sensing, in-situ sparse observations and Deep Learning, Arxiv, <http://arxiv.org/abs/2210.10767> (last access: 12 December 2023), 2022.
- Deuser, W. G. and Ross, E. H.: Seasonal change in the flux of organic carbon to the deep Sargasso Sea, *Nature* 283, 364–365, <https://doi.org/10.1038/283364a0>, 1980.
- Deuser, W. G., Muller-Karger, F. E., and Hemleben, C.: Temporal variations of particle fluxes in the deep subtropical and tropical North Atlantic: Eulerian versus Lagrangian effects, *J. Geophys. Res.*, 93, 6857–6862, <https://doi.org/10.1029/JC093iC06p06857>, 1988.
- Deuser, W. G., Muller-Karger, F. E., Evans, R. H., Brown, O. B., Esaias, W. E., and Feldman, G. C.: Surface-ocean color and deep-ocean carbon flux: how close a connection?, *Deep-Sea Res. Pt. A*, 37, 1331–1343, [https://doi.org/10.1016/0198-0149\(90\)90046-X](https://doi.org/10.1016/0198-0149(90)90046-X), 1990.
- Dever, M., Nicholson, D., Omand, M. M., and Mahadevan, A.: Size-Differentiated Export Flux in Different Dynamical Regimes in the Ocean, *Global Biogeochem. Cy.*, 35, e2020GB006764, <https://doi.org/10.1029/2020GB006764>, 2021.
- Dunne, J. P., Armstrong, R. A., Gnanadesikan, A., and Sarmiento, J. L.: Empirical and mechanistic models for the parti-

- cle export ratio, *Global Biogeochem. Cy.*, 19, GB4026, <https://doi.org/10.1029/2004GB002390>, 2005.
- Febvre, Q., Le Sommer, J., Ubelmann, C., and Fablet, R.: Training neural mapping schemes for satellite altimetry with simulation data, *J. Adv. Model. Earth Syst.*, 16, e2023MS003959, <https://doi.org/10.1029/2023MS003959>, 2024.
- Frigstad, H., Henson, S. A., Hartman, S. E., Omar, A. M., Jeansson, E., Cole, H., Pebody, C., and Lampitt, R. S.: Links between surface productivity and deep ocean particle flux at the Porcupine Abyssal Plain sustained observatory, *Biogeosciences*, 12, 5885–5897, <https://doi.org/10.5194/bg-12-5885-2015>, 2015.
- George, T. M., Manucharyan, G. E., and Thompson, A. F.: Deep learning to infer eddy heat fluxes from sea surface height patterns of mesoscale turbulence, *Nat. Commun.*, 12, 800, <https://doi.org/10.1038/s41467-020-20779-9>, 2021.
- Gula, J. and Collin, J.: Pyticles: a Python/Fortran hybrid parallelized code for 3D Lagrangian particles advection using ROMS/CROCO model data, Zenodo, <https://doi.org/10.5281/zenodo.4973786>, 2021.
- Hartman, S. E., Bett, B. J., Durden, J. M., Henson, S. A., Iversen, M., Jeffreys, R. M., Horton, T., Lampitt, R., and Gates, A. R.: Enduring science: Three decades of observing the Northeast Atlantic from the Porcupine Abyssal Plain Sustained Observatory (PAP-SO), *Prog. Oceanogr.*, 191, 102508, <https://doi.org/10.1016/j.pocean.2020.102508>, 2021.
- Haynes, K., Lagerquist, R., McGraw, M., Musgrave, K., and Ebert-Uphoff, I.: Creating and Evaluating Uncertainty Estimates with Neural Networks for Environmental-Science Applications, *Artificial Intelligence for the Earth Systems*, 2, 1–29, <https://doi.org/10.1175/aies-d-22-0061.1>, 2023.
- Henson, S. A., Laufkötter, C., Leung, S., Giering, S. L., Palevsky, H. I., and Cavan, E. L.: Uncertain response of ocean biological carbon export in a changing world, *Nat. Geosci.*, 15, 248–254, <https://doi.org/10.1038/s41561-022-00927-0>, 2022.
- Iversen, M. H. and Lampitt, R. S.: Size does not matter after all: No evidence for a size-sinking relationship for marine snow, *Prog. Oceanogr.*, 189, 102445, <https://doi.org/10.1016/j.pocean.2020.102445>, 2020.
- Jenkins, J., Paiement, A., Ourmières, Y., Sommer, J. L., Verron, J., Ubelmann, C., and Glotin, H.: A DNN Framework for Learning Lagrangian Drift With Uncertainty, Arxiv, <http://arxiv.org/abs/2204.05891> (last access: 24 May 2023), 2022.
- Jokulsdottir, T. and Archer, D.: A stochastic, Lagrangian model of sinking biogenic aggregates in the ocean (SLAMS 1.0): Model formulation, validation and sensitivity, *Geosci. Model Dev.*, 9, 1455–1476, <https://doi.org/10.5194/gmd-9-1455-2016>, 2016.
- Kingma, D. P. and Ba, J. L.: Adam: A method for stochastic optimization, in: 3rd International Conference on Learning Representations, ICLR 2015 – Conference Track Proceedings, San Diego, <https://doi.org/10.48550/arXiv.1412.6980>, 2015.
- Kostadinov, T. S., Siegel, D. A., and Maritorena, S.: Retrieval of the particle size distribution from satellite ocean color observations, *J. Geophys. Res.-Ocean.*, 114, C09015, <https://doi.org/10.1029/2009JC005303>, 2009.
- Lampitt, R. S., Achterberg, E. P., Anderson, T. R., Hughes, J. A., Iglesias-Rodriguez, M. D., Kelly-Gerrey, B. A., Lucas, M., Popova, E. E., Sanders, R., Shepherd, J. G., Smythe-Wright, D., and Yool, A.: Ocean fertilization: A potential means of geoengineering?, *Philos. T. R. Soc. A*, 366, 3919–3945, <https://doi.org/10.1098/rsta.2008.0139>, 2008.
- Lampitt, R. S., Briggs, N., Cael, B. B., Espinola, B., Hélaouët, P., Henson, S. A., Norrbin, F., Pebody, C. A., and Smeed, D.: Deep ocean particle flux in the Northeast Atlantic over the past 30 years: carbon sequestration is controlled by ecosystem structure in the upper ocean, *Front. Earth Sci.*, 11, 1176196, <https://doi.org/10.3389/feart.2023.1176196>, 2023.
- Laurenceau-Cornec, E. C., Le Moigne, F. A. C., Gallinari, M., Moriceau, B., Toullec, J., Iversen, M. H., Engel, A., and De La Rocha, C. L.: New guidelines for the application of Stokes' models to the sinking velocity of marine aggregates, *Limnol. Oceanogr.*, 65, 1264–1285, <https://doi.org/10.1002/limo.11388>, 2020.
- Le Cann, B.: Observed mean and mesoscale upper ocean circulation in the midlatitude northeast Atlantic, *J. Geophys. Res.*, 110, C07S05, <https://doi.org/10.1029/2004JC002768>, 2005.
- Le Corre, M., Gula, J., and Tréguier, A. M.: Barotropic vorticity balance of the North Atlantic subpolar gyre in an eddy-resolving model, *Ocean Sci.*, 16, 451–468, <https://doi.org/10.5194/os-16-451-2020>, 2020.
- Le Moigne, F. A., Henson, S. A., Sanders, R. J., and Madsen, E.: Global database of surface ocean particulate organic carbon export fluxes diagnosed from the 234Th technique, *Earth Syst. Sci. Data*, 5, 295–304, <https://doi.org/10.5194/essd-5-295-2013>, 2013.
- Le Moigne, F. A. C.: Pathways of Organic Carbon Downward Transport by the Oceanic Biological Carbon Pump, *Front. Mar. Sci.*, 6, 482488, <https://doi.org/10.3389/fmars.2019.00634>, 2019.
- Lecun, Y., Bengio, Y., and Hinton, G.: Deep learning, *Nature*, 521, 436–444, <https://doi.org/10.1038/nature14539>, 2015.
- Liu, G., Bracco, A., and Passow, U.: The influence of mesoscale and submesoscale circulation on sinking particles in the northern Gulf of Mexico, *Elementa*, 6, 36, <https://doi.org/10.1525/elementa.292>, 2018.
- Ma, W., Xiu, P., Chai, F., Ran, L., Wiesner, M. G., Xi, J., Yan, Y., and Fredj, E.: Impact of mesoscale eddies on the source funnel of sediment trap measurements in the South China Sea, *Prog. Oceanogr.*, 194, 102566, <https://doi.org/10.1016/j.pocean.2021.102566>, 2021.
- Manucharyan, G. E., Siegelman, L., and Klein, P.: A Deep Learning Approach to Spatiotemporal Sea Surface Height Interpolation and Estimation of Deep Currents in Geostrophic Ocean Turbulence, *J. Adv. Model. Earth Sy.*, 13, 1–17, <https://doi.org/10.1029/2019MS001965>, 2021.
- Passow, U. and Carlson, C. A.: The biological pump in a high CO<sub>2</sub> world, *Mar. Ecol. Prog. Ser.*, 470, 249–271, <https://doi.org/10.3354/meps09985>, 2012.
- Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., and Chintala, S.: PyTorch: An Imperative Style, High-Performance Deep Learning Library, *Adv. Neur. In.*, 32, 8024–8035, 2019.
- Pauthenet, E., Bachelot, L., Balem, K., Maze, G., Tréguier, A. M., Roquet, F., Fablet, R., and Tandeo, P.: Four-dimensional temperature, salinity and mixed-layer depth in the Gulf Stream, reconstructed from remote-sensing and in situ ob-

- servations with neural networks, *Ocean Sci.*, 18, 1221–1244, <https://doi.org/10.5194/os-18-1221-2022>, 2022.
- Picard, T.: SPARO: v1.0.0, Zenodo [code], <https://doi.org/10.5281/zenodo.10203352>, 2023.
- Picard, T., Gula, J., Fablet, R., Memery, L., and Collin, J.: Data for learning-based prediction of the particles catchment area of deep ocean sediment traps, SEANOE [data set], <https://doi.org/10.17882/97556>, 2023.
- Rembauville, M., Blain, S., Manno, C., Tarling, G., Thompson, A., Wolff, G., and Salter, I.: The role of diatom resting spores in pelagic–benthic coupling in the Southern Ocean, *Biogeosciences*, 15, 3071–3084, <https://doi.org/10.5194/bg-15-3071-2018>, 2018.
- Ronneberger, O., Fischer, P., and Brox, T.: U-Net: Convolutional Networks for Biomedical Image Segmentation, Arxiv, <http://arxiv.org/abs/1505.04597> (last access: 18 May 2015), 2015.
- Shchepetkin, A. F. and McWilliams, J. C.: The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model, *Ocean Model.*, 9, 347–404, <https://doi.org/10.1016/j.ocemod.2004.08.002>, 2005.
- Siegel, D. A., Granata, T. C., Michaels, A. F., and Dickey, T. D.: Mesoscale eddy diffusion, particle sinking, and the interpretation of sediment trap data, *J. Geophys. Res.*, 95, 5305–5311, <https://doi.org/10.1029/JC095iC04p05305>, 1990.
- Siegel, D. A., Fields, E., and Buesseler, K. O.: A bottom-up view of the biological pump: Modeling source funnels above ocean sediment traps, *Deep-Sea Res. Pt. I*, 55, 108–127, <https://doi.org/10.1016/j.dsr.2007.10.006>, 2008.
- Siegel, D. A., Buesseler, K. O., Doney, S. C., Sailley, S. F., Behrenfeld, M. J., and Boyd, P. W.: Global assessment of ocean carbon export by combining satellite observations and food-web models, *Global Biogeochem. Cy.*, 28, 181–196, <https://doi.org/10.1002/2013GB004743>, 2014.
- Soufflet, Y., Marchesiello, P., Lemarié, F., Jouanno, J., Capet, X., Debreu, L., and Benshila, R.: On effective resolution in ocean models, *Ocean Model.*, 98, 36–50, <https://doi.org/10.1016/j.ocemod.2015.12.004>, 2016.
- Turner, J. T.: Zooplankton fecal pellets, marine snow, phytodetritus and the ocean's biological pump, *Prog. Oceanogr.*, 130, 205–248, <https://doi.org/10.1016/j.pocean.2014.08.005>, 2015.
- Villa-Alfageme, M., de Soto, F., Le Moigne, F. A. C., Giering, S. L. C., Sanders, R., and García-Tenorio, R.: Observations and modeling of slow-sinking particles in the twilight zone, *Global Biogeochem. Cy.*, 28, 1327–1342, <https://doi.org/10.1002/2014GB004981>, 2014.
- Villa-Alfageme, M., de Soto, F. C., Ceballos, E., Giering, S. L. C., Le Moigne, F. A. C., Henson, S., Mas, J. L., and Sanders, R. J.: Geographical, seasonal, and depth variation in sinking particle speeds in the North Atlantic, *Geophys. Res. Lett.*, 43, 8609–8616, <https://doi.org/10.1002/2016GL069233>, 2016.
- Wang, L., Gula, J., Collin, J., and Mémery, L.: Effects of Mesoscale Dynamics on the Path of Fast-Sinking Particles to the Deep Ocean: A Modeling Study, *J. Geophys. Res.-Ocean.*, 127, 1–30, <https://doi.org/10.1029/2022JC018799>, 2022.
- Wekerle, C., Krumpen, T., Dinter, T., von Appen, W. J., Iversen, M. H., and Salter, I.: Properties of sediment trap catchment areas in fram strait: Results from Lagrangian modeling and remote sensing, *Front. Mar. Sci.*, 9, 407, <https://doi.org/10.3389/fmars.2018.00407>, 2018.
- Williams, R. G. and Follows, M. J.: Ocean Dynamics and the Carbon Cycle: Principles and Mechanisms, Cambridge University Press, <https://doi.org/10.1017/CBO9780511977817>, 2011.