

Disentangling Internal Tides from Balanced Motions with Deep Learning and Surface Field Synergy

Han Wang¹, Jeffrey Uncu², Kaushik Srinivasan³, Nicolas Grisouard²

¹Institute of Oceanography, University of Hamburg, 20146 Hamburg, Germany

²Department of Physics, University of Toronto, Toronto ON M5S 1A7, Canada

³University of California, Los Angeles, CA 90095, USA

Key Points:

- The imprints of incoherent internal tide and balanced motions on sea surface height are disentangled by a deep learning approach
- An algorithm simpler and computationally cheaper than previous work turns out equally effective if given a learning rate annealing scheme
- All fields act in synergy to improve the disentanglement performance and surface velocity is by far the most important field

Corresponding author: Han Wang, hannnwangus@gmail.com

Abstract

A fundamental challenge in ocean dynamics is the disentanglement of balanced motions and internal waves. Extracting internal tidal (IT) imprints on surface data is a central part of this challenge. For IT extraction, traditional harmonic analysis fails in the presence of strong incoherence and poor temporal sampling, as is common in global satellite observations. The advent of new wide-swath satellites, which provide two-dimensional spatial coverage, allows IT extraction to be reformulated as an image translation problem. Building on recent work where we developed a deep learning approach to extract IT signatures from sea surface height (SSH) in an idealized turbulent simulation, we show here that a simpler and computationally cheaper algorithm can perform equally well if the learning rate is annealed during training. Using this new, convenient algorithm, we experiment with different combinations of input surface fields – SSH, surface temperature, and surface velocity. All fields contribute synergistically to disentanglement, with surface velocity by far the most informative. These findings underscore the value of coordinated multi-platform observational campaigns and highlight the critical importance of surface velocity observations for separating balanced motions and internal waves. Additional insights into the behavior of deep learning algorithm emerge: both wave signature and scattering medium aids IT extraction, and to exploit large-scale information in the scattering medium, the algorithm must be highly non-local. Residual errors of our algorithm concentrate at small spatial scales near mode-2 tidal wavelengths, likely arising from artifacts introduced during data preparation (e.g., Doppler shifts) as well as imperfections in the deep learning architecture.

Plain Language Summary

The ocean hosts motions across a wide range of time and space scales. Some are long-lived, such as currents and eddies, while others are waves, such as internal tides generated when tidal currents flow over bottom topographic features. These different motions are very hard to separate. A new NASA satellite, SWOT, can measure the sea surface at fine spatial resolution, but because it passes the same place only every few weeks, traditional methods of extracting internal tides do not work well. We use machine learning to test whether combining different surface measurements—sea surface height, surface temperature, and surface currents—can help. Our results show that surface currents are the most important piece of information, but that combining all three fields works

best. These findings support future satellite missions designed to measure currents directly, and/or with concurrent measurements of different surface fields, which will improve our ability to track how tides move energy through the ocean. We also gain insights on the behavior of the machine learning algorithm and understand conceptually how some design considerations work.

1 Introduction

A fundamental challenge in ocean dynamics is the separation of balanced motions (hereafter “BMs”) and internal waves (hereafter “waves”). BMs arise when time-derivative terms are small relative to other terms such as effects of rotation, stratification, and advection in the dynamical equations. Their time scales span weeks to months. They include climate-scale circulations and submesoscale currents (McWilliams, 2016), taking up the bulk of the ocean’s kinetic energy (Ferrari & Wunsch, 2009). Also ubiquitous in the ocean are waves, which are unbalanced motions. Their frequencies are typically above the Coriolis frequency, with time scales of hours. They include, among others, storm-forced near-inertial waves and internal tides (hereafter “ITs”) that are generated as the astronomically forced tidal currents flow over underwater topography. BMs and waves have different impacts on tracer and energy transports, and an accurate BM-wave separation is crucial for representing dynamical processes like wave-driven mixing to create reliable maps of ocean circulation and to adequately represent the ocean’s role in the climate (Garrett & Munk, 1972; Polzin et al., 1997; Ferrari & Wunsch, 2009; Whalen et al., 2020).

This separation is difficult. Even with full temporal and spatial information (typically from simulations), BM-wave separation can be conceptually fraught and can only be made at the estimation level, as BMs can have overlapping temporal or spatial scales and strong dynamical interactions with waves (Vanneste, 2013; Barkan et al., 2024; Kar et al., 2025). A common approach is to assume time scale separation: slow motions are interpreted as BMs, and fast motions as waves (Kafiabad et al., 2019; Shakespeare et al., 2021; Jones et al., 2023). With observational data, where temporal/spatial information is limited, even this estimated separation becomes challenging. Global satellite observations of sea surface height (hereafter “SSH”) are a classic example of this limitation: repeat cycles of tens of days are far longer than wave periods, so simple frequency filtering is ineffective. For such altimetry observations, progress has been made under a

less ambitious objective: rather than separating the full wave spectrum from BMs, some efforts have focused on extracting ITs (Munk & Cartwright, 1966; Ray & Zaron, 2011; Zhao et al., 2012; Carrere et al., 2021; H. Wang et al., 2022b; Zhao, 2024), which constitute a significant portion of wave energy (Kantha & Tierney, 1997). ITs are more trackable than other wave components: under given background conditions, they concentrate around predictable frequencies and wavenumbers—a fact that is the basis of the classic approach of harmonic fitting, which matches observations to plane-wave-based models. Meaningful global maps of ITs are thus produced (Zhao et al., 2012; Carrere et al., 2021), illuminating mixing properties (Whalen et al., 2020) and climatological propagation properties of ITs (Zhao, 2016).

This progress must be interpreted with a commonly acknowledged caveat: IT harmonic fitting is inaccurate for coarsely sampled data in presence of strong incoherence. Incoherence—deviations from fixed-phase plane wave relationships—is ubiquitous in ITs. As ITs propagate, their propagation angles and spatial scales are constantly modulated by the turbulent BM and varying background conditions. Linear superposition of scattered ITs alone produces time- and space-dependent changes in phase and wavenumber, even without additional wave-wave interactions. In observations, IT incoherence is found universally across different types of data (Ray & Zaron, 2011; Löb et al., 2020; Caspar-Cohen et al., 2025). In simulations and analytical models, the causal effects of BMs and background conditions on IT incoherence are extensively demonstrated (Dunphy & Lamb, 2014; Ponte & Klein, 2015; Kelly et al., 2016; Caspar-Cohen et al., 2022; Lahaye et al., 2024; Uncu & Grisouard, 2024; Y. Wang & Legg, 2025). The approach of harmonic fitting is developed with incoherence in mind (Munk & Cartwright, 1966; Zhao, 2024) and can resolve incoherent ITs if sampling is sufficiently fine. But for coarsely sampled global satellite SSH, fine-scale incoherence cannot be reliably captured by standard plane-wave-based harmonic fitting schemes. Carrere et al. (2021) provides a quantified review of this issue. This can yield qualitatively incorrect results: for example, Buijsman et al. (2017) argues that the apparent IT damping in equatorial Pacific, inferred via harmonic fitting of decades of satellite SSH data, may reflect amplified fitting errors due to increased incoherence rather than true IT energy loss. Such systematic error hinders an accurate global picture of IT-related energy pathways.

The recently launched Surface Water and Ocean Topography (hereafter “SWOT”) mission shares the long repeat cycles of earlier single-satellite altimeters (about 21 days,

still much longer than tidal periods), so the limitations of harmonic fitting in the presence of incoherence persist. What SWOT fundamentally changes, however, is the spatial sampling: it delivers wide-swath SSH snapshots at unprecedented spatial resolution, creating an opening for methods that can exploit spatial information.

New approaches have sought to incorporate additional dynamical constraints to compensate for the lack of temporal data. These physics-based approaches range from potential vorticity inversion that leverages concurrent surface density and SSH data to reconstruct the BM (Ponte et al., 2017; C. Wang et al., 2025), to variational data assimilation that fit observations to coupled balanced and wave models (Bellemin-Laponnaz et al., 2025). Data-driven reduced-order methods have also emerged. For instance, Proper Orthogonal Decomposition based on prior knowledge of the BM’s structure isolates the IT component that is correlated with the eddy field (Maingonnat et al., 2025). For cases where relatively short revisiting periods are available, for example during the initial SWOT Cal/Val track, Dynamical Mode Decomposition can effectively perform a spatio-temporal filter (Lapo et al., 2025; Uchida et al., 2025). A comprehensive comparison of these methods, using robust benchmark data from simulations (where spatio-temporal data is fully accessible), is still lacking.

In parallel, deep learning has emerged as a promising alternative. Recent works treat flow separation as an image-to-image translation problem as opposed to filtering in time, and train neural networks on large simulated dataset where the BM and waves are known. H. Wang et al. (2022a) (hereafter “W22”) used a conditional Generative Adversarial Network (hereafter “cGAN”) to extract ITs from SSH snapshots in an idealized simulation where a baroclinic jet interacts with a mode-1 IT. Other work has targeted the BM: Gao et al. (2024) used a U-Net architecture with a custom loss function that heavily weights the SSH gradient error, thereby ensuring the recovered BM had realistic geostrophic currents. Xiao et al. (2023) also uses a U-Net to extract the BM, but instead implements a physics-informed approach by incorporating a soft constraint based on the quasi-geostrophic potential vorticity equation directly into the training loss function.

Here we expand the methodology of W22, which used solely SSH as the input. We broaden the scope of possible input fields to include surface horizontal velocities (hereafter “surface velocities”) and sea surface temperature (hereafter “SST”). These additional input fields are in principle observable in the ocean. The SST is commonly derived

from satellite observations using thermal infrared and passive microwave sensors, each with complementary characteristics, and global gap-free products spanning decades of observations already exist; Nielsen-Englyst et al. (2024) includes a recent review. Surface velocities has been traditionally measured by coastal high-frequency radars (Terrill et al., 2006), which already have overlaps with SWOT’s SSH observations (Kachelein et al., 2024), and demonstrably informative on incoherent ITs in the measured regions. In the open ocean, extensive measurements of $\mathbf{U}$ are relatively novel, made possible by recent developments of remote sensing instruments and data retrieval algorithms (Hauser et al., 2023); mission concepts such as ODYSEA (Larrañaga et al., 2025), SEASTAR (McCann et al., 2024) and Harmony (Theodosiou et al., 2021) are currently under evaluation/preparation; several regional campaigns for evaluation and proofs of concepts are already completed (Farrar et al., 2020; McCann et al., 2024). Our primary objective is to identify which fields provide the most valuable information and how they act in synergy. The results clarify the contribution of each field, highlight the critical role of surface velocity, and offer valuable insights for optimizing future observational and data preparation strategies to best disentangle the ocean’s complex, interacting balanced and unbalanced motions.

To facilitate experiments, we streamline the deep learning algorithm, discovering that with some care in learning-rate scheduling, a U-Net achieves performance on par with the more complex cGAN used in W22. Using data from a well-established Boussinesq simulation (Ponte & Klein, 2015; Dunphy et al., 2017) that features a mode-1 IT propagating through a turbulent BM, we test all combinations of sea surface height (H), surface velocities ($\mathbf{U} = (U, V)$), and surface temperature (T) as inputs. These are described in detail in §2. We conduct a systematic analysis in §3 on how different, concurrently measured surface fields contribute to the deep-learning-based extraction of IT. Along the way, we formulate a physics-motivated perspective explaining the different impacts from different input fields. Insights into some behaviors of the algorithm emerge; non-locality is important for the algorithm’s success (§3.5), and residue error centers around small spatial scales, possibly caused by spurious signals introduced during data preparation (§4). A summary and perspectives on future developments are offered in §5.

2 Methodology

2.1 Boussinesq Simulation

The training and testing data for our deep learning algorithms are outputs from an idealized simulation described in Ponte and Klein (2015) and Dunphy et al. (2017). This set of simulations is developed by members of the SWOT Science Team with the challenge of incoherent ITs in mind, and well established as a complex but idealized benchmark for studies of IT-BM interactions and disentanglement (Ponte et al., 2017; Le Guillou et al., 2021; Caspar-Cohen et al., 2022). The specific dataset we use is uploaded on Ponte et al. (2020). W22 used the same simulation data set. We briefly recapitulate some directly relevant aspects.

A three-dimensional, hydrostatic Boussinesq model is solved by the Regional Oceanic Modeling System (Shchepetkin & McWilliams, 2005), with horizontal resolution of 4 km. Equations are integrated on a β -plane centered around latitude 45° N. A snapshot of some surface fields plotted over the whole horizontal domain is shown in Figure 1. The model contains two essential dynamical components for our investigation: an incoherent IT, and a turbulent BM that scatters the IT. A continuously restored meridional density gradient in central latitudes induces baroclinic instability, forming a turbulent zonal jet with a characteristic meander width of ~ 800 km. This constitutes the turbulent BM. A wave-maker at the southern end of the domain radiates a mode-1 IT with a period of $P = 12$ hours. A northern sponge layer prevents reflections. The incoherence of ITs is induced by (a), the scatterings from the BM and (b), the modulations of peak wavelengths induced by spatially varying density profiles. To distinguish the performance of our deep learning algorithms at different locations relative to the turbulent jet, we partition the domain into up-jet, mid-jet and down-jet regions marked in Figure 1(a). The mid-jet region spans the latitudes with nonzero background meridional density gradients, and accordingly, the turbulent jet’s kinetic energy is stronger in the mid-jet region than the up/down-jet regions. The ITs are plane-wave-like in the up-jet region, which is close to the forcing region and not yet reaching the regions with strong BMs. As the BMs scatter ITs and introduce incoherence, ITs become strongly incoherent mid-jet (Figure 1, panel (b)). ITs in the down-jet region are strongly incoherent even though the BMs are not as active there; this is primarily because the BMs alter the directions of ITs, introducing non-local impacts downstream of the main propagation directions.

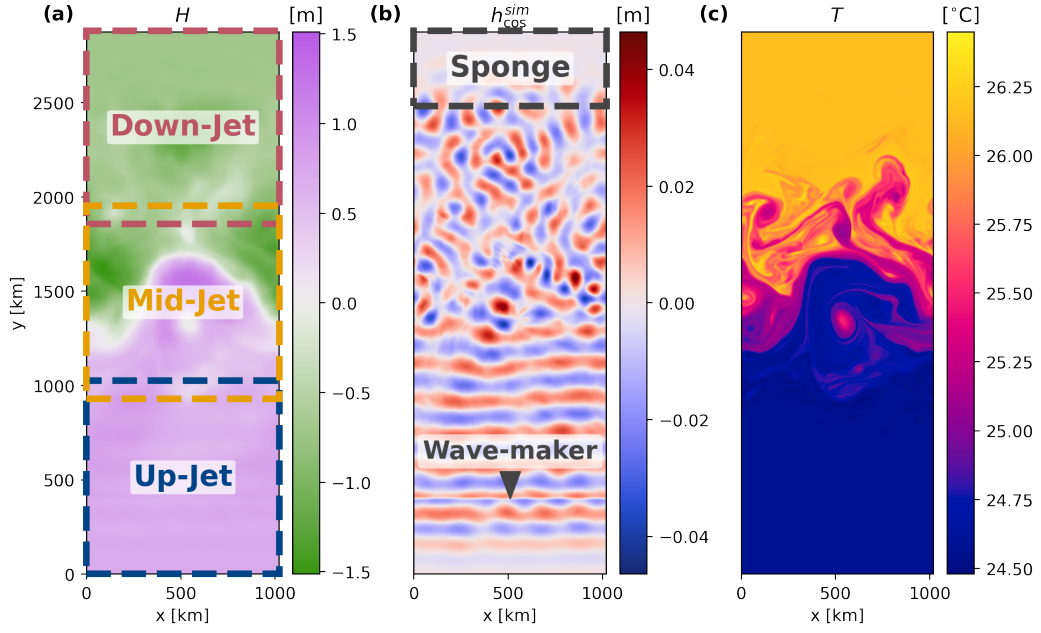


Figure 1. Snapshot of raw SSH H (panel (a)), IT signal h_{cos}^{sim} (panel (b)), and surface temperature T (panel (c)) from the T5 simulation at day 200. Dashed boxes mark the up-jet, mid-jet, and down-jet regions define for analysis. The Sponge region and the central latitude of the wave-maker are marked in panel (b).

The reference fields indicating IT imprints on SSH are defined following Ponte and Klein (2015):

$$h_{\cos}^{\text{sim}}(x, y, t) = \frac{1}{P} \int_{t-2P}^t H(x, y, t') \cos\left(\frac{2\pi t'}{P}\right) dt', \quad \text{and} \quad (1)$$

$$h_{\sin}^{\text{sim}}(x, y, t) = \frac{1}{P} \int_{t-2P}^t H(x, y, t') \sin\left(\frac{2\pi t'}{P}\right) dt', \quad (2)$$

where t denotes time in the Eulerian frame, and the superscript “sim” notes that the fields are produced by the Boussinesq “simulation” data. The components $h_{\cos}^{\text{sim}}$ and $h_{\sin}^{\text{sim}}$ constitute the reference fields that the network is trained to predict. (In W22, they were called “truth” fields.) In the reference data, equations (1) – (2) are evaluated at each (x, y) via harmonic fitting (least-squares method) from finely sampled time series with time intervals of $P/144$. Such temporal information makes harmonic fitting accurate even in the presence of strong incoherence, but—as explained in the Introduction—this refined temporal information is unavailable in global satellite data, which motivates our snapshot-based approach. Our deep learning algorithm only has access to snapshots captured with time intervals of two days ($4P$); under such coarse temporal sampling, harmonic analysis would yield nearly zero fields in the mid-jet and down-jet regions in our datasets (Aurélien Ponte, private communication). In the training and testing of our deep learning algorithm, all snapshots are randomly reshuffled, which further ensures that no temporal evolution is accessible.

We use simulation outputs under five different profiles of meridional density gradients. We refer to these five simulations as T1 to T5 in order of increasing density gradients. The variation of the density gradients affects the IT patterns in two ways. First, it adjusts the eddy kinetic energy budget and leads to five different levels of turbulent BM activity (hereafter “turbulence levels”). A larger density gradient leads to a higher turbulence level, which induces stronger scattering of ITs, making the IT signals more incoherent and the IT patterns more complex. This was established in Ponte and Klein (2015) and quantified in detail in W22. Second, the different density profiles affect the mode-1 IT wavelengths as a consequence of the mode-wise dispersion relationships (Gerkema & Zimmerman, 2008). From T1 to T5, the minimum of background mode-1 IT horizontal wavelength in the mid-jet regions decreases from 170 km to 140 km, as quantified in W22.

2.2 Deep Learning Algorithm

We cast the extraction of IT imprints as an image-to-image translation problem, and we apply convolutional neural networks to map *inputs*, which are different combinations of unfiltered surface fields, to *outputs*, which are the filtered IT signatures on SSH. In principle, we can also task our algorithm to learn IT signatures on other physical fields (e.g., on surface velocities); we stay focused on the IT in SSH as they are more relevant to SWOT’s SSH observations.

W22 employed a cGAN. Here we instead use a U-Net, which is simpler, more efficient, and—when trained properly—achieves comparable skill. The U-Net has a symmetric encoder–decoder structure with skip connections. Although the U-Net is widely used as a baseline in the literature (Azad et al., 2024), W22 excluded it because early tests showed weak performance. We now understand that our previous failure with the U-Net was due to an inadequate design of the learning rate scheme in the training of U-Nets: the learning rate was kept constant, following the original work that first proposed the U-Net architecture (Ronneberger et al., 2015). In this work, we vary learning rates periodically with respect to number of epochs (number of iterations of datasets during training). The rationale is that a deep learning algorithm’s training can be regarded as an optimization problem, with the loss function as the object to minimize. Small learning rates fine-tune the optimizer with higher accuracy and result in stabler convergence, but they slow the training, and may lead the optimizer to get trapped in local minima (Smith, 2017; Liu, 2022). By periodically amplifying the learning rates, the optimizer is encouraged to escape local minima, while still achieving accuracy and convergence when the learning rates become small again (Loshchilov & Hutter, 2016). With a periodically varying learning rate, our U-Net can now perform similarly to the cGAN; subtle differences are discussed in §5. The main results are run on a NVIDIA A100-SXM4-40GB GPU and take about 25 minutes to train the U-Net for 300 epochs. Convergence tends to happen around 150 epochs. The low computational cost of the U-Net allows us to more freely explore different combinations of input and outputs.

Figure 2 sketches the U-Net architecture we apply. Each input physical field is a separate input channel, and the U-Net’s topology remains the same for different combinations of input physical fields. The training loss function is the standard point-wise

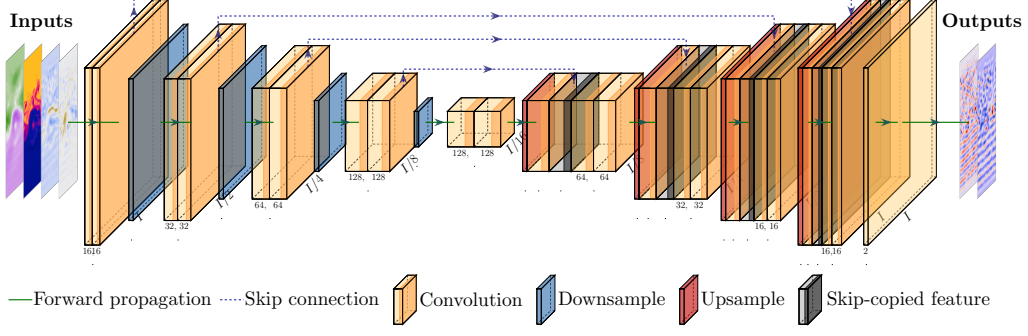


Figure 2. Schematic of the U-Net architecture employed in this work. The network takes combinations of surface fields in (H, U, T) as input and predicts two output channels corresponding to the IT imprints on SSH, $h_{\cos}^{\text{sim}}$ and $h_{\sin}^{\text{sim}}$. Labels below each block denote (a) the number of feature maps, i.e., channels (e.g., 16, 16) and (b) the dimension along one direction of each feature map (e.g., $I, I/2$). Example input and output snapshots are taken from the same snapshot as in Fig. 1. Plotting code is adapted from Iqbal (2018).

L_2 . More implementation details and design considerations are provided in Supporting Information Text S1 and the production codes are provided in H. Wang (2025).

For each combination of input data, we re-initialize the U-Net with random kernel values and re-train it ten times. The statistics recorded in Supplemental Information Text S2 show that for each Configuration of input data discussed in this work, the ten re-trained U-Nets show negligible variation. In the main text, for simplicity, under each configuration of input data, we arbitrarily choose one out of the ten re-trained U-Nets, and only report on the outputs from that one U-Net.

Like many other deep image-to-image translation algorithms, the U-Net does not apply a spatial filter in the same way as a transfer function. Instead, nonlinear activation layers enable it to learn complex, nonlinear mappings. A U-Net can exploit input fields (e.g., surface velocities) that do not contain the outputs (i.e., IT imprints on SSH) directly. This is the foundation for our explorations on the impacts of different input fields.

2.3 Inputs and outputs

We now describe the inputs and outputs of our U-Net. To clarify terminologies, we defined *inputs* and *outputs* in section 2.2, and sketched them in Figure 2. Both *train-*

ing and *testing* data consist of pairs of inputs and outputs. During training, the U-Net is exposed to pairs of inputs and outputs from the training data. After training, the U-Net is given inputs from the testing data and generates new outputs; the generated outputs are compared with the reference outputs from the testing data when we evaluate the performances.

As introduced in §2.1, our simulations are run at five different turbulence levels. W22 discussed different combinations of turbulence levels for the training and testing data. For example, in what they referred to as the “ET5 run”, the deep learning algorithm was trained on turbulence levels T1–T4, and tested on T5. Similarly, “ET1 run” uses T1 as the testing data. The training and testing data contained disjoint turbulence levels, which ensured that the ITs have different levels of incoherence and different characteristic spatial scales between the training and testing data, and tested the deep learning algorithm’s ability to generalize. In W22, among the five runs ET1–ET5, the ET5 run was identified as performing the worst, both quantitatively and qualitatively. The same behavior is found in experiments with our U-Net (not shown). The poorer performance in the ET5 run is not surprising, as the ITs and the BMs are more complex in the testing data. To stay focused on the most challenging case, hereafter we exclusively discuss ET5.

In the Boussinesq model, after the ITs are forced, there is a ramp-up period of around 100 days where the ITs propagate and fill up the simulation domain. W22 excluded these snapshots while here we include them. We also include the wave-maker and sponge regions that were excluded in W22. These choices introduce a slightly greater diversity of dynamic regimes present in the data. We include 150 snapshots with time intervals of two days from each turbulence level. This results in 600 snapshots from T1–T4 in the training data and 150 snapshots from T5 in the testing data.

The inputs of the model are different combinations of surface fields: the total SSH H , the surface temperature T , and the surface velocity fields $\mathbf{U} = (U, V)$, where U and V denote zonal and meridional velocities. These are fields represented in the Boussinesq simulations. A snapshot is shown in Figure 1 (panels (a,c)) and Figure 3 (panels (a,b)). We exhaustively test all possible combinations of these three fields, first inputting each field in isolation, then every two-member combination, and finally all three fields together. We refer to each combination of input fields as a “Configuration” followed by brackets

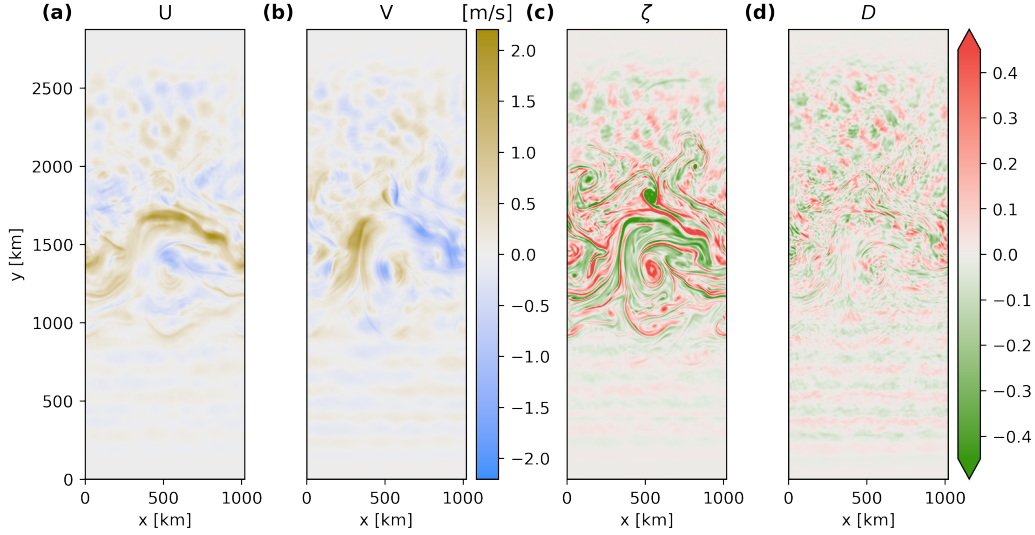


Figure 3. Surface zonal and meridional velocities (U, V) (panels (a,b)), derived surface vorticity ζ , and divergence D (panels (c,d)), taken at the same snapshot as in Figure 1. The fields ζ and D are computed by finite differencing the velocity field, and are non-dimensionalized by the Coriolis frequency evaluated at the central latitude of the domain (45° N).

specifying the input fields. For example, “Configuration $\{H, \mathbf{U}\}$ ” refers to deep learning experiments that take H and $\mathbf{U}$ as inputs. In W22, only Configuration $\{H\}$ was studied.

When we use $\mathbf{U}$ as an input, we include both zonal and meridional velocities, U and V , unless specified otherwise, as they are generally measured in tandem on ocean surface by the instruments (e.g., Doppler Scatterometers (Rodriguez, 2018) and coastal high-frequency radars (Kachelein et al., 2024)) we are aware of.

Under each configuration of inputs, the outputs that the U-Net tries to generate are fixed to be $h_{\cos}^{\text{sim}}$ and $h_{\sin}^{\text{sim}}$ defined in (1) – (2), formatted as two output channels. They are the IT imprints on SSH, not on $\mathbf{U}$ or other input fields (see §2.2, last paragraph, regarding why this works for U-Nets).

3 Impacts of different surface fields

3.1 Different inputs, different performances

We now generally compare the U-Net performance under different inputs. We denote the ITs generated from our U-Nets (hereafter “generated”) as $h_{\cos}^{\text{gen}}$ and $h_{\sin}^{\text{gen}}$, which are the U-Net’s attempted reconstructions of reference fields, $h_{\cos}^{\text{sim}}$ and $h_{\sin}^{\text{sim}}$, defined in (1) – (2); note the different superscripts, where “gen” and “sim” are shorthand for “generated” and “simulated” respectively. A snapshot in the mid-jet region generated from each Configuration compared against the reference are shown in Figure 5 for visual references. A visual inspection already suggests that some configurations perform better than others.

To quantify these impressions, here we evaluate two skill metrics between reference and generated fields:

- the correlation coefficient Υ , which rewards features in the generated images that vary linearly with the reference fields, and
- the coefficient of determination R_2 , which detects mismatches in overall magnitudes, or outliers where mismatches are large point-wise.

The ranges for Υ and R_2 are $[-1, 1]$ and $(-\infty, 1]$ respectively, with values closer to 1 indicating better alignment between the generated and reference fields.

At each t in the testing data, we apply a trained U-Net to the input snapshot to obtain $(h_{\cos}^{\text{gen}}(x, y, t), h_{\sin}^{\text{gen}}(x, y, t))$, which is then flattened into a one-dimensional array. We do the same for $(h_{\cos}^{\text{sim}}(x, y, t), h_{\sin}^{\text{sim}}(x, y, t))$. Υ and R_2 are computed between these two one-dimensional arrays at each t , and then averaged over all t .

In addition to the Υ and R_2 computed over all horizontal grid points in the simulated region, we also show Υ and R_2 computed over only the mid-jet regions. This is done by truncating $h_{\cos}^{\text{gen}}(x, y, t)$ and so on in y leaving only the mid-jet regions, before the arrays are flattened. In W22, the mid-jet regions were identified to be more challenging than the other regions, hence the focus here. (In §4, we will elaborate more on the nature of challenges in the mid-jet regions.) Table 1 shows Υ and R_2 for all configurations for the full regions and the mid-jet regions. These metrics turn out to improve in tandem when we compare between configurations; for example, a configuration with higher

Metric	T	U	H	H, T	U, T	H, U	H, U, T
$\Upsilon \times 100$	41/2.9	92/86	81/70	84/77	95/92	96/94	97/96
$R_2 \times 100$	17/0.0	85/74	65/49	71/59	91/85	93/88	95/92

Table 1. Mean Υ and R_2 for each Configuration. Values are reported as “full / mid-jet.” The headings note the inputs in the Configurations.

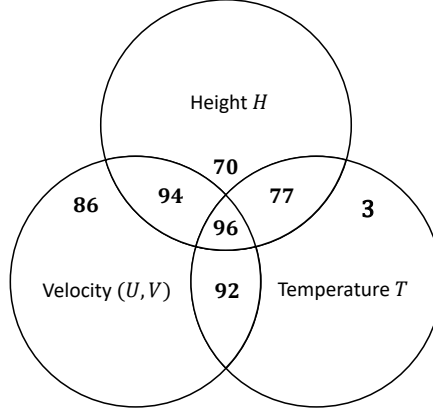


Figure 4. Summary of relative performances across input configurations. Each circle corresponds to one input field in (H, U, T) ; overlaps denote configurations that combine input fields. Numbers within circles show mean correlation (100Υ) averaged over the mid-jet region.

Υ over the full region also has higher R_2 over the mid-jet region. Therefore, we can identify which configuration performs better than another one with no ambiguities. Figure 4 is a concise summary on the relative performances between configurations; for brevity, only the mid-jet Υ is shown.

3.2 Wave signature and scattering medium

Clearly, having different input fields affect the performances. We discuss the origins of these differences as follows. Each input field can contribute in two ways: (a) by exhibiting a measurable signature of the ITs, which aids the identification of the phase and orientation; and (b) by reflecting properties of the BMs and background conditions that modulate the ITs, which helps to interpret the IT energy distribution and overall statistics in a time-averaged, non-phase-resolving sense. We refer to these two types of information as “wave signature” and “scattering medium”.

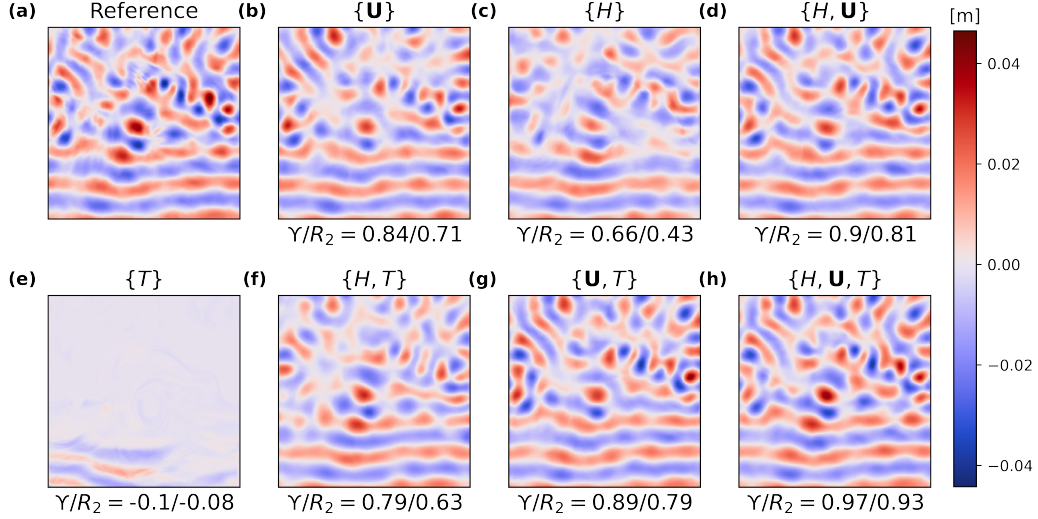


Figure 5. Comparison of reference IT signal $h_{\text{cos}}^{\text{sim}}$ (panel (a)) with U-Net reconstructions $h_{\text{cos}}^{\text{gen}}$ from different input configurations (panels (b-h)). Numbers below each panel in (b-h) report Υ and R_2 computed between the respective panels and panel (a). Only the mid-jet region is plotted, taken at the same snapshot as in Figure 1.

ITs induce SSH fluctuations, and so conversely, H contains wave signatures. It also provides information about the scattering medium. Part of the BMs that scatter ITs can be inferred from SSH. In our simulation data, the BM velocities are strongly correlated to geostrophic velocities proportional to the spatial gradients of SSH (Ponte et al., 2017). BM velocities that are not geostrophic also have a signature on SSH; for example, under the effective surface quasi-geostrophy model (Lapeyre & Klein, 2006) applicable to surface intensified eddies, SSH is related to geostrophic stream functions at all depths, which prescribes the entire BM dynamics. This is partly why deep learning approaches to extract BMs from SSH show promising performances (Xiao et al., 2023; Gao et al., 2024). More speculatively, if single modes of ITs behave as shallow-water waves (an approximation roughly applicable to our simulation data (Le Guillou et al., 2021)), the equivalent depth of each vertical mode is linked to SSH (Gill (2016), p161), which dictates propagation metrics such as group speeds. For Configuration $\{H\}$, the main challenge of the U-Net is to disentangle the waves and the scattering medium (e.g., BMs), utilizing the differences and connections between the two. One particular challenge is that the imprints of ITs on SSH are about two magnitudes smaller than the imprints from BMs in our simulation; in realistic ocean, internal wave imprints on SSH are also generally

small compared to BMs, as a consequence of the large air-sea density differences. Thus, under this Configuration, our U-Net needs to disentangle signals that are very uneven in magnitudes. This disentanglement is moderately successful, to the extent summarized in Figure 4, and to be further detailed in §4. We next discuss the impacts of the other input surface fields through the same perspective.

3.3 Surface velocities

From Figure 4, Configuration $\{\mathbf{U}\}$ outperforms Configuration $\{H\}$ or $\{T\}$ by a large margin. When $\mathbf{U}$ is added to other inputs, the improvements in performance are also significant. We now discuss why using $\mathbf{U}$ leads to such strong performances.

As mentioned before, we include both velocity components (U, V) in the inputs because they are usually measured together at the ocean surface. One might still worry that Configuration $\{\mathbf{U}\}$ performs better than $\{H\}$ or $\{T\}$ simply because it supplies two input channels instead of one, giving the network more flexibility regardless of any genuinely richer physical content. To separate this “number of channels” effect from the information content of the velocity field, we also trained networks with only a single velocity component as input, Configurations $\{U\}$ and $\{V\}$ (see Supporting Information Text S3). In these cases, the number of input channels is identical to Configurations $\{H\}$ and $\{T\}$, but the mid-jet mean R_2 values are 0.48 for $\{U\}$ and 0.60 for $\{V\}$, compared with 0.45 for $\{H\}$. Both single-component velocity configurations therefore outperform the SSH-only case despite having the same number of input channels. This indicates that the superior performance of Configuration $\{\mathbf{U}\}$ is not merely due to an advantage from providing more channels to the network.

There is an abundance of wave signatures and information on scattering medium in $\mathbf{U}$. Under linear polarization relations, the phases of IT imprints on H are related to the IT imprints on $\mathbf{U}$. The orientations of the IT imprints on $\mathbf{U}$ are guided by the phase velocity’s direction, which also guides the orientation of the IT imprints on H . Thus, $\mathbf{U}$ contains useful wave signature. On the other hand, BM velocities generate both local and far-reaching spatial variations of wave energies, as suggested by several works (Rainville & Pinkel, 2006; Wagner et al., 2017; Savva et al., 2021; Uncu & Grisouard, 2024). Hence, the imprints from BMs on $\mathbf{U}$ can provide important information on the scattering medium.

From this perspective, the information contained in $\mathbf{U}$ is abundant, and different from that in H . The exact reason why Configuration $\{\mathbf{U}\}$ outperforms Configuration $\{H\}$ is not completely clear. We speculate that one contributing factor is the stronger wave signatures in $\mathbf{U}$, as compared to H . Quantitatively, the root mean square (hereafter “rms”) of the IT amplitude of H , computed as $\sqrt{(h_{\cos}^{\text{sim}})^2 + (h_{\sin}^{\text{sim}})^2}$, is only 3.7% of the rms of H in the testing data. In comparison, if we compute the IT imprints on U and V in the same way as H (following (1)–(2)), the rms of the IT amplitudes of U and V are 20.8% and 17.8% of the rms of U and V respectively, which are about a magnitude larger than 3.7%. Theoretically, under a Helmholtz decomposition of surface velocities, linear waves dominate over BMs in the divergent component (e.g., Bühler et al. (2014)), which explains a source of the strong wave signatures in $\mathbf{U}$. We further speculate, from the perspective of the Helmholtz decomposition, that an additional advantage of $\mathbf{U}$ is a relatively clean conceptual pathway to disentangle the waves and BMs; this is discussed in §5.

The performance of Configuration $\{H, \mathbf{U}\}$ is higher than both Configuration $\{\mathbf{U}\}$ and Configuration $\{H\}$ by large margins. This reflects that the two variables provide complementary information, and adding one to the other improves the performance.

3.4 Surface Temperature

The surface temperature T provides mostly only information on the scattering medium. The imprint of the ITs on T is much weaker than the BMs in our simulation, as quantified in Ponte et al. (2017) (see e.g. their figure 5; relevant discussions on sea surface density in Ponte et al. (2017) can be generalized to T , as salinity is fixed in the simulation). Assuming that the evolution of SST is dominated by advections, Ponte et al. (2017) attributes the weak IT imprint on SST to the uneven scaling of horizontal advection terms. The weak IT imprints on SSTs are demonstrated in realistic global circulation oceanic models too (e.g., Torres et al. (2018), figure 4). Some theoretical arguments, taking into account ITs’ effects on vertical mixing or the surface thermal boundary layer, do predict strong imprints from ITs on SST, but whether these effects are significant under realistic ocean conditions remains under debate; a relatively recent perspective can be found in Farrar et al. (2007). By contrast, BMs leave strong imprints on T and are dynamically linked to T . For example, under surface quasi-geostrophy frameworks (Lapeyre & Klein, 2006), surface temperature (or density) anomalies determine a boundary condi-

tion for BM evolution. In our simulation, the strong correlation between T and potential vorticity at both the surface and the interior is verified in Ponte et al. (2017).

As T contains little wave signature, it is hard to infer the phases and directions of ITs from T alone. A snapshot generated by Configuration $\{T\}$, compared against the reference (Figure 5, panels (a,e)), shows that Configuration $\{T\}$ captures some structures near the up-jet region where ITs are relatively coherent (similar performance is found within the up-jet region, which is not shown), but misses the rest of the mid-jet region where the ITs are incoherent. As our snapshots are captured exactly every $4P$, the phases of coherent ITs are fixed between snapshots. For such ITs, phase information is less crucial, and the U-Net can extract some coherent IT signals from T alone. In contrast, strongly incoherent ITs have phases that vary strongly between snapshots in our dataset. Meanwhile, as BMs evolve slowly, the input $\{T\}$ varies little between snapshots. This makes the inference of incoherent ITs from $\{T\}$ an ill-posed problem where similar inputs lead to very different outputs, and the U-Net responds by outputting weak, near-zero signals to minimize the training loss. Consequently, Configuration $\{T\}$ performs poorly under all metrics, especially in the mid-jet region.

This highlights that unsurprisingly, under a dearth of phase-resolving information in the inputs, the U-Net performs badly at resolving variable phases. To further test this, we have run another configuration (not shown) using only the BM components of H and U , computed by top-hat averaging following Ponte et al. (2017)), as inputs. This configuration also performs poorly.

Nevertheless, T still provides useful information on the scattering medium. When combined with other inputs that contain abundant wave signatures, T helps the U-Net disentangle IT and BM imprints. Indeed, as seen in Figure 4, adding T improves the performance of all configurations. Notably, Configuration $\{H, U, T\}$ outperforms Configuration $\{H, U\}$, showing that T contains complementary information absent in U and H .

We also test the impacts of T with a low-pass filter applied, suppressing small-scale features to more realistically represent the spatial resolutions of typical gap-free, gridded satellite observations of T . Details are provided in Supporting Information Text S4. We find that reducing the effective resolution of T to 25 km has no statistically significant impact on performance, while reducing the effective resolution to 124 km only slightly

degrades accuracy. Therefore, configurations with the smoothed T as inputs still significantly outperform those without T . This supports the idea that T provides BM-related information on the scattering medium in our simulations, which is active primarily at large spatial scales and insensitive to smoothing.

In addition to SST, other surface fields with tracer-like behaviors may also provide complementary information on the scattering medium. In our simulations, the salinity is kept constant and no bio-geochemical activities are simulated; in more complex simulations, impacts of sea surface salinity or chlorophyll as additional inputs can be experimented on. These fields are in principle observable in the ocean; for instance, NASA’s Aquarius maps sea surface salinity, and the Plankton, Aerosol, Cloud, ocean Ecosystem (PACE) mission maps chlorophyll globally, partly overlapping with SWOT.

3.5 Non-locality

The background conditions and BMs that comprise the scattering medium can be active at spatial scales much larger than the ITs’ wavelengths. To harness information at large scales, neural networks of choice must be able to use non-local information. In our U-Net architecture, non-locality is affected by the depth of the network. Each encoding/decoding step increases the size of the receptive field, i.e., the spatial span of the region of the input image that influences a single grid point in the output.

In our main U-Net defined in §2.2, four encoding–decoding steps give a receptive field of about 200×200 grid points (see e.g., Araujo et al. (2019) for how the receptive field is computed). As our grid spacing is 4 km, this corresponds to a receptive field of $800 \times 800 \text{ km}^2$, which covers the scale of the meander width of the turbulent jet.

To test the effect of receptive field size, we construct a shallow U-Net, which has only two encoding/decoding layers, sketched in Figure 6. For brevity, we only discuss Configuration $\{H\}$ for the shallow U-Net. This shallow U-Net has a receptive field of 44×44 grid points, or $176 \times 176 \text{ km}^2$. The width of this receptive field is comparable to the IT wavelengths. To keep the number of trainable parameters comparable to the main U-Net, we increased the number of kernels in each layer of the shallow U-Net; the shallow U-Net has slightly more parameters than the main U-Net. All other aspects of the training are kept the same. As for our main U-Net, the shallow U-Net is also re-trained ten times and we find the performance to be insensitive to re-training (not shown).

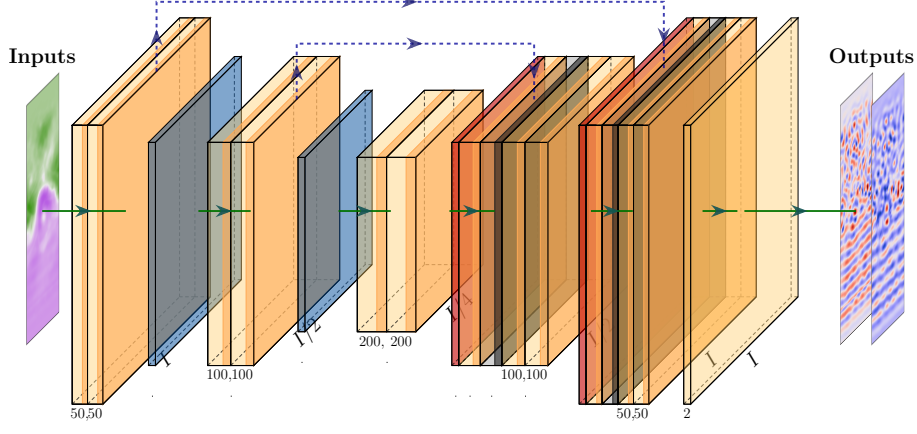


Figure 6. Architecture of the shallow U-Net with two encoding/decoding layers, used to test the role of nonlocal information. The labels are the same as in Figure 5. Further implementation details are provided in H. Wang (2025).

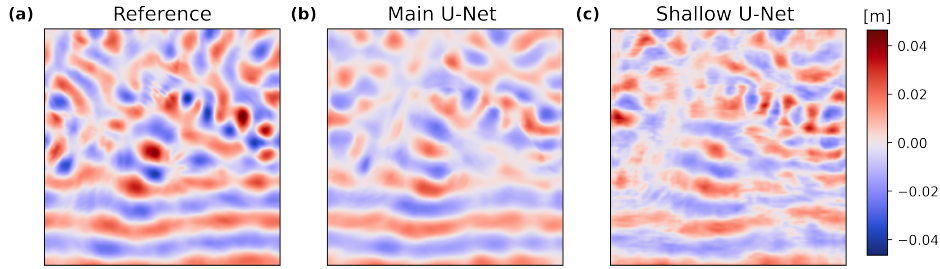


Figure 7. Comparison of the reference $h_{\cos}^{\text{sim}}$ (panel (a)) with $h_{\cos}^{\text{gen}}$ generated by the main U-Net (panel (b)) and shallow U-Net (panel (c)) under Configuration $\{H\}$. All for the mid-jet region at the same snapshot as in Figure 1.

The performance of this shallow U-Net is weaker than that of our main U-Net. The mean mid-jet Υ and R_2 of the shallow U-Net are 0.64 and 0.39 respectively, as opposed to 0.70 and 0.49 from the main U-Net. A snapshot (Figure 7) shows that compared to the main U-Net, the shallow U-Net's outputs contain more spurious signals at small spatial scales. This confirms that a large, mesoscale-reaching receptive field is critical for reconstructing ITs, consistent with the physical picture that scattering processes depend on broad background flow structures.

The receptive field ($176 \times 176 \text{ km}^2$) of our shallow U-Net is still large enough to include some features from the scattering medium, and wave signatures are active at smaller

scales. Thus, it’s not surprising that the shallow U-Net does not completely fail. In principle, a more rigorous experiment should be made with shallower, more local neural network architectures. This is currently computationally prohibitive. Feature maps stored in shallower layers are larger; for example, the feature maps in the bottleneck layers of our main and shallow U-Nets are of widths $I/16$ (Figure 2) and $I/4$ (Figure 6) respectively. Meanwhile, to keep the comparisons fair, we need to keep the number of free parameters comparable between shallower and deeper networks. Thus, shallower networks have more large feature maps, and are more demanding in computational memory. Our shallow U-Net approaches the limit of GPU memory we can access currently (i.e., 80 GB; admittedly, this quota can be larger if we learn to parallelize multiple GPUs). Thus, we can’t easily make it shallower without changing the network topography significantly. Nevertheless, the decreasing performance of our two-layer shallow U-Net already implies that the lack of a large receptive field is detrimental in our application.

For comparison, in a different context, learning subgrid closures for two-dimensional turbulence, Srinivasan et al. (2024) contends that the receptive field does not benefit from being very large: a receptive field of 9×9 grid points is found to give optimal performance. Some non-locality is still essential in the turbulence closure problem, as argued by Brolly (2025); this is consistent with Srinivasan et al. (2024), which argues that small-scale fluxes are physically related to neighboring grid points via (for example) terms involving spatial gradients, and weak non-locality is useful. In our problem, as the scattering medium is active at spatial scales much larger than the IT wavelengths, it makes sense physically that large, mesoscale-reaching receptive fields are helpful. In other words, IT–BM interactions operate over broader spatial ranges than the processes typically parameterized in turbulence closures. At the very least, we can say that the optimal receptive field of a neural network must depend on the nature of the problem at hand.

4 Captured and missed features

Judging by Υ and R_2 reported in §3, Configuration $\{H, \mathbf{U}, T\}$ achieves very high skill, with mean mid-jet Υ and R_2 close to 1. Within our set of experiments, this represents the best performance obtained when the network is given all three surface state variables. In this section, we examine more closely what Configuration $\{H, \mathbf{U}, T\}$ still misses. From §3, Configuration $\{H\}$ performs worse, and we also analyze what specific types of errors are worse in Configuration $\{H\}$.

Metric	$\{H\}$	$\{H, \mathbf{U}, T\}$
$\Upsilon \times 100$	79 (97/68/71)	97 (100/96/98)
$R_2 \times 100$	63 (94/46/50)	95 (99/92/96)

Table 2. Υ and R_2 under Configurations $\{H\}$ and $\{H, \mathbf{U}, T\}$. The headings denote the Configurations. The statistics are presented in the format “full (up-jet/mid-jet/down-jet)”.

We compute the mean Υ and R_2 over the up-jet/mid-jet/down-jet regions in the test data separately, and list them in Table 2. Performance in the up-jet is very high: even Configuration $\{H\}$ scores well, and Configuration $\{H, \mathbf{U}, T\}$ improves further. This is consistent with the snapshots in Figure 8, where visual agreements in the up-jet regions are excellent. As the ITs are more coherent in the up-jet regions, the IT patterns are more regular, posing fewer challenges to the U-Net.

In contrast, the performances in the mid-jet and down-jet regions are significantly worse than in the up-jet, with the mid-jet region giving the poorest metrics. As noted in §2.1, the mid-jet region has stronger meridional density gradients and more intense turbulent jet activity, which increase variations in background IT wavelengths and enhance scattering by BMs. With the increased incoherence, the IT imprints are more complex in the mid-jet regions, posing more challenge to the U-Net. This behavior was also demonstrated extensively in W22 for Configuration $\{H\}$, where a different neural network architecture (cGAN) was used; the increased challenge under stronger incoherence appears common between different neural networks.

To understand the scale-dependent distribution of error, we also inspect the wavenumber spectra. For brevity, we focus on the azimuthally averaged power spectra of ITs as in W22, defined as the averages of two-dimensional wavenumber spectra over the polar angles of wavenumber vectors. Figure 9 shows the azimuthally averaged wavenumber spectra (hereafter “spectra”) of ITs computed over mid-jet and down-jet regions. A Hanning window in both meridional and zonal directions is multiplied to each snapshot before the fast Fourier transforms are computed. The windowing in the zonal direction is called for even though our Boussinesq simulation is zonally doubly periodic: our U-Net is not enforced to recognize the periodicity, and outputs from our U-Net are not guaranteed to be zonally periodic.

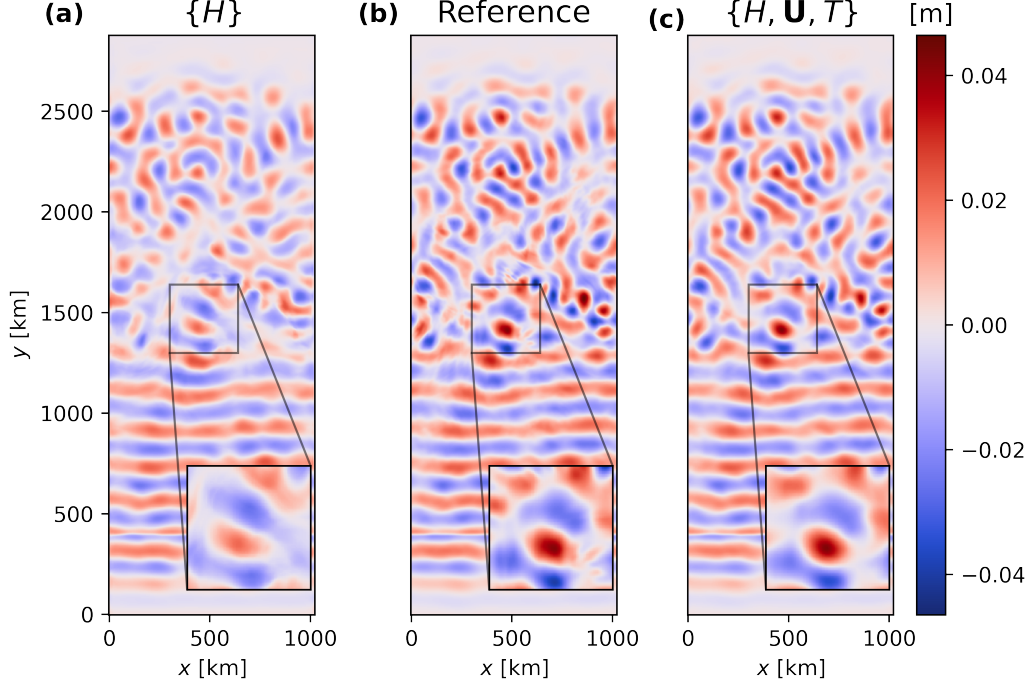


Figure 8. IT reconstructions $h_{\cos}^{\text{gen}}$ from Configuration $\{H\}$ (a) and Configuration $\{H, U, T\}$ (c), compared to the reference $h_{\cos}^{\text{sim}}$ (b). Taken at the same snapshot as in Figure 1.

In spectra from the reference fields, the locations of the peaks around mode-1 IT wavenumbers are different between the down-jet and mid-jet regions: the incoming mode-1 IT wavenumbers (the first vertical dashed line in each panel) are approximately $2\pi/185 \text{ km}^{-1}$ and $2\pi/140 \text{ km}^{-1}$ for mid-jet and down-jet. Around the incoming mode-1 IT wavenumbers, the spectra from Configuration $\{H, U, T\}$ align well with the reference field. The accurate reconstruction of the varying locations of the peaks near mode-1 wavenumbers implies that, under Configuration $\{H, U, T\}$, the U-Net recognizes variations in the dominant spatial scales of ITs.

The same cannot be stated for Configuration $\{H\}$, which shows tangible differences in the wavenumber spectra near the mode-1 peaks, especially in the mid-jet region. Descriptively, the errors of Configuration $\{H\}$ at spatial scales comparable to or larger than the mode-1 wavelengths can be divided into two types:

- (i) distortions in the scattering patterns, i.e. incorrect spatial structure and orientation, and
- (ii) systematic underestimation of amplitudes.

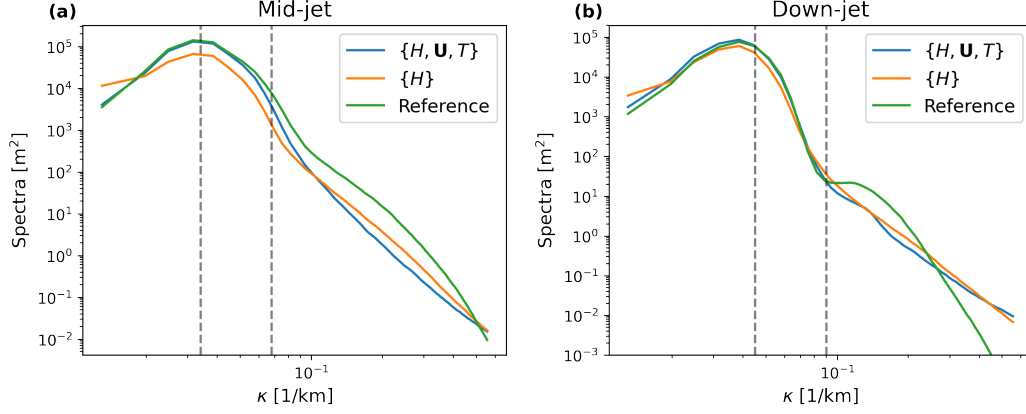


Figure 9. Azimuthally averaged wavenumber spectra of generated ITs $h_{\text{cos}}^{\text{gen}}$ (input configurations denoted in the legends) and reference ITs $h_{\text{cos}}^{\text{sim}}(x, y, t)$ (with legend label “Reference”). Panels (a,b) correspond to the mid-jet and down-jet regions, respectively. Vertical dashed lines mark mode-1 wavenumber and its harmonic ($2\times$), evaluated at the southern boundary of the mid-jet (panel (a)) or down-jet (panel (b))) regions. Spectra are averaged over the final 200 days (100 snapshots) of the testing set; values below 0.001 m^2 or at $\kappa > 0.56 \text{ km}^{-1}$ are omitted.

These two errors are visible in Figure 8 (comparing panels (a) and (b)), particularly in the zoomed-in regions at the center of turbulent jets. As errors of type (i) reflect disagreements in the trends of patterns, they affect Υ . As errors of type (ii) are mismatches of magnitudes, they affect R_2 .

Both errors (i) and (ii) are greatly alleviated in Configuration $\{H, \mathbf{U}, T\}$, as reflected in the example snapshot (Figure 8), the metrics (Table 2), and the spectra (Figure 9). Thus, adding $\mathbf{U}$ and T substantially alleviates errors of type (i) and (ii).

At wavenumbers larger than roughly twice the incoming IT wavenumber (the second vertical dashed line in each panel in Figure 9), even the maximally performing Configuration $\{H, \mathbf{U}, T\}$ exhibits much lower spectral energy than the reference. This corresponds to Figure 8, where generated ITs appear smoother than the reference. We describe this as a third type of error:

(iii) blurring of small-scale patterns.

Not all of this “error” necessarily reflects physical IT signals being missed. The ITs are computed in the reference fields through frequency filtering (equations (1) – (2)), which

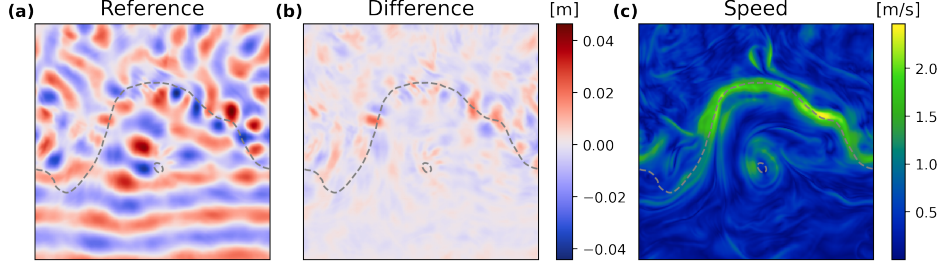


Figure 10. Panel (a): The reference $h_{\text{cos}}^{\text{sim}}$. Panel (b): Difference field $(h_{\text{cos}}^{\text{gen}} - h_{\text{cos}}^{\text{sim}})$ from Configuration $\{H, U, T\}$. Panels (a,b) share the same colorbar. Panel (c): Surface speed $|U|$. Gray dashed contours show $H = 0$ approximating the instantaneous jet axis. All are plotted over the mid-jet region at the same snapshot as in Figure 1.

assumes a separation in time scales between the ITs and BMs. However, the baroclinic jet in the BM contains energy near tidal frequencies, which is then artificially counted as ITs. Moreover, the frequency filtering (1)–(2) is based on time series at fixed Eulerian locations, and the BMs can Doppler-shift waves with other frequencies to the tidal frequency band. The simulation we use is designed so that waves are dominated by a single tidal frequency, so the wave energy that could be Doppler-shifted to the tidal frequencies is small in an averaged sense. Nevertheless, qualitative differences are reported in the behaviors of IT phases between Eulerian and Lagrangian recordings from simulations set up similarly to ours (Caspar-Cohen et al., 2022). Finally, the computations of (1) – (2) are realized by least-square fittings (detailed in Ponte and Klein (2015); Ponte et al. (2017)), which can contain numerical inaccuracies. The spurious signals induced from these mechanisms, which do not belong to ITs, may appear to be random noises to the U-Net, and are thus ignored by the U-Net.

Nevertheless, part of the missed small-scale patterns corresponds to genuine ITs, and it’s a failure of the U-Net to miss them. In particular, the secondary spectral peak near twice the incoming mode-1 wavenumber (seen in the down-jet region in Figure 9, panel (b)) arises from nonlinear scattering of mode-1 into mode-2. This physically meaningful signal is strongly damped in the U-Net reconstructions, even under Configuration $\{H, U, T\}$. Thus, error (iii) is partly a real limitation.

Figure 10 (panel (b)) shows a snapshot of the difference field $(h_{\text{cos}}^{\text{gen}} - h_{\text{cos}}^{\text{sim}})$, where $h_{\text{cos}}^{\text{gen}}$ is generated under Configuration $\{H, U, T\}$. This field is dominated by scales smaller

than the mode-1 wavelength, consistent with error (iii). Interestingly, the regions with the strongest differences do not align with the regions of stronger reference ITs (panel (a)), so the larger local errors are not due to stronger reference signals. Instead, the differences align with regions of large surface speed $|\mathbf{U}|$ (panel (c)). Two mechanisms can give rise to this. First, faster BM speeds and their gradients amplify spatially-local wave-mean interactions, leading to more complex patterns around them (Dunphy & Lamb, 2014; Uncu & Grisouard, 2024) that are harder to decipher for the U-Net. Second, BMs near tidal frequencies contaminate the reference ITs that are computed via frequency filtering; these spurious contributions, ignored by the U-Net due to their random nature, are stronger where BMs are more energetic (Caspar-Cohen et al., 2022).

For applications where behaviors of high-mode ITs are crucial, such as studies of energy dissipation pathways (Vic et al., 2019), this systematic suppression of small-scale signals can be problematic. We discuss possible approaches to address this in §5.

5 Summary and discussions

In this study, we have systematically evaluated the ability of a deep learning model—specifically a U-Net architecture—to extract IT signatures from different combinations of sea surface fields in an idealized Boussinesq simulation. We show that the U-Net’s performance depends critically on the physical information present in the inputs. A clear hierarchy of performance among the different input configurations is established, with the surface velocity field, $\mathbf{U}$, providing by far the most significant contribution to an accurate separation.

The relative success of each input field can be understood through the lens of the information they provide regarding the “wave signature” and the “scattering medium.”

- Surface velocity ($\mathbf{U}$) proves to be the most potent input. It contains a strong signature of the IT’s wave kinematics, as well as signatures from BM velocities that scatter ITs.
- Sea surface height (H) performs reasonably well. It contains the target wave signature by definition, but this signature is entangled with the much larger signal of the BM.
- Surface temperature (T) contains only a weak IT signature but provides valuable, unambiguous information on the scattering medium, as it is strongly related to

the BM that scatters the waves. Adding T to other inputs consistently improves performance by helping the network disambiguate the BM imprints from the total signal.

The full combination of SSH (H), velocity ($\mathbf{U}$), and temperature (T) yields the best performance, achieving a coefficient of determination (R_2) of 0.95 and a correlation (Υ) of 0.97 against the reference fields. Residual errors (the remaining 5% of unexplained variance) concentrate at small spatial scales. These residual errors reflect both genuine physical processes (e.g., nonlinear scattering from mode-1 to higher modes) and spurious signals in the reference data induced by the Eulerian frequency filtering used to compute the reference data (e.g., BM energy near tidal frequencies, Doppler-shifted non-tidal signals).

A central finding of this work is that surface velocity is the single most critical observable for IT-BM separation. Intriguingly, He and Mahadevan (2024), who focus on inferring interior vertical velocities, reach a similar conclusion regarding the importance of surface velocities. This apparent agreement on the critical role of surface velocities strongly advocates for the development and deployment of future satellite missions capable of measuring sea surface vector velocities, such as ODYSEA, SEASTAR, and Harmony.

Our results also highlight the immense value of coordinating multi-platform observational campaigns. For instance, aligning measurements from ship-based measurements, high-frequency radar, or surface drifters with SWOT and SST satellite tracks could achieve the synergistic benefits demonstrated in our $\{H, \mathbf{U}, T\}$ configuration.

Our investigations provide some broader lessons on neural networks:

- Architecture choices. Despite its relative simplicity, the U-Net performs on par with the more complex cGAN used in W22. The crucial element for the U-Net’s success is a learning rate that varies during training. The computational efficiency of U-Nets allows us to run many different experiments conveniently. Moreover, the architectural simplicity and the stable training behaviors of U-Nets allows us to understand more clearly what factors are conducive to its success. For other applications, a U-Net with a varying learning rate can be a good candidate for a baseline algorithm.

- Physical interpretability. The “wave signature vs. scattering medium” framework provides a clear lens for understanding why some inputs perform better than others. This perspective is transferable to other studies.
- Non-locality. The U-Net’s large receptive field ($800 \times 800 \text{ km}^2$) is key to its success in our application, consistent with the physical understanding that IT–BM interactions take place at large spatial scales. For other applications, the optimal receptive field depends on the problem at hand.

Careful readers may notice that under Configuration $\{H\}$, the performance metrics (Υ, R_2) from our U-Net are slightly stronger than those of the cGAN in W22; this difference arises primarily from small changes in data division/selection (§2.3 and Supporting Information Text S1). Qualitative differences between the U-Net and the cGAN do appear at small spatial scales (below the mode-1 tidal wavelengths): our U-Net tends to blur out the fine structures, whereas the cGAN in W22 tends to create random small-scale patterns that are deterministically incorrect, in an effort to mimic small-scale statistical behaviors. This behavior of the cGAN is consistent with the adversarial objective in its architectural design: if small-scale features are entirely absent from the generator’s output, it is easy for the discriminator to reject generated images, so the generator learns to add features. We leave a dedicated comparison of small-scale behaviors across architectures to future work, which should be part of the efforts to address the small-scale error reported in §4.

In general, we envision that the small-scale error may be mitigated under (a), a more careful data preparation, or (b), changes in the deep learning architecture. For (a), more accurate methods to separate the ITs and BMs can prevent the physical imprints of ITs from intermingling with spurious signals in the reference fields, which currently confuse our U-Net. Examples include time-filtering conducted in flow-following frames (Shakespeare et al., 2021; Kafiabad, 2022; Baker et al., 2025), and filtering methods that do not rely on the separation in time scales (Early et al., 2021; C. Wang et al., 2023). For (b), a dedicated comparison of small-scale behaviors between different architectures is required, and we speculate that in particular, attention mechanisms can help address the localized, velocity-associated errors observed in Figure 10, and mechanisms that penalize mismatches at small spatial scales generally, such as different loss functions, can lead the optimizers to be weighted more towards small scales.

We hypothesize that an additional, particular advantage of Configuration $\{U\}$ is that, in our data, ITs are the dominant source of divergent surface motion, while the BM is largely rotational. Internal waves with intrinsic frequencies above the Coriolis frequency project strongly onto divergence (e.g., Bühler et al. (2014)), whereas balanced, quasi-geostrophic flows (such as the turbulent jets in our simulations) that constitute the scattering background project weakly on the divergent component (e.g., Gill (2016)). The snapshot shown in Figure 3 demonstrates this visually: the divergence D is dominated by spatial scales around tidal wavelengths, with plane-wave-like patterns up-jet and scattered-wave-like patterns mid-/down-jet, consistent with the dominance of ITs, whereas the vorticity ζ visibly contains both imprints from ITs and large-scale structures contributed by BMs. Conceptually, this could provide the U-Net with a relatively clean pathway to distinguish (and hence use) the wave signature (in the divergence) from the imprints of scattering medium (largely in the vorticity). In some oceanic scenarios, unlike in our dataset, BMs can have strong imprints on divergent currents due to the presence of strong sub-mesoscale currents typical of the winter mixed layer (e.g., Barkan et al. (2019)). There, the disentanglement problem may be more complicated.

Out of theoretical curiosity, we also experiment with configurations where D and/or ζ are inputs. Supporting Information Text S5 reports details. Configuration $\{D\}$ outperforms any other configuration that has just one input channel, suggesting that the strong wave signatures in D is useful. Configuration $\{\zeta\}$ exhibits problematic behavior (i.e., counterintuitive responses to some training parameters we do not observe in other configurations); moreover, Configuration $\{\zeta, D\}$ performs slightly worse than Configuration $\{U\}$, even though conceptually, the separation between waves and BMs in (ζ, D) should be more straightforward than U . These counter-intuitive behaviors suggest that our U-Net is not optimal for inferring information from ζ and/or D . This may be because ζ and D are active at spatial scales smaller than U , and our U-Net may not focus sufficiently on extracting information from small scales.

In addition to new observational campaigns, another promising approach to obtain surface current velocities is through data-driven algorithms taking existing observed fields as inputs. For example, Zhou et al. (2025) infer surface velocities from SWOT’s SSH observations through deep learning; Fablet et al. (2024) infer the evolution of surface velocities from SSH and SST via a combination of deep learning and data assimilation, harnessing the dynamical links between the surface fields; and Lenain et al. (2025) infer sur-

face velocities from and surface heat flux observations, harnessing the kinematic relationship between velocities and that obey the advection-diffusion equation. Such approaches can be aided by gap-free data products constructed from along-track satellite observations (Martin et al., 2023), which provide more continuous inputs for training and application.

We choose to extract the IT imprints on SSH, rather than the BM imprints. In our datasets, the latter task would produce artificially high performance metrics because the input and output would be highly similar: the imprints from the BMs dominate the total SSH signal, taking up more than 90% of the total variance of SSH. In datasets where the BMs do not overwhelmingly dominate the signals (e.g., Gao et al. (2024)), it can still be meaningful to evaluate performance metrics for BM extraction.

From Figure 8 and other snapshots (not shown), the U-Net—under any Configuration we tested—appears to correctly recognize that ITs are more coherent in the up-jet region, as indicated by its generation of patterns resembling plane waves there. The same behavior is evident for the cGAN in W22. This suggests that neural networks may be used not only for extracting ITs but also for identifying regions of strong incoherence, which in turn could be exploited to quantify uncertainties in IT extractions. In this direction, neural networks that explicitly generate uncertainty metrics, such as Bayesian neural networks, appear especially promising.

Our work and its counterparts primarily concern new-generation satellite observations, which provide two-dimensional spatial coverage but suffer from poor temporal sampling. Mooring data represent the opposite case: they lack spatial coverage but provide higher temporal resolution, typically with sampling intervals on the order of an hour—well below tidal periods. Despite the superior temporal sampling, incoherence remains a major challenge for harmonic fitting of mooring data too, as phase shifts of ITs can still vary on timescales that the observations do not resolve. Recent works show that deep learning—particularly encoder-decoder architectures similar in spirit to our U-Net—has promise for addressing this challenge in time series data as well (Li et al., 2023).

While we provide a clear proof-of-concept using an idealized model, future work must progress toward more complex and realistic scenarios. The methodology should be tested on data from comprehensive general circulation models that include realistic bathymetry, a full spectrum of internal waves, and geographically varying stratification. The simpler,

computationally cheaper deep learning algorithm identified in this work contributes to efforts in this direction. A critical future step will be to improve the reference datasets used for training, potentially by employing more sophisticated separation techniques. Ultimately, the goal is to apply these trained models to actual satellite observations, a step that will require careful handling of instrument noise and data gaps. These explorations, guided by the findings herein, will be essential to fully leverage the wealth of new-era satellite data and make meaningful progress in understanding the ocean’s intricate multiscale dynamics.

Open Research Section

Production codes used for this paper are available on Zenodo (H. Wang, 2025). This includes scripts for the definitions and the training of our U-Net architecture, and for the analysis of outputs. Potential future updates of the codes can be accessed on GitHub https://github.com/hannnwang/Extract_internal_tides_with_UNet_and_surface_field_synergy. The data used for the training and testing of our U-Net is published on Scholars Portal Dataverse (Ponte et al., 2020).

Acknowledgments

J.U. and N.G. acknowledge the support of the Canadian Space Agency [14SUSWOTTO] and of the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2022-04560]. K.S. was supported by the Office of Naval Research (N00014-25-1-2183). This paper is a contribution to the projects W2, L2 and M2 of the Collaborative Research Centre TRR 181 “Energy Transfers in Atmosphere and Ocean” funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) - Projektnummer 274762653, which supports HW.

We thank Julien Le Sommer, Callum Shakespeare, Noé Lahaye and Belal Abdelhadi for helpful discussions. Simin Wang produced some results reported in the Supporting Information as part of her master thesis at University of Edinburgh. We thank Aurélien Ponte for helpful discussions, and for supplying simulation output data used in W22 and reused here.

References

Araujo, A., Norris, W., & Sim, J. (2019). Computing receptive fields of convolu-

- tional neural networks. *Distill.* (<https://distill.pub/2019/computing-receptive-fields>) doi: 10.23915/distill.00021
- Azad, R., Aghdam, E. K., Rauland, A., Jia, Y., Avval, A. H., Bozorgpour, A., ... Merhof, D. (2024). Medical image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
- Baker, L. E., Kafiabad, H. A., Maitland-Davies, C., & Vanneste, J. (2025). Lagrangian filtering for wave-mean flow decomposition. *Journal of Fluid Mechanics*, 1009, A40.
- Barkan, R., Molemaker, M. J., Srinivasan, K., McWilliams, J. C., & D’Asaro, E. A. (2019). The role of horizontal divergence in submesoscale frontogenesis. *Journal of Physical Oceanography*, 49(6), 1593–1618.
- Barkan, R., Srinivasan, K., & McWilliams, J. C. (2024). Eddy-internal wave interactions: stimulated cascades in cross-scale kinetic energy and enstrophy fluxes. *Journal of Physical Oceanography*, 54(6), 1309–1326.
- Bellemin-Laponnaz, V., Le Guillou, F., Clément, U., Eric, B., & Cosme, E. (2025). A variational method for reconstructing and separating balanced motions and internal tides from wide-swath altimetric sea surface height observations. *Authorea Preprints*.
- Brolly, M. T. (2025). Stochastic parameterization: The importance of nonlocality and memory. *Journal of Advances in Modeling Earth Systems*, 17(9), e2025MS005223.
- Bühler, O., Callies, J., & Ferrari, R. (2014). Wave-vortex decomposition of one-dimensional ship-track data. *Journal of Fluid Mechanics*, 756, 1007–1026.
- Buijsman, M. C., Arbic, B. K., Richman, J. G., Shriver, J. F., Wallcraft, A. J., & Zamudio, L. (2017). Semidiurnal internal tide incoherence in the equatorial Pacific. *Journal of Geophysical Research: Oceans*, 122(7), 5286–5305.
- Carrere, L., Arbic, B. K., Dushaw, B., Egbert, G., Erofeeva, S., Lyard, F., ... others (2021). Accuracy assessment of global internal-tide models using satellite altimetry. *Ocean Science*, 17(1), 147–180.
- Caspar-Cohen, Z., Ponte, A., Lahaye, N., Carton, X., Yu, X., & Gentil, S. L. (2022). Characterization of internal tide incoherence: Eulerian versus lagrangian perspectives. *Journal of Physical Oceanography*, 52(6), 1245–1259.
- Caspar-Cohen, Z., Ponte, A., Lahaye, N., Zaron, E. D., Arbic, B. K., Yu, X., ...

- Menemenlis, D. (2025). Combining surface drifters and high resolution global simulations enables the mapping of internal tide surface energy. *Scientific Reports*, 15(1), 10672.
- Dunphy, M., & Lamb, K. G. (2014, jan). Focusing and vertical mode scattering of the first mode internal tide by mesoscale eddy interaction. *Journal of Geophysical Research: Oceans*, 119(1), 523–536. Retrieved from <http://doi.wiley.com/10.1002/2013JC009293> doi: 10.1002/2013JC009293
- Dunphy, M., Ponte, A. L., Klein, P., & Le Gentil, S. (2017). Low-Mode Internal Tide Propagation in a Turbulent Eddy Field. *Journal of Physical Oceanography*, 47(3), 649–665. Retrieved from <http://journals.ametsoc.org/doi/10.1175/JPO-D-16-0099.1> doi: 10.1175/jpo-d-16-0099.1
- Early, J. J., Lelong, M. P., & Sundermeyer, M. A. (2021, apr). A generalized wave-vortex decomposition for rotating Boussinesq flows with arbitrary stratification. *Journal of Fluid Mechanics*, 912, A32. Retrieved from https://www.cambridge.org/core/product/identifier/S0022112020009957/type/journal_article doi: 10.1017/jfm.2020.995
- Fablet, R., Chapron, B., Le Sommer, J., & Sévellec, F. (2024). Inversion of sea surface currents from satellite-derived sst-ssh synergies with 4dvarnets. *Journal of Advances in Modeling Earth Systems*, 16(6), e2023MS003609.
- Farrar, J. T., D’Asaro, E., Rodriguez, E., Shcherbina, A., Czech, E., Matthias, P., ... others (2020). S-mode: The sub-mesoscale ocean dynamics experiment. In *Igarss 2020-2020 ieee international geoscience and remote sensing symposium* (pp. 3533–3536).
- Farrar, J. T., Zappa, C. J., Weller, R. A., & Jessup, A. T. (2007). Sea surface temperature signatures of oceanic internal waves in low winds. *Journal of Geophysical Research: Oceans*, 112(C6).
- Ferrari, R., & Wunsch, C. (2009, jan). Ocean circulation kinetic energy: Reservoirs, sources, and sinks. *Annual Review of Fluid Mechanics*, 41, 253–282. Retrieved from <http://www.annualreviews.org/doi/abs/10.1146/annurev.fluid.40.111406.102139> doi: 10.1146/annurev.fluid.40.111406.102139
- Gao, Z., Chapron, B., Ma, C., Fablet, R., Febvre, Q., Zhao, W., & Chen, G. (2024, apr). A Deep Learning Approach to Extract Balanced Motions From Sea Surface Height Snapshot. *Geophysical Research Letters*, 51(7). Retrieved from

- <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL106623>
doi: 10.1029/2023GL106623
- Garrett, C., & Munk, W. (1972). Oceanic mixing by breaking internal waves. In *Deep sea research and oceanographic abstracts* (Vol. 19, pp. 823–832).
- Gerkema, T., & Zimmerman, J. (2008). An introduction to internal waves. *Lecture Notes, Royal NIOZ, Texel, 207*, 207.
- Gill, A. E. (2016). *Atmosphere—ocean dynamics*. Elsevier.
- Hauser, D., Abdalla, S., Ardhuin, F., Bidlot, J.-R., Bourassa, M., Cotton, D., ... others (2023). Satellite remote sensing of surface winds, waves, and currents: Where are we now? *Surveys in Geophysics*, 44(5), 1357–1446.
- He, J., & Mahadevan, A. (2024, mar). Vertical Velocity Diagnosed From Surface Data With Machine Learning. *Geophysical Research Letters*, 51(6). Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2023GL104835> doi: 10.1029/2023GL104835
- Iqbal, H. (2018, December). *Harisigbal88/plotneuralnet v1.0.0*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.2526396> doi: 10.5281/zenodo.2526396
- Jones, C. S., Xiao, Q., Abernathey, R. P., & Smith, K. S. (2023). Using lagrangian filtering to remove waves from the ocean surface velocity field. *Journal of Advances in Modeling Earth Systems*, 15(4), e2022MS003220.
- Kachelein, L., Gille, S. T., Mazloff, M. R., & Cornuelle, B. D. (2024). Characterizing non-phase-locked tidal currents in the california current system using high-frequency radar. *Journal of Geophysical Research: Oceans*, 129(7), e2023JC020340.
- Kafiabad, H. A. (2022, jun). Grid-based calculation of the Lagrangian mean. *Journal of Fluid Mechanics*, 940, A21. Retrieved from https://www.cambridge.org/core/product/identifier/S0022112022002336/type/journal_article doi: 10.1017/jfm.2022.233
- Kafiabad, H. A., Savva, M. A., & Vanneste, J. (2019, jun). Diffusion of inertia-gravity waves by geostrophic turbulence. *Journal of Fluid Mechanics*, 869, R7. Retrieved from https://www.cambridge.org/core/product/identifier/S0022112019003008/type/journal_article doi: 10.1017/jfm.2019.300
- Kantha, L. H., & Tierney, C. C. (1997). Global baroclinic tides. *Progress in*

- Oceanography*, 40(1-4), 163–178.
- Kar, S., Barkan, R., McWilliams, J. C., & Molemaker, M. J. (2025). Spontaneous emission of internal waves by a radiative instability. *Journal of Physical Oceanography*, 55(5), 525–542.
- Kelly, S. M., Lermusiaux, P. F., Duda, T. F., & Haley Jr, P. J. (2016). A coupled-mode shallow-water model for tidal analysis: Internal tide reflection and refraction by the gulf stream. *Journal of Physical Oceanography*, 46(12), 3661–3679.
- Lahaye, N., Ponte, A., Le Sommer, J., & Albert, A. (2024). Internal tide surface signature and incoherence in the north atlantic. *Geophysical Research Letters*, 51(12), e2024GL108508.
- Lapeyre, G., & Klein, P. (2006). Dynamics of the upper oceanic layers in terms of surface quasigeostrophy theory. *Journal of physical oceanography*, 36(2), 165–176.
- Lapo, K., Ichinaga, S. M., & Kutz, J. N. (2025). A method for unsupervised learning of coherent spatiotemporal patterns in multiscale data. *Proceedings of the National Academy of Sciences*, 122(7), e2415786122.
- Larrañaga, M., Renault, L., Wineteer, A., Contreras, M., Arbic, B. K., Bourassa, M. A., & Rodriguez, E. (2025). Assessing the future odysea satellite mission for the estimation of ocean surface currents, wind stress, energy fluxes, and the mechanical coupling between the ocean and the atmosphere. *Remote Sensing*, 17(2), 302.
- Le Guillou, F., Lahaye, N., Ubelmann, C., Metref, S., Cosme, E., Ponte, A., ... Viard, A. (2021). Joint Estimation of Balanced Motions and Internal Tides From Future Wide-Swath Altimetry. *Journal of Advances in Modeling Earth Systems*, 13(12), 1–17. doi: 10.1029/2021MS002613
- Lenain, L., Srinivasan, K., Barkan, R., & Pizzo, N. (2025). An unprecedented view of ocean currents from geostationary satellites. *PREPRINT (Version 1) available at Research Square*. doi: 10.21203/rs.3.rs-7055642/v1
- Li, B., Wang, Y., Wei, Z., Pan, H., Xu, T., & Lv, X. (2023). Changing the unpredictable nature of internal tides through deep learning. *Geophysical Research Letters*, 50(8), e2022GL102227.
- Liu, Z. (2022). Super convergence cosine annealing with warm-up learning rate. In *Caibda 2022; 2nd international conference on artificial intelligence, big data*

- and algorithms (pp. 1–7).
- Löb, J., Köhler, J., Mertens, C., Walter, M., Li, Z., von Storch, J.-S., ... Rhein, M. (2020). Observations of the low-mode internal tide and its interaction with mesoscale flow south of the azores. *Journal of Geophysical Research: Oceans*, 125(11), e2019JC015879.
- Loshchilov, I., & Hutter, F. (2016). Sgdr: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*.
- Maingonnat, I., Tissot, G., & Lahaye, N. (2025). Coupled estimation of internal tides and turbulent motions via statistical modal decomposition. *Ocean Science*, 21(2), 807–827.
- Martin, S. A., Manucharyan, G. E., & Klein, P. (2023). Synthesizing Sea Surface Temperature and Satellite Altimetry Observations Using Deep Learning Improves the Accuracy and Resolution of Gridded Sea Surface Height Anomalies. *Journal of Advances in Modeling Earth Systems*, 15(5), 1–26. doi: 10.1029/2022MS003589
- McCann, D. L., Martin, A. C., de Macedo, K. A., Carrasco Alvarez, R., Horstmann, J., Marié, L., ... others (2024). A new airborne system for simultaneous high-resolution ocean vector current and wind mapping: first demonstration of the seastar mission concept in the macrotidal iroise sea. *Ocean Science*, 20(5), 1109–1122.
- McWilliams, J. C. (2016, may). Submesoscale currents in the ocean. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 472(2189), 20160117. Retrieved from <https://royalsocietypublishing.org/doi/10.1098/rspa.2016.0117> doi: 10.1098/rspa.2016.0117
- Munk, W. H., & Cartwright, D. E. (1966). Tidal spectroscopy and prediction. *Philosophical Transactions of the Royal Society of London. Series A, Mathematical and Physical Sciences*, 259(1105), 533–581.
- Nielsen-Englyst, P., Høyer, J. L., Karagali, I., Kolbe, W. M., Tonboe, R. T., & Pedersen, L. T. (2024). Impact of microwave observations on the estimation of arctic sea surface temperatures. *Remote Sensing of Environment*, 301, 113949.
- Polzin, K., Toole, J., Ledwell, J., & Schmitt, R. (1997). Spatial variability of turbulent mixing in the abyssal ocean. *Science*, 276(5309), 93–96.

- Ponte, A. L., & Klein, P. (2015, mar). Incoherent signature of internal tides on sea level in idealized numerical simulations. *Geophysical Research Letters*, 42(5), 1520–1526. Retrieved from <http://doi.wiley.com/10.1002/2014GL062583> doi: 10.1002/2014GL062583
- Ponte, A. L., Klein, P., Dunphy, M., & Le Gentil, S. (2017, mar). Low-mode internal tides and balanced dynamics disentanglement in altimetric observations: Synergy with surface density observations. *Journal of Geophysical Research: Oceans*, 122(3), 2143–2155. Retrieved from <http://doi.wiley.com/10.1002/2016JC012214> doi: 10.1002/2016JC012214
- Ponte, A. L., Le Gentil, S., & Grisouard, N. (2020). Internal tides and geostrophic turbulence in a Boussinesq re-entrant channel. *Borealis*. Retrieved from <https://doi.org/10.5683/SP2/HU58SG> doi: 10.5683/SP2/HU58SG
- Rainville, L., & Pinkel, R. (2006). Propagation of low-mode internal waves through the ocean. *Journal of Physical Oceanography*, 36(6), 1220–1236.
- Ray, R. D., & Zaron, E. D. (2011). Non-stationary internal tides observed with satellite altimetry. *Geophysical Research Letters*, 38(17).
- Rodriguez, E. (2018). On the optimal design of doppler scatterometers. *Remote Sensing*, 10(11), 1765.
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In *Medical image computing and computer-assisted intervention—miccai 2015: 18th international conference, munich, germany, october 5-9, 2015, proceedings, part iii 18* (pp. 234–241).
- Savva, M. A., Kafiabad, H. A., & Vanneste, J. (2021, jun). Inertia-gravity-wave scattering by three-dimensional geostrophic turbulence. *Journal of Fluid Mechanics*, 916, A6. Retrieved from https://www.cambridge.org/core/product/identifier/S0022112021002056/type/journal_article doi: 10.1017/jfm.2021.205
- Shakespeare, C. J., Gibson, A. H., Hogg, A. M., Bachman, S. D., Keating, S. R., & Velzeboer, N. (2021, oct). A New Open Source Implementation of Lagrangian Filtering: A Method to Identify Internal Waves in High-Resolution Simulations. *Journal of Advances in Modeling Earth Systems*, 13(10). Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2021MS002616> doi: 10.1029/2021MS002616

- Shchepetkin, A. F., & McWilliams, J. C. (2005, jan). The regional oceanic modeling system (ROMS): A split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean Modelling*, 9(4), 347–404. Retrieved from <https://linkinghub.elsevier.com/retrieve/pii/S1463500304000484> doi: 10.1016/j.ocemod.2004.08.002
- Smith, L. N. (2017). Cyclical learning rates for training neural networks. In *2017 IEEE winter conference on applications of computer vision (wacv)* (pp. 464–472).
- Srinivasan, K., Chekroun, M. D., & McWilliams, J. C. (2024). Turbulence closure with small, local neural networks: Forced two-dimensional and β -plane flows. *Journal of Advances in Modeling Earth Systems*, 16(4), e2023MS003795.
- Terrill, E., Otero, M., Hazard, L., Conlee, D., Harlan, J., Kohut, J., . . . Lindquist, K. (2006). Data management and real-time distribution in the hf-radar national network. In *Oceans 2006* (pp. 1–6).
- Theodosiou, A., Kleinherenbrink, M., & López-Dekker, P. (2021). Wide-swath ocean topography using formation flying under squinted geometries: The harmony mission case. In *2021 IEEE International Geoscience and Remote Sensing Symposium IGARSS* (p. 2134–2137). doi: 10.1109/IGARSS47720.2021.9554076
- Torres, H. S., Klein, P., Menemenlis, D., Qiu, B., Su, Z., Wang, J., . . . Fu, L. L. (2018). Partitioning Ocean Motions Into Balanced Motions and Internal Gravity Waves: A Modeling Study in Anticipation of Future Space Missions. *Journal of Geophysical Research: Oceans*, 123(11), 8084–8105. doi: 10.1029/2018JC014438
- Uchida, T., Yadidya, B., Lapo, K. E., Xu, X., Early, J. J., Arbic, B. K., . . . others (2025). Dynamic mode decomposition of geostrophically balanced motions from swot cal/val in the separated gulf stream. *Earth and Space Science*, 12(8), e2024EA004079.
- Uncu, J., & Grisouard, N. (2024). Wave scattering by an isolated cyclogeostrophic vortex. *Journal of Fluid Mechanics*, 999, A100.
- Vanneste, J. (2013). Balance and spontaneous wave generation in geophysical flows. *Annual Review of Fluid Mechanics*, 45(1), 147–172.
- Vic, C., Naveira Garabato, A. C., Green, J. M., Waterhouse, A. F., Zhao, Z., Melet, A., . . . Stephenson, G. R. (2019). Deep-ocean mixing driven by small-scale

- internal tides. *Nature communications*, 10(1), 2099.
- Wagner, G. L., Ferrando, G., & Young, W. R. (2017, oct). An asymptotic model for the propagation of oceanic internal tides through quasi-geostrophic flow. *Journal of Fluid Mechanics*, 828, 779–811. Retrieved from https://www.cambridge.org/core/product/identifier/S0022112017005092/type/journal_article doi: 10.1017/jfm.2017.509
- Wang, C., Liu, Z., & Lin, H. (2023). On dynamical decomposition of multiscale oceanic motions. *Journal of Advances in Modeling Earth Systems*, 15(3), e2022MS003556.
- Wang, C., Liu, Z., Lin, H., Chen, D., Yang, Q., & Ni, Q. (2025). A practical separation of oceanic vortical and wavy motions entangled in the swot measurements. *Geophysical Research Letters*, 52(12), e2024GL113995.
- Wang, H. (2025, November). *hannnwang/extract-internal-tides-with-unet-and-surface-field-synergy: For zenodo*. Zenodo. Retrieved from <https://doi.org/10.5281/zenodo.17525907> doi: 10.5281/zenodo.17525907
- Wang, H., Grisouard, N., Salehipour, H., Nuz, A., Poon, M., & Ponte, A. L. (2022a). A Deep Learning Approach to Extract Internal Tides Scattered by Geostrophic Turbulence. *Geophysical Research Letters*, 49(11). Retrieved from <https://doi.org/10.1002/essoar.10508849.1> doi: 10.1029/2022GL099400
- Wang, H., Grisouard, N., Salehipour, H., Nuz, A., Poon, M., & Ponte, A. L. (2022b, jun). A Deep Learning Approach to Extract Internal Tides Scattered by Geostrophic Turbulence. *Geophysical Research Letters*, 49(11). Retrieved from <https://agupubs.onlinelibrary.wiley.com/doi/10.1029/2022GL099400> doi: 10.1029/2022GL099400
- Wang, Y., & Legg, S. (2025). Agulhas rings locally enhance dissipation of internal tides. *Science Advances*, 11(20), eadq5963.
- Whalen, C. B., De Lavergne, C., Naveira Garabato, A. C., Klymak, J. M., MacKinnon, J. A., & Sheen, K. L. (2020). Internal wave-driven mixing: Governing processes and consequences for climate. *Nature Reviews Earth & Environment*, 1(11), 606–621.
- Xiao, Q., Balwada, D., Jones, C. S., Herrero-González, M., Smith, K. S., & Abernathy, R. (2023). Reconstruction of surface kinematics from sea surface

- height using neural networks. *Journal of Advances in Modeling Earth Systems*, 15(10), e2023MS003709.
- Zhao, Z. (2016). Internal tide oceanic tomography. *Geophysical Research Letters*, 43(17), 9157–9164.
- Zhao, Z. (2024). Internal tides from swot: A 75-day instantaneous mode-1 m 2 m_2 internal tide model. *Journal of Geophysical Research: Oceans*, 129(12), e2024JC021174.
- Zhao, Z., Alford, M. H., & Girton, J. B. (2012). Mapping low-mode internal tides from multisatellite altimetry. *Oceanography*, 25(2), 42–51.
- Zhou, S., Dong, J., Li, H., Xu, G., & Xu, F. (2025). A machine learning-based model infers the sea surface velocity of surface water and ocean topography (swot). *Geophysical Research Letters*, 52(9), e2024GL110731.