A deep learning approach to extract internal tides scattered by geostrophic turbulence

Han Wang[1], Nicolas Grisouard[1], Hesam Salehipour[2], Alice Nuz[1*], Michael Poon[1†], and Aurélien L. Ponte[3]

[1]Department of Physics, University of Toronto, Ontario, Canada
[2]Autodesk Research, Ontario, Canada
[3]Ifremer, Plouzané, France

Key Points:

- A deep conditional Generative Adversarial Network is trained to extract tidal components in SSH snapshots from an idealized model.
- The network can extract tidal signals accurately in a snapshot whose underlying dynamics are different from training data.
- Performance of the network degrades when extracting tidal signals entangled with higher turbulence energies.

*Current address, Tandon School of Engineering, New York University, New York, USA
†Current address, Department of Astronomy and Astrophysics, University of Toronto, Ontario, Canada

Corresponding author: Han Wang, hannnwangus@gmail.com

Abstract

A proper extraction of internal tidal signals is central to the interpretation of Sea Surface Height (SSH) data. The increased spatial resolution of future wide-swath satellite missions poses a challenge for traditional harmonic analysis, due to prominent and unsteady wave-mean interactions at finer scales. However, the wide swaths will also produce SSH snapshots that are spatially two-dimensional, which allows us to treat the tidal extraction as an image translation problem. We design and train a conditional Generative Adversarial Network, which, given a snapshot of raw SSH from an idealized numerical eddying simulation, generates a snapshot of the embedded tidal component. We test it on synthetic data whose dynamical regimes are different from the data provided during training. Despite the diversity and complexity of data, it accurately extracts tidal components in most individual snapshots considered and reproduces physically meaningful statistical properties. Predictably, TITE's performance decreases with the intensity of the turbulent flow.

Plain Language Summary

Wide-swath satellite observations of Sea Surface Height (SSH) data at high spatial resolutions will be available in abundance thanks to advances of instrumental technologies. Embedded in the observed SSH are internal tides, a dynamical component that plays a crucial role in ocean circulation. As they are entangled with background currents and eddies, such tidal signals are challenging to extract. Methods that worked with previous-generation altimeters will break down at the resolutions that the new generation promises. On the other hand, the wide satellite swaths provide new opportunities as they allow us to regard the observations as spatially two-dimensional. Here we treat the tidal extraction solely as an image translation problem. We train a deep neural net so that given a snapshot of a raw SSH signal, it produces a "fake" snapshot of the tidal SSH signal that is meant to reproduce the original. The data we use in this article is generated by idealized numerical simulations. Once adapted to realistic data, the network has the potential to become a new tidal extraction tool for satellite observations. More broadly, successes in our experiments can inspire other applications of generative networks to disentangle dynamical components in data where classical analysis may fail.

1 Introduction

Since the launch of TOPEX/Poseidon, oceanographers have used the geostrophic assumption to infer sea surface velocity from SSH. However, while an estimated 90% of the ocean's kinetic energy exists in the form of currents in quasigeostrophic balance (hereafter qualified as "balanced"; see Ferrari & Wunsch, 2009), one still must account for "unbalanced" flows such as internal tides, hereafter "ITs", for a refined inference of balanced currents (Fu & Ferrari, 2008). Furthermore, ITs play a crucial role in ocean mixing (Lien & Gregg, 2001; Whalen et al., 2020), and are helpful in detecting ocean temperature changes (Zhao, 2016). Therefore, whether ITs are considered "noise" (e.g., for inferring balanced flows) or "signal" (e.g., for inferring tidally induced mixing), their proper extraction from altimetry data is essential.

For decades, IT extraction has been conducted via harmonic analysis (Munk & Hasselmann, 1964), a method that relies on a close phase relationship (or coherence) between ITs and astronomical forcings. Departures from this condition are sometimes referred to as "incoherence" (Ponte & Klein, 2015; Zaron & Rocha, 2018). Current altimetry has a typical spatial resolution of $O(100)$ km (Ballarotta et al., 2019), which is sufficient to retrieve mode-1 and some of the mode-2 IT wavelengths of semidiurnal tides, along with the dominant turbulent balanced motions (hereafter "TBMs"; see Ray & Zaron, 2011). At these scales, the coupling between ITs and TBMs is usually weak and therefore sub-

stantial portions of the ITs are coherent (Egbert & Ray, 2000). Hence, harmonic analysis is in principle sufficient to retrieve the corresponding IT signal.

The next generation of satellite altimetry, in particular the Surface Water Ocean Topography (SWOT) satellite mission, aims to improve the spatial resolutions of the measurements to at least a few tens of km in wavelength (Morrow et al., 2019). A fundamental challenge arises at these smaller scales, namely, the potential inapplicability of traditional harmonic analysis. Indeed, ITs become incoherent (Dunphy et al., 2017; Ponte & Klein, 2015; Dunphy & Lamb, 2014) due to stronger couplings with the TBMs linked to the increased vorticity magnitude (Bühler, 2014). In the presence of strong incoherence, other separation techniques such as complex demodulation (Munk et al., 1965; Colosi & Munk, 2006) or Lagangarian filtering (Shakespeare & Hogg, 2017), while efficient when well-sampled time series are available, will not be readily applicable to SWOT, whose sampling period is 21 days during the science phase (Morrow et al., 2019).

Future altimeters will gather data along wide swaths (two 50 km-wide swaths, 20 km apart for SWOT) as opposed to current linear tracks and as a result they will produce spatially two-dimensional (2D) images. This has motivated the community to regard the extraction of IT signals as an operation on high-resolution 2D snapshots. Current methods rely on exploiting distinct spectral signatures of TBMs and internal waves (Torres et al., 2019), empirical modal mapping methods (Egbert & Erofeeva, 2021), dynamical relations to surface density fields (Ponte et al., 2017), or on data assimilation techniques (Metref et al., 2020; Le Guillou et al., 2021). A common element between these works and the present one is the circumvention of time series information, as motivated by the poor temporal resolution of SWOT.

In this work, we propose to regard the IT extraction solely as an image-to-image translation problem, conceiving and tackling the following challenge: *can we design an algorithm that extracts the SSH signature induced by IT from a raw, instantaneous SSH map*? To address it, we develop what we call the "Toronto Internal Tide Emulator" (TITE), a deep convolutional neural network that extracts IT signals from individual SSH snapshots. No physical knowledge, statistical properties, or temporal evolution are imparted to the network prior to the training. In general, we find TITE to perform well in most SSH snapshots generated from a set of idealized simulations. We present details about the dataset we use and the development of TITE in section 2, our experiments in section 3, and offer conclusions and discussions in section 4.

2 Methods

2.1 Idealized data supporting TITE's development

Data to support TITE's development are snapshots from a set of idealized numerical simulations, where mode-1 ITs are forced at a fixed tidal period T (12 hours) to propagate through TBMs created by a baroclinically unstable jet (Ponte & Klein, 2015; Ponte et al., 2020). The SSH signatures of TBMs in these simulations are generally larger than those induced by ITs, and exhibit a significant overlap in spatial scales at $O(100)$ km with ITs. Spatial filtering is thus difficult, an issue that is also faced by satellite altimetry in oceanic regions such as the Gulf Stream or Drake Passage, where powerful TBMs exist (Rocha et al., 2016; Richman et al., 2012).

We run the model under five different initial meridional density contrasts. With increasing contrast, the baroclinic jet becomes more unstable and creates a more vigorous baroclinic eddy field. The spectra induced by these eddies follow a geostrophic turbulence law (Ponte & Klein, 2015; Charney, 1971), and are thus identified as TBMs. In ascending order of stationary surface kinetic energy levels of TBM (hereafter referred to as "turbulence levels"), we label the five simulations as T1 to T5. The mean normalized vorticities (absolute values of surface vertical vorticities normalized by the local Cori-

olis frequency) over the jet width increase from 0.06 to 0.14 from T1 to T5. See Text S1 in the Supporting Information (SI) for more details on the numerical setup. IT snapshots are computed online via harmonic fits over time series that are $2T$ long and sampled every 300 seconds, or $T/144$. For simplicity, we only study $\eta_{\cos}^{(\mathrm{sim})}$, the cosine component of ITs from the simulations, defined in the same way as in Dunphy et al. (2017):

$$\eta_{\cos}^{(\mathrm{sim})}(x,y,t) = \frac{1}{T}\int_{t-2T}^{t} \eta(x,y,t')\cos\left(\frac{2\pi}{T}t'\right)\mathrm{d}t', \tag{1}$$

where x, y are the zonal and meridional coordinates, respectively, and η denotes raw SSH. For each snapshot, we cut out three square panels covering three fixed latitudinal bands, labeled as "down-jet", "mid-jet" and "up-jet" bands, as illustrated in Fig. 1. Here, the term "up-jet" denotes locations closer to the tidal forcing regions. One hundred snapshots are captured every $4T$ for each simulation in T1-5, resulting in 1500 pairs of $\left\{\eta,\ \eta_{\cos}^{(\mathrm{sim})}\right\}$ panels (5 runs, 3 latitudinal bands, and 100 snapshots) altogether.

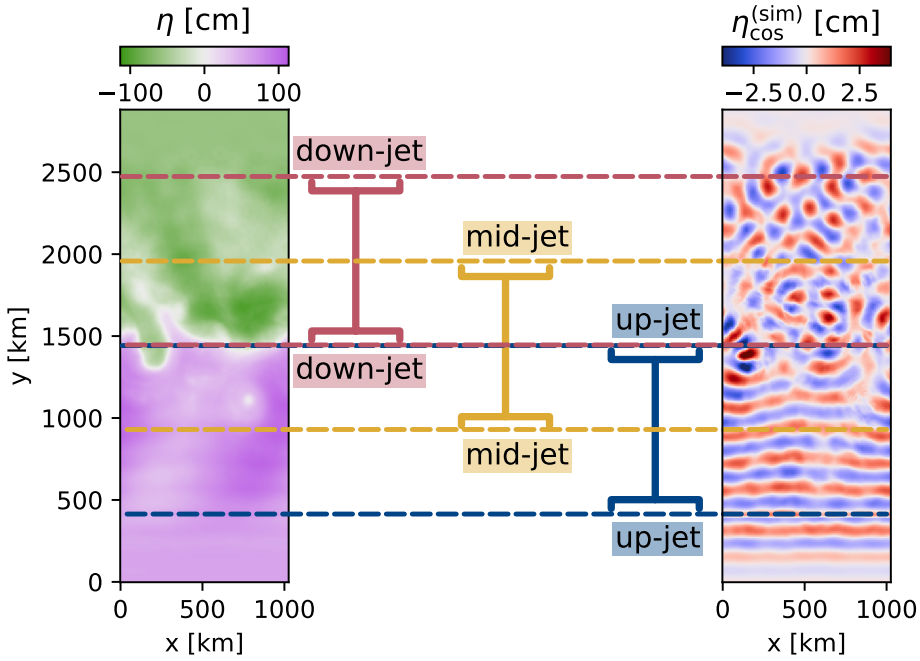

Figure 1. The "down-jet", "mid-jet" and "up-jet" bands illustrated over a snapshot of η (left) and $\eta_{\cos}^{(\mathrm{sim})}$ (right), sampled from T3 at day 2120. The "mid-jet" band is centred around the baroclinic jet. ITs are forced to the south of "up-jet" bands, and as the ITs propagates northward and loses coherence due to interactions with the TBM, the $\eta_{\cos}^{(\mathrm{sim})}$ patterns are less reminiscent of plane waves in the "down-jet" band than in the "up-jet" band.

2.2 Deep-learning algorithm designed to extract tidal signals

During the design of the TITE runs, we implicitly make four assumptions: (1) there is abundant spatial information, (2) all snapshots are statistically independent from each other, (3) a raw SSH functionally determines its IT component, but properties of the functional dependence are unknown, and (4) there exists abundant data where ITs are already extracted from the raw SSH. We discuss these assumptions at the end of this article.

TITE is based on a popular conditional Generative Adversarial Network (hereafter referred to as "cGAN"). As the name implies, a cGAN consists of two parts, namely, a conditional generator (hereafter "generator") that learns how to manufacture a "fake" image that's conditioned on an "input image", and a discriminator that tries to determine if an image is "genuine" (i.e., paired to the input image in the training data), or fake (i.e., created by the generator). Either part is on its own a convolutional neural network, and during training, the two parts compete against each other to co-evolve (Mirza & Osindero, 2014; Goodfellow et al., 2014). We denote the cosine IT panels generated from TITE as $\eta_{\cos}^{(\mathrm{gen})}$; following our notations, the *input* image would be η, the *genuine* image would be $\eta_{\cos}^{(\mathrm{sim})}$, and the *fake* image would be $\eta_{\cos}^{(\mathrm{gen})}$. As reflected in this general workflow, during training, other than the paired panels, no further information is given to TITE.

The particular cGAN TITE is derived from is called "pix2pix" (Isola et al., 2017), applications of which range from artistic creations (ml4a, 2017) to scientific problems such as remote sensing image classifications (Lebedev et al., 2018). Our codes are adapted from a TensorFlow Tutorial (Tensorflow, n.d.). We refer to Text S7 in the SI for details on the changes we made to the original code. Here, we mention a few relevant traits.

The generator and the discriminator have around 10^4 and 2000 convolutional layers respectively, each layer containing a 2-by-2 kernel to be learned during training. The considerable number of model parameters makes TITE a black box, as in the case of many deep learning algorithms.

Prior to each epoch, training images are randomly reshuffled in time, cropped, flipped, and rotated. Here, an epoch means the number of computations it takes for the cGAN to iterate over all data in the training set once. The random cropping, rotation and flipping are intended to introduce challenges that roughly mimic realistic situations where we don't have a priori knowledge of the observer's orientation/location about IT generation sites, direction of propagation, and exact boundary conditions. By randomly reshuffling in time, we enforce that every panel pair at every snapshot in the simulation be sequentially independent from the others. This means that any temporal information in the simulations is unknown to the pix2pix kernel, in line with our assumption (2) made previously in this section.

As the fully convolutional U-Net structure inherited from pix2pix in the generator can be applied to images of arbitrary sizes in principle, when producing Movies S1-5 in the SI, we directly apply the trained TITE onto rectangular input images, even though TITE is trained on square images illustrated in Fig. 1. This versatility on the shapes of input images would be useful for along-swath satellite products.

We systematically run our code with TensorFlow 2.3.0 under Python 3.7. One hundred training epochs with 960 pairs of $\left\{\eta, \eta_{\cos}^{(\mathrm{sim})}\right\}$ in the training set take about 1.5 hours with a NVIDIA GP100 GPU. For all the TITE runs in the article, we choose to present the results after 600 training epochs. Details on how we decide on the cut-off epoch are provided in Text S4 in the SI.

2.3 Division of data to training, testing and validation sets

As a first check on whether TITE's performancecould achieve any success at all, we randomly select 20% of all 1500 pairs of $\left\{\eta, \eta_{\cos}^{(\mathrm{sim})}\right\}$ panels from T1-5 to form a so-called validation set, and use the rest as the training set. During training, TITE has access to all pairs of $\left\{\eta, \eta_{\cos}^{(\mathrm{sim})}\right\}$ in the training set, but none from the validation set. After 600 epochs, the training phase is over, and we apply the trained TITE into snapshots in the validation set. The mean correlation between $\eta_{\cos}^{(\mathrm{sim})}$ and $\eta_{\cos}^{(\mathrm{gen})}$ in the validation set turns out to be 0.85, which suggests that the generated $\eta_{\cos}^{(\mathrm{gen})}$ reasonably resemble

the ground truths $\eta_{\cos}^{(\mathrm{sim})}$. However, under this division, the training set contains turbulence levels that are statistically similar to the validation set on which the trained TITE is applied, and the good correlation factors could be caused by *overfitting*. Here,"overfitting" refers to a commonly accepted definition in machine-learning contexts (Dietterich, 1995): a model is said to overfit when it tries to fit the training data so closely that it does not generalize well to new data. To address this possibility, we challenge TITE to extract $\eta_{\cos}^{(\mathrm{sim})}$ signals linked to a *different* turbulence level as those employed for its training.

Specifically, in what we refer to as the "ET1 run", we reserve a *test* set, which contains all 300 pairs of panels from the simulation T1 and *none* from T2, T3, T4 or T5. Among the remaining panels from T2-5, we randomly select 80% pairs for the training set, and reserve the other 20% for the validation set. The validation and test sets are both inaccessible to TITE during training, but crucially, in terms of average turbulence levels, the training set is similar to the validation set, yet *different* from the test set. This procedure is designed to limit, or at least detect, the occurrence of overfitting. Similarly, we carry out ET2-5 runs, following the same logic, where the test sets are panels from the simulations T2-5 respectively.

3 Performance of TITE

As a graphical illustration, in Movie S1 in the SI, we re-order all the shuffled test instances of ET1 in time. The reconstructed temporal information appears remarkable, considering that the snapshots were randomly shuffled at the time of their generation and hence that the temporal evolution of these images was unknown to TITE. However, as turbulence level in test sets increases from ET2 to ET5, the evolution of $\eta_{\cos}^{(\mathrm{gen})}$ bears less and less resemblance with $\eta_{\cos}^{(\mathrm{sim})}$ (Movies S2-5 in the SI). To find the underlying causes, in this section, we evaluate the performance of TITE with several statistical metrics and discuss the causes of relatively decreased performance when they arise. All metrics are computed using standard methods and detailed in Text S6 in the SI .

We first investigate how close $\eta_{\cos}^{(\mathrm{gen})}$ is to the ground truth $\eta_{\cos}^{(\mathrm{sim})}$ by measuring the correlation between the two, as in Torres et al. (2018). The mean correlation factors averaged over all test instances in the ET1-5 runs are 0.91, 0.89, 0.83, 0.80 and 0.70, respectively. The highly correlated predictions of TITE in the test set in ET1-4 are especially interesting, as turbulence levels of the test set are different from that of the training set. There is however a relatively sharper drop (from 0.80 to 0.70) in the mean correlation from ET4 to ET5 in the test sets. Such a sharp drop can also be observed in the histograms shown in the SI Fig. S7.

A few relevant snapshots from test sets are illustrated in Fig. 2. The first row presents the snapshot with the highest correlation among all test sets in ET1-5. It belongs to ET1 and has a correlation factor of 0.95. The second and third row present the snapshot with the lowest correlation among all test sets in ET5 and ET4, respectively. Judging by the correlations, the worst case in ET4 has a correlation factor of 0.68, which is a significant improvement over the correlation factor of 0.49 in the worst case in ET5. This agrees with visual comparisons between the second and third rows in Fig. 2. This observation, together with the sharp drop of mean correlation factors in ET5 noted above, suggest a possible categoric difference between ET5 and ET1-4.

To gain more insight about the relative failures in ET5, we conduct a spectral analysis that focuses on comparing ET4 and ET5. The wavenumber spectra for the down-jet and up-jet bands are computed separately for $\eta_{\cos}^{(\mathrm{sim})}$ and $\eta_{\cos}^{(\mathrm{gen})}$ in the test set of ET4 and ET5, and presented in Fig. 3. The spectra for mid-jet bands are omitted for readability here and attached in Text S2 in the SI .

Prominent bumps appear near the wavenumbers corresponding to mode-1 tidal wavelengths in all the spectra of $\eta_{\cos}^{(\mathrm{sim})}$ (Solid lines in Fig. 3). These bumps are somewhat broad,

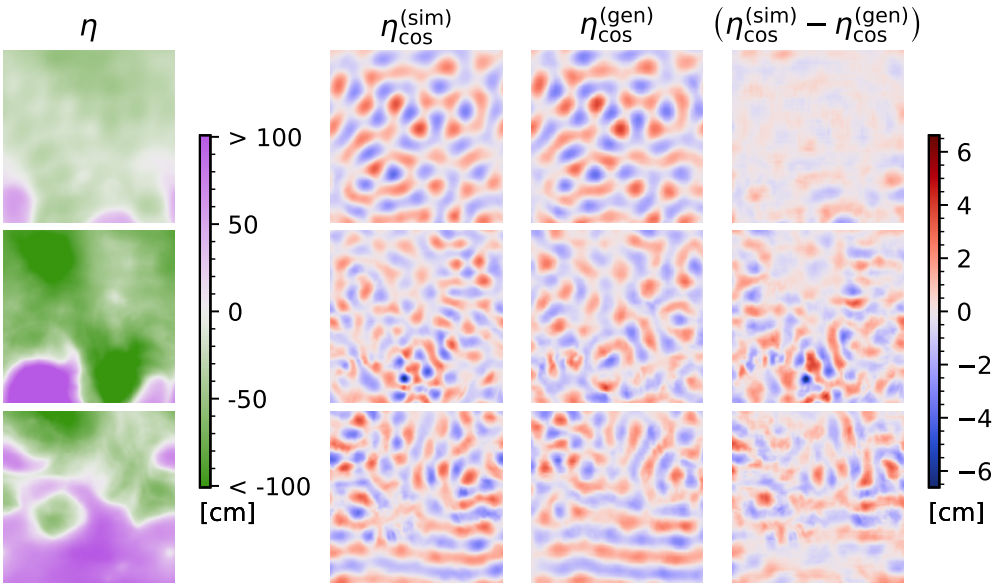

Figure 2. A few examples of individual tests. Top row: snapshot with the highest correlation among all sets (belongs to ET1). Middle and bottom rows: same for the lowest correlations among ET1-4, and ET1-5, respectively (belong to ET4 and 5, respectively). For legibility, we omit spatial axis labels, noting that the first, second, and third rows correspond to down-jet, down-jet, and mid-jet bands respectively; see Fig. 1 for their definitions.

and their locations are noticeably different between the down-jet and up-jet bands. This is expected, as the density profiles and the Coriolis parameter both vary with latitude, which modulates the mode-1 tidal wavelength (See Text S1 and Fig. S1 in the SI). Such variations can also be found in satellite observations (Ray & Zaron, 2011).

A qualitative difference can be found in the spectral behaviours between ET4 and ET5. In ET4, the locations of spectral bumps in the $\eta^{(gen)}_{cos}$ spectra vary between the down-jet and up-jet bands, in a manner such that they *closely* overlap with bumps of the $\eta^{(sim)}_{cos}$ spectra at both bands. This implies that in the ET4 run, the trained TITE identifies the dominant wavelength even as it varies. In other words, TITE can identify patterns at varying spatial scales. However, in the ET5 run, the $\eta^{(gen)}_{cos}$ spectra *fail* to trace the location of the bumps in the down-jet bands (dashed and solid red lines in lower panel of Fig. 3). The performance in up-jet bands appears as good as ET4, which may be attributed to the fact that the mode-1 tidal wavelengths to the south of the jets are the same in all five simulations.

What causes the decrease of performance in ET5? We can formulate two possible hypotheses: (i) the inherent difficulty of predicting increasingly complex patterns at higher turbulence levels, and (ii) overfitting (see section 2.3 for its definition). Overall, we believe factor (i) affects TITE's performance *more* than factor (ii), as we now argue.

In support of hypothesis (i), as explained in Text S1 in the SI, stronger scatterings of ITs from TBMs induce more longitudinal variations as well as small-scale features in the IT components due to increased incoherence. In addition, the tidal wavelengths vary

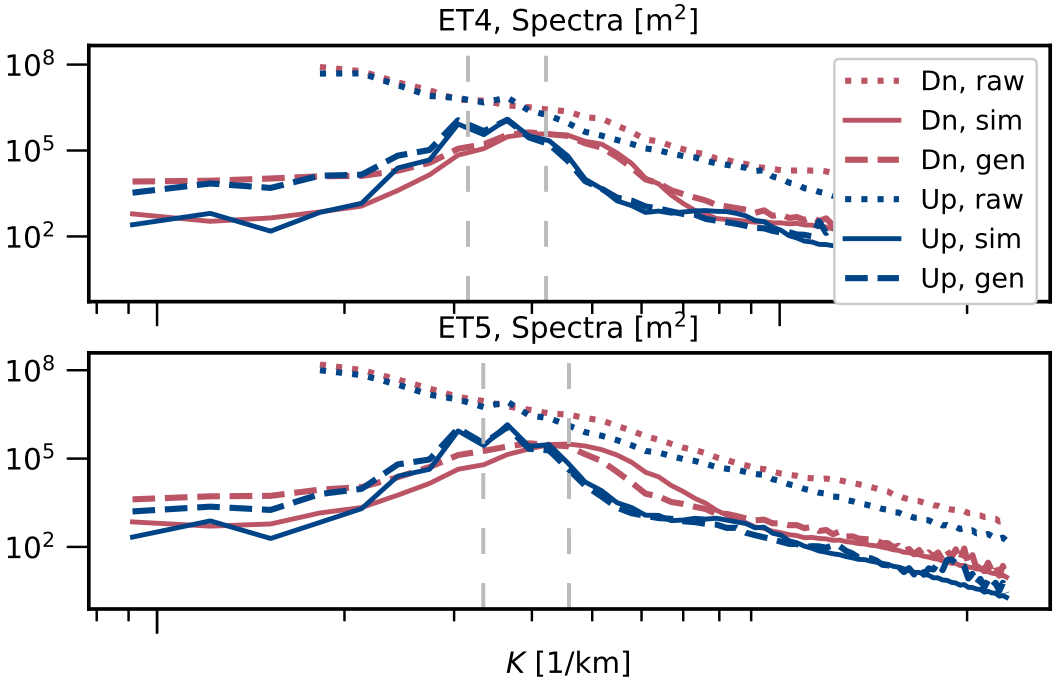

Figure 3. Spectra for the down-jet and up-jet bands in ET4 and ET5 test set. In the legends "Up", "Dn" denote the down-jet and up-jet bands respectively. "raw", "sim", and "gen" denote spectra computed from panels of $\eta, \eta_{\cos}^{(\mathrm{sim})}$, and $\eta_{\cos}^{(\mathrm{gen})}$, respectively. "$K$" denotes the horizontal wavenumber magnitude. The vertical dashed lines mark the largest and smallest mode-1 tidal wavenumbers over the simulation domain at initial time, following Fig.S1 in the SI. Raw spectra higher than $2 \times 10^8 \mathrm{m}^2$ at large scales are omitted. Higher wavenumbers are omitted.

more latitudinally due to increased density gradients, which increases the diversity of dominant spatial scales of IT signals across the domain and time. Both factors add complexities to the η and $\eta_{\cos}^{(\mathrm{sim})}$ patterns, and therefore pose more challenges to our algorithm. In Text S5 of the SI, we show that a generically defined metric of pattern complexities introduced by Bagrov et al. (2020) generally increases under stronger TBMs as we expected. As the turbulence levels increase from T1 to T5, the testing instances become increasingly challenging from ET1 to ET5, which explains the decreased performance in ET5 in correlations and spectral behaviors.

The difficulty associated with vigorous turbulence levels is also reflected in the relatively worse performance of TITE in the mid-jet bands centered around the turbulence. In the histograms shown in Fig. S7 of the SI, or last three columns in Table S2 of the SI, the correlations for the down-jet, mid-jet and up-jet bands are presented separately for the test sets in ET1-5. Within each of ET1-5, the up-jet bands have a higher mean correlation than the mid-jet bands. As the turbulence level increases, this difference gets more pronounced. The degraded performance at mid-jet bands is also reflected from the "square coherences" in Text S2 in the SI.

Turning our attention over to hypothesis (ii), in our experiments, overfitting may provide a partial explanation for the performance degradation from ET4 to ET5. In support of this hypothesis, Table S1 in the SI shows that the kinetic energy and normalized vorticity for the TBM and IT all increase from T1 to T5. Moreover, in terms of these dynamical metrics, the training set of ET5 is less diverse compared to, say, the train-

ing set of ET4 that spans a wider range of these metrics. Therefore, in ET5, TITE may be fittingly closely for turbulence levels in T1-4, and fails at T5 due to its mismatch with T1-4, whereas in ET4, both T3 and T5 are included during training, and a combination of T3 and T5 may impart some knowledge about T4 during testing.

However, several behaviours of TITE are inconsistent with the overfitting hypothesis. First, if overfitting was the only factor, then during the test phase, TITE should perform worse in ET1 than in ET2, as the range of dynamical metrics in ET1 is narrower than in ET2. This does not turn out to be the case. In fact, the ET1 run produces the best mean correlation in the test set among ET1-5; in Text S2 in the SI, we show that the ET1 test set also demonstrates excellent spectral behaviours. Second, compared to test sets, validation sets are closer to training sets, and an overfitting model should perform better in the validation sets than in the test sets. However, the mean correlations produced by TITE in the test sets are *higher* than the validation sets by about 0.05 in ET1 and ET2 (second and third columns in SI Table S2).

On the other hand, hypothesis (i) provides better explanations for these disagreements. In the ET1 test set, the turbulence levels are lower, and TITE performs well despite the possible impacts from overfitting. In the ET1 and ET2 runs, the test data are at a lower turbulence level than the validation data, which leads TITE to generate better predictions too. Therefore, we believe that TITE's performance is more influenced by hypothesis (i) than hypothesis (ii) (note that they may happen simultaneously).

We note that despite the relative lack of prediction accuracy under higher turbulence levels, in our data, TITE would still outperform simple spatial filtering methods that would break down due to the strong TBMs superimposing the ITs around tidal wavelengths (Text S2 in the SI), or harmonic analysis that would not work due to the strong incoherence and the temporal interval of $4T$. To improve TITE's performance at higher turbulence levels, a possible solution is to feed TITE with "overly turbulent" simulations during training, as motivated by the better performances we have found in ET4, as compared to ET5.

4 Conclusions and Discussions

We designed a novel technique based on a deep neural network algorithm to extract internal tides that are entangled with geostrophic turbulence. We trained and validated TITE using randomly shuffled simulation snapshots that were categorically different from the dynamic regime of the testing data. The testing data sets are designed in a way that classical methods such as harmonic fits or spectral filtering could not extract tidal signals accurately, and yet in most test cases, TITE can still (1) extract IT signals that agree well with ground truths in a deterministic sense, and (2) capture the dominant tidal energy in the wavenumber spectra, even when it varies temporally and latitudinally. When TITE does not perform as well, the main cause seems to be the high complexities of the patterns linked to stronger turbulent motions. Overall, we believe that this work provides a fresh angle on how to disentangle dynamical components from two-dimensional data via a deep learning approach. Some discussions are offered below.

Although we make no claim about TITE or cGANs in general as being the best possible algorithms to specifically achieve our goal, we found it superior to other deep learning methods we investigated, which include several types of decision trees regressors, long short-term memory networks, and U-Net structures without a discriminator. Our implementation of TITE may not be the most efficient or accurate type of cGAN for our problem, as We did not attempt to optimize model parameters such as numbers of layers or learning rates, among others. More recent variations of pix2pix such as pix2pixHD (Park et al., 2019) could also improveoutperform our current implementation. Moreover, as mentioned in section 3, the generated images always contain spurious signals outside

the dominant tidal bump, which remains to be resolved. We leave these as thoughts for future work.

In this work, TITE only extracts the cosine IT signals. The generalization to the sinusoidal IT signals, which are defined by replacing $\cos(2\pi t'/T)$ in equation (1) with $\sin(2\pi t'/T)$, should be straightforward. With both cosine and sinusoidal IT signals, phase information at any given time can be retrieved. One may also study the performance of TITE for extractions of signals at higher tidal frequencies that correspond to smaller spatial scales. Pix2pix has been observed to be capable of capturing fine features in images (Isola et al., 2017), and smaller scales don't necessarily make the problem more challenging to TITE.

So far, TITE has only been developed by applied to the idealized simulations T1-5 with a single baroclinic jet and single tidal frequency, simplistic boundary conditions, flat topography, an absence of atmospheric forcings. As an ongoing work, we are investigating the effects of including snapshots from a global ocean GCM.

We used correlation factors as a main quantitative metric in our evaluations of TITE's performance, similar to Torres et al. (2018) and Ponte et al. (2017). However, no direct physical meanings are attached to correlation factors, and rigorous hypothesis tests are challenging due to our lack of knowledge on the probability distributions of the IT signals. Wavenumber spectra, another main metric we use, does have a clearer physical interpretation, but a direct link between it and practical applications such as mapping or forecasting is challenging to establish. Our community has yet to design a more illustrative, rigorous and widely-adopted skill metric to assess the accuracy of 2D tidal extraction.

With SWOT in mind, we may reassess the four assumptions stated in the first paragraph of section 2.2. All images used in this work have a 4 km horizontal resolution that resolves the tides adequately, addressing assumption (1). In preparation for satellite data that will suffer from measurement error and more limited resolutions, we may coarse-grain and augment the training data with the type of error expected in SWOT (Gaultier et al., 2016) to investigate their impacts. Assumption (2), motivated by the incoherence of ITs and the relatively long sampling intervals of SWOT, is satisfied by the design of the TITE architecture, and by the frequent random shuffling of snapshots during training. However, complete statistical independence between ITs and TBMs can be overly strict for several reasons, ranging from a higher temporal sampling at high latitude, to the possibility of "filling in the time gaps" with other sources of data such as those from assimilated models or in-situ instruments (d'Ovidio et al., 2019). From the overall satisfactory performance of TITE, the assumption (3) appears to be satisfied in our simulation outputs, perhaps due to simulation design choices such as a perfectly harmonic incoming IT signal, or simple boundary conditions. Under more realistic configurations, a functional dependence might not be guaranteed. On the other hand, the assumption (3) can also be overly strict, considering recent progress in the theory of IT/TBM interactions (Torres et al., 2018; Savva & Vanneste, 2018; Savva et al., 2021; Wagner et al., 2017). Assumption (4) relies on the premise that there will be pre-processed training data (presumably from highly skilled model outputs) that mimic the dynamics to be sampled by SWOT. Productions of such data are receiving significant attention within the modelling communities (Zaron & Rocha, 2018; Rocha et al., 2016; Arbic et al., 2010; Savage et al., 2017). Overall, to make TITE eventually applicable to SWOT and other satellite missions in the future, more work is required, especially in coordination with different communities and perhaps combination with complementary methods.

Acknowledgments

We thank two anonymous reviewers for numerous constructive advice. H.W. and N.G. acknowledge the financial support of the Canadian Space Agency [14SUSWOTTO] and of the Natural Sciences and Engineering Research Council of Canada (NSERC) [RGPIN-2015-03684]. M.P. acknowledges support from the Summer Undergraduate Research Program of the Department of Physics at the University of Toronto. A.L.P. was supported by ANR project number 17-CE01-0006-01 and by project DIEGO of the CNES/TOSCA SWOT program. This work was made possible by the facilities of the Shared Hierarchical Academic Research Computing Network (SHARCNET: www.sharcnet.ca) and Compute/Calcul Canada. The work has benefited from scientific discussions with Laure Zanna, Brian Arbic, Michael W. Lever, Jason Summers, and Ilia A. Iakovlev.

Simulation results corresponding to snapshots in T1-5 used in this study to train and test TITE are published on Scholars Portal Dataverse (Ponte et al., 2020). Codes defining the architecture of TITE are available on Github via link `https://github.com/hannnwang/Pix2PixTITEexamples`.

References

Arbic, B. K., Wallcraft, A. J., & Metzger, E. J. (2010). Concurrent simulation of the eddying general circulation and tides in a global ocean model. *Ocean Modelling*, *32*(3-4), 175–187.

Bagrov, A. A., Iakovlev, I. A., Iliasov, A. A., Katsnelson, M. I., & Mazurenko, V. V. (2020). Multiscale structural complexity of natural patterns. *Proceedings of the National Academy of Sciences*, *117*(48), 30241–30251.

Ballarotta, M., Ubelmann, C., Pujol, M.-I., Taburet, G., Fournier, F., Legeais, J.-F., ... others (2019). On the resolutions of ocean altimetry maps. *Ocean Science*, *15*(4), 1091–1109.

Bühler, O. (2014). *Waves and Mean Flows* (2nd ed.). Cambridge: Cambridge University Press. Retrieved from `http://ebooks.cambridge.org/ref/id/CBO9781107478701` doi: 10.1017/CBO9781107478701

Charney, J. G. (1971). Geostrophic turbulence. *Journal of the Atmospheric Sciences*, *28*(6), 1087–1095.

Colosi, J. A., & Munk, W. (2006). Tales of the venerable honolulu tide gauge. *Journal of physical oceanography*, *36*(6), 967–996.

Dietterich, T. (1995). Overfitting and undercomputing in machine learning. *ACM computing surveys (CSUR)*, *27*(3), 326–327.

Dunphy, M., & Lamb, K. G. (2014). Focusing and vertical mode scattering of the first mode internal tide by mesoscale eddy interaction. *Journal of Geophysical Research: Oceans*, *119*(1), 523–536.

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.

d'Ovidio, F., Pascual, A., Wang, J., Doglioli, A. M., Jing, Z., Moreau, S., ... others (2019). Frontiers in fine-scale in situ studies: Opportunities during the swot fast sampling phase. *Frontiers in Marine Science*, *6*, 168.

Egbert, G., & Erofeeva, S. (2021). An approach to empirical mapping of incoherent internal tides with altimetry data. *Geophysical Research Letters*, e2021GL095863.

Egbert, G., & Ray, R. (2000). Significant dissipation of tidal energy in the deep ocean inferred from satellite altimeter data. *Nature*, *405*(6788), 775–778.

Ferrari, R., & Wunsch, C. (2009). Ocean circulation kinetic energy: Reservoirs, sources, and sinks. *Annual Review of Fluid Mechanics*, *41*, 253–282.

Fu, L.-L., & Ferrari, R. (2008). Observing oceanic submesoscale processes from space. *Eos, Transactions American Geophysical Union*, *89*(48), 488–488.

Gaultier, L., Ubelmann, C., & Fu, L.-L. (2016). The challenge of using future swot

data for oceanic field reconstruction. *Journal of Atmospheric and Oceanic Technology*, *33*(1), 119–126.

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., ... Bengio, Y. (2014). Generative adversarial nets. *Advances in neural information processing systems*, *27*.

Isola, P., Zhu, J.-Y., Zhou, T., & Efros, A. A. (2017). Image-to-image translation with conditional adversarial networks. In *Proceedings of the ieee conference on computer vision and pattern recognition* (pp. 1125–1134).

Lebedev, M., Vizilter, Y. V., Vygolov, O., Knyaz, V., & Rubis, A. Y. (2018). Change detection in remote sensing images using conditional adversarial networks. *International Archives of the Photogrammetry, Remote Sensing & Spatial Information Sciences*, *42*(2).

Le Guillou, F., Lahaye, N., Ubelmann, C., Metref, S., Cosme, E., Ponte, A., ... Vidard, A. (2021). Joint estimation of balanced motions and internal tides from future wide-swath altimetry. *Journal of Advances in Modeling Earth Systems*, e2021MS002613.

Lien, R.-C., & Gregg, M. (2001). Observations of turbulence in a tidal beam and across a coastal ridge. *Journal of Geophysical Research: Oceans*, *106*(C3), 4575–4591.

Metref, S., Cosme, E., Le Guillou, F., Le Sommer, J., Brankart, J.-M., & Verron, J. (2020). Wide-swath altimetric satellite data assimilation with correlated-error reduction. *Frontiers in Marine Science*, *6*, 822.

Mirza, M., & Osindero, S. (2014). Conditional generative adversarial nets. *arXiv preprint arXiv:1411.1784*.

ml4a. (2017). *Machine Learning for Artists.* Retrieved from `https://ml4a.github.io/guides/Pix2Pix/`

Morrow, R., Fu, L.-L., Ardhuin, F., Benkiran, M., Chapron, B., Cosme, E., ... others (2019). Global observations of fine-scale ocean surface topography with the surface water and ocean topography (swot) mission. *Frontiers in Marine Science*, *6*, 232.

Munk, W., & Hasselmann, K. (1964). Super-resolution of tides. *Studies on oceanography*, 339–344.

Munk, W., Zetler, B., & Groves, G. (1965). Tidal cusps. *Geophysical Journal International*, *10*(2), 211–219.

Park, T., Liu, M.-Y., Wang, T.-C., & Zhu, J.-Y. (2019). Semantic image synthesis with spatially-adaptive normalization. In *Proceedings of the ieee/cvf conference on computer vision and pattern recognition* (pp. 2337–2346).

Ponte, A. L., & Klein, P. (2015). Incoherent signature of internal tides on sea level in idealized numerical simulations. *Geophysical Research Letters*, *42*(5), 1520–1526.

Ponte, A. L., Klein, P., Dunphy, M., & Le Gentil, S. (2017). 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.

Ponte, A. L., Le Gentil, S., & Grisouard, N. (2020). *Internal tides and geostrophic turbulence in a Boussinesq re-entrant channel.* Scholars Portal Dataverse. Retrieved from `https://doi.org/10.5683/SP2/HU58SG` doi: 10.5683/SP2/HU58SG

Ray, R. D., & Zaron, E. D. (2011). Non-stationary internal tides observed with satellite altimetry. *Geophysical Research Letters*, *38*(17).

Richman, J. G., Arbic, B. K., Shriver, J. F., Metzger, E. J., & Wallcraft, A. J. (2012). Inferring dynamics from the wavenumber spectra of an eddying global ocean model with embedded tides. *Journal of Geophysical Research: Oceans*, *117*(C12).

Rocha, C. B., Chereskin, T. K., Gille, S. T., & Menemenlis, D. (2016). Mesoscale to submesoscale wavenumber spectra in drake passage. *Journal of Physical Oceanography*, *46*(2), 601–620.

Savage, A. C., Arbic, B. K., Alford, M. H., Ansong, J. K., Farrar, J. T., Menemenlis, D., ... others (2017). Spectral decomposition of internal gravity wave sea surface height in global models. *Journal of Geophysical Research: Oceans*, *122*(10), 7803–7821.

Savva, M. A., Kafiabad, H. A., & Vanneste, J. (2021, jun). Inertia-gravity-wave scattering by three-dimensional geostrophic turbulence. *J. Fluid Mech.*, *916*, A6. Retrieved from `https://www.cambridge.org/core/product/identifier/S0022112021002056/type/journal{\_}article` doi: 10.1017/jfm.2021.205

Savva, M. A., & Vanneste, J. (2018). Scattering of internal tides by barotropic quasigeostrophic flows. *Journal of Fluid Mechanics*, *856*, 504–530.

Shakespeare, C. J., & Hogg, A. M. (2017). Spontaneous surface generation and interior amplification of internal waves in a regional-scale ocean model. *Journal of Physical Oceanography*, *47*(4), 811–826.

Tensorflow. (n.d.). *pix2pix: Image-to-image translation with a conditional GAN.* Retrieved from `https://www.tensorflow.org/tutorials/generative/pix2pix` (Accessed on Sep.20, 2020.)

Torres, H., 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.

Torres, H., Klein, P., Siegelman, L., Qiu, B., Chen, S., Ubelmann, C., ... Fu, L.-L. (2019). Diagnosing ocean-wave-turbulence interactions from space. *Geophysical Research Letters*, *46*(15), 8933–8942.

Wagner, G. L., Ferrando, G., & Young, W. R. (2017). An asymptotic model for the propagation of oceanic internal tides through quasi-geostrophic flow. *Journal of Fluid Mechanics*, *828*, 779–811.

Whalen, C. B., de Lavergne, C., Garabato, A. C. N., 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.

Zaron, E. D., & Rocha, C. B. (2018). Internal gravity waves and meso/submesoscale currents in the ocean: anticipating high-resolution observations from the swot swath altimeter mission. *Bulletin of the American Meteorological Society*, *99*(9), ES155–ES157.

Zhao, Z. (2016). Internal tide oceanic tomography. *Geophysical Research Letters*, *43*(17), 9157–9164.