JGR Oceans

RESEARCH ARTICLE
10.1029/2021JC017588

Special Section:
Air-Sea Coupling and its Effect on Electromagnetic Propagation in the Marine Atmospheric Boundary Layer

An Empirical Evaluation of Turbulence Closure Models in the Coastal Ocean

I. B. Savelyev[1], P. J. Martin[2], Y. Fan[2], D. K. Savidge[3], R. K. Shearman[4], T. Haack[5], T. de Paolo[6], E. J. Terrill[6], and Q. Wang[7]

[1]Remote Sensing Division, U.S. Naval Research Laboratory, Washington, DC, USA, [2]Oceanography Division, U.S. Naval Research Laboratory, Stennis Space Center, Bay St. Louis, MS, USA, [3]Skidaway Institute of Oceanography, Savannah, GA, USA, [4]College of Earth, Ocean, and Atmospheric Sciences, Oregon State University, Corvallis, OR, USA, [5]Marine Meteorology Division, U.S. Naval Research Laboratory, Monterey, CA, USA, [6]Scripps Institute of Oceanography, La Jolla, CA, USA, [7]Department of Meteorology, U.S. Naval Postgraduate School, Monterey, CA, USA

Martin and Haack are retired.

Key Points:

- A comprehensive coastal ocean small-scale turbulence data set was collected
- Turbulence closure models were tested against the in-situ data set
- Model performance was evaluated and remaining modeling errors quantified

Correspondence to:

I. B. Savelyev,
ivan.savelyev@nrl.navy.mil

Citation:

Savelyev, I. B., Martin, P. J., Fan, Y., Savidge, D. K., Shearman, R. K., Haack, T., et al. (2022). An empirical evaluation of turbulence closure models in the coastal ocean. *Journal of Geophysical Research: Oceans*, *127*, e2021JC017588. https://doi.org/10.1029/2021JC017588

Received 18 MAY 2021
Accepted 31 MAR 2022

Author Contributions:

Conceptualization: I. B. Savelyev
Data curation: I. B. Savelyev, D. K. Savidge, T. de Paolo, Q. Wang
Methodology: I. B. Savelyev, P. J. Martin, Y. Fan, D. K. Savidge, R. K. Shearman, T. Haack, T. de Paolo, E. J. Terrill, Q. Wang
Resources: I. B. Savelyev, D. K. Savidge, Q. Wang
Validation: I. B. Savelyev, P. J. Martin, Y. Fan
Writing – original draft: I. B. Savelyev

Abstract A comprehensive in situ data set was obtained for the purpose of testing upper ocean turbulence models. The data set was collected in 38 m deep water over the North Carolina shelf. Available time series of surface meteorological forcing, including waves, winds, and heat fluxes, as well as underwater profiles of temperature, salinity, and horizontal velocity, were found suitable for constraining and testing numerical models in realistic environmental scenarios of the coastal ocean. The Navy Coastal Ocean Model was tested in a vertical 1-D mode with a suite of previously incorporated subgrid turbulence closure models. Modeled output of turbulent quantities, such as the turbulent kinetic energy and its dissipation rate, were evaluated against comparable turbulence observations from a bottom mounted acoustic velocity profiler and a glider mounted turbulent shear probe. The results demonstrate a steady incremental skill increase among available turbulence closure models over several decades of their development, however the correlation with observed turbulence remains weak. The analysis of remaining model errors identified a need to add wave-dependence to the air-sea momentum flux formulation to account for waves that are out of balance or misaligned with the wind.

Plain Language Summary Ocean circulation contains an interconnected cascade of fluid motions with lengthscales ranging from the basin scales (~1,000 km), down to the smallest dissipative scales (~10 mm). Resolving all these motions is prohibitively computationally expensive for a numerical ocean circulation model, even on the world's fastest supercomputers. Therefore, the model grid usually covers the basin with spacing only down to ~1 km. Smaller scale turbulent processes are averaged within each grid cell and are parameterized and solved for by means of a simplified subgrid turbulence model. This study takes on the challenge of testing such subgrid model against real ocean observations. The turbulent quantities computed by the model, such as the amount of small-scale turbulent kinetic energy and the rate of its dissipation, were evaluated against similar quantities observed at sea. The results demonstrate a steady improvement in model performance from earlier versions a few decades ago, to the modern version. However, a recommendation is made to modify the formulation of air-sea boundary condition to achieve further improvements in the model performance. In summary, this study confirms the usefulness of subgrid models to simplify ocean modeling, and offers a direction and an empirical testbed for further development.

1. Introduction

1.1. The Origins of Turbulence Closure Models for the Upper Ocean

Small-scale boundary layer turbulent mixing processes, taking place in the top few tens of meters of the ocean, control its mixed layer structure, including the depth of the thermocline. This, in turn, has an impact on larger scale air-sea interaction processes, which ultimately impact ocean circulation and climate. Therefore, a large scale ocean circulation model has to be able to account for the upper ocean turbulent mixing taking place on scales well below its typical grid resolution. This motivated the development of dedicated turbulence closure models, such as the popular Mellor and Yamada (1982) level 2.5 turbulence closure model, further MY82. The core function of such model is to prognostically solve a 1-D vertical column turbulent kinetic energy (TKE) budget, including production, dissipation, advection, and other terms relevant to TKE and to turbulent length scale variables. This, in turn, allows estimating vertical turbulent fluxes of momentum and tracers, which are passed along to the larger

Writing – review & editing: I. B. Savelyev

scale ocean circulation model. By assuming quasi steady state and horizontal homogeneity, the flux profiles are allowed to work on existing background stratification throughout a single grid cell and time step.

The development of a working subgrid model, such as MY82, was a major undertaking. It encompassed a large set of physical mechanisms and had to rely on existing literature to appropriately define relevant terms, which often came with empirical constants. As a result, MY82 requires an input of eight parameters, such as A_1, B_1, A_2, B_2, among others. Initially, the values of some of these parameters came from dedicated channel flow or pipe flow laboratory boundary layer experiments (e.g., setting $B_1 = 16.6$). But as was pointed out in MY82, it is often unclear what these values should be in application to the upper ocean boundary layer. Indeed, later field observations (e.g., Terray et al., 1996) discovered a far more energized boundary layer in the vicinity of a wavy surface, motivating the implementation of appropriate corrections to the MY82 model (Craig & Banner, 1994). However, much uncertainty in the choices of empirical parameters still remain, while other parameters do not have an explicit empirical guidance at all. Facing the lack of adequate empirical data, the values of these parameters were finalized by means of a "numerical training" exercise.

In a numerical training exercise, the entire model is tested against known ocean observations of the mixed layer. This provides not only the overall model performance evaluation, but also enables calibration for parameters that remained without direct empirical guidance. For example, long term observations of the mixed layer structure and of atmospheric forcing are available for Ocean Weather Station (OWS) Papa. The offshore location of this station in the Northeast Pacific (145° W, 50° N) is ideal, as it is far from any bathymetric variability or other sources of small-scale oceanographic or meteorological disturbances, hence presenting a somewhat steady state and horizontally homogeneous scenario typical for the open ocean. Numerous model tests were performed at OWS Papa (e.g., Denman & Miyake, 1973; Large, 1996, etc.). Among them, one of the more recent comparisons between a turbulence closure model output and OWS Papa observations can be found in Martin et al. (2013). They tested the Navy Coastal Ocean Model (NCOM) with a modified MY82 subgrid model in 1-D setup, forced by the meteorological conditions encountered at the station for the duration of a year. The resulting mixed layer structure output from NCOM was compared and evaluated against sea surface temperature (SST) and mixed layer depth (MLD) time series observed at the station. Assuming SST and MLD variability was a result of local mixing, this constitutes a valid "black box" performance evaluation of MY82. There are, however, several notable deficiencies in such evaluation, or in any numerical training exercise based on such comparison. First, even one year of these observations poorly constrain all uncertain parameters and other elements within the NCOM-MY82 system. Second, with its multiple assumptions for steady state, homogeneity, stability, sea state maturity, among others, MY82 was not designed or field tested for a far more complex range of environments that we know exist in the ocean. This second consideration, in part, motivates the present study. To narrow down our specific objectives, next we overview the development progress that took place in the last few decades, leading up to the modern version of MY82.

1.2. Wave Forcing in Turbulence Closure Models

Further development of MY82 in the past two or three decades primarily focused on wave induced effects. These can include a number of turbulence production mechanisms attributed to the wave Stokes drift, of which the most widely known are Langmuir circulations (Langmuir, 1938). It was becoming evident (e.g., Smith, 1992) that somehow waves are responsible for mixed layer deepening, unaccounted for by the original MY82 formulation. However, nonbreaking waves are irrotational and cannot cause mixing, whereas breaking waves do not inject TKE deep enough into the water column, and therefore cannot account for MLD deepening at tens of meters. The theory of Langmuir circulations provided a much needed explanation. It suggested a mechanism by which surface wave forcing would create pairs of large horizontal counter-rotating vortices (Craik & Leibovich, 1976). These vortices are relatively weak in terms of TKE, but can extend from the surface all the way down to the thermocline (Gargett & Wells, 2007), hence providing an efficient vertical scalar transport mechanism and contributing to MLD deepening.

The problem, or rather the challenge with the mechanism proposed by Craik and Leibovich (further CL2 mechanism), is that it is notoriously difficult to verify and quantify in the field. The existence of Langmuir circulations was, of course, well documented by Langmuir himself and by many others. The circulations are usually expressed by surface streaks or windrows of Sargassum seaweed or other floating debris, often observed in the open ocean, as well as in the shallow coastal environment (see literature review by Thorpe, 2004). However,

quantifying their subsurface structure, energetics, growth rates, or other useable parameters, as a function of a range of wind and wave conditions proved to be nearly impossible. One of the more substantial efforts dedicated to obtaining such data set was undertaken by Weller et al. (1985), which serves as a good illustration of why a direct approach presents such a challenging task. Later, other occasional attempts were made (e.g., Smith, 2001; Gargett & Wells, 2007, among others), but the results were mostly qualitative, episodic, or indirect and could not amount to a systematic and quantitative data set needed for a conclusive model validation.

The next breakthrough came not from the field or even laboratory experiments, but from idealized Large Eddy numerical Simulations (LES). Most notably, a pioneering study by McWilliams et al. (1997), further MW97, implemented the CL2 mechanism into the system of primitive equations of motion. By simulating a small segment of the ocean mixed layer with sufficiently high grid resolution, they were able to demonstrate "Langmuir turbulence" (LT), which is the boundary layer turbulence with the addition of Langmuir effects brought about by CL2. In the vacuum of appropriate empirical datasets, MW97-type idealized numerical experiments took on the role of ground truth for consequent turbulent closure model development and calibration in various environmental scenarios. Among MW97 findings was a substantial increase in vertical mixing rates in response to wave forcing, presenting an opportunity for the much needed correction in turbulence closure models. The theoretical foundation and formulation explaining the mechanism for increased TKE production by nonbreaking waves was further developed by Ardhuin and Jenkins (2006) and Ardhuin et al. (2008). Their work properly explained how the Stokes drift shear, which is a result of irrotational wave motion, can none-the-less act to produce turbulence, similar to the conventional turbulent shear production. Based on these theoretical and LES findings, a modified MY82 turbulence closure model was developed by Kantha and Clayson (2004), further KC04. Martin et al. (2013) implemented KC04 into NCOM ocean circulation model and demonstrated its performance against OWS Papa data set. As indirect as it was, this last step presents a rare instance of a quantitative empirical check of Langmuir effects in a numerical model. Other similar notable efforts relating LT modeling to observations were made by Smyth et al. (2002), Kukulka et al. (2009, 2010) and by Kantha et al. (2010). This body of work laid foundation for the development of a variety of modern turbulent closure models, including second moment closures and KPP-type schemes. A recent intercomparison paper by Li et al. (2019) takes stock of some of the most commonly used schemes (with or without wave forcing), and subjects them to a detailed comparison between each other, as well as LES simulations.

Following MY82 and KC04, the latest step in the MY82-type turbulence closure model development was described by Harcourt (2015), (further H15), which was also one of the models evaluated by Li et al. (2019). Notable improvements in H15 included a more physically consistent formulation of stability functions, and also the reliance on LES that were "scaled" with real field observations (Harcourt & D'Asaro, 2008). The LES model used vertical velocity variance measurements from a Lagrangian float in various wind-wave condition to constrain the amount of TKE in the model. Following H15 publication, Martin and Savelyev (2017), further MS17, proceeded with H15 implementation into NCOM and consequent 1-D model tests against OWS Papa in a manner similar to KC04 and NCOM tests done in Martin et al. (2013). MS17 demonstrated a substantial performance improvement in H15 over KC04 when subjected to the exact same test case at the OWS Papa. More specifically, H15 offered a better correction for the well-known lack of mixing in MY82, and hence resulted in a closer agreement with MLD observed at OWS Papa.

1.3. Objectives of This Study

Motivated by the success of H15 at OWS Papa, this study seeks to further expand the body of H15 performance evaluation evidence to more complex environmental scenarios. The challenges this study takes on are to assemble necessary state-of-the-art observational capabilities, to design and execute an appropriate field test, and to use the resulting data set for a critical model evaluation. Of a particular interest are non-trivial environmental scenarios with misaligned wind and waves, rapidly time varying atmospheric forcing, and complex background stratification—all fairly common in some parts of the ocean, such as near a coast. Another focus of this study is on the impact of waves on near-surface turbulence and on the evaluation of the ability of existing wave-dependent modules within turbulence closure models to capture that impact.

2. Methodology

2.1. Overall Model Evaluation Strategy

The evaluation of a turbulent mixing model in a 1-D vertical mode against long-term observations of slowly varying background mixed layer structure can only work if these variations are caused by local mixing. In more complex scenarios, a non-locally produced structure could advect into the sampling volume and compete with, if not dominate over the local mixing processes. The minimal non-locally produced variability was both the advantage, but also the shortcoming of the OWS Papa data set location. Here we take a different approach - instead of being the compared variable, measured background structure will be fed into the model. That is, at every time step, modeled vertical profiles will be relaxed to the observed profiles. The evaluated variables, instead, will be small-scale turbulent quantities, such as TKE and TKE dissipation rate, which evolve on much shorter time scales, and therefore can be assumed to be the result of local mixing processes. The following methodology subsections provide detailed description of measured variables (mean for model forcing and turbulent for model evaluation) and of the tested model setup.

2.2. Field Experiment Overview

The field experiment described here is a subset of measurements taken during a larger field campaign "CASPER-East" (Wang et al., 2017). The experiment was approximately 1 month long, taking place over the North Carolina shelf in the Fall of 2015. It included multiple research aircraft, research vessels, in-water autonomous vehicles, moorings, and a shore station. This study focuses on one of the air-sea interaction super sites established by CASPER-East, namely ASI-2. The location of ASI-2 (N36.1837, W75.0451) was approximately 63 km offshore, directly eastward from FRF Duck pier, approximately 2/3 of the way to the shelf break. The average depth at ASI-2 is 38 m, surrounded by smooth sandy bottom gently sloping away from the shore. Note, some of the atmospheric, underwater, and airborne remote sensing measurements collected at ASI-2 during CASPER-East have been published in our earlier analysis (Fan et al., 2020; Savelyev et al., 2018). Sections 2.2.1 and 2.2.2 below will briefly reiterate the methodology behind some of these previously used measurements. Additionally, Section 2.2.3 will describe turbulent underwater profiles collected at ASI-2, which are being presented here for the first time.

2.2.1. Atmospheric Forcing

The atmospheric forcing parameters most relevant to this study are near-surface winds, waves, and air-sea heat fluxes. Directional wave spectra were measured continuously by a dedicated Scripps miniature wave buoy, moored at ASI-2 for the duration of the experiment. The wave spectra were used to compute Stokes drift current (SDC) profiles, as well as the significant wave height H_s, dominant wave period T_p and direction θ time series. Detailed measurements of near-surface momentum and scalar fluxes were collected onboard R/V Atlantic Explorer. A multi-level instrument array on a bow mast of the vessel measured mean atmospheric quantities used for bulk formula air-sea flux estimates (COARE 3.5). Also, occasionally the vessel would point into the wind to take direct measurements of turbulent fluxes. In this study we are using results of bulk formula outputs, and of a particular interest are the air-sea momentum flux τ and its direction, as well as the four components of the air-sea heat flux: Sensible and latent heat fluxes across the interface, downwelling solar radiation, and upwelling long-wave infrared radiation. A summary of the atmospheric forcing parameters described above is presented in Figures 1–3. In these figures some of the observed variables are compared to a model output, as descrived further below.

The red curves in Figures 2 and 3 correspond to R/V Atlantic Explorer measurements. The R/V was tasked to continuously traverse East-West line from FRF Duck to the shelf break, therefore it was not always near the ASI-2 location. In these figures we include R/V measurements whenever it was within 1/3 longitude degrees of ASI-2. A continuous time series of the evolving atmospheric forcing was provided by the Navy's high-resolution mesoscale numerical weather prediction model, the Coupled Ocean-Atmosphere Mesoscale Predication System (COAMPS® is a registered trademark of the Naval Research Laboratory). COAMPS consists of an atmospheric model coupled to NCOM ocean model with exchange of heat and momentum fluxes at every oceanic time step. Both models were run in real-time at 2 km horizontal grid spacing for the duration of the 30 days CASPER East campaign yielding hourly forecasts of atmospheric, oceanic, and air-sea coupling parameters. The models were updated with local observations ingested every 6 hr (exclusive of any CASPER-East measurements), and were

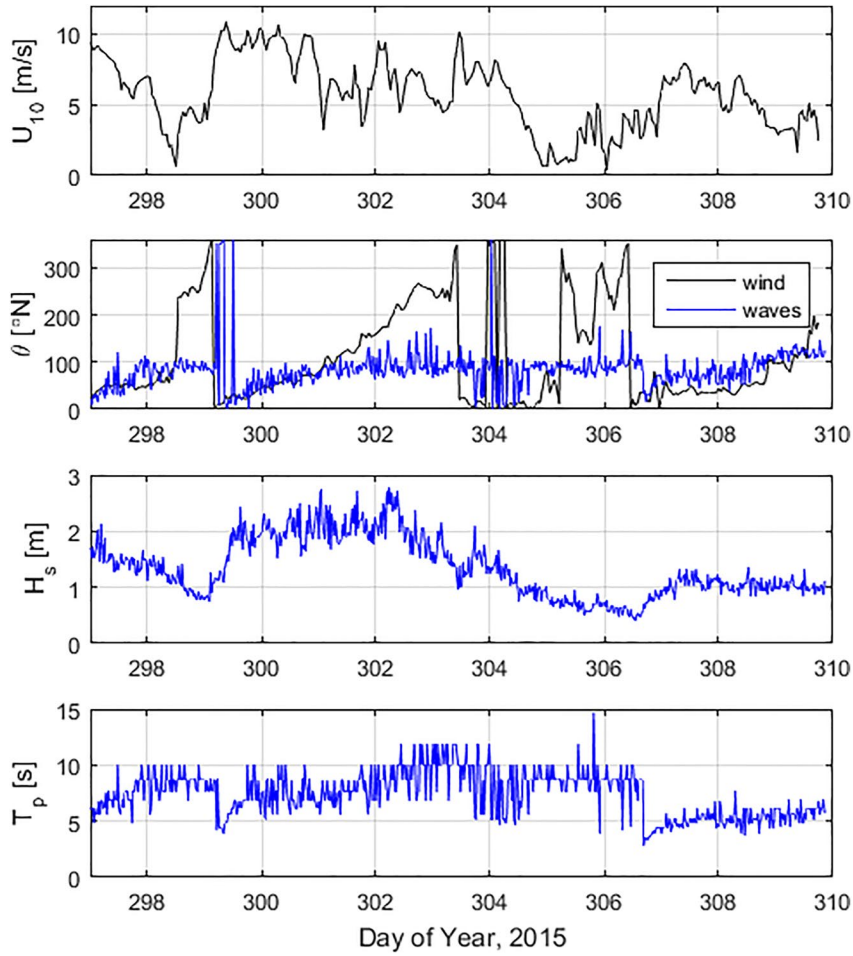

Figure 1. Time series of the wind speed at 10 m height U_{10}, dominant wave and wind directions θ, significant wave height H_s, and dominant wave period T_p. Black curves correspond to Coupled Ocean-Atmosphere Mesosacle Predication System output, blue curves are derived from wave buoy measurements.

fed boundary conditions from global operational models (i.e., HYCOM and NAVGEM). In comparison to the available R/V observations, the COAMPS air-sea momentum and heat fluxes subplots are in close agreement giving confidence that the model data can be used as a "virtual" weather station in this study. Here we utilize the model data at the grid point corresponding to the ASI-2 to provide atmospheric forcing for the ensuing NCOM 1D ocean model simulations. Since COAMPS did not include a wave model, wave information was derived purely from wave buoy observations. That included dominant wave parameters, such as significant wave height, dominant wave period and direction. Additionally, the turbulent closure model requires time series of the vertical SDC profile as an input. It was calculated based on linear wave theory as a superposition of all wave frequencies within the frequency spectrum, then given a single dominant wave direction. The time series of the shallowest SDC profile point u_{st} (t) at depth $z = 0$ is given in Figure 3.

Furthermore, Figure 4 gives some common higher level wind-wave forcing parameters, which are derived from observations given in the first three figures. Among them are the turbulent Langmuir number La_t, calculated as $La_t = (u_{*w}/u_{st})^{1/2}$, where u_{*w} is the wind friction velocity. The dominant wave steepness was calculated as $ak = 2H_s\pi^2/gT_p^2$, and the inverse wave age U_{10}/C_p was calculated as the ratio between the wind speed at 10 m height and the dominant wave phase speed.

Next, we review the general meteorological conditions encountered during the experiment, based on the examination of Figures 1–4. First, the wind speed U_{10}, was mostly moderate with a few weather systems coming through the region, causing fluctuations from calm to slightly rough (10 m/s) conditions. The wind direction time series indicate that these changes in wind speed were also accompanied by a full rotation of its direction, associated with

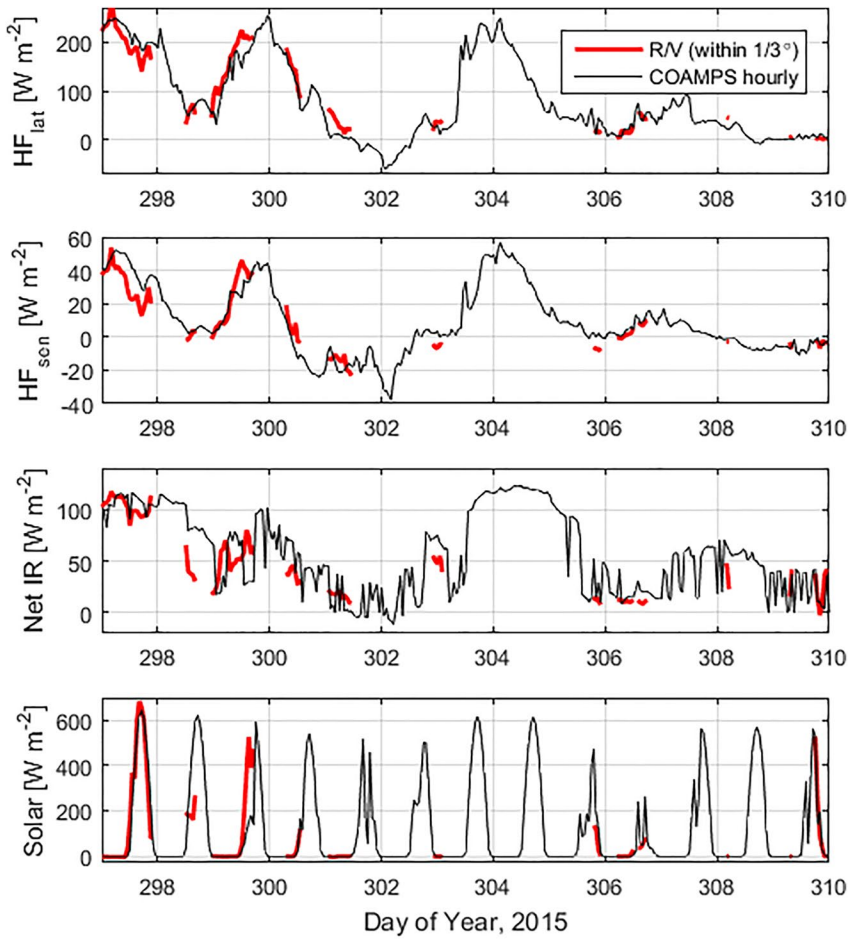

Figure 2. Time series of heat fluxes: latent heat flux HF_{lat}, sensible heat flux HF_{sen}, long-wave infrared radiation Net IR, and short-wave solar radiation Solar. Black curves are Coupled Ocean-Atmosphere Mesosacle Predication System (COAMPS) outputs, red curves are measurements collected by the R/V. Conventionally, the directions of heat fluxes are such that the solar flux acts to warm the ocean, whereas the other three fluxes are acting to cool it.

the passage of the weather systems. Nevertheless, the dominant wave direction remained more steady, primarily out of East or Northeast. This result indicates that the seas were swell dominated, with the swell propagating into the study area from non-local storms offshore. It is also supported by the wave age sublot (Figure 4), indicating frequent occurrence of mature or over-mature seas that must have a non-local origin. For more in-depth discussion, Hanley et al. (2010) provides a global wave age climatology, quantifying and explaining larger scale wave age patterns in this and other similar regions. Various components of the air-sea heat flux are behaving as expected for the Fall season in this region. As can be seen from the daily (dashed) curve on the total heating subplot in Figure 3, the total value tends to be slightly negative, indicating heat loss from the ocean into the atmosphere, as expected in the Fall and Winter seasons in this region. However, the total heat flux value is still weak in the beginning of the cooling season, and therefore we do not expect the convection to play a significant role in the upper ocean mixing, leaving the dominance for wind or wave induced mixing.

2.2.2. Mean Underwater Profiles

A Slocum underwater glider, equipped with Sea-Bird SBE41CP temperature and conductivity sensors, sampling at 1 Hz, was operated near the ASI-2 location for the duration of the CASPER-East experiment. The glider was able to sample a full vertical profile (minus few meters near the bottom and the surface) every few minutes, while typically keeping its location within ~3 km radius around ASI-2. Figure 5 gives time series of temperature and salinity profiles measured by the glider, as well as the Brunt–Väisälä buoyancy frequency N^2 derived from these profiles.

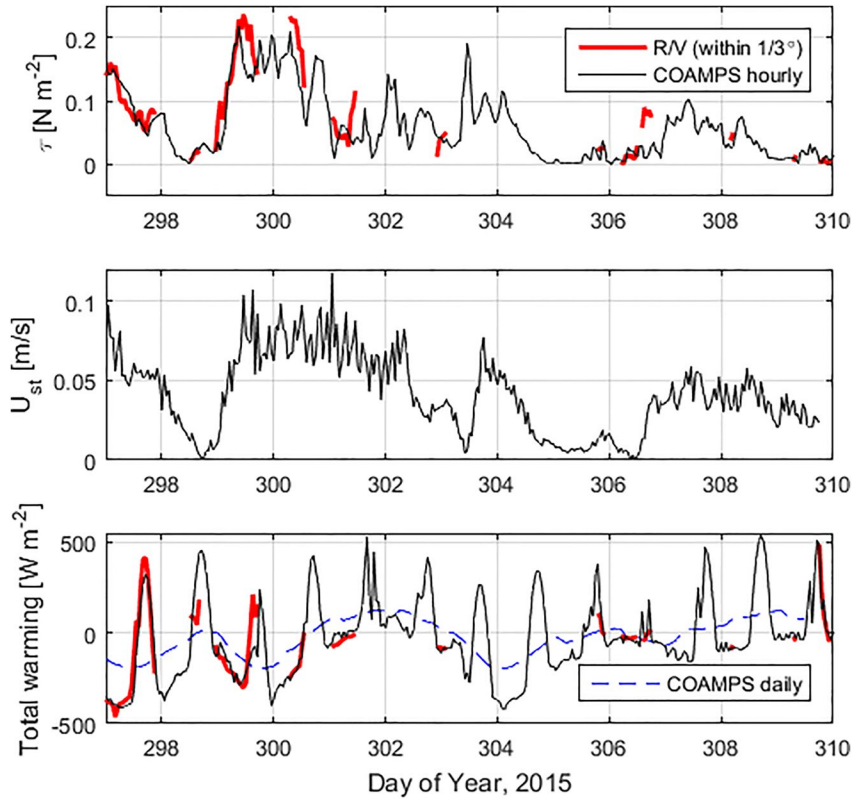

Figure 3. Time series of the air-sea momentum flux τ, Stokes drift current surface drift U_{st}, and the total ocean warming rate (a sum of all four heat fluxes in Figure 2). Black curves are Coupled Ocean-Atmosphere Mesoscale Predication System (COAMPS) outputs, blue dashed curve is the same hourly COAMPS output subjected to 24 hr running window average, red curves are measurements collected by the R/V.

An upward looking Nortek Signature 500 kHz 5-beam acoustic doppler current profiler (ADCP) was deployed in a bottom mount on the bottom at the ASI-2 location for the duration of the CASPER-East experiment. While carefully leveled at deployment, the passage of Hurricane Joaquin through the region shortly before CASPER-East resulted in several degrees of pitch and roll being introduced prior to the measurements discussed here. The raw data was collected at 4 Hz sampling frequency, averaged to 1 Hz, corrected for pitch and roll using all 5 beams and a bin-depth matching routine then smoothed with an hourly Hanning filter and subsampled to half-hour intervals to obtain mean profiles of horizontal velocities shown in Figure 6. Because of sidelobe interference and the induced pitch and roll, the figure shows profiles resolved from the ADCP depth up to ~10 m depth with 1 m bin size increments, eliminating data between 0 and 10 m depth. Note, this restriction is less severe for the fifth vertical beam alone, as explained further in Section 2.2.3. Figure 6 also shows the time series of the vertical shear profile, which is derived from the vertical profiles of the mean horizontal velocity.

Figures 5 and 6 give an overview of the ocean mixed layer mean background profile structure at the ASI-2 location during the CASPER-East experiment. The dominant variability in the current magnitude and direction (Figure 6) correspond to the semidiurnal tidal frequency. Although the currents are relatively slow, they do have the capacity to generate a significant amount of turbulence due to the bottom friction. For this reason, for the purposes of this study an important aspect of the stratification profile (Figure 5) is the existence of some stratification in the first place. This is significant, because it ensures the separation between the wind-driven mixed layer turbulence, which is the subject of this study, and the bottom boundary layer turbulence, which would introduce unwanted interference during the model evaluation phase of the study.

As seen in Figure 5, the ocean is mostly layered except for a short period between days 300 and 302. During this two-day period, stronger wind and wave forcing (as seen in Figure 1) was able to eliminate all stratification and establish a well-mixed profile to the bottom. However, shortly after the high wind weather event passed, some stratification is re-established within the profile.

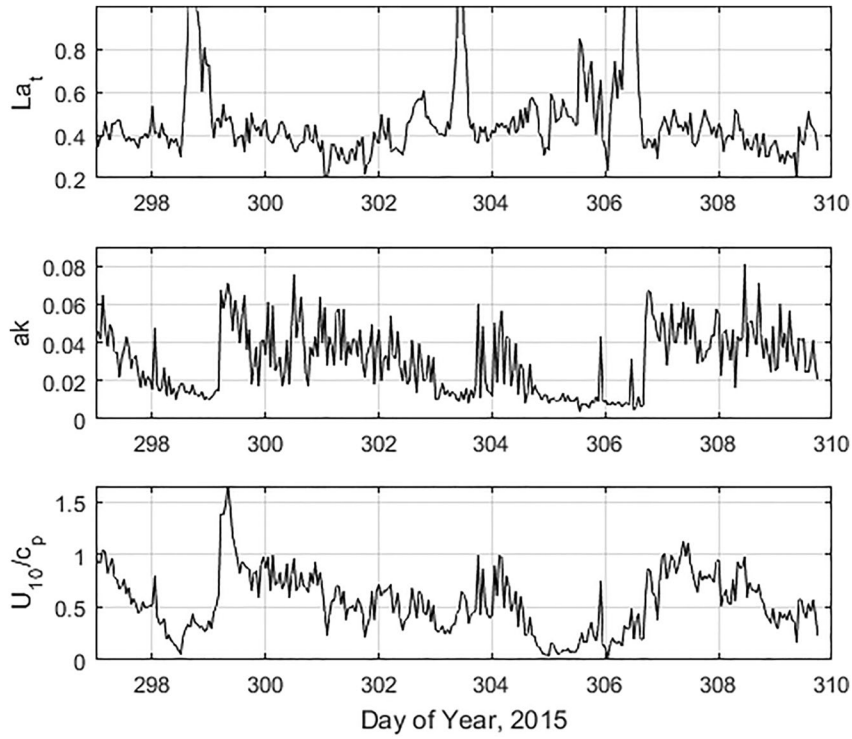

Figure 4. Time series of turbulent Langmuir number La_t, wave steepness ak, and inverse wave age U_{10}/Cp.

In addition, a strong stratification layer is seen starting on day 306 in the upper 0–10 m (Figure 5). It is primarily driven by relatively low values of salinity, consistent with expected values of Mid-Atlantic Bight shelf water in fall (Savidge et al., 2013), which normally reflect some influence from Chesapeake Bay waters. Strong density gradient layers like these can be effective barriers for the vertical transport of momentum, causing significant current differences across the layer. Indeed, in the vertical current shear subplot in Figure 6 we see an elevated shear near the top of the profile during this time period. Unfortunately, the ADCP is unable to resolve shallower depths, where most of the current shear presumably resides alongside the density gradient. Other areas with elevated vertical shear can be seen along the bottom of the profiles, for example, in the first two days, or during the 303–304 time frame. These instances indicate that there might be a significant TKE production in these areas, not caused by the surface forcing.

Based on the above observations and considerations, we conclude that this data set is suitable for a purely surface forcing driven turbulence study. But the turbulence measurements have to be taken as shallow as possible in order to allow a stratification and depth cushion to separate the bottom boundary layer turbulence, and also to amplify near-surface turbulence, which presumably decays with depth.

2.2.3. Underwater Turbulence

2.2.3.1. Underwater Turbulent Profiles Measured by ADCP

The best ADCP measurement for the purposes of turbulent profile extraction comes from the nominally vertical fifth beam, which because of it's smaller angle to the vertical than the slant beams, requires fewer near-surface bins be discarded due to sidelobe interference. Specifically, the shallowest uncontaminated bin depth of the fifth beam was found to be at 4 m, whereas the shallowest depth for the rest of the beams was at 13 m.

Ideally, a well leveled fifth beam is useful for the purposes of separating the mean and the turbulent components of the current. One potential difficulty with using even a well-leveled vertical beam is that turbulence may be biased by wave associated velocities. For this data set, a frequency gap exists between the wave dominated band and the turbulence band, so this source of bias has been removed with a 20 s Hanning filter. However the O(9deg) angle of the vertical beam (introduced by the hurricane Joaquin) means that velocities sampled by that beam include both vertical and horizontal components, so potentially energetic submesoscale turbulence or tidal variability in

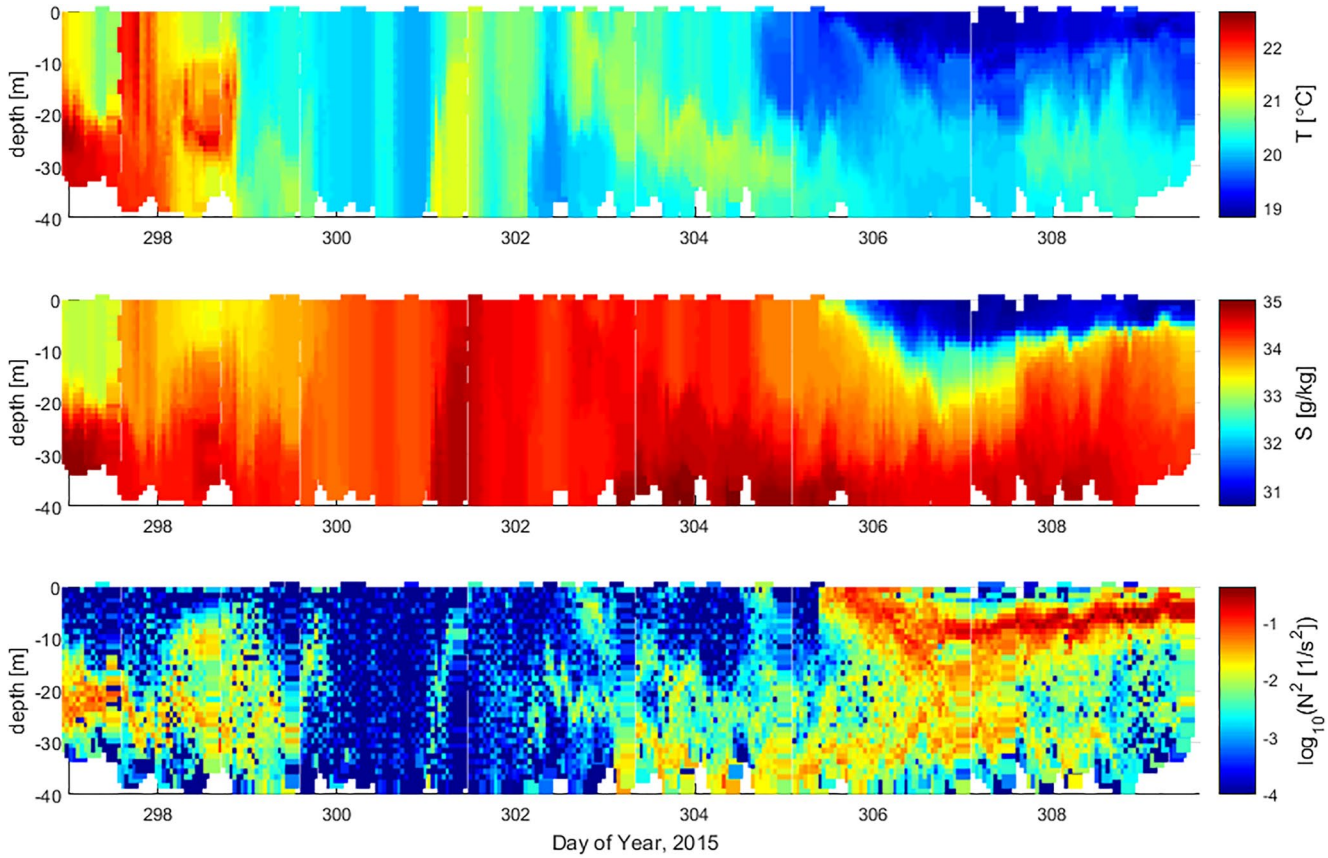

Figure 5. Temperature T, salinity S, and buoyancy frequency N time series derived from Slocum glider measurements.

the horizontal velocities may bias the vertical velocity estimates, which should contain no such variability. This was addressed by filtering out lower frequencies where these processes may dominate in the horizontal, and assuming the resulting band-passed beam velocities contain only isotropic turbulence. This filtering procedure was applied equally to all five beams. Specifically, the 4 Hz beam velocities were first averaged to 1 Hz, then depth matched to uniform depths by interpolation along the beams. While the shape of the velocity spectrum (not shown) dictated 20 s to be a good cutoff for the high frequency noise, the choice of the low frequency cutoff was less obvious. In agreement with the observations of the airborne survey (Savelyev et al., 2018), frequencies corresponding to submesoscales were found to be far more energetic, and therefore a slight change in the low frequency cutoff choice could easily double the total amount of TKE in the leftover signal. The same was observed for the fifth beam of the tilted instrument: the amount of low frequency TKE dominated over the useful higher frequency signal, although not as much as in the side beams. In the end, the choice of low frequency cutoff was set for 10 min, to allow most of the Langmuir circulation energy, but cutoff larger submesoscale turbulence. Therefore, the bandpass filtering limits on all five beams were set equally between 20 s and 10 min. Next, it had to be recognized that each of the 5 beams was looking at its own sampling volume in the horizontal, despite depth matching to identical vertical depths. Slant beam spread can be large in the upper half of the water column, so the assumption of homogenous turbulence in opposing beams is likely violated, preventing velocity vector reconstruction on these short time scales. Therefore, the next step operates on the time-filtered velocity measurements from 5 beams (u_1, u_2, u_3, u_4, u_5) separately.

To arrive at TKE time series, first each beam's velocity component was squared and then a one hour running average window was applied to it (i.e., $<u_1^2>$, $<u_2^2>$, $<u_3^2>$, $<u_4^2>$, $<u_5^2>$). The comparison of five resulting time series (not shown) indicates that while at a given instance they may vary significantly (by a factor of up to 2), the amount of TKE in each of the components averaged for the entire experiment was very similar (within 10%). This finding is used as a justification for assuming turbulent isotropy used in the next step. Under this assumption, the

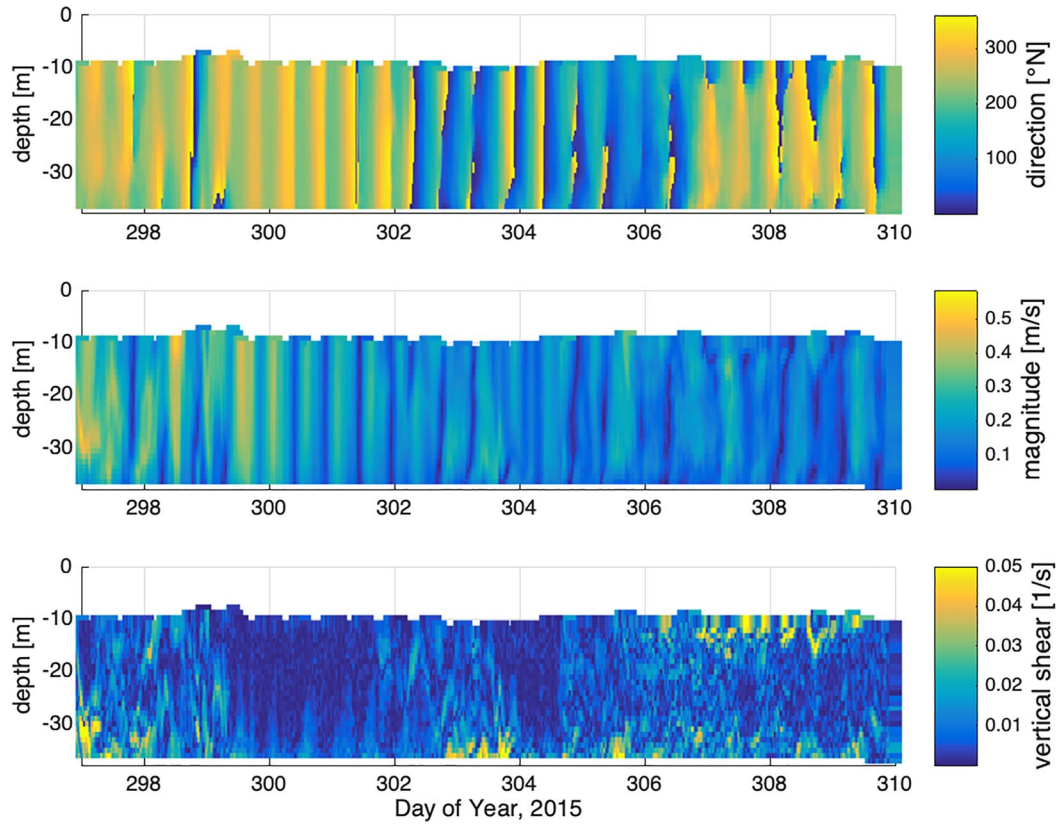

Figure 6. Time series of vertical profiles of mean horizontal velocity (its direction, magnitude, and vertical shear), measured by the bottom mounted acoustic Doppler current profiler.

three-dimensional TKE can be calculated as a triple of TKE measured in any given direction. To further increase confidence, at 13 m depth all five measurements were combined into a single TKE estimate, calculated as

$$q^2 = 3\left(< u_1^2 > + < u_2^2 > + < u_3^2 > + < u_4^2 > + < u_5^2 >\right)/5,$$

where $TKE \equiv q^2/2$. At 4 m depth only the fifth beam was available, simplifying this calculation down to

$$q^2 = 3\left(< u_5^2 >\right).$$

The observed TKE values at 4 and 13 m depths are presented in Figure 7 with red curves. These results and the overlaid model comparisons are described in detail in Section 3.

2.2.3.2. Turbulent Underwater Profiles Measured by a Glider

In addition to standard sensors on the Slocum glider described earlier (Section 2.2.2), it was also equipped with the MicroRider instrument manufactured by Rockland Scientific. In this setup, the glider provides a long-term autonomous sampling platform, moving through water smoothly and slowly. This type of motion is uncommon among ocean platforms - usually there is a vibrating motor or a tow cable under tension. Yet, it is exactly the kind of motion desired for the delicate turbulent velocity shear measurements with a highly sensitive airfoil-type shear probe on the MicroRider. This glider technique is becoming popular in recent years (e.g., Scheifele et al., 2018), providing an alternative to the more conventional labor-intensive free fall profiling. One potential shortcoming of glider-based turbulence measurements comes from the fact that unlike the vertical sampling of a free falling instrument, the measurements are taken along a gently sloping glider track. Interpreting the resulting measurements as a 1-D vertical profile requires an additional assumption of local horizontal homogeneity and isotropy of turbulent properties.

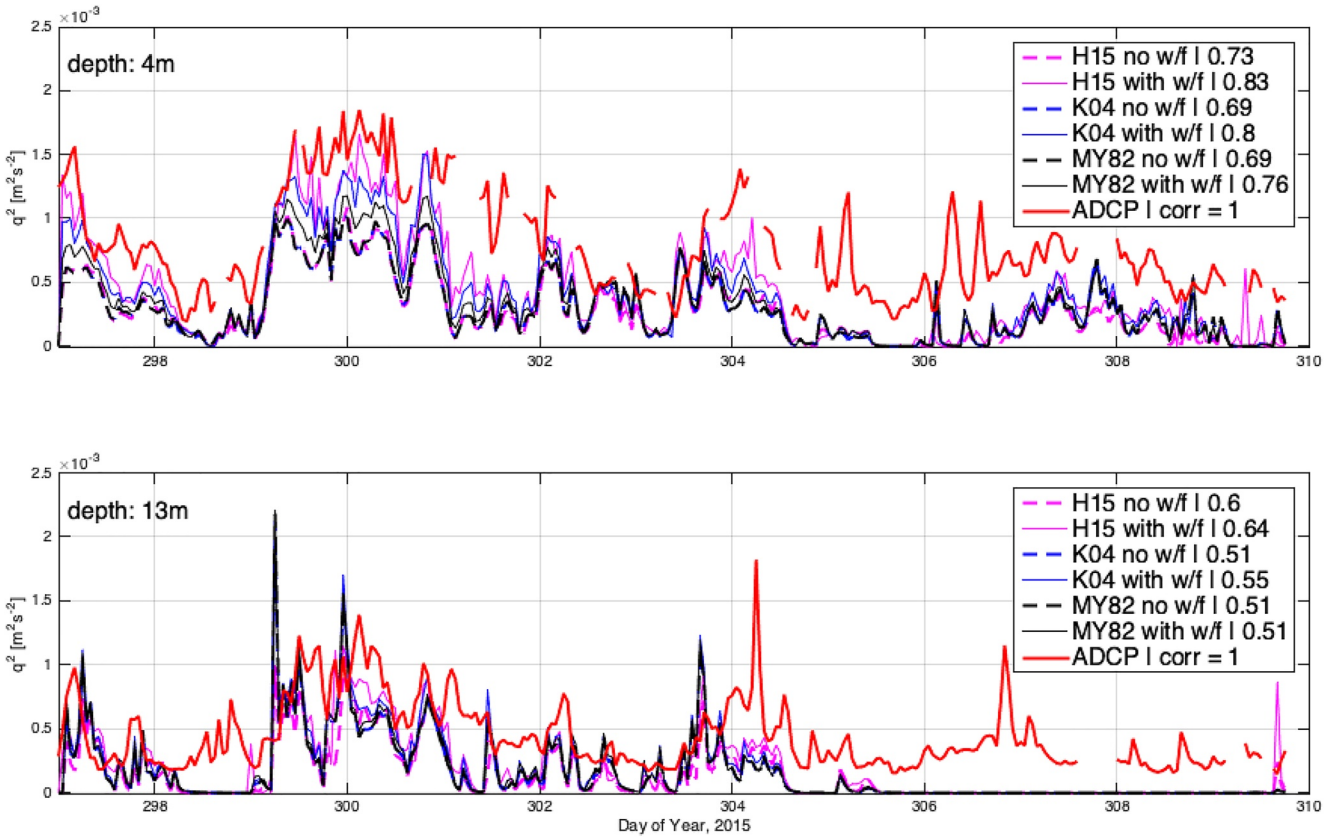

Figure 7. Turbulent kinetic energy time series at 4 and 13 m depth. In-water acoustic doppler current profiler measurements are compared to Navy Coastal Ocean Model model outputs with various turbulent closure schemes. The legend gives the correlation coefficient *R* between the acoustic Doppler current profiler measurement and each of the tested models.

An extensive and robust method for extracting the TKE dissipation rate (ϵ) from the turbulent velocity shear probe was developed by Lueck (2016). It is based on fitting a Nasmyth (1970) dissipation rate dependent turbulent shear spectrum function through the high frequency tail of the measured shear spectrum. Rockland Scientific developed and made available a set of Matlab functions enabling a near automated computation of TKE dissipation rate profiles starting from raw shear probe time series. In this study, the dissipation rate was estimated in 1 m depth bin increments along each glider dive. Figure 8 shows a subset of these measurements taken at 4 and 13 m depths to match ADCP time series in Figure 7. Each data point in Figure 8 represents an average of all dives sampled within one hour. This was done for consistency with ADCP and model output time intervals.

2.2.3.3. Airborne Remote Sensing of the Upper Ocean Turbulence

Measurements of the upper ocean turbulence by means of airborne remote sensing (Savelyev et al., 2018) were also a part of the effort to resolve underwater turbulence at the ASI-2 location during CASPER-East. Unlike a continuous measurement from an instrument in a fixed location, the aircraft samples large swaths of the ocean surface during each flight. But each flight lasts only a few hours and is nearly instantaneous in the context of a month-long survey. Savelyev et al. (2018) analyzed data from five flights over ASI-2 during year days 296, 302, 303, 308, and 309. The airborne sampling during each flight was done by means of tracking dye plume deformations, infrared mapping of spatial SST structures, and surface velocity mapping by means of a synthetic aperture radar. Some analysis was also done to relate the airborne turbulence measurements to the abovementioned ADCP turbulence measurements.

Findings of the airborne survey overwhelmingly pointed out the dominance of submesoscale turbulence in the study area. Both in terms of turbulent structures and energetics, submesoscale variability seemed to overshadow boundary layer turbulence or Langmuir turbulence processes or structures. These observations, in part, motivated the overall model evaluation strategy choice in the present study, that is, the choice to measure the background

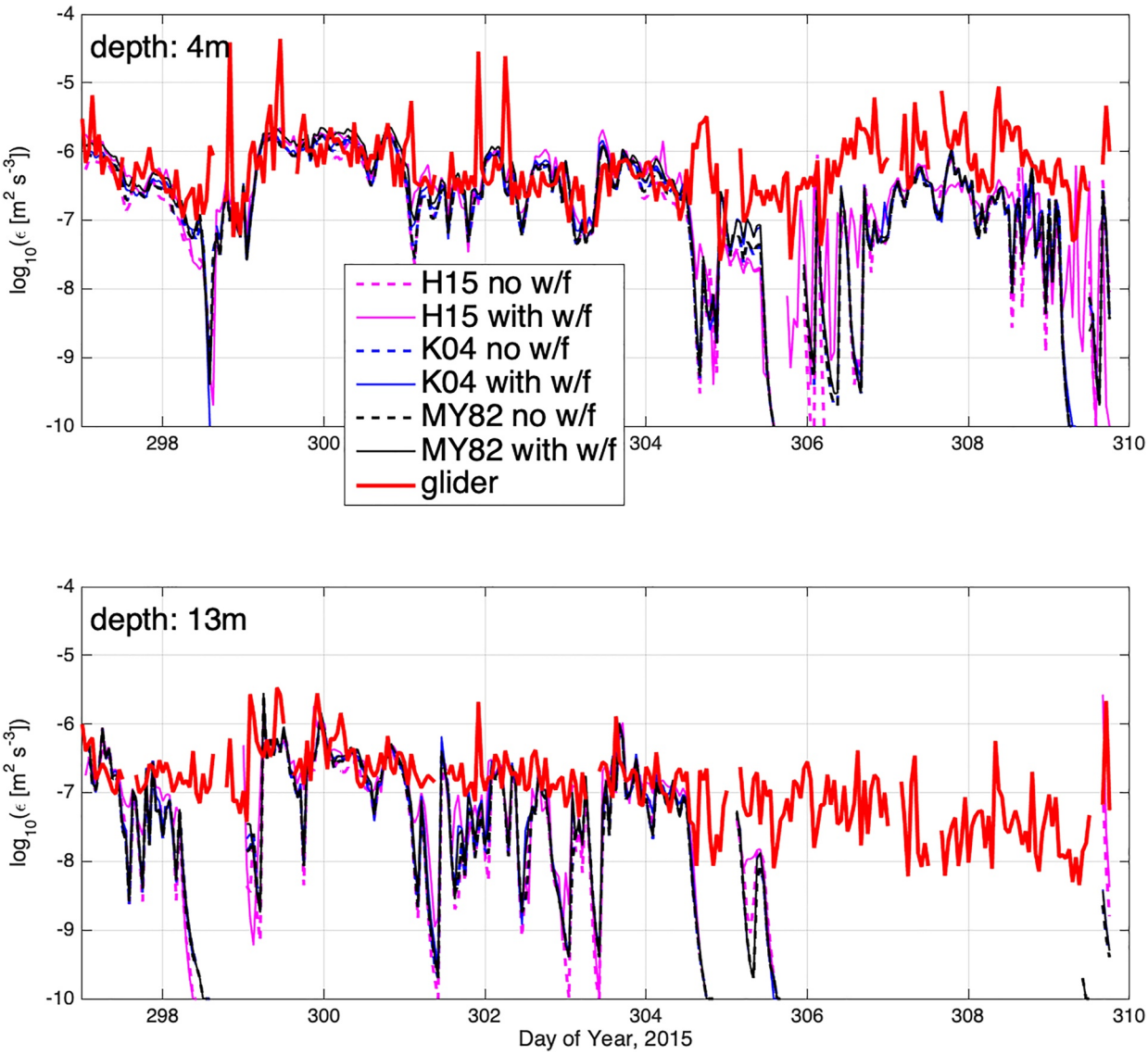

Figure 8. Time series of the turbulent kinetic energy dissipation rate at 4 and 13 m depth. In-water glider measurements are compared to Navy Coastal Ocean Model outputs with various turbulent closure schemes. Note, the values of modeled dissipation rate that fall below figure limits usually equal 0.

stratification structure (including its submesoscale fluctuations) and to relax the model profiles to it. But other than that, the results of the airborne surveys are not used further in this study in any quantitative manner.

2.3. NCOM Model Setup

All evaluated turbulent closure models used the same setup and forcing procedures. The model domain extended from the surface to the bottom at 38 m depth. It has 51 vertical layers, with grid spacing starting at 0.3 m near the surface, and increasing exponentially toward the bottom. At the surface, the atmospheric forcing (Figures 1–3) was used as a time-dependent boundary condition. More specifically, wind stress vector, air-sea heat flux, and SDC profile measurements were passed to the model. The SDC profile was calculated by integrating Stokes drift current contribution from each frequency across the measured wave frequency spectrum. The resulting SDC was non-directional, thus a potential misalignment between the wind and the SDC was not accounted for. Surface TKE flux boundary condition was calculated as

$$K_q \frac{\partial \left(q^2\right)}{\partial z} = m u_*^3,$$

where K_q is eddy viscosity, determined by q and l (length scale) variables, and m is a constant. The mean vertical profiles of temperature and salinity within the model were continuously relaxed to the vertical profiles measured by the glider (Figure 5) with a 30 min e-folding time scale. This particular time scale was chosen to allow averaging over at least several glider dives. In the horizontal direction, the model domain has the minimal possible 2×2 grid point domain, with doubly periodic boundary conditions. Model output was generated at 1 hr intervals.

While originally intended to be used in the same way as temperature and salinity, mean velocity profiles measured by ADCP were not used by the model. It is primarily due to the large measurement gap in the 0–10 m depth range. This is where much of the wind-driven current variability is expected to occur, making any extrapolation highly uncertain. Instead, the mean current was allowed to develop freely within the model in response to the surface forcing. It is certainly unfortunate that we had to make this simplification, as it can introduce a difference between measured and modeled turbulence due to the turbulence produced by non-local disturbances in background current shear. This simplification also likely causes an underestimation of bottom friction effects within the model, further emphasizing the need to stay closer to the surface and away from the bottom for a valid comparison between modeled and observed turbulence.

One of the simulations forced using the above procedure was an LES model described in Fan et al. (2020). In their study, similar to Large et al. (2019, 2021), they have expanded from an MW97-type idealized model to real ocean applications. While the model still uses doubly periodic boundary conditions, and solves turbulence in high resolution using the CL2 formulation, it was forced by realistic atmospheric and wave conditions and relaxed to realistic underwater profiles. Hence, presumably the small-scale turbulence generated in that model was similar to the turbulence that occurred in the ocean at that time and location. Fan et al. (2020) provides a detailed description of the numerical method, as well as a discussion of the phenomenology seen in the LES model output. In this paper we put LES work aside, leaving its comparison to turbulence measurements for a dedicated future publication. Instead, we will focus on other models forced in a similar way.

The turbulence closure parameterizations considered in this study are contained within NCOM. The model was setup in "1-D mode" with the smallest possible domain grid of 2×2 with doubly periodic boundary conditions. This method is similar to that used in Martin et al. (2013) and in MS17. Also, similarly, NCOM was forced by observed surface meteorology time series, which in the case of this study included measured waves. A significantly different method used here is the relaxation of the vertical temperature and salinity profiles within NCOM to the observed profiles, in a manner similar to the LES study by Fan et al. (2020). The three types of NCOM's turbulence closures evaluated are MY82, KC04, and H15. Note, these schemes were chosen among others because they were already implemented into NCOM as a part of earlier efforts. The comparison and contrast between these and other existing schemes was summarized above in Section 1.2 and the details of their implementation into NCOM were described in detail in Martin et al. (2013) and MS17. In this study we take the next logical step and critically evaluate the performance of these models against observed turbulence. The modeled variables here that are compared to observations are q^2 and ϵ time series taken at 4 and 13 m depths (Figures 7 and 8). Both variables are solved for by the turbulent kinetic energy budget equation within the turbulence closure model. TKE represented by q^2 is one of the primary variables, and the dissipation rate is one of the budget terms $\epsilon = 2 * q^3 / (B_1 * l)$, where the turbulent length scale l is the other primary variable solved for by the turbulent closure model. The output of q^2 and ϵ time series at 4 and 13 m depths from all model runs is shown alongside comparable observations in Figures 7 and 8. An output from each model is shown with and without the wave forcing (w/f) formulation enabled.

3. Model Evaluation Results

Figures 7 and 8 give an initial overview of the comparison between various models and the observations. At the first glance, there appears to be a good agreement, with noticeable correlation between fluctuations in various curves. But upon closer examination, many features predicted by models do not find support in observations. The overall observed q^2 and ϵ levels appear to be consistent, or at times higher than the corresponding output from any model. Note, the overall offset between modeled and observed TKE levels is not a meaningful comparison metric, because, as was mentioned in Section 2.2.3.3, it is strongly controlled by the arbitrary choice of the high pass filter cutoff frequency. Instead, the comparison should focus on relative variability, such as long or short term periods of increased wind forcing and consequent relaxation. One of the most obvious disagreements is the

widening model underperformance in the second half of the experiment. This is especially evident at the deeper 13 m level, where both modeled TKE and dissipation are near zero, whereas the observations show substantial levels of turbulence. Notwithstanding the supposed improvements from older to newer model versions, all models were found unable to account for this turbulent activity. The agreement appears to be closer during the windier first half of the experiment. However, even here the variability predicted largely in unison by model does not quite match the variability seen in observations: the peaks and troughs are often misaligned, the amplitude of TKE fluctuations appears to be larger in models, and the turbulence decay after the high wind event (days 301–303) appears to be faster in models. Here, the latest tested model (H15 with wave forcing) does appear to improve and somewhat slow down that decay. Much of the following discussion is dedicated to understanding and quantifying these differences, pointing out modeling successes, as well as sources of remaining errors.

As the initial quantitative metric for the model agreement with observations we chose the correlation coefficient R. The resulting values of R, calculated between the model output of each model and the ADCP time series, are given in the legend of Figure 7. Similar calculations were attempted for the dissipation rate time series, but were found unreliable. The behavior of the dissipation rate signal is highly noisy and intermittent, sometimes jumping by orders of magnitude from one point to the next (note, Figure 8 gives dissipation rate on a logarithmic scale). While qualitatively the curves compare well, we found that any kind of quantitative comparison metric was highly dependent on arbitrary filtering choices, and also on the periods of extremely low dissipation when the model drops to zero. Averaging over time or depth to boost confidence offers little help due to the highly nonlinear nature of the signal. For these reasons we leave quantitative dissipation comparison at this stage and proceed further only with TKE.

The values of R for various models at both depths are orderly. Each new generation of the turbulence closure model offers an incremental performance improvement, and enabling the wave forcing formulation within each generation results in a slight increase of correlation as well. For example, as seen at 4 m depth, the model performance advanced from $R = 0.69$ for the original no-wave MY82 model to $R = 0.83$ for the latest wave-dependent H15 model. We see similar orderly improvement at 13 m depth, but correlation coefficients are lower across the board. This is, presumably, because the surface forcing and consequent small-scale turbulence are less prominent at this deeper level, giving rise to relative importance of other turbulence production mechanisms not included in the models, such as bottom effects. Based on the findings so far, in the following analysis we focus on 4 m deep TKE as the purest measurement, and on H15 wave-dependent model, as the best performing model.

Overall, the results indicate steady and orderly improvements in model performance from older to newer generations. However, the output of all models is grouped closely together, offering only small incremental improvements, without closing much larger gaps between modeled and observed values. The next section is dedicated to a more in-depth analysis of the remaining differences, on identifying missing physics and quantifying remaining errors as functions of secondary sensitivity parameters.

4. Remaining Error Analysis and Discussion

In the following discussion we provide a more in-depth analysis of the remaining problems and limitations, particularly sources of remaining model errors, as well as measurement strategy shortcomings. We push the extent to which the model output and the observations at the ASI-2 can be compared and evaluated side-by-side, and what can be learned from that comparison. One of our primary interests here is to test some secondary parameters (e.g., given in Figure 4), to gauge their ability to close the remaining modeling errors, and thus indicate the most pressing areas for further turbulent closure model development.

The obstacles preventing further increase in the correlation between the model and observations can be thought of in three categories. First is the observational uncertainty, which could be the lack of sufficiently high resolution and quality data, the instrument error or noise, noise filtering imperfections, the somewhat arbitrary separation of mean and turbulent components of the current, the lack of sufficient statistical significance within each ensemble, or other assumptions made about turbulence behavior, such as isotropy. The second category is the part of the measured turbulence that was generated by mechanisms other than the local surface forcing, and hence not represented in the model. These could include non-locally generated turbulence that is advected into the sampling volume, turbulence generated within the column by additional current shear (e.g., the high shear associated with the high-density gradient during 306–310 days), or the turbulence generated near the bottom, which could rise

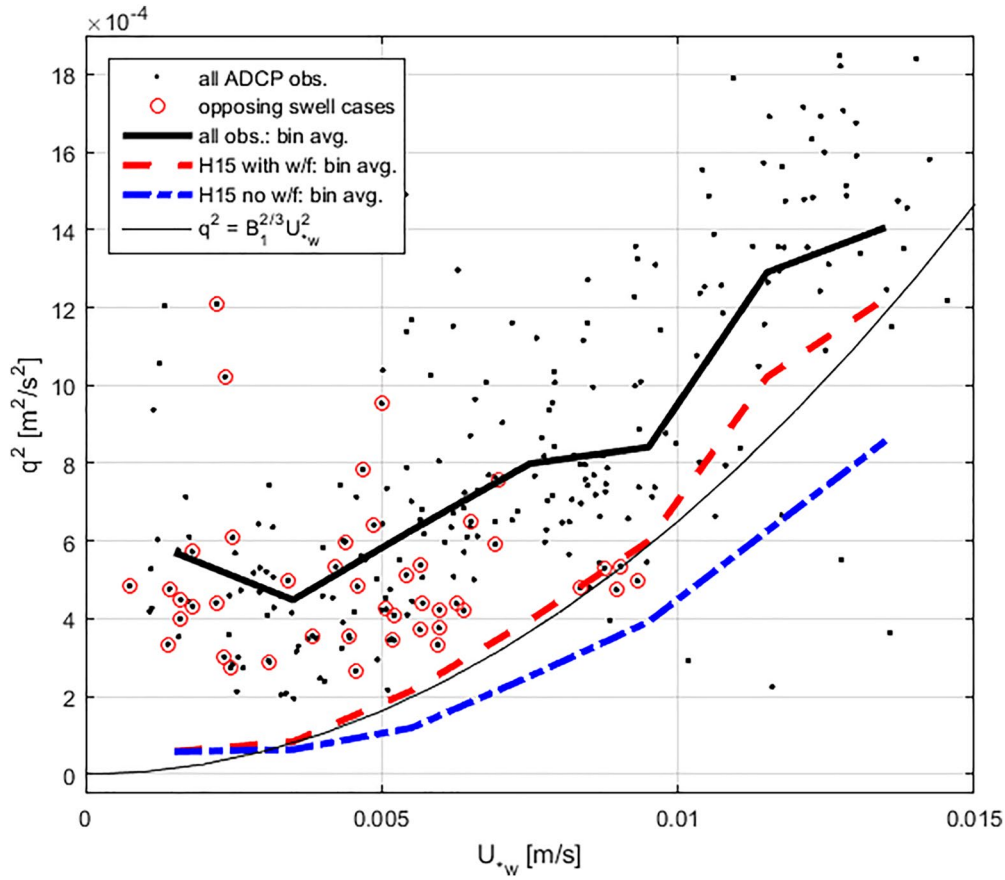

Figure 9. Turbulent kinetic energy dependence on the wind friction velocity. Individual hourly observations are shown with black dots (cases with opposing swell conditions highlighted with red circles). Black curve represents bin-averaged observations, red dashed and blue dot-dashed curves are similar bin averages calculated for hourly model output points (not shown) for H15 model with and without wave forcing, respectively. The thin black line shows a commonly used turbulent kinetic energy surface boundary condition, defined in the legend.

close to the surface in the absence of stratification (e.g., days 300–302). Both the first and the second categories are observational or sampling strategy problems, which even a perfect model is not expected to be able to address. Here, the disagreement seen at 4 m depth between model and observations during the second half of the experiment both in Figures 7 and 8 deserves a special explanation. Since the model was relaxed only to the temperature and salinity profiles associated with what appears to be a low salinity plume, the impact that this plume might have had on the vertical profile of mean current was not accounted for (models were not relaxed to observed velocity profiles). While the velocity profile measurements were not available in the top 10 m, it is likely that the large density gradient was associated with a strong current shear. Some indication of this elevated shear can be seen below 10 m in Figure 6. Since the existence of this elevated shear was not passed to the model, in could explain the lack of elevated TKE and ϵ levels (Figures 7 and 8) in the model output.

The third and final category of the disagreement between the model and the observation is the remaining model deficiency. The model architecture, of course, contains a multitude of assumptions and simplifications. Most of them, presumably, are entirely valid and their use does not deteriorate the model performance in a measurable way. But perhaps there are some that do. Below we look at the model error as a function of some surface forcing parameters, in an attempt to reveal which physical processes are captured well by the model and which are not.

4.1. Wind Speed Dependence

As expected, the dominant source of TKE appears to be the wind forcing. In Figure 9 all TKE data is sorted according to the water side wind friction velocity u_{w*}. It is related to the wind speed and the wind stress as

$\tau = \rho_w u_{w*}^2 = \rho_a C_D U_{10}^2$, where ρ_w and ρ_a are water and air densities, and C_D is the drag coefficient. The figure shows all hourly TKE observations with black dots and the thick solid black curve is the result of their bin-averaging. Then the black curve is compared to identically computed red dashed and blue dot-dashed curves, which used hourly model outputs instead of observations (note, the figure does not show individual hourly model output data points to avoid over-crowding). The two models shown here are H15 with and without the wave forcing enabled, respectively. The thin solid black line corresponds to the original MY82 TKE surface boundary condition: $q^2 = B_1^{2/3} u_{*w}^2$.

When comparing observations to other curves in Figure 9, the most striking feature is the significant positive offset, especially in low wind speeds. According to any surface forcing driven turbulence model, or common sense, there should be near zero TKE in calm wind conditions. Yet the observations tell a different story. This offset clearly indicates the large role that the errors of the first two categories (measurement and sampling strategy imperfections) play in the overall model-observations disagreement. Clearly, some other sources of TKE are present in the sampling volume, which the model was not equipped to reproduce.

Taking the wind forcing as the primary parameter, next we search for the secondary parameter, which will best capture remaining model-observations disagreements. For this, we fit a second order polynomial function through the bin-averaged curve, then subtract its value from each individual data point. The remaining amount of TKE is called Δq^2 and is expressed as a percentage of the bin-averaged value for that wind speed. The value of Δq^2 was calculated for every data point, both observed and modeled. It essentially represents the relative deviation of TKE from its expected value for a given wind speed. This way, the wind speed influence is removed and all data points are normalized, enabling an unobstructed view of secondary dependencies.

4.2. Wave Forcing

With the wind speed influence removed, next we look at the secondary influence that waves might have on TKE, independently of the wind. Among a number of possible wave-dependent parameters, here we show two non-dimensional numbers: the turbulent Langmuir number, defined as $La_t = (u_{*w}/u_{st})^{1/2}$, and the dominant wave steepness, defined as $ak = 2H_s\pi^2/gT_p^2$. Time series of both parameters are shown in Figure 4. Figure 10 shows the dependence of TKE on these parameters. Because the search for secondary dependencies is inherently associated with degrading statistical confidence, the figure shows only the simplest linear fit through the available data. Even this fit is highly uncertain, and should be used not as a quantitative guidance, but as a qualitative assessment of a positive or negative correlation. Solid and dashed thick black lines show such linear fits through observed and modeled data points, respectively. Note, because Δq^2 is a normalized quantity, here we are primarily interested only in the slope of the resulting fit, as an indication of TKE's secondary dependence on the parameter of choice.

Both the observations and the model show remarkably similar and strong dependencies on wave steepness, ak, as the secondary parameter. The black curves in the figure demonstrate that outside of wind dependence, TKE varies within a ~40% range, depending on the wave conditions. This finding qualitatively supports the expected response to wave forcing, where at higher ak more TKE is produced, resulting in a positive TKE ~ ak correlation. The wide range of the wave dependence found here also points to the overall importance of including wave forcing physics into turbulence closure models.

An identical calculation, done with La_t instead of ak, also supports the wave forcing concept, although not as strongly (also shown in Figure 10). According to the definition, lower La_t means relative importance of wave forcing over wind forcing, and, indeed, for a given wind forcing, data points with enhanced wave forcing show higher TKE. However, the total change of the TKE fit across the range of measured La_t is only ~10%, indicating that La_t is not as effective as ak for the purposes of capturing wave-dependent TKE change. Moreover, most of La_t data points were found to be tightly grouped around the value of 0.4, offering little variation in TKE for all but the most extreme cases.

Another commonly used non-dimensional parameter used to describe wind-wave forcing is the inverse wave age, U_{10}/C_p (see its time series in Figure 4). As the name implies, its value indicates if the waves are still young and have room to grow under the given wind, or if they are in mature age, where the little remaining speed difference between waves and wind is unable to support further growth. The wave age record shows that waves were either mature or over-mature during most of the CASPER-East experiment. This means that for the most part,

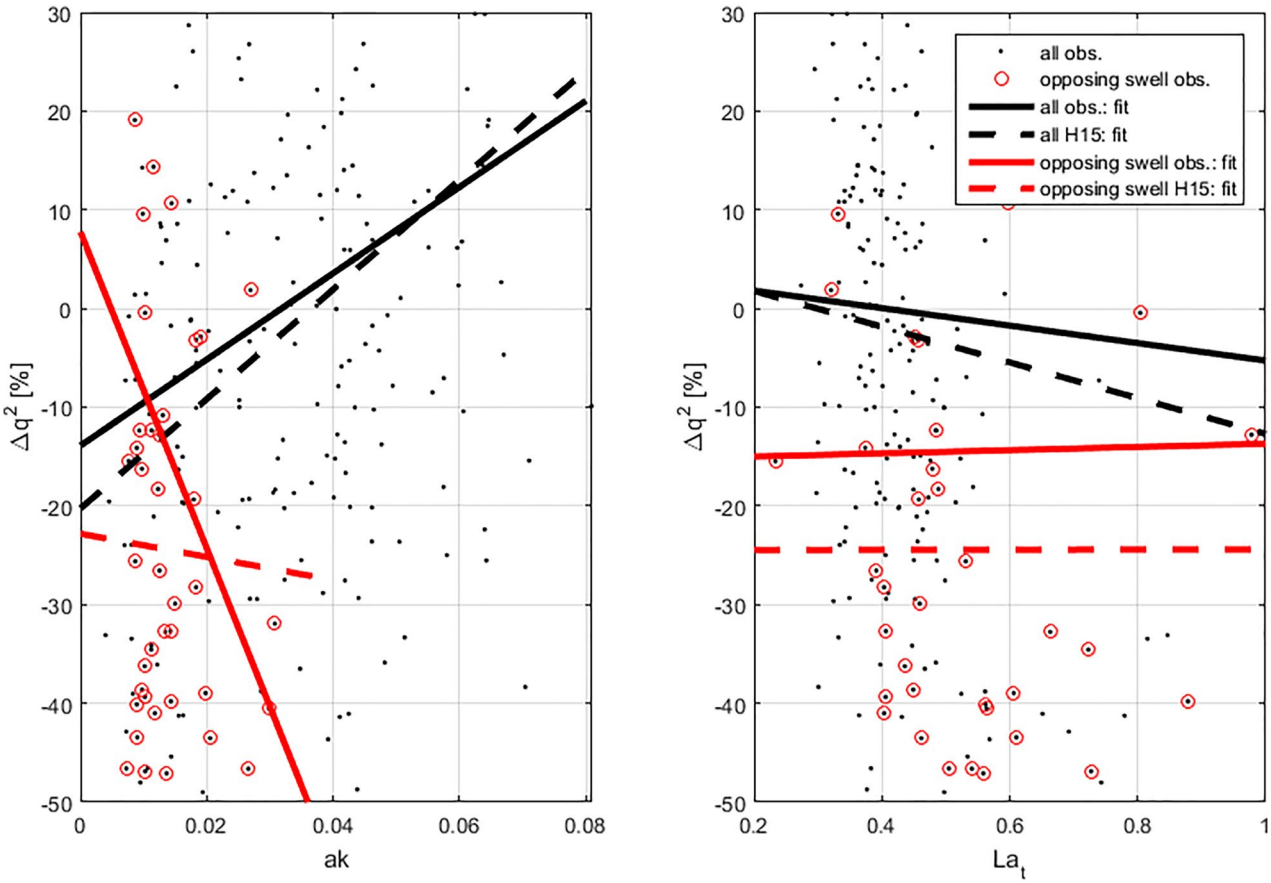

Figure 10. Relative turbulent kinetic energy change as a function of two control parameters: Dominant wave steepness *ak*, and turbulent Langmuir number La_t. Black dots are individual observations, solid black lines are their linear fits. Dashed black lines are identically obtained fits based on H15 model output points (not shown). Red circles highlight data points for which swell was opposing wind direction. Red solid and dashed curves are the same as black curves, calculated for the highlighted subset.

the dominant part of the wave spectrum was not generated locally, but rather was a result of a distant high wind weather event (see wave age climatology maps in Hanley et al., 2010). The same conclusion can be reached by examining dominant wave and wind direction differences (Figure 1). While the wind direction was highly variable, waves remained relatively steady out of East or Northeast, presumably propagating as swell waves from distant offshore storms in the North Atlantic.

This imbalance and misalignment between the winds and the dominant waves is common for the coastal mid-Atlantic region. In fact, it is one of the desired complicating features that sets this region apart from the more balanced and aligned open ocean wind-wave conditions, such as around OWS Papa. This prompts us to take a closer look at the role of this misalignment, at how it modifies wind-wave forcing, and impacts the TKE. More specifically, a subset of data points was chosen with opposing wind and dominant waves, defined as instances with at least 135° difference between the two directions. First, this subset of data points is highlighted with red circles in Figure 9. Right away, it can be seen that for the most part these points tend to fall below the bin-averaged values. This tendency becomes particularly distinct as the wind speed rises, see the cluster of 5 highest wind cases (around $u_{w*} = 0.009$ m/s) well below their respective bin-averaged value. The subset of opposing swell points is carried over to the next analysis step, where the secondary wave parameter analysis is applied to it, the same way it was applied to the full data set earlier. The resulting linear fits for the observations and for the model output are shown in Figure 10 with thick red solid and dashed lines, respectively. The observed TKE for this subset shows a strong negative dependence on *ak*, which is opposing the trend seen for the overall data set (black lines).

To summarize, the analysis of the opposing swell subset shows a reduction of TKE, which amplifies particularly in stronger wind and in higher wave steepness conditions. This latter finding contradicts the behavior expected

from increased wave forcing, which expects more TKE under steeper waves, in agreement with the rising black curves in Figure 10. Moreover, this time the dashed red curve, representing the model output, does not reproduce the severe TKE drop-off seen in observations (solid red curve). All this evidence suggests that there might be another counteracting wave-dependent physical mechanism taking place in these conditions, which is also a wave forcing mechanism, but which is not included in the H15 turbulence closure model.

We hypothesize that the airflow separation behind wave crests can be that secondary counteracting wave-dependent mechanism, responsible for TKE reduction. The existence of such mechanism is consistent with the findings of an air-sea momentum flux reduction in wind-opposing-swell cases, as reported from field observations by García-Nava et al. (2012). This wave-driven effect is thought to be caused by an increased likelihood of the airflow separation behind wave crests, which effectively shelters wave trough regions of the water surface from the full impact of wind forcing. Similar negative impact on surface TKE was observed in laboratory conditions by Savelyev et al. (2020), where the airflow separation caused by increased wave steepness was found to reduce surface TKE driven by steady winds. None of the models evaluated in this study are equipped to account for airflow separation effects, and hence are unable to reproduce the observed TKE reduction. Meanwhile, as Figure 10 demonstrates, this overlooked phenomenon can cause an up to a factor of 2 model error in some scenarios. We hypothesize that this error was previously overlooked, because model evaluation and calibration was primarily based either on idealized LES outputs (e.g., MW97, Harcourt & D'Asaro, 2008), or on weather station Papa data (e.g., MS17), neither of which provide a range of wind-wave misalignment scenarios. Meanwhile, in coastal environments, such as the one analyzed here, out of balance and misaligned seas are commonplace. Therefore, to enable accurate modeling of the coastal ocean, a more complete air-sea flux boundary condition needs to be developed, capturing a wider range of wind-wave forcing variability. Additionally, perhaps accounting for wind-wave direction difference in the definition of La_t, as suggested by Kantha et al. (2010), will mitigate at least some of these errors. This task, however, comes with a complication of choosing the appropriate wave direction in mixed seas. It can be especially ambiguous in situations with significant swell coming from an off-wind direction (see discussion on the topic by Kantha et al., 2014).

4.3. Air-Sea Heat Flux

In calm wind and wave forcing conditions, the destabilizing effect of an upward air-sea heat flux can emerge as another dominant mechanism of TKE production in the water column. Similar to the wave forcing, here we test the total air-sea heat flux, that is, the rate of ocean warming (see time series in Figure 3), as a secondary parameter to evaluate its impact on TKE. As such, the relative change of TKE (with wind speed dependence removed) is given as a function of the ocean warming in Figure 11. The figure shows individual observations with black dots and their liner fit with the solid black line, whereas the dashed dotted line shows a similarly obtained fit through the H15 hourly model output realizations. Both the observed and the modeled lines are sloping as expected, suggesting more TKE in case of ocean cooling (i.e., convective forcing) and less TKE in case of warming (i.e., stabilizing stratification). This effect appears to be present even in the wind and wave dominant conditions of the CASPER-East experiment, and the H15 model appears to capture its impact on TKE correctly, at least within the statistical uncertainty limits. To investigate this effect to a fuller extent, a different season within this region, or a different region altogether with a wider air-sea heat flux range would have been more appropriate. In the case of this late fall experiment, the intense summer warming had recently switched to winter cooling, hence the heat flux values are relatively small. But the performance of H15 that we do see for the range of ocean warming observed here appears to be in order and does not motivate a more detailed dedicated investigation.

4.4. Submesoscale Turbulence

One of the core assumptions in this study is the separation of the slowly varying background stratification from the small-scale boundary layer turbulence. It is assumed that the glider can capture this background stratification and the model can be slowly relaxed to it with a 30 min e-folding time scale. However, as was shown with a variety of airborne remote sensing methods by Savelyev et al. (2018), CASPER-East study area around the ASI-2 location was rich with multi-scale submesoscale activity. The submesoscale structures were often associated with sharp fronts, which can introduce abrupt change in the stratification profile. This source of error is amplified by the fact the the glider is not always perfectly able to maintain station, and also has to move horizontally during each dive. Some evidence of sharp transition events does exist in the glider and ADCP data used in this study,

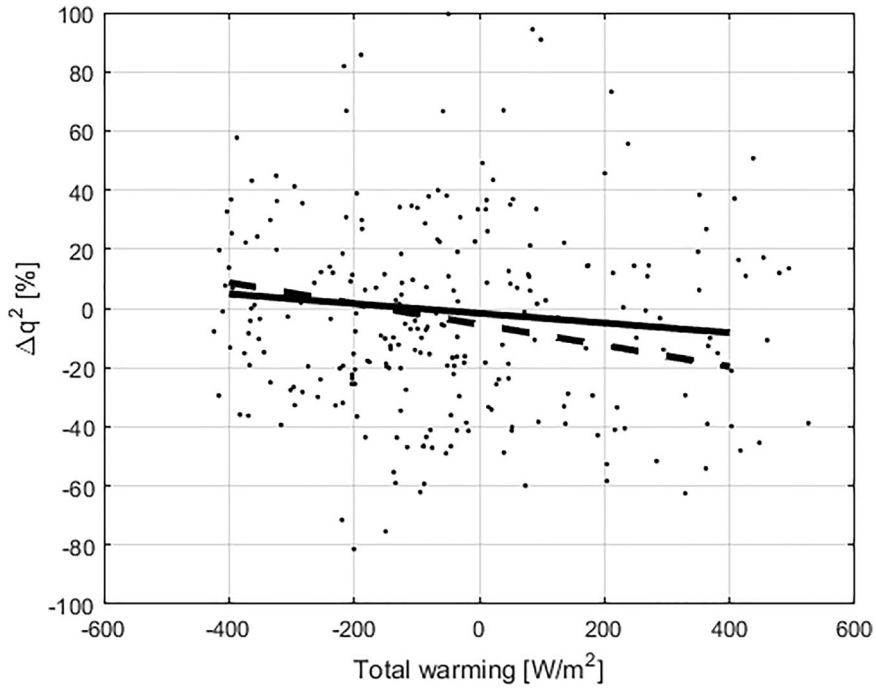

Figure 11. Similar to Figure 10, here relative turbulent kinetic energy change is shown as a function of the total air-sea heat flux. Black dots are individual observations, solid black line is their linear fit. Dashed black line is an identical fit through H15 model output points.

although they are hard to distinguish from measurement errors. The impact of these events is felt most strongly in an LES model (Fan et al., 2020), whereas NCOM is equipped with a set of numerical safeguards and stabilizers preventing unrealistically abrupt behavior. But whether overreacting or underreacting, none of the models appear to be equipped with the physics necessary to address the submesoscale turbulence challenge.

This, primarily, speaks not to the deficiency of the models, but to the limitations of the assumptions behind the entire model test methodology used in this and other turbulence closure model development efforts. The fundamental assumption is that the background variables are changing slowly and that their horizontal variability can be assumed homogeneous and steady-state on the small scales relevant to the boundary layer turbulence. But the reality is that this assumption does not always hold. As was shown in submesoscale spectrum obtained from airborne imagery of dye plumes (Figure 13 in Savelyev et al., 2018), there is no spectral energy gap between boundary layer turbulence scales and submesoscale ocean turbulence. It is entirely possible that these processes interact by means of a turbulent cascade energy flux or through more intricate mechanisms described by the submesoscale frontal dynamics theory (e.g., McWilliams, 2016). Whatever the nature of this interaction is, it violates the basic assumption of steady state and horizontal homogeneity, and hence indicating the applicability limit for the 1-D vertical turbulence closure formulation paradigm. Perhaps the greatest challenge before the next generation turbulence closure models would be to understand the impact of submesoscales on the vertical mixing, to find an effective and efficient way to model it, and to merge these submesoscale-induced mixing models with the existing turbulence closure model framework.

5. Conclusions

By using state-of-the-art in-situ capabilities, this study collected a comprehensive coastal ocean sample data set necessary to constrain and evaluate the performance of small-scale upper ocean turbulence models. Turbulent closure schemes by Mellor and Yamada (1982), Kantha and Clayson (2004), and Harcourt (2015) were evaluated within the Navy Coastal Ocean Model in a 1-D vertical test mode. The model was forced by measured surface meteorology and relaxed to measured vertical stratification time series. The output of the turbulence

closure schemes, particularly the turbulent kinetic energy, was compared to observed values to evaluate the model performance.

The results of the tests demonstrate a steady incremental improvement of the turbulence closure models with each new generation introduced over the past few decades of their development. The correlation with measured turbulence went from 0.69 for the earliest tested model formulation to 0.83 in the latest version. Based on these results, along with other positive test results in the open ocean (Martin & Savelyev, 2017), it appears the latest version offers a significant measurable improvement over the previous or any earlier versions. However, the overall correlation between the best performing model and observations remain weak, motivating further development of turbulent closure schemes.

The analysis of the remaining errors revealed three potential sources of errors. The first two are non-modeling errors, having to do with the sampling instrumentation and strategy, with the third source lying within the remaining model imperfections. More specifically, a wave-dependent error in the air-sea turbulent kinetic energy boundary condition was identified and quantified. The error becomes substantial (factor of 2) in cases with winds opposing swell and in waves with high steepness. The observations report reduced amount of turbulent kinetic energy in these scenarios, whereas the model does not. This is hypothesized to be due to the airflow separation physics, not accounted for in the model. Another source of modeling error likely stems from the presence of submesoscale fronts, which were commonplace in the study area, yet their physics is not accounted for in the model. A detailed quantitative analysis on the likelihood and magnitude of this error is left for future studies.

Data Availability Statement

Data used in this manuscript (*.mat), along with the Matlab script (generate_figures.m) to generate all figures above can be downloaded here: https://doi.org/10.5281/zenodo.6454602.

Acknowledgments
This research was funded by ONR-NRL program element 61153N (WUs BE023-01-41-1C04, NRL/JA/7320-19-4508, and 72-1R25) and by ONR grant N0001418WX01087. Funding for Savidge was provided by NSF grant OCE-1540648.

References

Ardhuin, F., & Jenkins, A. D. (2006). On the interaction of surface waves and upper ocean turbulence. *Journal of Physical Oceanography*, *36*(3), 551–557. https://doi.org/10.1175/jpo2862.1

Ardhuin, F., Rascle, N., & Belibassakis, K. A. (2008). Explicit wave-averaged primitive equations using a generalized Lagrangian mean. *Ocean Modelling*, *20*(1), 35–60. https://doi.org/10.1016/j.ocemod.2007.07.001

Craig, P. D., & Banner, M. L. (1994). Modeling wave-enhanced turbulence in the ocean surface layer. *Journal of Physical Oceanography*, *24*(12), 2546–2559. https://doi.org/10.1175/1520-0485(1994)024<2546:mwetit>2.0.co;2

Craik, A. D., & Leibovich, S. (1976). A rational model for Langmuir circulations. *Journal of Fluid Mechanics*, *73*(3), 401–426. https://doi.org/10.1017/s0022112076001420

Denman, K. L., & Miyake, M. (1973). Upper layer modification at ocean station Papa: Observations and simulation. *Journal of Physical Oceanography*, *3*(2), 185–196. https://doi.org/10.1175/1520-0485(1973)003<0185:ulmaos>2.0.co;2

Fan, Y., Yu, Z., Savelyev, I., Sullivan, P. P., Liang, J.-H., Haack, T., et al. (2020). The effect of Langmuir turbulence under complex real oceanic and meteorological forcing. *Ocean Modelling*, *149*, 101601. https://doi.org/10.1016/j.ocemod.2020.101601

García-Nava, H., Ocampo-Torres, F. J., Hwang, P. A., & Osuna, P. (2012). Reduction of wind stress due to swell at high wind conditions. *Journal of Geophysical Research*, *117*(C11). https://doi.org/10.1029/2011jc007833

Gargett, A. E., & Wells, J. R. (2007). Langmuir turbulence in shallow water. Part 1. Observations. *Journal of Fluid Mechanics*, *576*, 27–61. https://doi.org/10.1017/s0022112006004575

Hanley, K. E., Belcher, S. E., & Sullivan, P. P. (2010). A global climatology of wind–wave interaction. *Journal of Physical Oceanography*, *40*(6), 1263–1282. https://doi.org/10.1175/2010jpo4377.1

Harcourt, R. R. (2015). An improved second-moment closure model of Langmuir turbulence. *Journal of Physical Oceanography*, *45*(1), 84–103. https://doi.org/10.1175/jpo-d-14-0046.1

Harcourt, R. R., & D'Asaro, E. A. (2008). Large-eddy simulation of Langmuir turbulence in pure wind seas. *Journal of Physical Oceanography*, *38*(7), 1542–1562. https://doi.org/10.1175/2007jpo3842.1

Kantha, L., Lass, H. U., & Prandke, H. (2010). A note on Stokes production of turbulence kinetic energy in the oceanic mixed layer: Observations in the Baltic Sea. *Ocean Dynamics*, *60*(1), 171–180. https://doi.org/10.1007/s10236-009-0257-7

Kantha, L., Tamura, H., & Miyazawa, Y. (2014). Comment on "Wave-turbulence interaction and its induced mixing in the upper ocean" by Huang and Qiao. *Journal of Geophysical Research: Oceans*, *119*(2), 1510–1515. https://doi.org/10.1002/2013jc009318

Kantha, L. H., & Clayson, C. A. (2004). On the effect of surface gravity waves on mixing in the oceanic mixed layer. *Ocean Modelling*, *6*(2), 101–124. https://doi.org/10.1016/s1463-5003(02)00062-8

Kukulka, T., Plueddemann, A. J., Trowbridge, J. H., & Sullivan, P. P. (2010). Rapid mixed layer deepening by the combination of Langmuir and shear instabilities: A case study. *Journal of Physical Oceanography*, *40*(11), 2381–2400. https://doi.org/10.1175/2010JPO4403.1

Kukulka, T., Plueddemann, A. J., Trowbridge, J. H., & Sullivanet, P. P. (2009). Significance of Langmuir circulation in upper ocean mixing: Comparison of observations and simulations. *Geophysical Research Letters*, *36*(10), L10603. https://doi.org/10.1029/2009GL037620

Langmuir, I. (1938). Surface motion of water induced by wind. *Science*, *87*(2250), 119–123. https://doi.org/10.1126/science.87.2250.119

Large, W. G. (1996). An observational and numerical investigation of the climatological heat and salt balances at OWS Papa. *Journal of Climate*, *9*(8), 1856–1876. https://doi.org/10.1175/1520-0442(1996)009<1856:aoanio>2.0.co;2

Large, W. G., Patton, E. G., DuVivier, A. K., Sullivan, P. P., & Romero, L. (2019). Similarity theory in the surface layer of large-eddy simulations of the wind-wave-and buoyancy-forced southern ocean. *Journal of Physical Oceanography*, *49*(8), 2165–2187. https://doi.org/10.1175/jpo-d-18-0066.1

Large, W. G., Patton, E. G., & Sullivan, P. P. (2021). The diurnal cycle of entrainment and detrainment in LES of the southern ocean driven by observed surface fluxes and waves. *Journal of Physical Oceanography*, *51*(10), 3253–3278. https://doi.org/10.1175/jpo-d-20-0308.1

Li, Q., Reichl, B. G., Fox-Kemper, B., Adcroft, A. J., Belcher, S. E., Danabasoglu, G., et al. (2019). Comparing ocean surface boundary vertical mixing schemes including Langmuir turbulence. *Journal of Advances in Modeling Earth Systems*, *11*(11), 3545–3592. https://doi.org/10.1029/2019ms001810

Lueck, R. G. (2016). *Calculating the rate of dissipation of turbulent kinetic energy*. Rockland Scientific International Technical Note. TN-028 18.

Martin, P. J., Rogers, E., Allard, R. A., Dykes, J. D., & Hogan, P. J. (2013). *Tests of parameterized Langmuir circulation mixing in the ocean's surface mixed layer*. NRL Report No. NRL/MR/7322-12-9444Naval Research Lab. 47.

Martin, P. J., & Savelyev, I. B. (2017). *Tests of parameterized Langmuir circulation mixing in the ocean's surface mixed layer II*. NRL Report No. NRL/MR/7320-17-9738 Naval Research Lab. 49.

McWilliams, J. C. (2016). Submesoscale currents in the ocean. *Proceedings of the Royal Society A: Mathematical, Physical & Engineering Sciences*, *472*(2189), 20160117. https://doi.org/10.1098/rspa.2016.0117

McWilliams, J. C., Sullivan, P. P., & Moeng, C. H. (1997). Langmuir turbulence in the ocean. *Journal of Fluid Mechanics*, *334*, 1–30. https://doi.org/10.1017/s0022112096004375

Mellor, G. L., & Yamada, T. (1982). Development of a turbulence closure model for geophysical fluid problems. *Reviews of Geophysics*, *20*(4), 851–875. https://doi.org/10.1029/rg020i004p00851

Nasmyth, P. W. (1970). *Ocean Turbulence. PhD thesis*. University of Bristish Columbia.

Savelyev, I., Miller, W. D., Sletten, M., Smith, G. B., Savidge, D. K., Frick, G., et al. (2018). Airborne remote sensing of the upper ocean turbulence during CASPER-East. *Remote Sensing*, *10*(8), 1224. https://doi.org/10.3390/rs10081224

Savelyev, I. B., Buckley, M. P., & Haus, B. K. (2020). The impact of nonbreaking waves on wind-driven ocean surface turbulence. *Journal of Geophysical Research: Oceans*, *125*(1), e2019JC015573. https://doi.org/10.1029/2019jc015573

Savidge, D. K., Austin, J. A., & Blanton, B. O. (2013). Variation in the Hatteras front density and velocity structure Part 2: Historical setting. *Continental Shelf Research*, *V54*, 106–116. https://doi.org/10.1016/j.csr.2012.11.006

Scheifele, B., Waterman, S., Merckelbach, L., & Carpenter, J. R. (2018). Measuring the dissipation rate of turbulent kinetic energy in strongly stratified, low-energy environments: A case study from the Arctic Ocean. *Journal of Geophysical Research: Oceans*, *123*(8), 5459–5480. https://doi.org/10.1029/2017jc013731

Smith, J. A. (1992). Observed growth of Langmuir circulation. *Journal of Geophysical Research*, *97*(C4), 5651–5664. https://doi.org/10.1029/91jc03118

Smith, J. A. (2001). Observations and theories of Langmuir circulation: A story of mixing. In *Fluid mechanics and the environment: Dynamical approaches* (pp. 295–314). Springer.

Smyth, W. D., Skyllingstad, E. D., Crawford, G. B., & Wijesekera, H. (2002). Nonlocal fluxes and Stokes drift effects in the K-profile parameterization. *Ocean Dynamics*, *52*(3), 104–115. https://doi.org/10.1007/s10236-002-0012-9

Terray, E., Eugene, A., Donelan, M. A., Agrawal, Y. C., Drennan, W. M., Williams, A., et al. (1996). Estimates of kinetic energy dissipation under breaking waves. *Journal of Physical Oceanography*, *26*(5), 792–807. https://doi.org/10.1175/1520-0485(1996)026<0792:eokedu>2.0.co;2

Thorpe, S. A. (2004). Langmuir circulation. *Annual Review of Fluid Mechanics*, *36*(1), 55–79. https://doi.org/10.1146/annurev.fluid.36.052203.071431

Wang, Q., Alappattu, D. P., Billingsley, S., Blomquist, B., Burkholder, R. J., Christman, A. J., et al. (2017). CASPER: Coupled air-sea processes and electromagnetic (EM) ducting research. *Bulletin of the American Meteorological Society*, *99*(7), 1449–1471. https://doi.org/10.1175/BAMS-D-16-0046.1

Weller, R. A., Dean, J. P., Price, J. F., Francis, E. A., Marra, J., & Boardman, D. C. (1985). Three-dimensional flow in the upper ocean. *Science*, *227*(4694), 1552–1556. https://doi.org/10.1126/science.227.4694.1552