Deep ocean learning of wave-induced turbulence

Ali Mashayek[1], Fangming Zhai[1], Nick Reynard[1], Adam Jelley[2], Colm Cauldfield[3], Alberto Naveira Garabato[4]

[1]Imperial College London, UK
[2]University of Edinburgh, UK
[3]University of Cambridge, UK
[4]University of Southampton, UK

Key Points:

- Machine learning can be used to infer ocean turbulent mixing from basic seawater and geometric properties.
- The machine learning models are trained based on limited available direct turbulence measurements.
- The trained models can be applied to data from global observational programs which do not sample turbulence directly.

Corresponding author: Ali Mashayek, mashayek@ic.ac.uk

Abstract

Turbulent mixing at centimetre scales is an essential component of the ocean's meridional overturning circulation and its associated global redistribution of heat, carbon, nutrients, pollutants and other tracers. Whereas direct turbulence observations in the ocean interior are limited to a modest collection of field programs, basic information such as temperature, salinity and depth (T, S, Z) is available globally. Here, we show that supervised (deep) machine learning algorithms, informed by physical understanding, can be trained on the existing turbulence data to develop skillful predictions of the key properties of turbulence from T, S, Z and topographic data. This constitutes a promising first step toward a hybrid physics - artificial intelligence approach to parameterize turbulent mixing in climate-scale ocean models.

Plain Language Summary

Ocean turbulence plays an important role in sustaining the general ocean circulation and in the mixing of heat, carbon, nutrients, and other processes within the ocean interior. Turbulent mixing is technically challenging to measure and is often inferred from measurable quantities using parameterizations that are based on numerous simplifying assumptions about the physics of turbulence. In this work, we show that Artificial Intelligence (in form of Machine Learning) can be successfully employed to infer turbulent mixing from quantities measured routinely by global observational programs.

Introduction

Turbulent mixing across density surfaces (i.e. diapycnal mixing) in the ocean interior is key to sustaining the meridional overturning circulation and its global regulation of heat, carbon and nutrient distributions, as well as other climatically and environmentally important tracers (Talley et al., 2016). Such turbulence is primarily excited at the ocean surface by winds, or at the bottom boundary via flow impingement on topography (Garabato & Meredith, 2022). To illustrate, Fig. 1c shows a global estimate of the power provided to ocean interior turbulence from some of the leading energy sources. Figure 1a shows a snapshot of the ensuing, spatio-temporally intermittent turbulent field in a highly dynamic ocean region (the Drake Passage, where all primary energy sources are in play). When and where the vertical gradients of horizontal velocity (known as shear) become sufficiently intense, strong patches of turbulence are generated that mix temperature, salinity and other tracer vertically (i.e. to leading order across density surfaces or diapycnally.

The spatio-temporal variability of turbulence makes its measurement especially challenging. However, turbulence can leave an imprint on vertical temperature (T) and salinity (S) profiles obtained from hydrographic surveys. T, S and depth (Z) are regularly sampled through global international programs, such as ship-based efforts like WOCE (Gouretski & Koltermann, 2004), GO-SHIP (GO-SHIP, 2018), GEOTRACES(GEOTRACERS, 2019) or globally-distributed floats deployed by the Argo Program (Argo, 2000) (see Supplementary Materials for a visual summary, and Davis *et al*. (2019) for a review (Davis et al., 2019)). While turbulence characteristics may be inferred from these T, S, Z data(Polzin et al., 2014; Whalen et al., 2012), such estimates involve many assumptions and uncertainties.

The gold standard in measuring turbulence in the ocean interior is represented by ship-deployed microstructure profiler observations, which are limited in number due to their technical complexity and cost(Shroyer et al., 2018). Figure 1b shows microstructure measurements taken in the abyssal Samoan Passage, an important chokepoint in the flow of anthropogenic heat- and carbon-rich Antarctic Bottom Water into the North Pacific pool, the largest 'storage room' of the deep oceans(Alford et al., 2013). Measure-

ments such as these provide invaluable data of turbulent variables (primarily, the rate of dissipation of turbulent kinetic energy ε) concurrent to T, S and Z.

In this work, we utilize artificial intelligence (AI) algorithms (specifically, supervised deep learning) that are trained on a unique collection of observations from microstructure field programs(MacKinnon et al., 2017; Waterhouse et al., 2014) to predict turbulence characteristics successfully across the global ocean. These predictions are based on widely-available T, S, Z and topographic data, rendering our approach applicable to major global surveys that do not measure turbulence. The training phase is informed by our current physical understanding of turbulence. Our results demonstrate a first step toward a physically-informed, AI-based parameterization of oceanic turbulence in climate models.

Microphysics of Turbulence

A key property of density-stratified ocean turbulence is the significantly enhanced rate at which it fluxes density and tracers in the vertical. The flux is directly proportional to $\Gamma \times \varepsilon$, where Γ is a coefficient that determines the fraction of the energy available to turbulence (from breaking internal waves or, more generally, hydrodynamic instabilities) that contributes to the turbulent density flux, with the remainder of the energy being dissipated via the viscosity of seawater (Osborn, 1980; Mashayek et al., 2013). It has been shown that Γ can be accurately quantified using the ratio of two length scales, L_T and L_O (Mashayek, Caulfield, & Alford, 2021). The Thorpe scale L_T is a geometrical scale characterising the vertical displacement of notional fluid parcels within an overturning turbulent patch (crudely, a measure of the size of the overturn), and can be calculated from profiles of T, S with depth (Thorpe, 2005). The Ozmidov scale L_O is the size of the largest turbulent structures forming within the breaking wave. Figure 1d shows the co-evolution of these two scales during the life cycle of a typical breaking wave. As the wave grows, initially L_T rises, and upon sufficient growth, the wave transitions to turbulence (i.e. breaks) through a zoo of hydrodynamic instabilities(Mashayek & Peltier, 2012a, 2012b) that feed on the energy stored in the parent wave. L_O, the upper bound on the vertical size of such turbulent structures, increases as turbulence feeds on the energy stored in the wave, draining L_T. Both scales eventually vanish as turbulence decays. The ratio L_O/L_T has proven a good predictor of Γ at any stage of turbulence (Mashayek, Caulfield, & Alford, 2021). L_O is defined as $\sqrt{(\varepsilon/N^3)}$, where the buoyancy frequency $N = \sqrt{-[g/\rho_0]\partial\rho/\partial z}$ can be directly inferred from T, S and Z through the construction of the vertical density gradient, a characteristic density ρ_0, and the gravitational acceleration g (whereas ε cannot).

Given the appropriateness of L_O/L_T for quantification of the turbulent density flux and also that ε can be directly calculated from L_O if density stratification (and hence N) is known, we choose $L_O = f_\theta(T, S, Z, L_T)$ as the parameter of interest, where f is a function relating the inputs to output. f is to be parameterized with a machine learning algorithm with parameters θ. The key concept to appreciate is that the predictors are inferrable from all major global hydrographic surveys, whereas L_O, an emergent turbulence measure, is not. Once L_O is predicted, it can be used to calculate ε and Γ and, thereby, the turbulent density flux. We also add height above the bottom (Hab), to the list of predictors, as it determines the distance from the bottom boundary of the ocean in the same way as Z is the distance from the ocean's top boundary, noting that both boundaries are crucial turbulence generation sites. Knowledge of Hab requires topographic data, which has become increasingly more accurate in recent decades thanks to advanced satellite-based gravity measurements and deep-ocean echo-sounding records (Sandwell et al., 2014) (see Supplementary Materials Fig S1).

The Training Dataset

We employ a global dataset of microstructure profiles compiled by the Climate Process Team on internal wave–driven ocean mixing(MacKinnon et al., 2017). Figure 2 shows the location of the field measurements that spanned a wide range of geographic locations, depths, and turbulence-inducing physical processes. The figure also provides the list of the field experiments, and the fraction of the total data associated with each program. The data is available at `https://microstructure.ucsd.edu/`, and data description and relevant references may be found in Waterhouse *et al.* (2014)(Waterhouse et al., 2014). The same dataset was employed by Cael & Mashayek (2021)(Cael & Mashayek, 2021) to show that the data 'collapses' on a seemingly universal log-skew-normal statistical distribution. This result motivated the present study by suggesting that such universality might be detectable through data-driven methods. Importantly, they showed that a collapse occurs for a certain ratio of ε and N^2, thus further guiding our approach to consider L_O as the desirable output parameter. Together, the experiments in Fig. 2 contain over 700 full-depth microstructure profiles, binned into 10 m vertical bins (amounting to $\sim 2\times10^5$ data points). The concurrent measurements of ε, T, S, Z in this dataset allow for the construction of the aforementioned predictor list as well as L_O. Jointly, these provide the ingredients for training the AI algorithms. Given that the logarithm of L_O is much better constrained than L_O, and that $\log_{10}(L_O/L_T)$ has a universal well-bounded distribution(Thorpe, 2005; Mashayek, Barry, et al., 2021), we will use $\log_{10}(L_T), \log_{10}(L_O)$ hereafter.

Classification And Regression Trees (CART)

We employ CART, one of the most common machine learning predictive models (Wu et al., 2008; Breiman et al., 1984). The method uses a decision tree to connect observations of a parameter of interest (represented in the branches) to predictions about its value (represented in the leaves). When applied to target variables that take continuous values (such as L_O in this study), such decision trees are referred to as regression trees. Additionally, we employ an ensemble method, bootstrap aggregating, to improve the stability and accuracy of the decision tree algorithm, reduce variance, and avoid overfitting. Bootstrap aggregated decision trees (hereafter bagging trees) construct multiple trees by repeatedly re-sampling the training data with replacements, and voting the trees for a consensus prediction (Breiman, 1996).

Figure 3a shows the application of the bagging tree to the training microstructure dataset (from which $\log_{10}(L_O)$, our prediction target, is calculated). The model was trained based on 50 cross-validation k-folds of all data across 13 field experiments. This method involves splitting the dataset into equally sized 'k' number of groups, or 'folds', and taking it in turn to use each group as the test data while the rest of the data is used to train the model, with an average of the results being adopted. A k-fold validation approach is useful when input data is limited, and ensures that every data point is used within the training and test dataset, hence reducing bias when compared to other methods. The fit is satisfactory, with a coefficient of determination (R^2) of 0.77. R^2 is a statistical metric of how well the regression predictions approximate the real data, and so is a measure of the goodness of fit of a model. To analyse further the quality of the agreement between predictions and data shown in Fig. 3c we display the cumulative contribution of various predictors to R^2. The first four parameters (T, S, Z, and Hab) contribute 0.65 to R^2, with the rest (local overturn estimates based on density and temperature [i.e. $\log_{10}(L_T^\rho)$, $\log_{10}(L_T^T)$], density, and vertical gradients of temperature and density), adding another 0.1. Figure 3d shows the associated error in prediction, which diminishes as more features are included. Importantly and consistently with the contributions to R^2, T, S, Z, and Hab reduce the mean-square error to less than 10%.

The above agreement implies that knowledge of parameters most basic to turbulence, i.e. density stratification (T, S), distance from turbulence-generating boundaries (Z, Hab), and localized vertical gradients of T, S, which arguably are measures of local turbulent overturns or wave activity, suffice to obtain an estimate of the turbulence intensity accurate at least to within an order of magnitude. Given that indirect inferences of turbulent mixing do not always possess such accuracy, as –unlike our AI-driven approach– they rely on many assumptions related to the physical processes involved(Polzin et al., 2014), this is a significant finding and underscores the potential for a future role of AI in parameterizing oceanic turbulence. We emphasize, however, that the success of our approach is rooted in our physical understanding of turbulence, which led us to select $\log_{10}(L_O)$ as the most appropriate and robust prediction, as well as the choice of the straightforwardly available associated predicting features (i.e. the 'predictors').

It is worth noting that we also tried another standard choice, namely the Least Squares Boost (LSBoost) algorithm, as an alternative ensemble learning method. LSBoost is a gradient boosting method in which the mean squared error is chosen as the cost function (Breiman et al., 1984). While we found LSBoost to outperform bagging tree for a smaller number of features (up to 3), bagging tree was superior for the number of features employed herein, and thus is our method of choice.

Neural Networks

As an entirely different approach, we also train neural networks with the same data (Fig. 2). Specifically, we use a fully-connected feed-forward neural network (FNN) with an input layer, an output layer and multiple hidden layers in between (Goodfellow et al., 2016). Each hidden layer combines the (learned) features of the previous layer to build up a non-linear transformation of the input predictors to predict turbulence properties (L_O) in the output layer. Adding additional layers to make the network deeper incorporates more parameters to be learned, which allows for a more flexible mapping between the easily measurable predictors and the less widely available turbulence properties, but requires more data to learn an effective generalisable mapping without overfitting (see Supplementary Materials for the model architecture).

We also tried an initial implementation of a Convolutional Neural Network (CNN), which is another class of artificial neural networks, commonly applied to image analyses (LeCun et al., 1995). CNNs were originally inspired by their resemblance to the connectivity pattern between neurons and the organization of the visual cortex in humans and animals (Fukushima & Miyake, 1982), and consequently are typically employed where there is relative information between nearby pixels in an image. Our motivation for including convolutional layers was the existence of such relative information, from a physical perspective, between neighbouring profiles of various research cruises. However, since initial performance was comparable with the FNN, we save a full investigation of the effectiveness of convolutional architectures for future work. Details of the networks are provided in the Supplementary Materials (see figures S2,S6,S7 and their description).

Fig. 3b shows outcome of applying the deep neural networks to the data shown in Fig. 2. In this case, validation is performed by making the prediction for 10% of the data randomly selected from across the 13 experiments and then set aside, with the network trained on the other 90% of the data (to avoid overfitting). The agreement is reasonable, but as shown in Figure 3, the error is larger than for the bagging tree algorithm. As discussed above, deep learning algorithms require less human intervention compared to more traditional machine learning algorithms (e.g. the bagging tree), and so generally have larger data requirements and their performance increases more strongly with the size of data. This makes the agreement in Fig. 3b particularly promising, given the limited nature of the training data compared to data sizes typically employed in deep learning. Thus,

investment in extending the training data through a community effort appears worthwhile.

Discussion & Outlook

The primary message of this paper is encapsulated in Fig. 3. AI can successfully be employed to use measured quantities widely sampled by global observational programs to predict turbulence characteristics that are not observed by such programs yet are of great importance to ocean circulation. However, to our knowledge, this is a first such attempt, and there is clear scope for improvement. Figure 4 breaks down the analysis to different field programs and compares the predictions of two distinctly different machine learning algorithms with the observational data. Both methods can successfully predict the patterns and, in most cases, the order of magnitude adequately, bearing in mind that the observational data itself has uncertainties that can add up to an order of magnitude(Thorpe, 2005). The agreement is notable given that, for each of the panels in Fig. 4, the models were trained excluding the data from that particular experiment to demonstrate successful generalisation. To further illustrate this agreement, Figures S3 and S4 in Supplementary Materials show analogues of Fig. 3a,b for each of the 13 field experiments.

There are numerous factors that can contribute to the misfits. Three important ones are: (I) the percentage of the training and validation data can vary significantly between the experiments (Fig. 2); (II) the relevance of the underlying physics in each experiment to the rest of the data used for training might be limited; (III) ocean mixing is not entirely 'local' in nature, e.g., waves generated thousands of kilometres away can contribute to mixing, and no such information was included in our training by construct(de Lavergne et al., 2019). Factor (I) can only be addressed through application of AI to larger datasets. In particular, the success of deep learning directly scales with the data size, and what was achieved in this work lies at the lower bound of the data volume required. Our analyses show that while the bagging tree method converges to the optimal performance once a few hundreds of profiles are considered, the FNN approach does not show such convergence and retains a large standard deviation even when all profiles are included (see Supplementary Materials Fig. S5). Thus, further community efforts are required to pull turbulence datasets together and subject them to the same levels of quality control and grid interpolations. Furthermore, adding microstructure sensors to global observational endeavors (such as the Argo float program), while ambitious, is within reach and conceivable in the coming decades (Roemmich et al., 2019). Factor (II) will naturally advance as our physical understanding of ocean turbulence keeps progressing. A conscious effort towards connecting such physical understanding to data-driven parameterizations is required. Addressing factor (III) is more readily achievable in the near future, as it will require inclusion of theoretical estimates of local and non-local energy injected to internal waves from various sources (winds, tides, etc.) in training algorithms. In summary, AI provides a valuable tool to harness our observational, theoretical and statistical knowledge of ocean turbulence to direct the development of a next-generation 'smart' turbulence parameterization for climate models.

Acknowledgments

The authors report no conflict of interests. The authors thank Lois Baker for constructive comments.

Data Availability Statement

The microstructure data employed for the training of the Machine Learning algorithms may be obtained from `https://microstructure.ucsd.edu/` by locating the names

of the experiments in Figure 2; also see (Waterhouse et al., 2014) and (MacKinnon et al., 2017) for further information. The AI algorithms will be shared via an online depository upon acceptance of the manuscript.

References

Alford, M. H. (2020). Global calculations of local and remote near-inertial-wave dissipation. *Journal of Physical Oceanography*, *50*(11), 3157–3164.

Alford, M. H., Girton, J. B., Voet, G., Carter, G. S., Mickett, J. B., & Klymak, J. M. (2013). Turbulent mixing and hydraulic control of abyssal water in the Samoan Passage. *Geophysical Research Letters*, *40*(17), 4668–4674.

Argo, G. (2000). Argo float data and metadata from global data assembly centre (argo gdac). *SEANOE*.

Breiman, L. (1996). Bagging predictors. *Machine learning*, *24*(2), 123–140.

Breiman, L., Friedman, J., Olshen, R., & Stone, C. (1984). Classification and regression trees. wadsworth int. *Group*, *37*(15), 237–251.

Cael, B., & Mashayek, A. (2021). Log-skew-normality of ocean turbulence. *Physical Review Letters*, *126*(22), 224502.

Davis, R. E., Talley, L. D., Roemmich, D., Owens, W. B., Rudnick, D. L., Toole, J., ... Barth, J. A. (2019). 100 years of progress in ocean observing systems. *Meteorological Monographs*, *59*, 3–1.

de Lavergne, C., Falahat, S., Madec, G., Roquet, F., Nycander, J., & Vic, C. (2019, 5). Toward global maps of internal tide energy sinks. *Ocean Modelling*, *137*, 52–75. Retrieved from `https://linkinghub.elsevier.com/retrieve/pii/S1463500318302890` doi: 10.1016/j.ocemod.2019.03.010

de Lavergne, C., Vic, C., Madec, G., Roquet, F., Waterhouse, A. F., Whalen, C., ... Hibiya, T. (2020). A parameterization of local and remote tidal mixing. *Journal of Advances in Modeling Earth Systems*, *12*(5), e2020MS002065.

Fukushima, K., & Miyake, S. (1982). Neocognitron: A self-organizing neural network model for a mechanism of visual pattern recognition. In *Competition and cooperation in neural nets* (pp. 267–285). Springer.

Garabato, A. N., & Meredith, M. (2022). Ocean mixing: oceanography at a watershed. In *Ocean mixing* (pp. 1–4). Elsevier.

GEOTRACERS. (2019). Geotracers. *https://www.geotraces.org/*. Retrieved from `https://www.geotraces.org/`

Goodfellow, I., Bengio, Y., & Courville, A. (2016). *Deep learning*. MIT Press. (`http://www.deeplearningbook.org`)

GO-SHIP. (2018). Go-ship. *http://www.go-ship.org/*. Retrieved from `http://www.go-ship.org/`

Gouretski, V., & Koltermann, K. P. (2004). {WOCE} global hydrographic climatology. *Berichte des BSH*, *35*, 1–52.

LeCun, Y., Bengio, Y., et al. (1995). Convolutional networks for images, speech, and time series. *The handbook of brain theory and neural networks*, *3361*(10), 1995.

MacKinnon, J. A., Zhao, Z., Whalen, C. B., Waterhouse, A. F., Trossman, D. S., Sun, O. M., ... others (2017). Climate process team on internal wave–driven ocean mixing. *Bulletin of the American Meteorological Society*, *98*(11), 2429–2454.

Mashayek, A., Barry, B., Alford, M., Cimoli, L., & Caulfield, C.-c. (2021). Shear coincidence: implications of the statistics of ocean turbulence microphysics for global diapycnal mixing. *Earth and Space Science Open Archive*, 20. Retrieved from `https://doi.org/10.1002/essoar.10508227.1` doi: 10.1002/essoar.10508227.1

Mashayek, A., Caulfield, C., & Alford, M. (2021). Goldilocks mixing in oceanic shear-induced turbulent overturns. *Journal of Fluid Mechanics*, *928*.

Mashayek, A., Caulfield, C. P., & Peltier, W. R. (2013). Time-dependent, non-monotonic mixing in stratified turbulent shear flows: implications for oceanographic estimates of buoyancy flux. *Journal of Fluid Mechanics*, *736*, 570–593.

Mashayek, A., Ferrari, R., Merrifield, S., Ledwell, J. R., St Laurent, L., & Garabato, A. N. (2017). Topographic enhancement of vertical turbulent mixing in the Southern Ocean. *Nature communications*, *8*, 14197.

Mashayek, A., & Peltier, W. (2012a). The 'zoo'of secondary instabilities precursory to stratified shear flow transition. part 1 shear aligned convection, pairing, and braid instabilities. *Journal of Fluid Mechanics*, *708*, 5–44.

Mashayek, A., & Peltier, W. (2012b). The 'zoo'of secondary instabilities precursory to stratified shear flow transition. part 2 the influence of stratification. *Journal of fluid mechanics*, *708*, 45–70.

Nikurashin, M., & Ferrari, R. (2013). Overturning Circulation Driven by Breaking Internal Waves. *Journal of Physical Oceanography*, DOI: 10.1002/grl.50542.

Osborn, T. R. (1980). Estimates of the local rate of vertical diffusion from dissipation measurements. *Journal of Physical Oceanography*, *10*, 83–89.

Polzin, K. L., Garabato, A. C. N., Huussen, T. N., Sloyan, B. M., & Waterman, S. (2014). Finescale parameterizations of turbulent dissipation. *Journal of Geophysical Research: Oceans*, *119*(2), 1383–1419.

Roemmich, D., Alford, M., Claustre, H., Johnson, K., King, B., Moum, J., et al. (2019). On the future of argo: an enhanced global array of physical and biogeochemical sensing floats. front. *Mar. Sci*, *6*, 439.

Sandwell, D. T., Müller, R. D., Smith, W. H., Garcia, E., & Francis, R. (2014). New global marine gravity model from cryosat-2 and jason-1 reveals buried tectonic structure. *Science*, *346*(6205), 65–67.

Shroyer, E. L., Nash, J. D., Waterhouse, A. F., & Moum, J. N. (2018). Measuring Ocean Turbulence. In (pp. 99–122). Springer, Cham. Retrieved from `https://link.springer.com/chapter/10.1007/978-3-319-66493-4_6` doi: 10.1007/978-3-319-66493-4{_}6

Talley, L. D., Feely, R. A., Sloyan, B. M., Wanninkhof, R., Baringer, M. O., Bullister, J. L., ... Zhang, J. Z. (2016, 1). Changes in Ocean Heat, Carbon Content, and Ventilation: A Review of the First Decade of GO-SHIP Global Repeat Hydrography. *Annual Review of Marine Science*, *8*, 185–215. Retrieved from `www.go-ship.org` doi: 10.1146/annurev-marine-052915-100829

Thorpe, S. A. (2005). The turbulent ocean. *Cambridge University Press*.

Waterhouse, A. F., MacKinnon, J. A., Nash, J. D., Alford, M. H., Kunze, E., Simmons, H. L., ... others (2014). Global patterns of diapycnal mixing from measurements of the turbulent dissipation rate. *Journal of Physical Oceanography*, *44*(7), 1854–1872.

Whalen, C., Talley, L., & MacKinnon, J. (2012). Spatial and temporal variability of global ocean mixing inferred from argo profiles. *Geophysical Research Letters*, *39*(18).

Wu, X., Kumar, V., Quinlan, J. R., Ghosh, J., Yang, Q., Motoda, H., ... others (2008). Top 10 algorithms in data mining. *Knowledge and information systems*, *14*(1), 1–37.

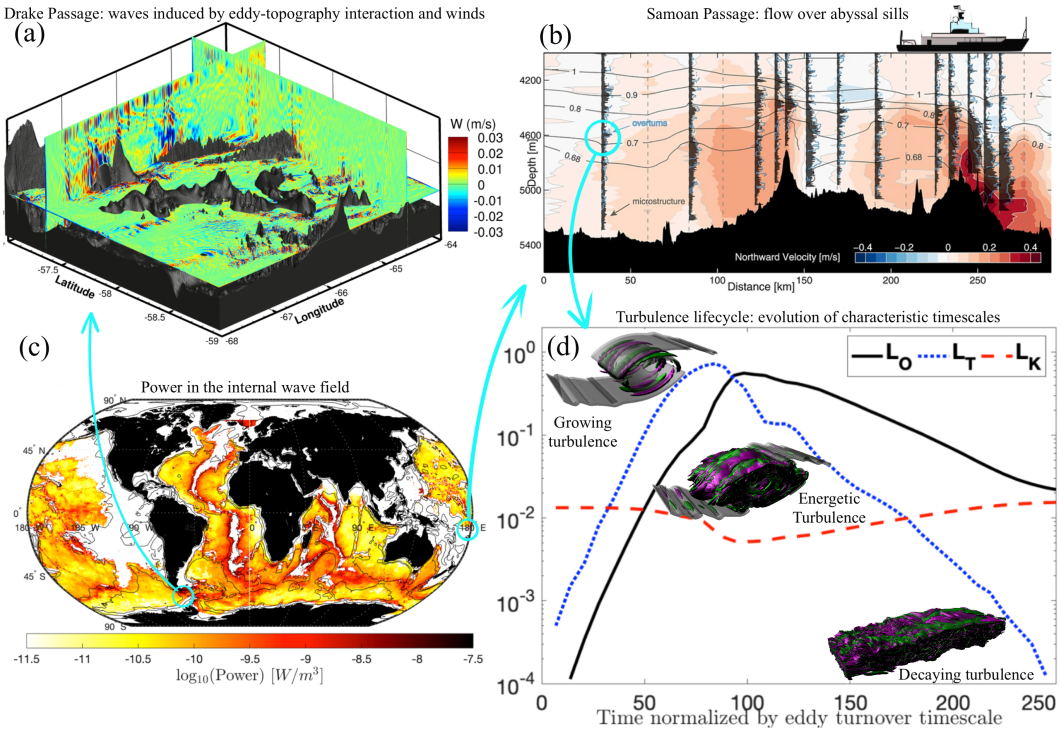

Figure 1. High spatio-temporal intermittency of ocean turbulence makes it challenging to measure turbulence directly and to connect it with physical understanding. (a) A snapshot of the internal wave field in the Southern Ocean's Drake Passage, generated by intense wind forcing and flow over rough topography, from an observationally-constrained and verified high-resolution simulation (Mashayek et al., 2017). (b) Turbulence in the deep Samoan Passage: northward velocity (color), potential temperature (black contours), and turbulent dissipation rate measured with a microstructure profiler (grey horizontal bars plotted along vertical profiles)– from Alford *et al.* (2013)(Alford et al., 2013). (c) Power available to small-scale turbulence from the internal wave field, constructed based on estimates of waves generated by winds(Alford, 2020) and interaction of tides(de Lavergne et al., 2020) and eddies(Nikurashin & Ferrari, 2013) with rough topography. Power is plotted on the neutral density level 1028.1 kg/m^3 with a mean global depth of 3170 m. (d) Time evolution of turbulence length scales: the Thorpe scale L_T, characterising the overturning scale; the Ozmidov scale L_O, characterising the largest turbulence structures within a breaking wave, and the Kolmogorov microscale L_K, the scale below which viscosity dissipates the energy in the turbulent field – all for a canonical breaking wave(Mashayek et al., 2013). The insets show the turbulence breakdown of the wave by hydrodynamic instabilities, which form a turbulent cascade that transfers energy from the energy-containing scale (L_T) to the smallest scales (L_K)(Mashayek & Peltier, 2012a, 2012b).

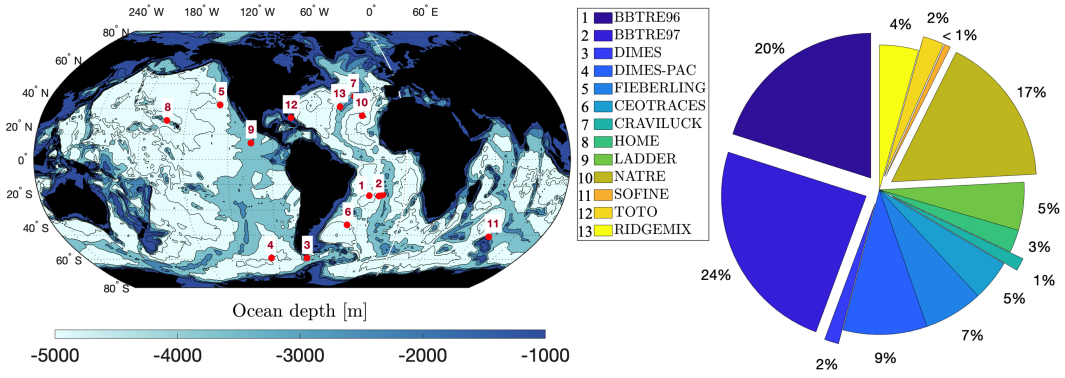

Figure 2. Limited direct turbulence observations exist to be used for training machine and deep learning algorithms. Left: location of the field programs that include direct measurements of turbulence (specifically, turbulent kinetic energy dissipation rate ε using microstructure profilers) along with co-located temperature, salinity and depth sampling; the experiments are listed in the legend, with more details available at `https://microstructure.ucsd.edu/`(Waterhouse et al., 2014) and in the Supplementary Materials. The data contains a total of ∼700 profiles, with ε binned into 10 m vertical bins. Right: the percentage of data from each experiment.

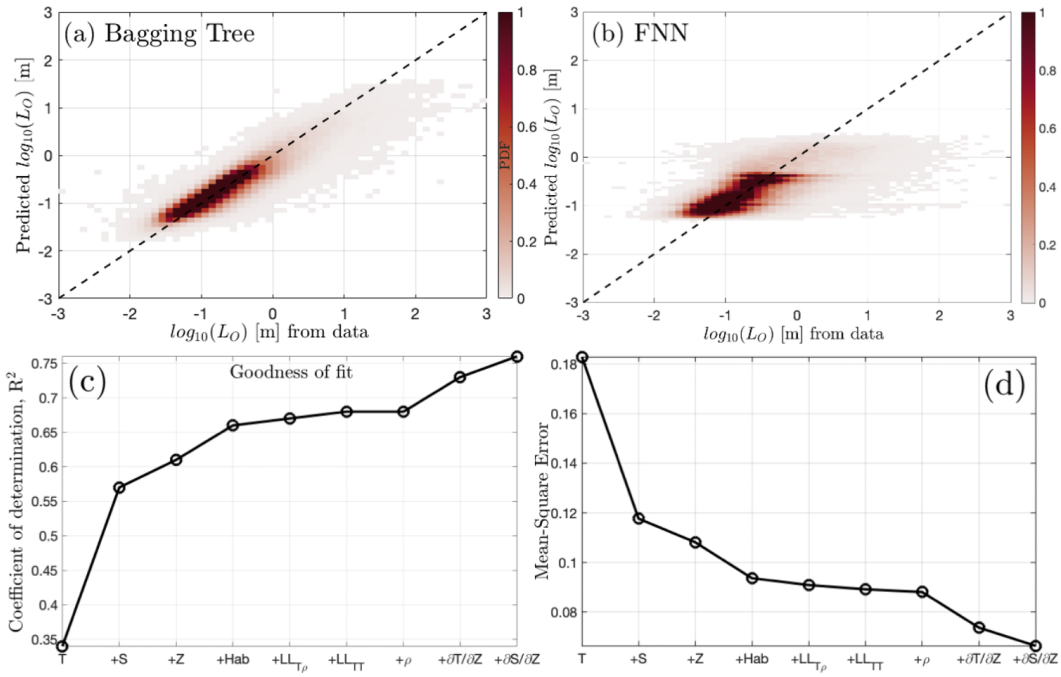

Figure 3. Both machine learning and deep learning models can successfully fit the global microstructure data based on few predicting features. Bivariate histograms (in the form of a probability density function, PDF) of predicted turbulence overturn scale versus the actual data from machine learning in (a), and from deep learning in (b). In panel *a*, the model is validated using K-fold validation with 50 folds to avoid overfitting (see main text). In *b*, the model is trained based on 90% of the entire dataset of Fig. 2 (i.e. for all 13 experiments) and then applied to the remaining 10% (randomly selected from across all datasets) for validation– hence no overfitting. (c) Contributions of various predicting features to the goodness of the fit (R^2) based on the validation data (i.e. the same data as in panels a, b); a similar trend follows when applied to the training data or to the entire data. (d) The Mean Square Error (MSE) associated with panel *c*. Machine learning was used to identify the appropriate predictors (as per panels c, d) that were then employed by the deep learning approach. For histogram analogues of panels *a* and *b*, but for individual experiments, see Supplementary Materials, Figures S3,S4.

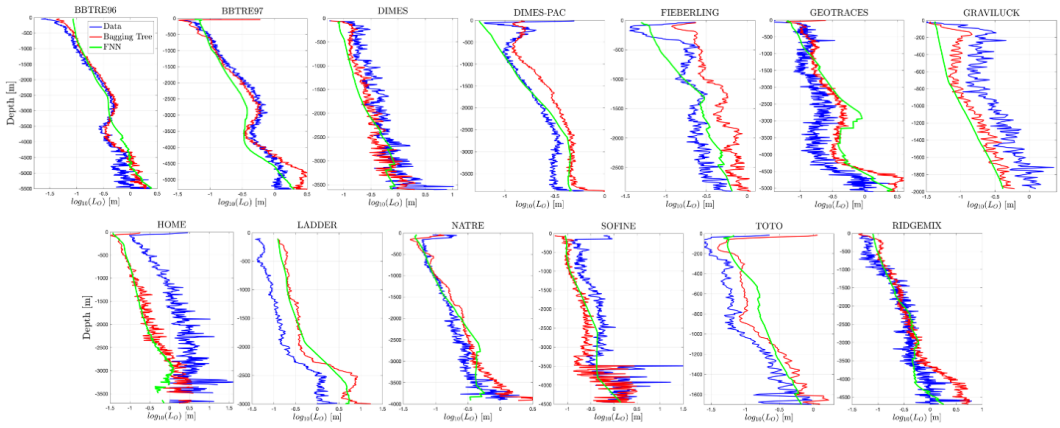

Figure 4. Predictions for individual field programs also show promise. Comparison of the prediction of the machine learning approach (bagging tree) and deep learning (convolutional neural network) to each of the 13 field programs introduced in Fig. 2. For each case, the profiles are the mean over all the profiles in that experiment. Specifically, the algorithms were used to make individual predictions for each profile of each experiment, before the mean of the individual predictions was taken per experiment. Note that for each experiment, the models (both bagging tree and FNN) were trained based on the data from all other 12 experiments, **excluding** the data from the given experiment itself, to avoid overfitting.