Make Our Planet Great Again.

Annual rapport on Post-Doctoral research by Anna Sommer.
Period: 09/05/2019-31/12/2019.
Employer: CEA/LSCE/LOCEAN.

Title of the project:
Improvement of mesoscale processes representation in ocean models using Machine Learning techniques.

Introduction and motivation.

The role of mesoscale eddies is crucial for ocean circulation and its energy budget. At scales of 10 to 300 km, the mesoscale eddies transfer hydrographic properties and energy at different spatial and temporal scales, hence contributing to equilibrating large scale ocean dynamics and thermodynamics, which is paramount for long-term climate modelling [Olbers et al., 2012]. They also affect biogeochemical tracers, which in return influence ocean thermodynamics (through light penetration), climate and ecosystems, hence representing correctly their effect in ocean models is of greatest importance.
Running long term climate simulations, to achieve quasi-steady equilibrium, still impedes to use coarse resolution ocean models that do not represent explicitly mesoscale processes. Indeed, higher-resolution ocean models (so called eddy-permitting) not only require a lot of computing resources (CPU time and storage) but also neglect Reynolds constraints on eddy Reynolds stress, and require a high viscosity and dissipation to maintain numerical stability [Zanna et al., 2017].
Thus, it is important to better represent mesoscale processes in ocean models with a lower computational cost.
Machine learning (ML), and in particular deep learning, have great potential for solving problems such as classification, processing and reconstruction of large amounts of data (Big Data). By extracting the information from the existing data, this approach makes it possible to reconstruct the information in regions where the data availability is mediocre and at different spatio-temporal scales.
The important mesoscale processes, which can not be captured by satellite and are not well represented by numerical models that do not have eddy resolution, can now be provided by machine learning methods. A deep neural network is used to represent all subgrid atmospheric processes in a climate model and successively replaces traditional subgrid parameterizations in a global general circulation model [Rasp et al., 2018].
The use of machine learning in models opens new horizons for improving the representation of environmental phenomena in modeling.
One of the driving mechanisms for the emergence of mesoscales features is the baroclinic instability, especially in winter [Boccaletti, et al., 2007; Capet et al., 2008; Fox-Kemper and Ferrari, 2008; Mensa et al., 2013; Oiu et al., 2014; Sasaki et al., 2014] . The baroclinic instability can be quantified through the term of ***u****'b'*, the so called eddy buoyancy flux.

This work is the first step in the development of a method to reconstruct values of ***u****'b'* from the large-scale ocean parameters. Further, it can be used as a parameterization of the effect of sub-grid-scale processes in coarse resolution ocean models.

What was done during this contract.

Pre-processing.

Data preparation.

The high-resolution simulation eNATL60 is used as a base for the future Neural Network model. It is a basin-scale configuration of NEMO from about 6°N up to the polar circle. The horizontal resolution is 1/60° with Δx between 0.8 and 1.6 km and 300 vertical levels, hence it can be assumed to be resolving explicitly all mesoscale processes, which this study focuses upon. The outputs have an hourly time resolution and 300 vertical levels.
The variables provided by eNATL60 are salinity (S), temperature (T), 3 components of ocean current velocity and sea surface height.

The use of eNATL60 started our collaboration with MEOM team at LGGE (Grenoble), particularly with Julien Le Sommer and Aurelie Albert.

The first step of the pre-processing is estimation of the buoyancy *b* from the eNATL60 data. *b* is the function of ocean temperature (T) and salinity (S).

$b=-g\frac{\rho}{\rho_0}$, with $\rho=\rho(T,S)$ (Jackett and McDougall, 1994).

$\rho(T,S)=(A*S+B*\sqrt{S}+C)*S+D-\rho_0$

A=4.8314 x 10^{-4}

B=(-1.6546 x 10^{-6}*T+1.0227 x 10^{-4})*T-5.72466 x 10^{-3}

C=(((5.3875 x 10^{-9}*T-8.2467 x 10^{-7})*T+7.6438 x 10^{-5})*T-4.0899 x 10^{-3})*T+
+0.824493

D=((((6.536332 x 10 $^{-9}$*T-1.120083 x 10^{-6})*T+1.001685 x 10^{-4})*T -
-9.095290 x 10^{-3})*T+6.793952 x 10^{-2})*T+999.842594

The second step is the spatial filtering of the eNATL60 outputs and estimated *b*. The filtering identifies large scale processes (b_f, u_f) removing which gives the scale of target processes at mesoscale (b', u'). The
A low-pass spatial filter with Hanning window is used to estimate turbulent part of ***u*** and *b*, where ***u*** = *(u,v,w)*. After several tests the optimal filter parameters are: the window size is 30 grid points and cut-off frequency is about 20km (Fig. 1). The choice of parameters is based on the spectrum of surface vorticity (made by Aurelie Albert, IGE).

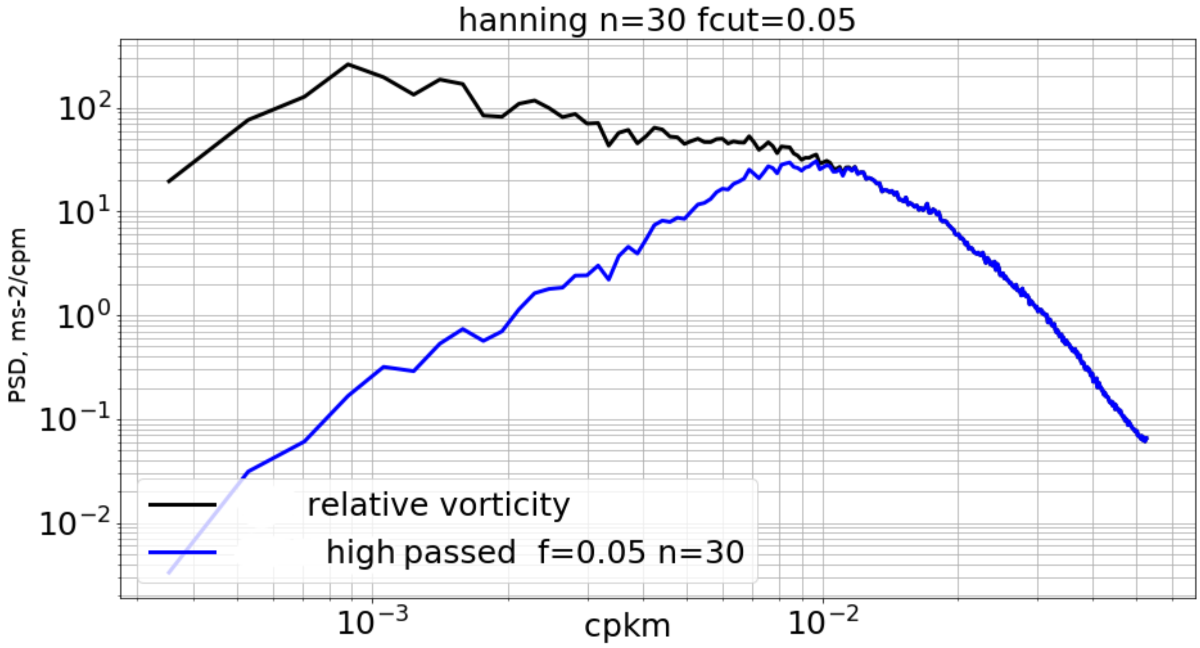

Figure 1. eNATL60 relative vorticity in the North Atlantic before (black) and after (blue) removing of a large-scale part estimated by low-pass spatial Hanning filter

The third step is the spatial and temporal average of the filtered eNATL60 data (b_f, u_f) and its turbulent part (b', u'). As the aim of our work is to create a NN model that can reconstruct a buoyancy flux from large scale variables the idea of this step is to prepare the target data and predictors data on the coarse resolution. The target of this average is the ORCA resolution: daily data with 75 vertical levels and spatial resolution 1°.
The target for the NN is u'b', v'b', w'b' (the bar here is the daily average on 1°x1° that is different from the previous bar used for filter). At first we will focus our analysis on the w'b'. The lateral components are different from w'b' and have other pythical dynamics and order of magnitude.
The drivers (or predictors) for the NN are u_f, v_f, w_f, T_f, S_f and their gradients.
At this step we should check the effective resolution (the resolution of scale separation) of eNATL60 model and how it works with grid target (ORCA resolution 1°). Details are presented further.

Choice of regions.

To simplify the problem, we will avoid coastal regions and regions with islands (limits for latitude and longitude). The island grid points can cause problems when we will apply ML approaches for such regions, particularly CNN will fill these points with data that do not have a physical meaning.
We will also avoid the possible effect of the bathymetry and stop at a certain depth before reaching a shallower point (limits for z). The bathymetry can cause the same problems as islands (see before) at certain depth. Moreover, the bathymetry in high-resolution simulation is different from the one in low-resolution simulation, that will need a coarsening technique that is able to preserve the spectrum (horizontal extrapolation, filtering).

For the first tests we choose two regions: in the Gulf Stream (Fig. 3, 5) (33.5°-39.5°N, 70.6°-60°W, min depth = 840.646 m) and the Labrador Sea (Fig. 2, 4) (56°-61°N, 55°-50°W, min depth = 2955.2122 m).

The Gulf Stream region gives an opportunity to test the CNN approach in the region with strong meso-scale activity.
The region in the Labrador Sea has a rich seasonal dynamic of water properties at meso-scales.

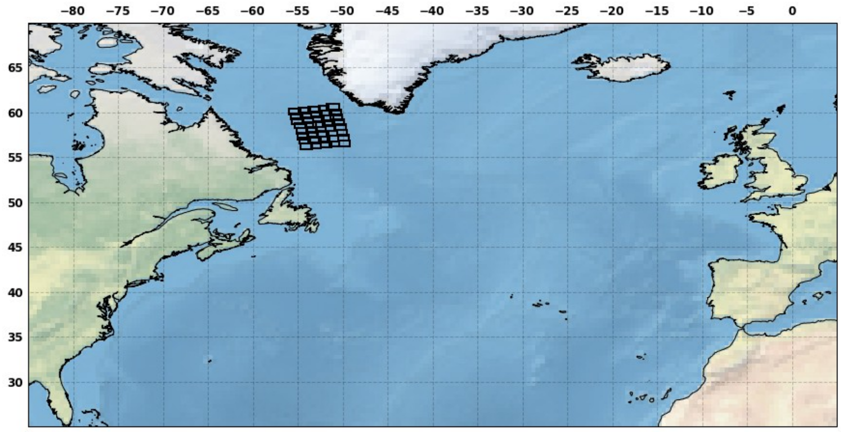

Figure 2. Labrador Sea region, ORCA1 grid boxes applied to eNATL60 grid

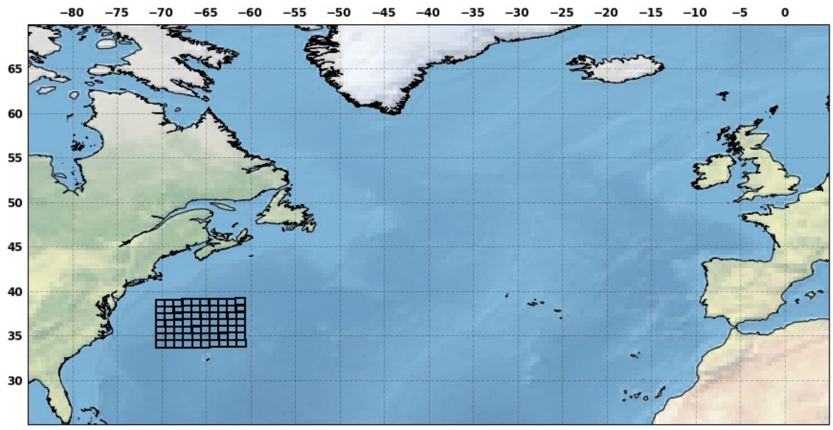

Figure 3. Gulf Stream region, ORCA1 grid boxes applied to eNATL60 grid

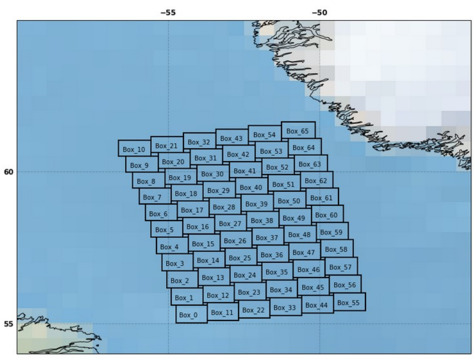

Figure 4. Zoom of the Labrador Sea region

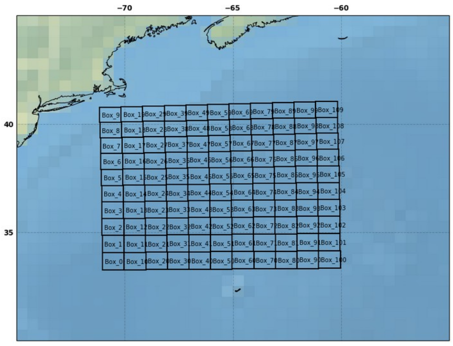

Figure 5. Zoom of the Gulf Stream region

The region in the middle of the North Atlantic will be added too. It is a region with a week eddy activity that allow us to test a model in simplified conditions.

<u>Comparison of the profiles.</u>

As mentioned before, at this stage the important step is the comparison of eNATL60 filtered and averaged profiles with ORCA1 vertical profiles in regions of study. We started from the Labrador Sea region, other regions are still in the data-preparation stage.
The Fig. 6 shows the comparison of temperature vertical profiles averaged over the Labrador Sea region and 100 days started from 01/07/2009 between ORCA1 data (blue) and filtered and averaged eNATL60 data. We see that both products indicate a presence of the mixed layer in first few hundred meters and further thermocline. However, there is an evident quantitative difference in the thermocline. To better understand it we provided Power Density Functions (PDFs) over 100 days (Fig. 7, 8).

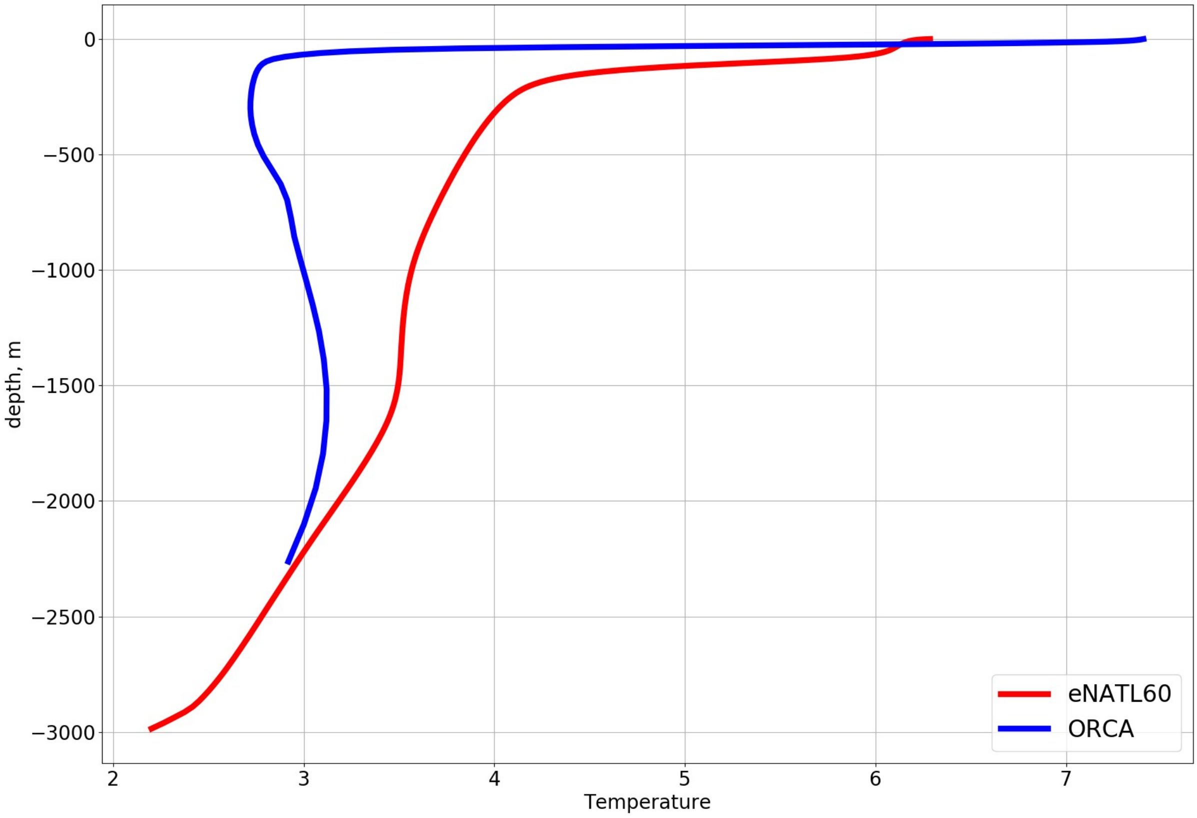

Figure 6. Comparison of temperature vertical profile averaged over the Labrador Sea region and 100 days (started from 01/07/2009) between ORCA1 (blue) and eNATL60 filtered and averaged data (red)

Fig. 7 and 8 show the PDFs of temperature vertical profile over 100 days in the Labrador Sea region for ORCA1 and eNATL60 respectively. eNATL60 shows a strong restratification from the surface to 1000m depth that is not presented in ORCA1. This is an effect of meso-scale activity that we want to bring in the coarse ocean model. However, this difference between the data used for training and data used for further prediction can be crucial.

These results were presented in form of an oral presentation at the AGU annual meeting in December 2019 in San-Francisco, US.

During my collaboration with David Ferreira at the University of Reading (Reading, UK), who specialises on the processes at the meso-scale, the method of classification from [Maze et al., 2017] was proposed. It will allow to create classes based on the eNATL60 data and apply this classes to ORCA1. This comparison will show us at what point the difference between vertical profiles of these two products is significant.

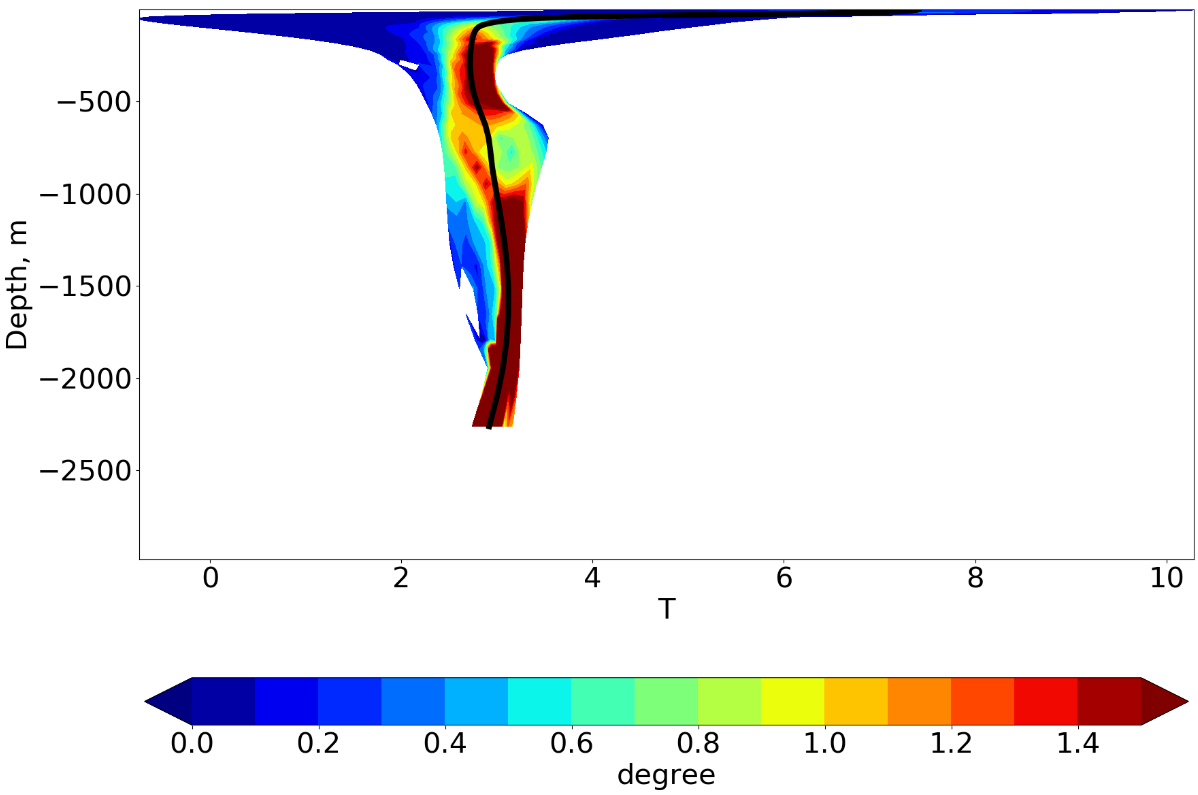

Figure 7. PDF of temperature vertical profile from ORCA1 over 100 days in the Labrador Sea region

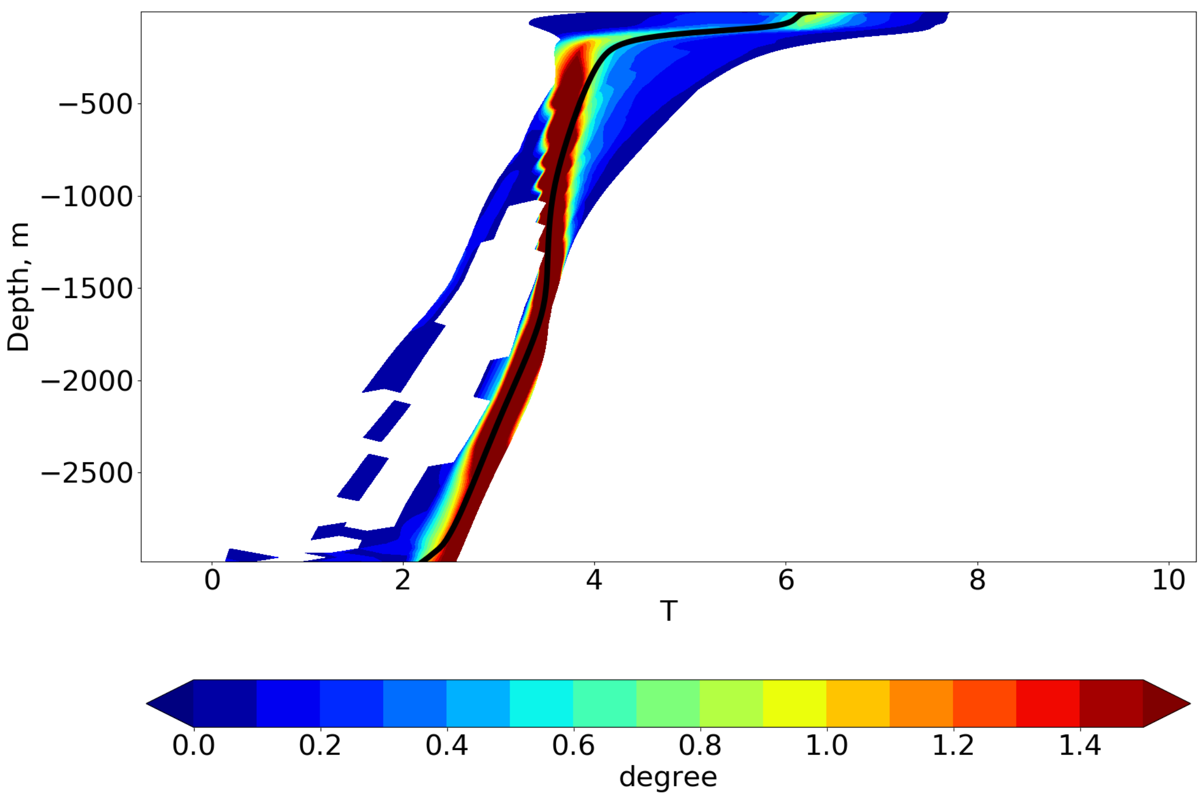

Figure 8. PDF of temperature vertical profile from eNATL60 over 100 days in the Labrador Sea region

Conclusion.

During my Post-Doctoral research in the frame of “Make Our Planet Great Again” grant I (1) defined better the project dedicated to the better representation of meso-scale processes in the ocean models.
Two collaboration were created: with MEOM team (3) in Grenoble (J. Le Sommer, A. Albert), who specialises on the high-resolution ocean modelling, and with University of Reading (4) in Reading (UK) (David Ferreira), where I collaborated with specialists on meso-scale ocean dynamics.

First results were presented (5) at the AGU annual meeting in San-Francisco (US) in December 2019: RECONSTRUCTION OF SUB-GRID-SCALE BUOYANCY FLUXES FROM LARGE-SCALE OCEAN VARIABLES.

The method of classification [Maze et al., 2017] was proposed and its application is in process (6). We plan to publish obtained results of this classification (7).

REFERENCES

1. Boccaletti, G., Ferrari, R., and Fox-Kemper, B. Mixed layer instabilities and restrati cation, Journal of Physical Oceanography, 37(9): 2228-50, doi: 10.1175/JPO3101.1, 2007.
2. Capet, X., Campos, E.J., and Paiva, M. Submesoscale activity over the Argentinian shelf, Geophysical Research Letters, 35, 2-6, doi:10.1029/2008GL034736, 2008.
3. Fox-Kemper, B., and Ferrari, R. Parameterization of mixed layer eddies. Part II: prognosis and impact, Journal of Physical Oceanography, 38, 1166-79, doi: 10.1175/2007JPO3788.1, 2008.
4. Maze,G., H. Mercier, R.Fablet, P. Tandeo, M.L. Radcenco, P. Lenca, C. Feucher, C.L . Goff. Coherent heat patterns revealed by unsupervised classification of Argo temperature profiles in the north Atlantic ocean. Prog. Oceanogr., http://dx.doi.org/10.1016/j.pocean.2016.12.008, 2017
5. Mensa, J.A., Z Garraffo, Z., Griffa, A., Ozg okmen, T.M., Haza, A., and Veneziani, M. Seasonality of the submesoscale dynamics in the Gulf Stream region, Ocean Dynamics, 63, 923-41, doi: https://doi.org/10.1007/s10236-013-0633-1, 2013.
6. Olbers, D., Willebrand, J., Eden, C. Ocean Dynamics, Springer, Heidelberg, 2012.
7. Qiu, B., Klein, P., and Chen, S. Seasonal mesoscale and submesoscale eddy variability along the North Paciffic Subtropical Countercurrent, Journal of Physical Oceanography, 44(12), 3079-98, doi: 10.1175/JPO-D-14-0071.1, 2014.
8. Rasp, S., Pritchard, M.S., and Gentine, P. Deep learning to represent subgrid processes in climate models, PNAS 115 (39), 9684-9689, https://doi.org/10.1073/pnas.1810286115, 2018.
9. Sasaki, H., Klein,P., Qiu B., and Sasa, Yi.. Impact of oceanic scale interactions on the seasonal modulation of ocean dynamics by the atmosphere, Nature Communications, 5, 5636 https://doi.org/10.1038/ncomms6636, 2014
10. Zanna, L., P. P. Mana, J. Anstey, T. David, and T. Bolton. Scale-aware deterministic and stochastic parametrizations of eddy-mean flow interaction, Ocean Modelling, 111, 66–80, https://doi.org/10.1016/j.ocemod.2017.01.004, 2017.