World Ocean Circulation

Algorithm Theoretical Basis Document for North Atlantic data-driven 2D Currents and and 3D Currents and Tracers products WOC-L4-CUR-NATL2D_REP-1D, WOC-L4-CUR-NATL3D_REP-1D (Theme 2)

customer	ESA/ESRIN
ESA contract	ESA Contract No. 4000130730/20/I-NB
document reference	WOC-ESA-ODL-NR-009_T2_WOC-L4-CUR-NATL2D_3D_REP-1D_V2.0
Version/Rev	**2.0**
Date of issue	**13/05/2022**

Issued by	Bruno Buongiorno-Nardelli (CNR) Daniele Ciani (CNR)
Approved by	Gilles Larnicol (ODL/Magellium)
Approved by	Marie-Hélène Rio (ESA)

Distribution List

	Name	Organization	Nb. copies
Sent to :	M.H. Rio	ESA/ESRIN	ESA-STAR
Internal copy :	Project Manager	OceanDatalab	1 (digital copy)

Document evolution sheet

Ed.	Rev.	Date	Purpose evolution	Comments
1	0	10/06/2020	Creation of document	Created by B.Buongiorno Nardelli and D.Ciani
2	0	13/05/2022	Updated	modified by B.Buongiorno Nardelli and D.Ciani

Contents

List Of Images

List Of Tables

- --

1 Introduction

1.1 Products summary

This document provides the algorithm theoretical basis for the two products developed by the Consiglio Nazionale delle Ricerche – Istituto di Scienze Marine (CNR-ISMAR) in the framework of the ESA World Ocean Circulation project.
The first product (WOC-L4-CUR-NATL2D_REP-1D) provides a level 4 (L4) gap free reconstruction of 2D ocean surface currents obtained through the combination of altimeter-derived geostrophic currents and satellite sea surface temperature (SST) L4 data.
The second product (WOC-L4-CUR-NATL3D_REP-1D) provides a 3D reconstruction of the horizontal and vertical currents, as well as of temperature and salinity, over 75 unevenly spaced vertical levels (denser close to the surface), between the surface and 1500 m depth.
Both products are computed on a 1/10° x 1/10° horizontal resolution grid, with a daily resolution, covering the 2010-2018 period, over a wide portion of the North Atlantic Ocean (20°N-50°N, 76°W-6°W).

1.2 Scope & Objectives

This document is the Algorithm Theoretical Basis Document (ATBD) which is dedicated to the description and justification of the algorithms used in the generation of:

- WOC-L4-CUR-NATL2D_REP-1D
- WOC-L4-CUR-NATL3D_REP-1D

products.

1.3 Document structure

In addition to this introduction, this document includes the following chapters:

Chapter 2: North Atlantic data-driven 2D Currents (WOC-L4-CUR-NATL2D_REP-1D)
Chapter 3: North Atlantic data-driven 3D Currents and Tracers (WOC-L4-CUR-NATL3D_REP-1D)

1.4 Applicable & Reference documents

- [RD-1] ESA WOC2019: http://woc2019.esa.int/index.php

- [RD-2] Synthesis of the WOC2019 User Consultation Meeting recommendations http://woc2019.esa.int/files/WOC2019_summary_synthesis.pdf

- [RD-3] Piterbarg, L.I. A simple method for computing velocities from tracer observations and a model output. *Appl. Math. Model.* 2009, *33*, 3693–3704.

- [RD-4] Rio, M.H.; Santoleri, R. Improved global surface currents from the merging of altimetry and Sea Surface Temperature data. *Remote. Sens. Environ.* 2018

- [RD-5] Ciani, D.; Rio, M.H.; Nardelli, B.B.; Etienne, H.; Santoleri, R. Improving the Altimeter-Derived Surface Currents Using Sea Surface Temperature (SST) Data: A Sensitivity Study to SST Products. Remote Sensing 2020, 12, 1601.

- [RD-6] Roberts-Jones, J., Fiedler, E. K., & Martin, M. J. (2012). Daily, global, high-r solution SST and sea ice reanalysis for 1985–2007 using the OSTIA system. *Journal of Climate*, *25*(18), 6215-6232

- [RD-7] Shchepetkin, A. F., & McWilliams, J. C. (2005). The regional oceanic modeling system (ROMS): a split-explicit, free-surface, topography-following-coordinate oceanic model. *Ocean modelling*, *9*(4), 347-40

- [RD-8] Vallis, G.K. *Atmospheric and Oceanic Fluid Dynamics*; Cambridge University Press: Cambridge, U.K., 2006; p. 745.

- [RD-9] Buongiorno Nardelli, B.; Guinehut, S.; Verbrugge, N.; Cotroneo, Y.; Zambianchi, E.; Iudicone, D. Southern ocean mixed-layer seasonal and interannual variations from combined satellite and in situ data. J. Geophys. Res. Oceans 2017, 122, 10042–10060, doi:10.1002/2017JC013314.

- [RD-10] Buongiorno Nardelli, B.; Mulet, S.; Iudicone, D. Three Dimensional Ageostrophic Motion and Water Mass Subduction in the Southern Ocean. J. Geophys. Res. Oceans 2018, 23, 1533–1562, doi:10.1002/2017jc013316.

- [RD-11] Buongiorno Nardelli, B. A Multi-Year Timeseries of Observation-Based 3D Horizontal and Vertical Quasi-Geostrophic Global Ocean Currents. 2020, No. April. https://doi.org/10.5194/essd-2020-73.

- [RD-12] Guinehut, S., Dhomps, a.-L., Larnicol, G. and Le Traon, P.-Y.: High resolution 3-D temperature and salinity fields derived from in situ and satellite observations, Ocean Sci., 8(5), 845–857, doi:10.5194/os-8-845-2012, 2012.

- [RD-13] Dee, D. P., Uppala, S. M., Simmons, A. J., Berrisford, P., Poli, P., Kobayashi, S., Andrae, U., Balmaseda, M. A., Balsamo, G., Bauer, P., Bechtold, P., Beljaars, A. C. M., Berg, L. Van De, Bidlot, J., Bormann, N., Delsol, C., Dragani, R., Fuentes, M., Geer, A. J. and Dee, D. P.: The ERA-Interim reanalysis : configuration and performance of the data assimilation system, , (April), 553–597, doi:10.1002/qj.828, 2011.

- [RD-14] Buongiorno Nardelli, B. A Deep Learning Network to Retrieve Ocean Hydrographic Profiles from Combined Satellite and In Situ Measurements. Remote Sens. 2020, 12 (19), 3151. https://doi.org/10.3390/rs12193151.

- [RD-15] Rio, M.; Mulet, S.; Picot, N. Beyond GOCE for the Ocean Circulation Estimate: Synergetic Use of Altimetry, Gravimetry, and in Situ Data Provides New Insight into Geostrophic and Ekman Currents. Geophys. Res. Lett. 2014, 41, 8918–8925. https://doi.org/10.1002/2014GL061773.

- [RD-16] Droghei, R.; Buongiorno Nardelli, B.; Santoleri, R. Combining in-situ and satellite observations to retrieve salinity and density at the ocean surface. J. Atmos. Ocean. Technol. 2016, 33, 1211–1223, doi:10.1175/JTECH-D-15-0194.1.

- [RD-17] Szekely, T.; Gourrion, J.; Pouliquen, S.; Reverdin, G. The CORA 5.2 dataset for global in situ temperature and salinity measurements: Data description and validation. Ocean Sci. 2019, 15, 1601–1614, doi:10.5194/os-15-1601-2019.

- [RD-18] Droghei, R., Buongiorno Nardelli, B. and Santoleri, R.: A New Global Sea Surface Salinity and Density Dataset From Multivariate Observations (1993–2016), Front. Mar. Sci., 5(March), 1–13, doi:10.3389/fmars.2018.00084, 2018.

- [RD-19] Locarnini, R.A.; Mishonov, A.V.; Antonov, J.I.; Boyer, T.P.; Garcia, H.E.; Baranova, O.K.; Zweng, M.M.; Paver, C.R.; Reagan, J.R.; Johnson, D.R.; et al. World Ocean Atlas 2013. Volume 1: Temperature; Levitus, S., Mishonov, A., Eds.; NODC: Silver Spring, MD, USA, 2013; Volume 73, p. 40.

- [RD-20] Zweng, M.M.; Reagan, J.R.; Antonov, J.I.; Mishonov, A.V.; Boyer, T.P.; Garcia, H.E.; Baranova, O.K.; Johnson, D.R.; Seidov, D.; Bidlle, M.M. World Ocean Atlas 2013, Volume 2: Salinity; NODC: Silver Spring, MD, USA, 2013; Volume 119, pp. 227–237.

- [RD-21] Hersbach H, Bell B, Berrisford P, Hurihara S, Horányi A et al. (2020). The ERA5 global reanalysis.

- [RD-22] Giordani, H., Prieur, L. and Caniaux, G.: Advanced insights into sources of vertical velocity in the ocean, Ocean Dyn., 56(5–6), 513–524, doi:10.1007/s10236-005-0050-1, 2006.

- [RD-23] Smyth, W. D., Skyllingstad, E. D., Crawford, G. B. and Wijesekera, H.: Nonlocal fluxes and Stokes drift effects in the K-profile parameterization, Ocean Dyn., 52(3), 104–115, doi:10.1007/s10236-002-0012-9, 2002.

- [RD-24] Roach, C. J.; Phillips, H. E.; Bindoff, N. L.; Rintoul, S. R. Detecting and Characterizing Ekman Currents in the Southern Ocean. J. Phys. Oceanogr., 2015, 45 (5), 1205–1223. https://doi.org/10.1175/JPO-D-14-0115.1.

- [RD-25] Nagai, T.; Tandon, A.; Rudnick, D. L. Two-Dimensional Ageostrophic Secondary Circulation at Ocean Fronts Due to Vertical Mixing and Large-Scale Deformation. J. Geophys. Res. 2006, 111 (C9), C09038. https://doi.org/10.1029/2005JC002964.

- [RD-26] Kalnay de Rivas, E.: On the use of nonuniform grids in finite-difference equations, J. Comput. Phys., 10, 202–210, 1972.

- [RD-27] Sundqvist, H. and Veronis, G.: A simple finite-difference grid with non-constant intervals, Tellus, XXII, 1–6 [online] Available from: http://onlinelibrary.wiley.com/doi/10.1111/j.2153-3490.1970.tb01933.x/abstract (Accessed 26 August 2013), 1970.

- [RD-28] Baker, A. H., Jessup, E. R. and Manteuffel, T.: A technique for accelerating the convergence of restarted gmres, SIAM J. Matrix Anal. Appl., 26(4), 962–984, doi:10.1137/S0895479803422014, 2005.

- [RD-29] Virtanen, P., Gommers, R., Oliphant, T. E., Haberland, M., Reddy, T., Cournapeau, D., Burovski, E., Peterson, P., Weckesser, W., Bright, J., van der Walt, S. J., Brett, M., Wilson, J., Millman, K. J., Mayorov, N., Nelson, A. R. J., Jones, E., Kern, R., Larson, E., Carey, C., Polat, İ., Feng, Y., Moore, E. W., VanderPlas, J., Laxalde, D., Perktold, J., Cimrman, R., Henriksen, I., Quintero, E. A., Harris, C. R., Archibald, A. M., Ribeiro, A. H., Pedregosa, F., van Mulbregt, P. and Contributors, S. 1. 0: SciPy 1.0--Fundamental Algorithms for Scientific Computing in Python, , 1–22 [online] Available from: http://arxiv.org/abs/1907.10121, 2019.

- [RD-30] Lumpkin, R., Grodsky, S. A., Centurioni, L., Rio, M. H., Carton, J. A. and Lee, D.: Removing spurious low-frequency variability in drifter velocities, J. Atmos. Ocean. Technol., 30(2), 353–360, doi:10.1175/JTECH-D-12-00139.1, 2013.

- [RD-31] Lumpkin, R., Özgökmen, T. and Centurioni, L.: Advances in the Application of Surface Drifters, Ann. Rev. Mar. Sci., 9(1), 59–81, doi:10.1146/annurev-marine-010816-060641, 2017.

1.5 Terminology

ACCUA	Analisi della dinamica della Corrente CircUmpolare Antartica
AMSR2	Advanced Microwave Scanning Radiometer 2
ADT	Absolute Dynamic Topography
AI	Artificial Intelligence
AIL	Action Items List
AIS	Automatic Identification System
ASAR	Advanced Synthetic Aperture Radar
ASCAT	Advanced SCATterometer
ATBD	Algorithm Theoretical Basis Document
AVHRR	Advanced Very High Resolution Radiometer
CCD	Contract Closure Document
CCI	Climate Change Initiative
CCMP	Cross-Calibrated Multi-Platform

CFOSAT	Chinese-French Oceanography Satellite
CIESM	Mediterranean Science Commission
CMEMS	Copernicus Marine Environment Monitoring Service
CNES	Centre National d'Etudes Spatiales
CNR	Consiglio Nazionale delle Ricerche
CTD	Conductivity, Temperature and Depth
DP	Data Pool
DTU	Danmarks Tekniske Universitet
DUACS	Data Unification and Altimeter Combination System
EBUS	Eastern Boundary Upwelling System
ECCO	Estimating the Circulation & Climate of the Ocean
ECMWF	European Centre for Medium-Range Weather Forecasts
EFARO	European Fisheries and Aquaculture Research Organisations
EMB	European Marine Board
ENVISAT	Environmental Satellite
EO	Earth Observation
EPB	European Polar Board
EOEP-5	5th Earth Observation Envelope Programme (2017-2021)
ERA	ECMWF Reanalysis
ESA	European Space Agency
ESF	European Science Foundation
EU	European Union
EuroGOOS	European Global Ocean Observing System
FAO	Food and Agriculture Organization of the United Nations
FR	Final Report
FSLE	Finite Size Lyapunov Exponent
GCM	Global Circulation Model
GOCE	Gravity Field and Steady-State Ocean Circulation Explorer
GOOS	Global Ocean Observing System
GMI	Global precipitation monitoring Microwave Imager
GMM	Gaussian Mixture Models
HYCOM	Hybrid Coordinate Ocean Model
IAR	Impact Assessment Report
ICCAT	International Commission for the Conservation of Atlantic Tunas
ICSU	International Council for Science
IGPB	International Geosphere-Biosphere Programme
IUGG	International Union of Geodesy and Geophysics
IFREMER	Institut Français de Recherche pour l'Exploitation de la Mer
IOC	Intergovernmental Oceanographic Commission
ITCZ	InterTropical Convergence Zone

mEOF-r	Multivariate Empirical Orthogonal Functions reconstruction
MoM	Minutes of Meeting
NATL3D	North Atlantic 3D Ocean Currents
NEMO	Nucleus for European Modelling of the Ocean
NOAA	National Oceanic and Atmospheric Administration
NCC	Norwegian Coastal Current
NwAFC	Norwegian Atlantic Front Current
NwASC	Norwegian Atlantic Slope Current
OLCI	Ocean and Land Color Imager
OSSE	Observing System Simulation Experiment
OC	Ocean color
OSCAT	Oceansat-2 SCATterometer
PD	Product Delivery
PM	Project Manager
PMP	Project Management Plan
PUB	Publication
PUM	Product User Manual
QUID	Quality Information Document
RB	Requirement Baseline
REMSS	Remote Sensing Systems
ROMS	Regional Oceanic Modeling System
RTOFS	Real-Time Ocean Forecast System
S3	Sentinel 3
SAR	Synthetic Aperture Radar
SEVIRI	Spinning Enhanced Visible and InfraRed Imager
SCOR	Scientific Committee on Oceanic Research
SIED	Single Image Edge Detection
SKIM	Sea surface KInematics Multiscale monitoring
SLSTR	Sea & Land Surface Temperature Radiometer
SOCIB	Sistema d'observació i predicció costaner de les Illes Balears
SODA	Simple Ocean Data Assimilation Ocean/sea ice reanalysis
SSH	Sea Surface Height
SSS	Sea Surface Salinity
SST	Sea Surface Temperature
SVP	Surface Velocity Program
SoW	Statement of Work
TN	Technical Note
TOPAZ	Tracers of Phytoplankton with Allometric Zooplankton
TUOC	Total Upper Ocean Currents
UCL	Use Case Library

UCM	User Consultation Meeting
UCPC	Upper-layer ocean Circulation Processes e-Catalogue
UI	Upwelling Index
UN	United Nations
URD	User Requirement Document
VR	Validation Report
VT	Visualization Tool
WBS	Work Breakdown Structure
WOC	World Ocean Circulation

2 North Atlantic data-driven 2D Currents (WOC-L4-CUR-NATL2D_REP-1D)

2.1 Overview

The WOC-L4-CUR-NATL2D_REP-1D product has been developed by CNR-ISMAR applying a methodology to optimally combine altimeter-derived geostrophic currents and sequences of L4 SST based on the results of [RD-3] (also referred to as Piterbarg Method), [RD-4] and [RD-5]. This technique enables to reconstruct the ocean surface circulation from the altimeter-derived geostrophic currents and sequences of satellite L4 SST observations taking into account the source/sink terms for the SST evolution equation (related to the large scale interactions with the atmospheric boundary layer).
The idea behind the methodology is to infer the two components of the marine surface circulation using the geostrophic currents as a first guess estimate and inverting the SST evolution equation (1) to infer the correction factors to compute an optimized surface current estimate:

$$\frac{\partial SST}{\partial t} + u\frac{\partial SST}{\partial x} + v\frac{\partial SST}{\partial y} = F \qquad (1)$$

where (u,v) are respectively the zonal and meridional components of the ocean surface flow, (x,y) are the zonal and the meridional directions and F is the forcing term (here approximated as the low pass spatial filtering of the SST temporal derivatives, choosing the 1000 km filtering scale).

The method implementation extensively relies on the computation of the SST spatial and temporal derivatives, and requires the following input data:

1. daily maps of gap-free (L4) satellite SST;
2. daily maps of altimeter-derived L4 absolute geostrophic currents;
3. a static map describing the uncertainty on the geostrophic currents (computed via comparison with SVP drifting buoys over 25 years), also necessitating an empirical calibration;
4. a static map describing the uncertainty on the SST forcing (the source and sink terms, F in equation 1), which is computed via comparison with SVP drifting buoys over 25 years), also necessitating an empirical calibration.

The input data provided at points 3 and 4 were firstly identified by [RD-4] as crucial to optimize and calibrate the reconstruction method. This synergistic approach, applied to OSTIA L4 SST allowed to improve the global sea surface currents estimates with respect to the altimeter system [RD-5].

2.2 Step/module 1: input data preparation

2.2.1 Sea surface temperature

The SST used in input is the L4 multi-year reprocessed Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) developed by the U.K. Met Office and distributed by CMEMS (CMEMS product ID: SST_GLO_SST_L4_REP_OBSERVATIONS_010_011). OSTIA combines and interpolates the reprocessed ESA SST CCI, C3S, EUMETSAT, REMSS and OSPO satellite data, and in situ data from HadIOD. OSTIA L4 provides daily foundation SST values (i.e., values that are not affected by the diurnal cycle). The processing is based on an optimal interpolation (OI) algorithm, run over a 1/20° regular grid [RD-6]. Here, the OSTIA SST was sub-sampled to 1/10° resolution, and the resulting grid is taken as the final grid used for the pre-processing of the other surface datasets.

2.2.2 Background geostrophic currents

The input geostrophic currents are obtained from CMEMS, (CMEMS product ID: SEALEVEL_GLO_PHY_L4_REP_OBSERVATIONS_008_047). Such currents are a multi-satelllite mission product and are derived merging observations from all available altimeters, e.g., Jason-3, Sentinel-3A, HY-2A, Saral/AltiKa, Cryosat-2, Jason-2, Jason-1, T/P, ENVISAT, GFO, ERS1/2. The altimeter measurements are used to produce global maps of absolute dynamic topography at the L4 processing level, thus allowing to derive L4 surface currents velocities via the geostrophic approximation. The geostrophic currents are originally provided on a 1/4° regular grid, thus necessitating an upsize to the final 1/10° grid to be merged with SST data through the method described in [RD-3] and [RD-4].

2.2.3 Additional input data for the Piterbarg algorithm

2.2.3.1 Error on the geostrophic currents

To determine the error on the geostrophic currents is useful to understand where the Altimeter currents are less accurate and to weight differently the corrections introduced by the tracer (SST) observations through the PIT09 and RS18 methods. This has been computed interpolating the Altimeter derived geostrophic currents (zonal and meridional) over the trajectories of the NOAA/AOML SVP drogued drifting buoys during 1993-2018. Then, the RMS error of the geostrophic currents is evaluated by binning all the interpolated observations in 4°x4° boxes, using the SVP currents as a benchmark. A 500 km smoothing is finally applied to mitigate sharp spatial variations due to the 4°x4° binning procedure. An example for the zonal currents error is provided in Figure 1.

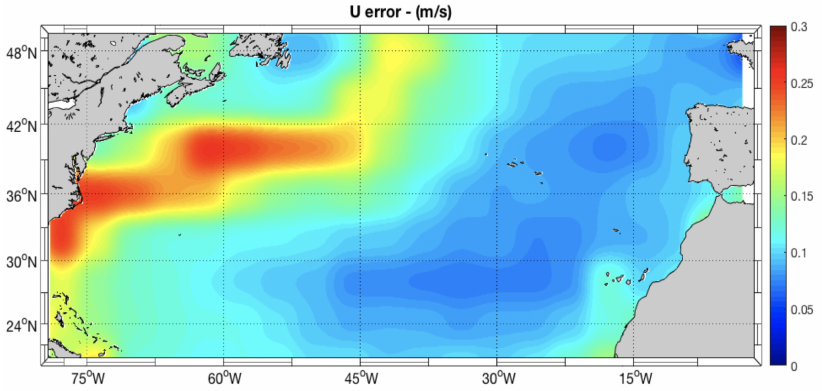

Figure 1. Uncertainty on the zonal geostrophic currents (identified as σu in equation (4)).

2.2.3.2 Error on the forcing term

In our work, following [RD-4], the Forcing term is approximated via low-pass spatial filtering of the daily OSTIA SST temporal derivatives

$$F = \partial tSST|_{L}$$

with the subscript t standing for temporal derivative and L indicating the low pass filtering scale, initially identified with 500 km by [RD-4].

The error on such an approximation of the F term is also evaluated relying on observations from the drogued SVP drifting buoys. Such lagrangian platforms allow to have an in-situ estimate of the forcing term according to Equation 1. The RMS error of the F term is thus evaluated binning the satellite and in-situ derived F term in 4°x4° boxes during the 1993-2018 period. The forcing term error field is provided in Figure 2.

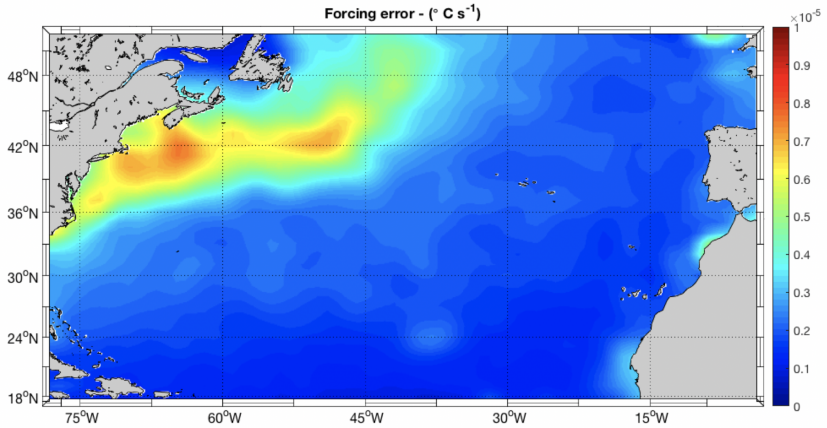

Figure 2. Uncertainty on the Forcing term (identified as "h" in equation (4)).

In the present study, relying on the in-situ measured F term provided by the drifting buoys, we performed a dedicated analysis to assess the optimal filtering scale to approximate the F term from the OSTIA ∂_tSSTs. Focusing on the 1993-2018 period, we created daily time series of

∂tSSTs. Such ∂_tSSTs were low-pass filtered at scales of 200, 300, 500, 1000, 1500 and 2000 km. For each low-pass filtered ∂_tSST time serie we built six static error maps (one for each ∂tSST filtering scale) as illustrated at the beginning of this section. The resulting mean error value is given in Figure 3 as a function of the ∂_tSST filtering scale.

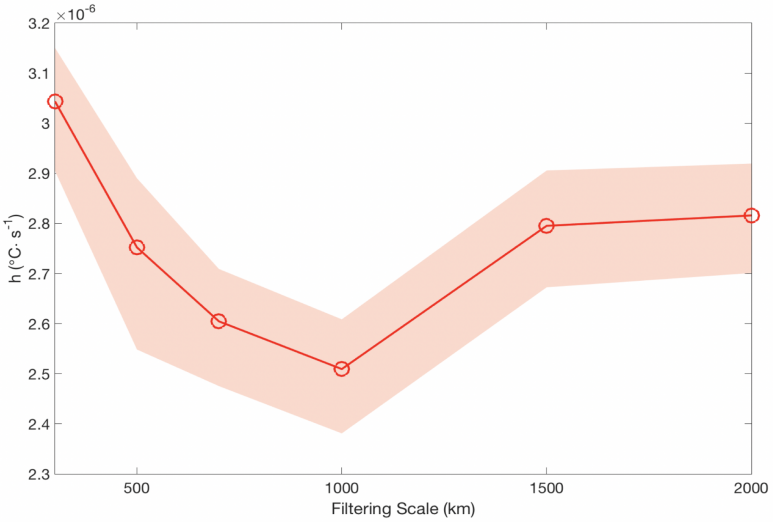

Figure 3. Mean error on the Forcing term as a function of the ∂_tSST spatial low-pass filtering scale. The shadowed area indicates the uncertainty on the mean value evaluated via bootstrap technique.

The analyses showed that the 1000 km filtering minimises the discrepancies between the satellite and the in-situ measured F term. Therefore, the NATL_2D surface currents exploit the RS18 method, with modified forcing term based on a ∂_tSST low passed filtered at 1000 km. The corresponding static error map is the one sketched in Figure 2.

2.3 Step/module 2: Computation of the Optimal Currents

The computation of the OPtimal Currents (OPC) is carried out by applying zonal and meridional correction factors (U_{CORR}, V_{CORR}, respectively) to the Altimeter geostrophic (GEO) currents.

$$\begin{aligned} u_{CORR} &= u_0 \sin\phi + v_0 \cos\phi \\ v_{CORR} &= -u_0 \cos\phi + v_0 \sin\phi \end{aligned} \qquad (2)$$

where

$$\begin{aligned} u_0 &= \frac{f(\min(\beta, q)) - f(\max(\alpha, -q))}{g(\min(\beta, q)) - g(\max(\alpha, -q))} \\ v_0 &= p u_0 \\ \phi &= \operatorname{atan}(A/B) \end{aligned} \qquad (3)$$

and

$$
\begin{aligned}
&f(\gamma) = -2(q^2-\gamma^2)^{3/2}/3;\\
&g(\gamma) = x(q^2-\gamma^2) + q^2 \mathrm{asin}(\gamma/q);\\
&p = \sin\phi\cos\phi(\sigma^2{}_v - \sigma^2{}_u)q^{-2};\\
&q = \sqrt{\sigma^2{}_u \sin^2\phi + \sigma^2{}_v \cos^2\phi};\\
&\alpha = (Au_{GEO} + Bv_{GEO} + E - h) / \sqrt{A^2 + B^2};\\
&\beta = (Au_{GEO} + Bv_{GEO} + E + h) / \sqrt{A^2 + B^2};
\end{aligned}
\qquad (4)
$$

with

- A=∂xSST
- B=∂ySST
- E=∂tSST-F (with F= the forcing term appearing in (1)
- the subscripts (x,y,t) stand for derivative with respect to longitude, latitude and time
- $\sigma u/v$ stands for the error on the background zonal/meridional geostrophic currents
- u_{GEO},v_{GEO} stand for zonal and meridional altimeter-derived geostrophic currents, respectively
- h is the error on the forcing term

The optimal currents are finally given by

$$(U,V)_{OPC}=(U,V)_{GEO}+(U,V)_{CORR} \qquad (5)$$

The correction factors are thus strongly dependent on the SST spatial temporal derivatives as well as the SST source/sink terms. After [RD-4] and [RD-5], the computation of the optimal currents requires a fine tuning of the PIT09 algorithm. In order to maximise the improvements with respect to the altimeter system, the error maps presented in Section 2.2.3 need to be empirically calibrated relying on in-situ data provided by the lagrangian SVP drogued drifters. Such an operation will impact on the set of equation (4), thus modulating the value of the correction factors.

The empirical calibration of $\sigma u/v$ and h has been carried out computing three year long timeseries of optimal currents (during 2014-2016) using equations (2),...,(5) and applying empirical calibration factors going from 1 to 4 at steps of 0.5 to both $\sigma u/v$ and h fields. This resulted in 49 different reconstructions, for each combination of the calibration factors. For each reconstruction, the average improvement (percentage of improvement, PI) with respect to standard altimetry has been computed via equation (6) (using the technique detailed in Section 2.4)

$$\mathrm{PI}_{U,V} = 100\left[1 - \left(\frac{\mathrm{RMSE}^{OPC}_{U,V}}{\mathrm{RMSE}^{GEO}_{U,V}}\right)^2\right]$$

(6)

In Equation (6), the subscripts U,V stand for zonal and meridional PI, respectively. The PI indicates the improvements of the Altimeter-derived currents after the optimal combination with the satellite SST data. The $RMSE^{OPC/GEO}$ computation relies on the knowledge of a reference surface current, provided by the SVP drifters estimates. Equation (6) allowed us to evaluate the mean intensity of the Percentage of Improvement (PI) and the areas in which the PI was larger than zero, the latter expressed in percentage with respect to the entire NATL study area. The calibration factors that enabled to maximise the areas of improvement as well as its mean value were 2.5 and 3 for $\sigma u/v$ and h, respectively, in agreement with [RD-4].

2.4 Product validation

An example of OPC reconstruction is provided in Figure 4. The structure depicted by both the Altimeter "GEO" currents and the OSTIA SSTs (remapped at the 1/10° resolution) corresponds to a cyclonic mesoscale eddy in the North Atlantic, most likely generated by the pinching off of a Gulf Stream meander. The Altimeter-derived GEO currents are not in full agreement with the underlying higher resolution SST field, exhibiting cross-SST gradient circulations in the northern and southernmost areas of the eddy. Merging the SST gradients with the lower effective resolution Altimeter estimates, the surface circulation is optimized and better resolves the structure found in the high-resolution satellite SST image. This is graphically emphasized in green in Figure 4: the OPC reconstruction rotates the GEO currents by constraining them using the information contained in the satellite SST.

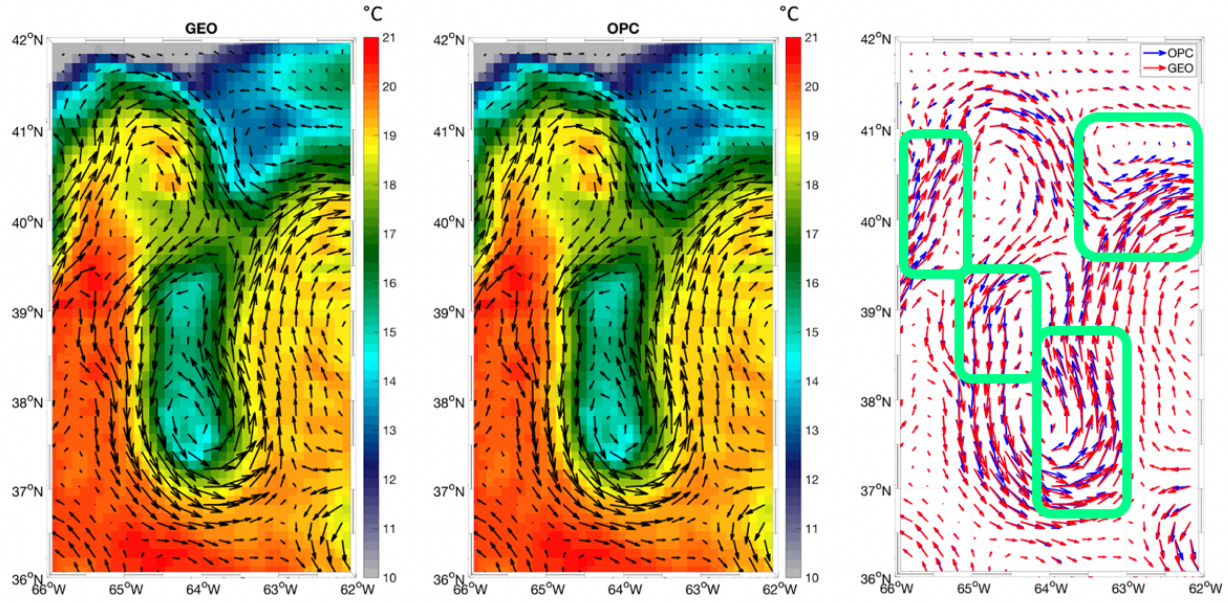

Figure 4. left) Vectors: CMEMS Altimeter-derived GEO currents (upsized to the 1/10° WOC grid). Contours: OSTIA SST, 1/10°; middle) Vectors: Optimal currents (OPC), obtained by merging the Altimeter currents with the OSTIA SSTs. Contours: OSTIA SST; right) direct comparison of the GEO and OPC currents estimate. The fields refer to March 24th, 2015.

An additional illustration based on the use of the WOC Visualisation portal (https://woc.oceandatalab.com) is also provided and is accessible via the following share link: https://odl.bzh/ngvDR7bm

It shows the 2D surface circulation on July 11th 2018, described by three different approaches:

- The CMEMS Globcurrent Geostrophic Currents (white streamlines)
- The 2D surface currents from the WOC NATL2D product (dark streamlines)
- The evolution of a Coriolis CMEMS drogued drifter, whose position on July 11th 2018 is given by the white diamond (the drifter is moving northward)

The underlying SST field is a 3 day composite from Sentinel 3 SLSTR-A observations (accounting for observations from July 10th to July 12th 2018). The circulation feature presented here shows a warm meander characterised by a ~+2K temperature anomaly with respect to the surrounding environment and located between 41°W to 40°W and 44°N to 45°N (well depicted by the SLSTR observations).

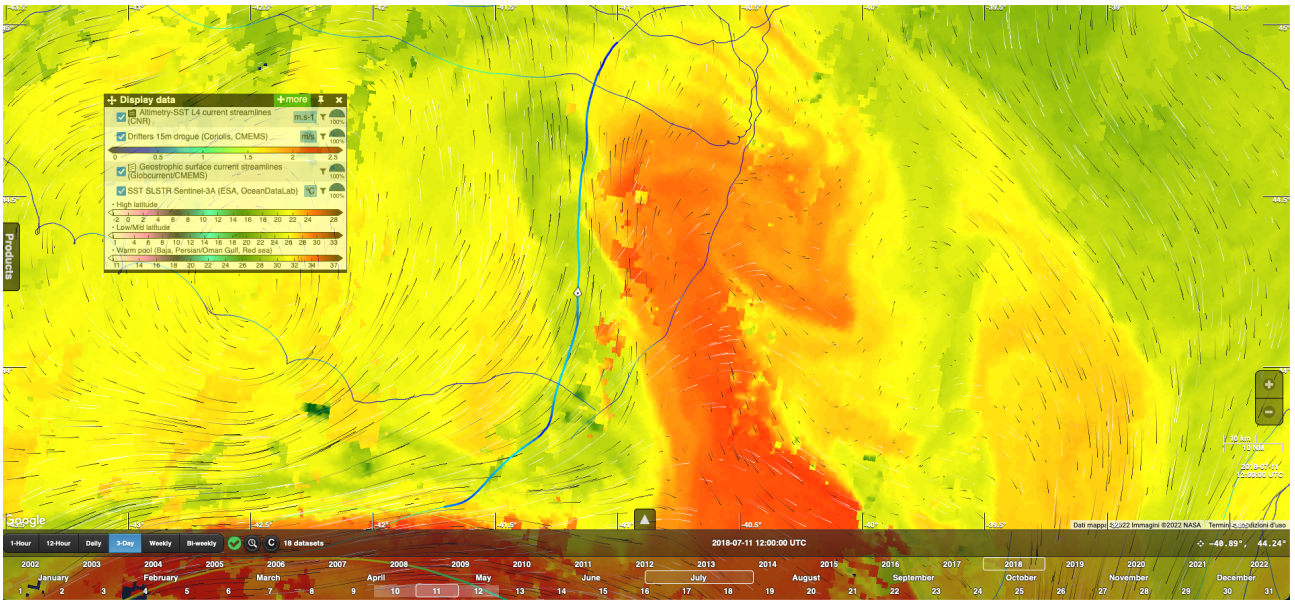

Figure 5. Intercomparison of CMEMS geostrophic currents (white streamlines), WOC NATL2D currents (dark streamlines) on top of the Sentinel 3 SLSTR-A derived sea surface temperature (July 11th, 2018). The map also shows the trajectory of a Coriolis drifter (continuous line).

The CMEMS geostrophic currents entirely cross the meander and exhibit a larger angle than the NATL2D product with respect to the drifter trajectory. Moreover, the NATL2D product streamlines (dark green) are in a good agreement with the head-shaped feature in the northern section of the meander, suggesting a nearly closed anticyclonic circulation, which is not found in the CMEMS geostrophic current streamlines.

While the last results must be interpreted as a qualitative validation of the OPC, a quantitative validation of the OPC was carried out computing the percentage of improvement (PI) with respect to standard altimetry via equation (6). In the 2010-2019 period, we interpolated the daily altimeter-derived and optimal currents maps onto the trajectories of the SVP drogued drifting buoys. Then, we computed RMS errors of the GEO and OPC currents estimates using the in-situ measured currents as a benchmark. In the end, the PI is computed in 2° x 2° boxes, binning all available observations and using equation (6).

The resulting PI is provided in Figure 6. In a fair agreement with past studies (e.g.[RD-4], [RD-5]) our reconstruction methodology brings larger improvements for the meridional component of the surface circulation, with PIs that locally reach 20% and an enhanced improvement area located along the axis of the Gulf Stream. Moreover, 67% of the analysed boxes exhibit PIs larger than zero.

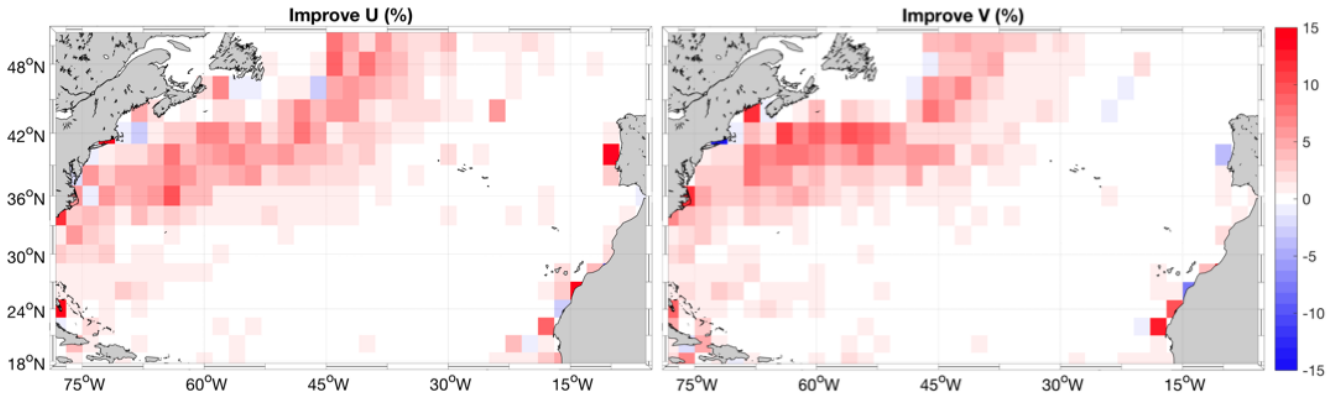

Figure 6. left) percentage of improvement computed via equation (6) in 2° x 2° boxes, for the zonal optimal currents. right) same as left, for the meridional currents. The statistics are computed over the 2010-2019 period.

Regarding the zonal component of the motion, the local improvements are in the 5-13% range and the boxes where the PI is larger than zero cover 69% of the study area. For both components of the surface circulation, occasional degradations may occur, although they never exceed -2%. We also notice that the larger improvements are obtained in proximity of the Gulf Stream main axis, i.e. where the corrections brought by our methodology are generally larger due to an intensification of the SST gradients magnitude (thus resulting in an enhancement of the correction factors intensity).

2.5 Computation of ADT for ingestion in the WOC-L4-CUR-NATL3D_REP-1D product

Using the dataset of optimal currents described in section 2.4, we built an absolute dynamic topography (ADT) to be ingested in the algorithm used to compute the WOC-L4-CUR-NATL3D dataset (described in section 3). Such ADT is computed relying on software tools developed within the Regional Oceanic Modeling System (ROMS, [RD-7]), that enable to compute the

stream function associated with a given surface current field. In our case, the surface currents field is represented by the WOC-L4-CUR-NATL2D currents.

The algorithm that computes the stream function firstly derives the surface relative vorticity ξ,

$$\xi = \partial_x v - \partial_y u \quad (7)$$

where (u,v) are the zonal and meridional currents and the subscripts (x,y) stand for derivative with respect to the zonal and meridional directions.

Then, the algorithm inverts equation (8) to obtain the stream function ψ,

$$\Delta\psi = \xi \quad (8)$$

where Δ is the laplacian operator. Finally, the ADT is obtained from the stream function relying on the framework of the geostrophic approximation [RD-8], e.g. using equation (9)

$$\eta = \psi \cdot \frac{f}{g} \quad (9)$$

where f is Coriolis parameter and g is the gravity acceleration.

An example of ADT derived from the Optimal Currents is provided in Figure 7.

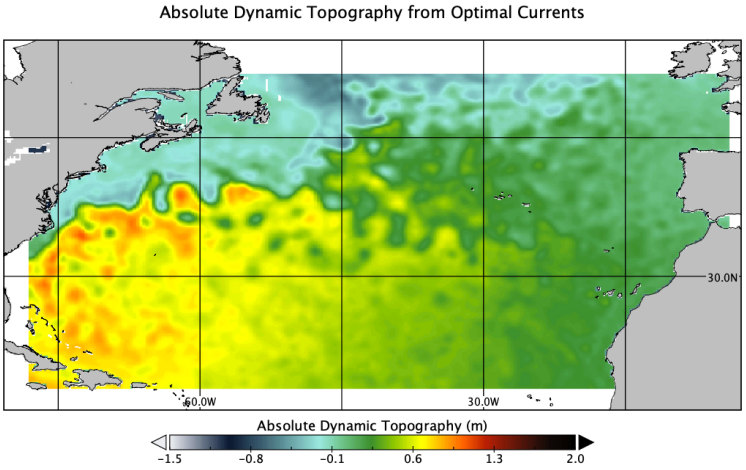

Figure 7. ADT derived from the Optimal currents (March 24th 2015)

3 North Atlantic data-driven 3D Currents and Tracers (WOC-L4-CUR-NATL3D_REP-1D)

3.1 Overview

The WOC-L4-CUR-NATL3D_REP-1D product has been developed by CNR starting from the Omega diagnostic model initially set-up during the ACCUA (Analisi della dinamica della Corrente CircUmpolare Antartica) project [RD9-10], that has been successively implemented over the global oceans within the Copernicus Marine Environment Monitoring Service (CMEMS) at 1/4°x1/4°, weekly resolution (CMEMS product ID: MULTIOBS_GLO_PHY_W_REP_015_007, [RD-11]).
The CMEMS Omega diagnostic tool requires in input the 3D fields of density and geostrophic currents, as well as momentum and heat fluxes at the air-sea interface and it is fully documented in [RD-10]. Within the CMEMS processing, the 3D fields of density are taken from the ARMOR3D processing [RD-12], which is based on a synergistic approach to combine remotely sensed surface measurements of sea surface temperature (SST) and salinity (SSS), absolute dynamic topography and in situ profiles of temperature (T) and salinity (S). CMEMS processing then includes surface forcings obtained from ERA-Interim atmospheric re-analysis [RD-13].
Within WOC, several modifications/improvements have been implemented to the original processing chain: 1) input surface fields have been taken from existing or specific newly-developed high resolution data (1/10°, daily); 2) CMEMS ARMOR3D 2D-to-3D processing has been substituted by a novel deep learning technique [RD-14], tuned to optimize the reconstruction of 3D temperature and salinity fields (from which density is obtained through standard UNESCO formula) in the study area; 3) the Omega processing has been adapted to the new output grid. It was also modified to improve the reconstruction of the diabatic components of the ageostrophic motion, by taking advantage of the modelled horizontal Ekman currents developed during ESA-Globcurrent project and operationally provided by CMEMS (product ID: MULTIOBS_GLO_PHY_REP_015_004, [RD-15]) and including the atmospheric forcings derived from the ERA5 reanalysis.

3.2 Step/module 1: Input data preparation

3.2.1 Sea Surface Temperature

The reader is referred to Section 2.2.1.

3.2.2 Sea Surface Salinity

Sea surface salinity (SSS) data were obtained by adapting to the 1/10° North Atlantic grid the multidimensional optimal interpolation algorithm used within the CMEMS to retrieve the global SSS product (CMEMS product_id: MULTIOBS_GLO_PHY_S_SURFACE_MYNRT_015_013). This algorithm interpolates SMOS observations and in situ SSS observations using a space-time-thermal decorrelation function, which extracts information on water mass distribution from high-pass filtered daily SST data [RD-16]. Within WOC, we ingested as input data: the L3OS 2Q debiased daily valid ocean salinity values product from SMOS satellite, produced and disseminated by the Centre Aval de Traitement des Données SMOS (CATDS, 2017), the sub-sampled OSTIA SST data and CMEMS-CORA5.2 surface salinity values (CMEMS product_id: INSITU_GLO_TS_REP_OBSERVATIONS_013_001_b, doi: 10.17882/46219TS1, [RD-17]). CMEMS weekly SSS fields were used to build the background (linearly interpolating it in time between the two closest analysis dates, and upsizing to the 1/10° grid through a cubic spline). All other interpolation parameters have been set as in [RD-18].

3.2.3 Absolute Dynamic Topography

The reader is referred to Section 2.5 for details on the ADT processing.

Before entering the 3D reconstruction, ADT data are further processed to make them consistent with in situ steric heights. The adjustment was carried out as in [RD-9], namely by regressing steric heights and co-located ADT data in the neighbourhood of each grid point, considering matchups within a temporal window of ± 10 days.

3.2.4 In Situ Vertical Profiles

The vertical hydrographic profiles were taken from the quality controlled Argo and CTD profiles produced by CMEMS CORA 5.2 (CMEMS product_id: INSITU_GLO_TS_REP_OBSERVATIONS_013_001_b) [RD-17]. The data considered here were to the 2010–2018 period, and were interpolated through a spline on a regularly spaced vertical grid (with 10 m intervals). Steric heights were computed taking 1500 m as reference level.

3.2.5 In situ Climatology

Temperature and salinity monthly climatological fields computed by the World Ocean Atlas 2013 were used to convert all daily observations to anomaly fields. These climatologies are estimated on a 1/4° × 1/4° grid by applying an objective analysis algorithm [RD-19-20]. The values in the first 1500 m, provided on 125 levels, were interpolated through a spline on a regularly spaced vertical grid (with 10 m intervals), and upsized to the 1/10° NATL3D grid through a cubic spline.

3.2.6 Surface fluxes

Surface air-sea fluxes are included in the ERA5 global atmospheric reanalysis by the European Centre For Medium-Range Weather Forecasts (ECMWF). ERA5 is a global reanalysis for the atmosphere, ocean and land surface that provides outputs of a range of variables with temporal and spatial continuity over the period 1979 to present. A description of the ERA5 1979-present system is provided in [RD-21]. ERA5 data processing is carried out using ECMWF Earth System model IFS, cycle 41r2 (see IFS Documentation CY41R2). IFS assimilates several observations of upper-air atmospheric variables (e.g. satellite radiances, temperature, wind vectors, specific humidity, and ozone). NATL3D diabatic forcings take in input the mean daily

fields of the zonal and meridional components of the turbulent surface stress, the surface latent and heat flux, the surface net solar and thermal radiation, as well as total precipitation and evaporation (needed to estimate the equivalent surface salinity flux). Daily averaged fields have been mapped to NATL3D grid through a cubic spline.

3.2.6 Ekman currents

CMEMS provides a multi-year L4 of global total velocity field at 0 m and 15 m, including zonal and meridional velocity at a 3h frequency over 1/4° regular grid (product ID: MULTIOBS_GLO_PHY_REP_015_004. These data are obtained by combining geostrophic surface currents estimated from satellite altimetry and Ekman currents modelled through the method developed by [RD-15], using ECMWF ERA5 wind stress in input. This product has been developed during the Globcurrent project (funded by the ESA User Element Program) and it is now distributed by CMEMS. Daily averaged fields have been upsized to NATL3D grid through a cubic spline.

3.3 Step/module 2: 2D-to-3D projection

The algorithm used to retrieve the 3D temperature and salinity from surface data has been developed during WOC. It has been fully documented in [RD-14]. It is based on a Long Short-Term Memory Network (LSTM) and its main characteristics and assumptions are only summarized in the following sub-section.

The LSTM code has been released under the terms of the GNU General Public Licence v3 and is available at the following address: https://github.com/bbuong/3Drec.

3.3.1 Long Short-Term Memory Network with Monte-Carlo Dropout (LSTM3D)

Long Short-Time Memory (LSTM) networks are a particular type of Recurrent Neural Networks that are particularly fit to model ordered sequences of data. They are based on a particular cell design, able to keep information from very far cells in a sequence. They can include a single layer of LSTM cells or multiple LSTM layers stacked one on top of the other. The number of cells in each layer matches the length of the sequence by definition, but the number of layers and hidden units in the LSTM cell needs to be optimized through empirical tuning.

Here, the sequential information in input is provided by the multivariate output state vector comprising temperature, salinity and steric height anomaly profiles (all anomalies are computed with respect to WOA13 climatology). In practice, each cell in the sequence considers in input the same values (i.e., the anomalies of SST, SSS and adjusted ADT, plus latitude, longitude and cyclic day), but takes the output values at increasing depths (with depth “acting” as time in more standard applications of LSTM). All vectors are scaled within the 0–1 range before feeding the network. The best performance was obtained with a 2-layer stacked network, including 35 hidden units in each LSTM layer (hereafter, LSTM (35–35)), as depicted in Figure 8.

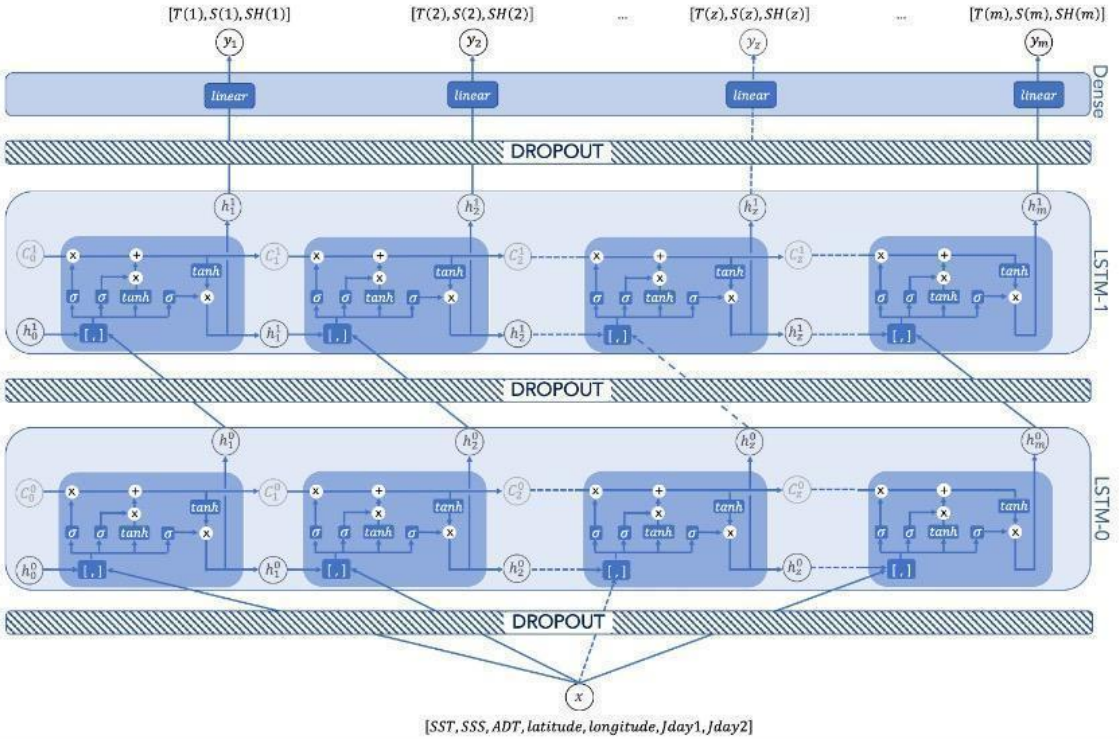

Figure 8. Diagram showing the elements of the stacked LSTM model for the reconstruction of vertical hydrographic profiles.

The LSTM3D network was trained using a particular regularization strategy known as dropout. It consists in randomly excluding a percentage of units during network training. Dropout was applied here both during training and testing. When applied during testing/prediction, it provides an extremely simple and powerful approach to quantify a neural network uncertainty. In fact, running a regression neural network several times with dropout during testing generates different outputs for the same input, equivalent to a Montecarlo sampling, and ensemble mean and variance thus provide the network's output values and related uncertainty, respectively. Here, a dropout rate of 20% of the units has been used.

3.4 Step/module 3: 3D currents retrieval

The reconstruction of the 3D horizontal and vertical currents is based on the solution of a diabatic Q-vector formulation of the QG Omega equation [RD-10, RD-11, RD-22]. The numerical implementation of the Omega solution is exactly the same as that used within CMEMS to deliver the OMEGA3D product, detailed in [RD-11], and will thus be only briefly recalled in the following.

Of course, the CMEMS Omega solution tool was first adapted to the NATL3D spatial grid and to the forcings derived from the WOC 2D/3D input data. Then, the Omega diabatic forcing has been modified with respect to the configuration used in OMEGA3D, which was based on a modified version of the K-Profile Parameterization (KPP, in the formulation by [RD-23]), by introducing a background empirical ageostrophic shear term based on the horizontal Ekman

currents developed during ESA-Globcurrent project, by removing the non-local flux of momentum, and by further adjusting the viscosity within the Ekman layer to reduce the differences between the background and retrieved horizontal ageostrophic velocities (details are provided in section 3.4.2).

3.4.1 Omega equation

The following diabatic Q-vector formulation of the quasi-geostrophic Omega equation [RD-10, RD-11, RD-22] is solved to get the NATL3D vertical velocity fields:

$$\nabla^2{}_h(N^2 w) + f^2 \frac{\partial^2 w}{\partial z^2} = \nabla_h \cdot \boldsymbol{Q} \tag{10}$$

In Eq. (10), w represents the vertical velocity (positive upwards), N^2 is the Brunt–Väisälä frequency, f is the Coriolis parameter, h indicates the horizontal components, and the $\boldsymbol{Q}$ vector includes three components reflecting different processes (kinematic deformation, twg, turbulent buoyancy, th, and turbulent momentum, dm), as defined below:

$$\boldsymbol{Q} = 2\boldsymbol{Q_{twg}} + \boldsymbol{Q_{th}} + \boldsymbol{Q_{dm}}$$

$$\mathbf{Q_{twg}} = \frac{g}{\rho_0}\left(\frac{\partial u_g}{\partial x}\frac{\partial \rho}{\partial x} + \frac{\partial v_g}{\partial x}\frac{\partial \rho}{\partial y}, \frac{\partial u_g}{\partial y}\frac{\partial \rho}{\partial x} + \frac{\partial v_g}{\partial y}\frac{\partial \rho}{\partial y}\right)$$

$$\mathbf{Q_{dm\_woc}} = \frac{f}{\rho_0}\left(\frac{\partial^2}{\partial z^2}\left[\rho K_m\left(\frac{\partial v_g}{\partial z} + \frac{\partial v_{Ekman}}{\partial z}\right)\right], -\frac{\partial^2}{\partial z^2}\left[\rho K_m\left(\frac{\partial u_g}{\partial z} + \frac{\partial u_{Ekman}}{\partial z}\right)\right]\right)$$

$$\mathbf{Q_{th}} = -\frac{g}{\rho_0}\nabla_h\left(\frac{\partial}{\partial z}\left[K_\rho\left(\frac{\partial \rho}{\partial z} - \gamma_\rho\right)\right]\right) = \nabla_h\left(\frac{\partial}{\partial z}\left[K_\rho\left(N^2 + \frac{g}{\rho_0}\gamma_\rho\right)\right]\right)$$

In the above definitions, ρ indicates the potential density, g is the gravitational acceleration, (u_g,v_g) and (u_a,v_a) represent the geostrophic and ageostrophic horizontal velocities, while turbulent terms are defined through classical non-local effective gradient, γ_ρ, and viscosity/diffusivity, K_x.

Once the equation is solved for w, horizontal ageostrophic components can be estimated by integrating two expressions that are obtained during the analytical derivation of the omega equation:

$$u_a(z) = \int_{REF}^{z} \frac{\partial u_a}{\partial z} dz = \frac{1}{f^2} \int_{REF}^{z} \left(\frac{\partial}{\partial x}(N^2 w) - \frac{f}{\rho_0} \frac{\partial}{\partial z} \left(\frac{\partial \tau_{yz}}{\partial z} \right) - 2 \frac{g}{\rho_0} \left(\frac{\partial u_g}{\partial x} \frac{\partial \rho}{\partial x} + \frac{\partial v_g}{\partial x} \frac{\partial \rho}{\partial y} \right) - \frac{g}{\rho_0} \frac{\partial}{\partial x} \left(\frac{\partial \mathrm{F}_{\rho z}}{\partial z} \right) \right) dz$$

$$v_a(z) = \int_{REF}^{z} \frac{\partial v_a}{\partial z} dz = \frac{1}{f^2} \int_{REF}^{z} \left(\frac{\partial}{\partial y}(N^2 w) + \frac{f}{\rho_0} \frac{\partial}{\partial z} \left(\frac{\partial \tau_{xz}}{\partial z} \right) - 2 \frac{g}{\rho_0} \left(\frac{\partial u_g}{\partial y} \frac{\partial \rho}{\partial x} + \frac{\partial v_g}{\partial y} \frac{\partial \rho}{\partial y} \right) - \frac{g}{\rho_0} \frac{\partial}{\partial y} \left(\frac{\partial \mathrm{F}_{\rho z}}{\partial z} \right) \right) dz$$

where we have defined:

$$\tau_{xz} = \rho K_m \left(\frac{\partial u}{\partial z} \right)$$

$$\tau_{yz} = \rho K_m \left(\frac{\partial v}{\partial z} \right)$$

$$\mathrm{F}_{\rho z} = K_\rho \left(\frac{\partial \rho}{\partial z} - \gamma_\rho \right)$$

and assumed that the ageostrophic velocities can be neglected at the reference layer considered in the integral (here taken as the deepest level).

3.4.2 Diabatic forcing term estimation

The function used in CMEMS OMEGA3D product to estimate vertical mixing coefficients (adapted from [RD-23]) includes an amplification of classical KPP turbulent velocity scales to account for Langmuir cells mixing, a nonlocal momentum flux term and a parameterization of Stokes drift effects. In the CMEMS processing, forcing terms were computed from ARMOR3D potential density and geostrophic velocity fields and ERA-interim atmospheric re-analyses.

Here, we used the LSTM3D reconstruction of potential density and geostrophic velocity fields and ERA5 surface fluxes, as well as the empirical estimation of the horizontal Ekman velocity provided by Copernicus/Globcurrent product [RD-14]. We removed the nonlocal flux of momentum from the formulation of the upper layer mixing parameterization and further constrained the viscosity values not to exceed a consistent empirical estimate. This allowed to correct some dynamically inconsistent/inaccurate values retrieved by CMEMS algorithm within the surface layer (as shown by the assessment of pre-existing products preliminarily carried out within WOC).

Specifically, we assumed that the background Ekman velocity can be approximated through an analytical fit of the ageostrophic currents (provided at 0 m and 15 m by the empirical reconstruction) to a compressed Ekman spiral:

$$u_{Ekman}(z) = e^{\frac{z}{D_{amp}}} [\, u_0 \cos(z/D_{rot}) - \, v_0 \sin(z/D_{rot})]$$

$$v_{Ekman}(z) = e^{\frac{z}{D_{amp}}} [\, u_0 \sin(z/D_{rot}) + \, v_0 \cos(z/D_{rot})]$$

where (u_0, v_0) are the components of the empirical Ekman current at 0 m, and D_{amp} *and* D_{rot} are the Ekman depth estimates obtained from the amplitude decay and vector rotation between 0 m and 15 m depth, respectively (see also [RD-24]).

The maximum viscosity, K_{max}, allowed within the Ekman layer was then derived from the local Ekman amplitude decay scale:

$$K_{max} = \frac{f D_{amp}^2}{2}$$

imposing an analytical profile within the boundary layer, chosen as in [RD-25]:

$$K_m(z) = K_{max}\left[1 + \tanh\left(\frac{z - D_{amp}}{\delta}\right)\right]$$

where δ represents the thickness of the transition layer (here set to 40 m, as in [RD-25]).

3.4.3 Numerical solution

All equations used for the NATL3D retrieval are solved numerically as in [RD-10-11]. In the interior domain, the Omega equation is re-written substituting derivatives with central finite differences, considering a non-staggered grid. Vertical derivatives are computed considering a variable grid spacing, increasing as the square of depth [RD-26], and adopting a second-order accuracy finite difference scheme [RD-27].

Dirichelet conditions are imposed at the surface and topographical boundaries and Neumann conditions at the bottom and lateral boundaries. These latter are imposed through forward/backward finite schemes and make the solution not suited to model current topography interactions along the coasts. Resulting linear system in *w* is solved through a matrix inversion. To make the inversion computationally feasible, the original grid is split in tiles (with a horizontal dimension of 75 grid points), which overlap by one third. The inversion is carried out sequentially on these sub-domains, imposing the vertical velocity values that resulted from the previous step as lateral boundary conditions to the subsequent calculations. The algorithm used for the matrix inversion is Loose Generalized Minimum Residual, a.k.a., LGMRES [RD-28] with incomplete lower-upper (LU) preconditioning (as implemented in python sparse linear algebra package scipy.sparse.linalg [RD-29], imposing a tolerance for convergence of 10^{-7}). A simple trapezoidal rule is used to compute the integral providing the ageostrophic horizontal velocities (see section 3.4.1).

3.5 Product validation

3.5.1 Data & Methods

3.5.1.1 NATL3D Tracer data

The assessment of the LSTM3D reconstruction technique was performed by estimating temperature and salinity root mean squared differences (RMSD) with respect to randomly selected independent test in situ profile data. To this aim, 15% of the 35,344 Argo profiles collected in the area (i.e. 5125 profiles) were excluded from the network training.

3.5.1.2 NATL3D Velocity data

Due to their small magnitude (around 1-100 md^{-1}), vertical velocities cannot be directly measured in the open ocean. As such, NATL3D vertical velocities cannot be directly validated (no reference dataset is available). However, the algorithm used to retrieve NATL3D horizontal velocities requires the vertical velocity in input, and improvements in quasi-geostrophic

horizontal components with respect to standard geostrophic velocities would necessarily imply that vertical velocity is reliable. The NATL3D algorithm validation has thus been based on the comparison between retrieved daily horizontal velocities over one test year (2018) and concurrent daily averaged Surface Velocity Program (SVP) data [RD-30] from NOAA Global Drifter Program (covering the period 1993-2018 and freely available at https://www.aoml.noaa.gov/phod/gdp/). SVP drifters are drogued with a 7 m long holey-sock centered at 15 m depth and their velocity estimates are representative of currents at 15 m depth [RD-31]. Conversely, undrogued SVP data (i.e. data coming from SVP instruments that lost their drogue by accident) can be used to validate currents in the surface layer. Part of the differences will have to be attributed anyway to representativeness errors, i.e. to the distinct information provided by velocities estimated from drifter displacements along a Lagrangian trajectory (including averaging and filtering over space/ time) and by Eulerian velocities retrieved in NATL3D.

3.5.2 Results

3.3.1.1 Assessment of the tracer reconstruction

The LSTM3D reconstruction has been compared to the multivariate Empirical Orthogonal Function reconstruction (mEOF-r), to a two-layer feedforward network and to WOA13 climatology. Full details on the validation exercise are given in [RD-14] and we only recall here that the stacked LSTM (35–35) significantly improves the reconstruction all along the water column, with the LSTM temperature RMSD never exceeding 1 °C, and attaining below 0.75 °C already at a 200 m depth, and the LSTM salinity reducing the RMSD to almost one half with respect to the climatology, further improving in the 200–800 m layer (Figure 9).

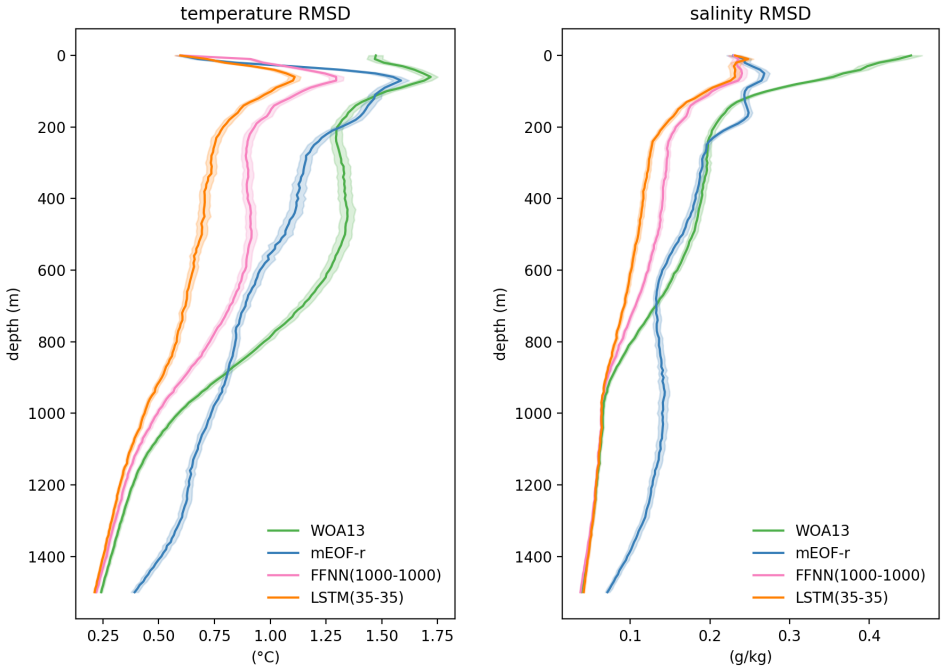

Figure 9. RMSD between temperature (a) and salinity (b) climatological and reconstructed profiles, estimated from independent test data. RMSD confidence intervals (one , displayed here as shadowed areas) have been estimated through a Monte Carlo approach—i.e., as the standard deviation of the statistics computed from 1000 resampling with replacement.

3.3.1.1 Assessment of the velocity retrieval

The 3D current velocity accuracy depends on both the input data and on the theoretical
The 3D current velocity accuracy depends on both the input data and on the theoretical
limits of the diagnostic models and parameterizations used. As such, the NATL3D product is basically suited for open ocean applications only, and coastal data must be looked at with caution. Retrieved vertical velocity patterns are very similar to these found in CMEMS OMEGA3D product, but typical velocity values are generally more intense, (still attaining around O(10-100 m/day)), reflecting the NATL3D ability to reveal the stronger spatial gradients (figure 10).

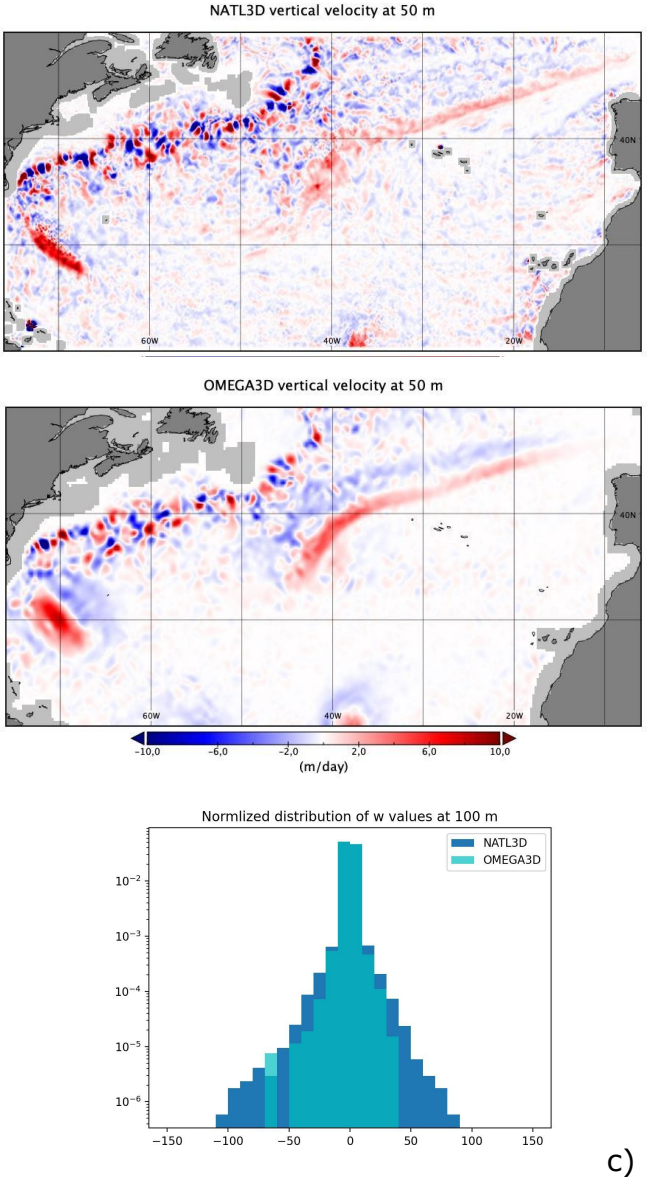

Figure 10. NATL3D vertical velocity at 50 m depth on 2018-09-12 (a); OMEGA3D vertical velocity at 50 m depth on 2018-09-12 (b); Histogram of the differences between NATL3D and OMEGA3D vertical velocities at 100 m depth on 2018-09-12 .

As the horizontal ageostrophic components are estimated only once the vertical velocities are known, total horizontal and geostrophic components (obtained as the difference between the total and the ageostrophic components) have been compared with independent estimates of the ocean currents retrieved from drifting buoys. The Root Mean Square Differences (RMSD) of the total horizontal velocities vs SVP velocities are 0.119±0.001 m/s at 0 m and 0.113±0.001 m/s at 15 m depth, respectively, slightly improving with respect to simple geostrophic estimates, which show a RMSD of 0.145±0.001 m/s at 0 m and 0.117±0.001 m/s at 15 m depth. The number of matchups used to compute the statistics is 311735 for undrogued SVP velocities (at 0) m and 243742 for drogued SVP data (at 15 m). The largest RMSD values are clearly found along the Gulf Stream, while improvements with respect to geostrophic velocities are found all over the domain, with only a few negative values in limited areas along the Gulf Stream (figure 11).

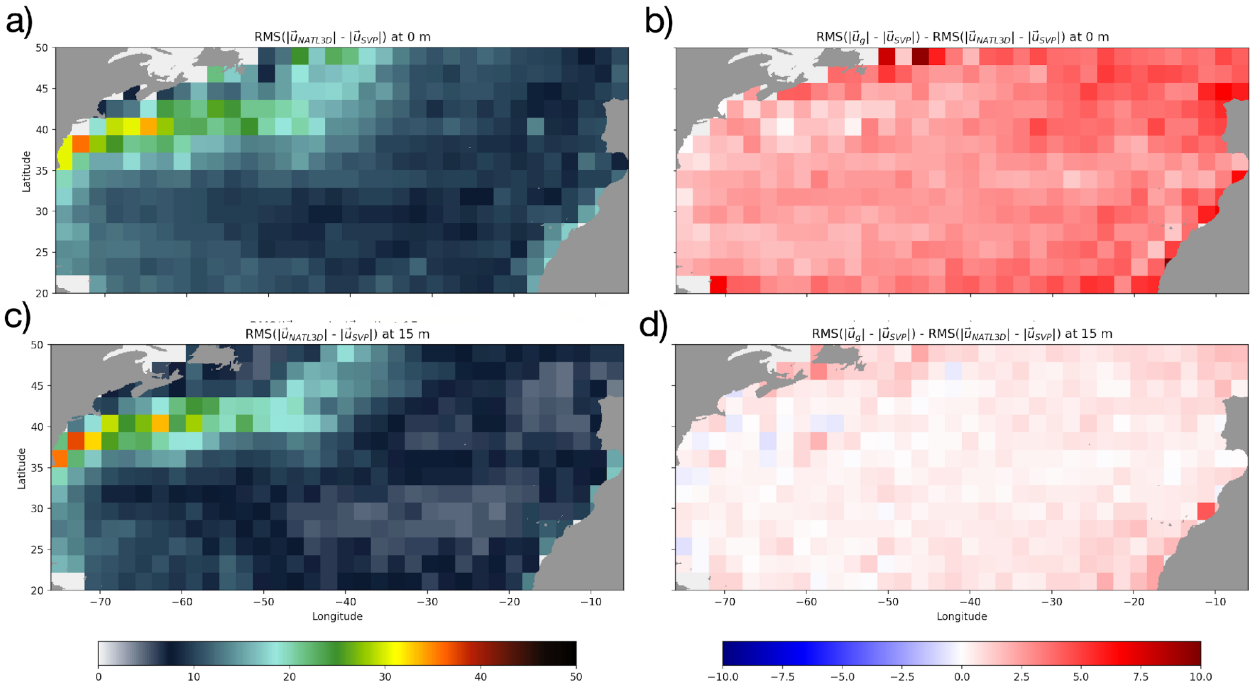

Figure 11. RMSD between NATL3D total horizontal currents and co-located SVP velocities at 0 m (a) and 15 m (b) depth, estimated in 2°x2° bins. Differences between RMSD of NATLD3D total horizontal velocities and geostrophic velocities vs drifters in 2°x2° bins, at 0 m (c) and 15 m (d) depth, respectively. Positive values indicate an improvement with respect to geostrophy.

The analysis has been extended to assess the temporal variations of the validation metrics over the entire time series. NATL3D horizontal current MBE and RMSD have thus been estimated on yearly basis and compared to geostrophic estimates, showing consistent improvements over time both at 0 m (figure 12) and 15 m (figure 13).

As a final additional test, we estimated the correction to the currents driven by the adiabatic component alone, thus looking separately at the impact of the flow deformation

(kinematic term in the Omega equation). The assessment consistently shows a positive, though relatively small, improvement along the Gulf Stream (figure 14).

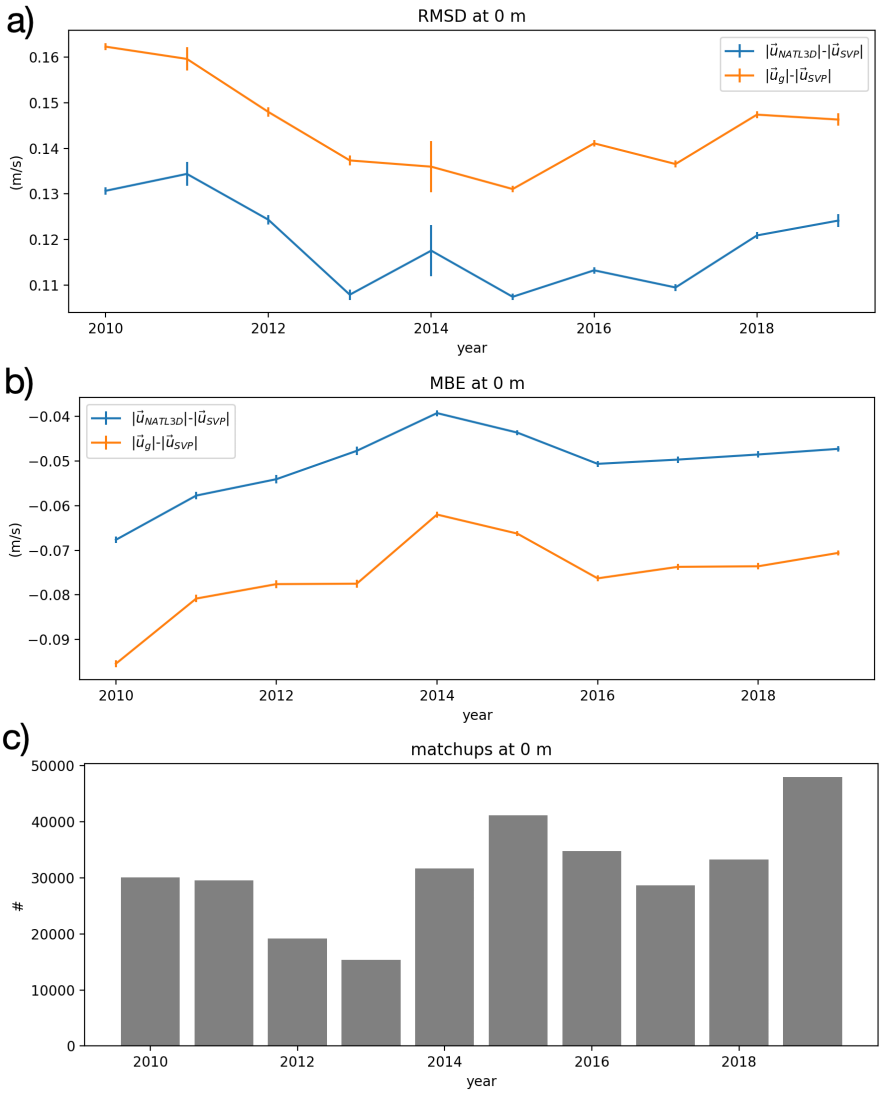

Figure 12. Timeseries of yearly RMSD (a) and MBE (b) between NATL3D total horizontal currents and co-located SVP velocities at 0 m depth. (c) Number of yearly matchup points.

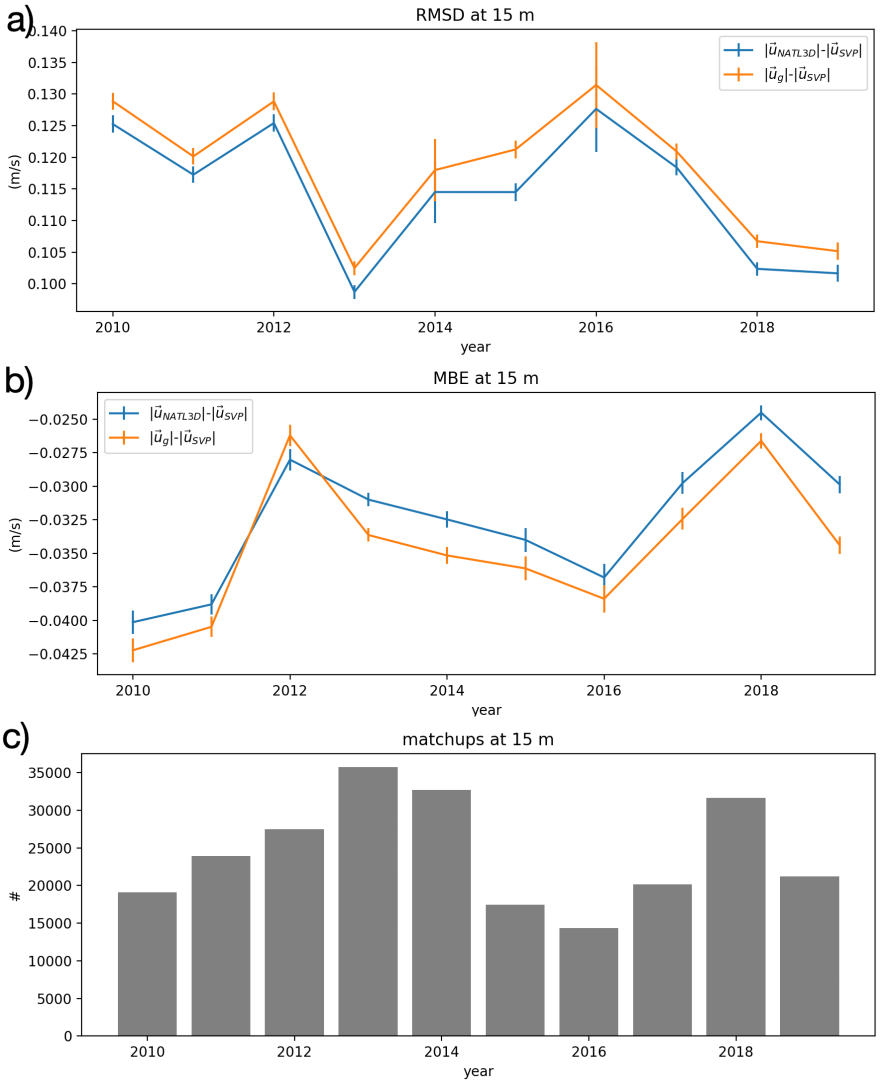

Figure 13. Timeseries of yearly RMSD (a) and MBE (b) between NATL3D total horizontal currents and co-located SVP velocities at 15 m depth. (c) Number of yearly matchup points.

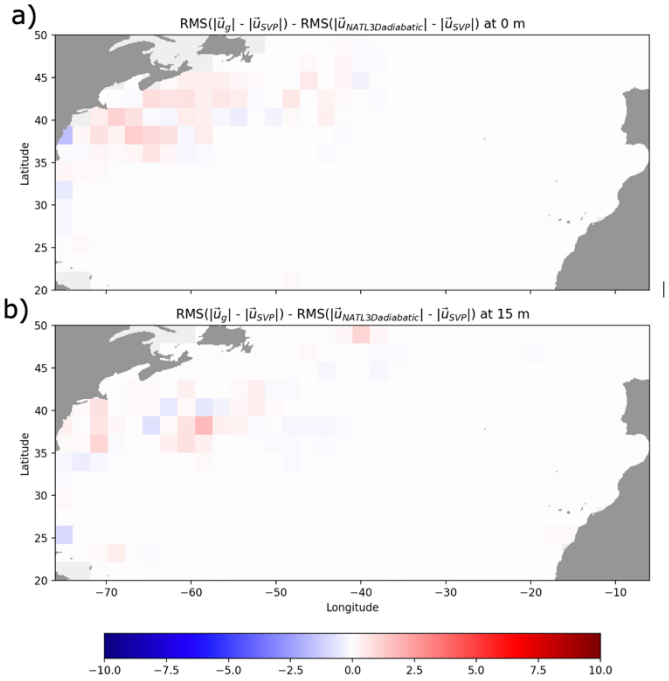

Figure 14. Differences between RMSD of NATLD3D adiabatic components of the horizontal velocities and geostrophic velocities vs drifters in 2°x2° bins, at 0 m (c) and 15 m (d) depth, respectively. Positive values indicate an improvement with respect to geostrophy.