High-resolution downscaling of CMEMS oceanographic reanalysis in the area of the Tuscany Archipelago (Italy)

Michele Bendoni[1] · Maria Fattorini[1,2] · Stefano Taddei[1] · Carlo Brandini[1,2]

Abstract

A native nested configuration of the ROMS model is implemented on the marine area between the Ligurian and Tyrrhenian basins, which includes the Tuscany Archipelago. Initial and boundary conditions are provided by the CMEMS Mediterranean Sea Physical Reanalysis product (1/16°), feeding the parent ROMS model (BLUE, 1/72°), in which a high-resolution grid is nested (PURPLE, 1/216°). Atmospheric forcing comes from a downscaled version of ERA5 reanalysis. Temperature and salinity profiles from gliders and floats, and HF-radar-derived surface currents, are compared to model outputs within the high-resolution area for the whole year 2017. Results show the downscaling procedure is able to reduce model errors for temperature profiles, whereas errors in salinity profiles remain comparable. However, the downscaled model cannot recover large errors inherited from the parent one. The mean bias largest values found in both temperature and salinity profiles may be explained by a model underestimation of the depth of stable stratification limit with respect to field data. Errors in surface currents are reduced for the downscaled dynamics and appear to be uncorrelated to the original CMEMS product, being surface dynamics less affected by initial condition than by atmospheric forcing. A simple scalar metric, to quantify the error in the surface current vector fields from observations and models, is proposed. The novel metric allows to better quantify the improvement gained by the downscaling procedure.

Keywords Downscaling · CMEMS · High-resolution · Vector field metric

1 Introduction

The need for high-resolution oceanographic data, and in particular for reanalysis products, is rising due to the increasing importance gained by studies aimed at understanding the effect of climatic trends along the coast and the analysis of local oceanographic conditions both to manage productive activities, such as aquaculture and fisheries, and to support navigation. Furthermore, the production of hazard and risk maps is crucial for the evaluation of impacts caused by different environmental threats, including marine litter and various contaminants. All these aspects strongly affect social and economic challenges worldwide, both now and in the future (Visbeck 2018).

In coastal areas, characterized by complex coastline and variable bathymetry, high-resolution numerical models become important to better grab the interaction of the sea dynamics with the solid boundaries (e.g. when the circulation is constrained by topography), and a higher spatial and temporal variability of atmospheric forcing. Downscaling Ocean General Circulation Models (OGCMs), to regional and sub-regional scales, is the principal mechanism to describe ocean dynamics in more detail for specific locations. Such a procedure consists of a "parent" model used to transfer information toward a "child" model, having a higher resolution grid. The exchange of information can be sequential (offline) (Mason et al. 2010), or simultaneous, both from the parent to the child (one-way) and mutual (two-way) (Blayo and Debreu 2006).

In ocean sciences, several studies tried to understand whether the downscaling procedure is worthy anyway, and identify weaknesses and strengths with application in different areas of the world: Lorente et al. (2019) in the

---

✉ Michele Bendoni
   bendoni@lamma.toscana.it

[1] LaMMA Consortium, Via Madonna del Piano 10, Sesto Fiorentino 50019, Italy

[2] CNR-IBE, Institute of BioEconomy, Via Madonna del Piano 10, Sesto Fiorentino 50019, Italy

Ⓐ Springer

Iberia-Biscay-Ireland system, Bricheno et al. (2014) in Great Britain, Onken et al. (2020) in the Baltic Sea, Katavouta and Thompson (2016) in the Gulf of Maine. The improvement in model capability to reproduce real physical patterns is indeed counterbalanced by the introduction of a higher variability, which may not be in phase with observations.

Concerning the Mediterranean Sea, the direct downscaling of the Copernicus - Marine Environment Monitoring Service (CMEMS) products has been carried out with horizontal resolution ranging from 3.0 to 1.5 km (Olita et al. 2013; Onken 2017; Aguiar et al. 2020), whereas the use of multi-level nested configuration allowed to reach resolution up to $O(100)$ m (Sorgente et al. 2016; Trotta et al. 2017), and even to $O(10)$ m by means of of unstructured grids (Cucco et al. 2012; Trotta et al. 2016). More specifically, the area of the Tuscany Archipelago has been analysed by Sorgente et al. (2016) to evaluate the effect of assimilating CTD data measurements into a model aimed at forecasting of drifters trajectories, in oil spill emergencies, and by Trotta et al. (2016) to test the skill of a relocatable multiple-level nested model platform. In both cases, the resolution of the grids covering the whole Archipelago was at most around 1200 m.

The Tuscany Archipelago is a complex shelf area characterized by the presence of several islands (from north Gorgona, Capraia, Elba, Pianosa, Montecristo, Giglio, and Giannutri) which make the Corsica Channel the main pathway connecting Ligurian and Tyrrhenian basins. The area is bounded westward by Eastern Corsica Current, which is characterized by a strong seasonal variability, and merges to the Western Corsica Current, farther north, to form the so-called Northern Mediterranean Current (Astraldi and Gasparini 1992; Millot 1999). At the southern edge of the shelf, the intermediate layer water coming from the south veers westward following the bottom topography (Millot 1999). The mean circulation allows the relatively warmer waters from the Northern Tyrrenian Sea to enter the colder waters of the Ligurian Sea, affecting not only the distribution of marine species (Astraldi et al. 1995), but also the dispersion of pollutants, such as floating plastic debris, which tend to accumulate in the area of recurrent anticyclonic circulation corresponding to the so-called Capraia Gyre (Onken et al. 2005; Fossi et al. 2017; Iacono and Napolitano 2020).

The objective of the present work is to test the skill and reliability of a model configuration which covers the whole shelf area separating the Ligurian and Tyrrhenian basin, at sufficient resolution to analyse sub-mesoscale processes $O(1-10)$ km, at least for the upper portion of their spatial range. The model configuration is built by downscaling the CMEMS Mediterranean Sea Physical Reanalysis product via a one-level, one-way nest in the ROMS model, implemented on the area between the Ligurian and the Tyrrhenian Seas.

Atmospheric forcing is provided by a high-resolution downscaling of the ERA5 reanalysis. Both the CMEMS reanalysis product and the downscaled model outputs are systematically tested against in situ temperature and salinity profiles, and surface currents derived from HF-radars, to evaluate the improvement in model skill gained by the downscaling procedure.

The second section describes the model configuration, and the third one is dedicated to the available field observations employed to test model skills, which are assessed in the fourth section. The fifth section is dedicated to the discussion of the results, analysing the source of errors, and proposing a simple metric to compare vector fields and finally conclusions are drawn.

## 2 Model configuration and validation

2.1 One-way nesting in ROMS

The Regional Oceanic Modelling System (ROMS) is a primitive equation, terrain-following coordinates, hydrostatic model (Shchepetkin and McWilliams 2003, 2005), which is widely used for several oceanographic applications (López et al. 2020). Both one-way and two-way native nesting are available in ROMS (Warner et al. 2010). However, the one-way configuration is considered sufficiently accurate to allow the downscaling of prognostic model variables from low to high resolution, saving a significant amount of computational effort, and is therefore employed for the present application.

The parent model, called BLUE, extends from -0.4 to $15.9°$ longitude and from 39.6 to $44.5°$ latitude, covering the North-Western Mediterranean Sea (NWM), at 1/72° resolution (roughly 1.2 km), and 30 sigma-layers. Nested model is called PURPLE and runs simultaneously to the parent one. It covers the area of the Tuscany Archipelago and part of the Ligurian Sea, from 8.96 to 12.29° longitude and from 41.95 to 44.41° latitude, at 1/216° (roughly 0.45 km), with a child to parent grid ratio equal to 3, and 30 sigma-layers (Fig. 1a and c). The sigma coordinates distribution along the vertical is determined by setting $v_{tr} = 2$, $v_{str} = 4$, $\theta_s = 5$ and $\theta_b = 0.4$ for both BLUE and PURPLE models. Horizontal and vertical advections are handled with third-order upstream and fourth-order centered schemes, respectively. Harmonic horizontal diffusion of momentum along s-surfaces, and of tracers along geopotential surfaces, is adopted: turbulent horizontal eddy viscosity coefficients v are set equal to 3.0 and 1.0 m²/s, whereas the eddy diffusivity coefficients κ_T for both temperature T and salinity S are set equal to 0.3 and 0.1 m²/s for the BLUE and PURPLE models, respectively.

The Generic Length Scale turbulence closure scheme (Umlauf and Burchard 2003) with the $k - \epsilon$ parameters is

Fig. 1 **a** View of the extension of BLUE and PURPLE domains. Flow fields at the resolution of NEMO model (blue arrows for a portion of model domain) (**b**), HF-radar (black arrows) (**c**), PURPLE model (red arrows for a portion of model domain) (**d**). Black and red squares in subfigure **b** represent the enlarged area in subfigures **c** and **d**, respectively. To facilitate the reading, flow field versors have been reported
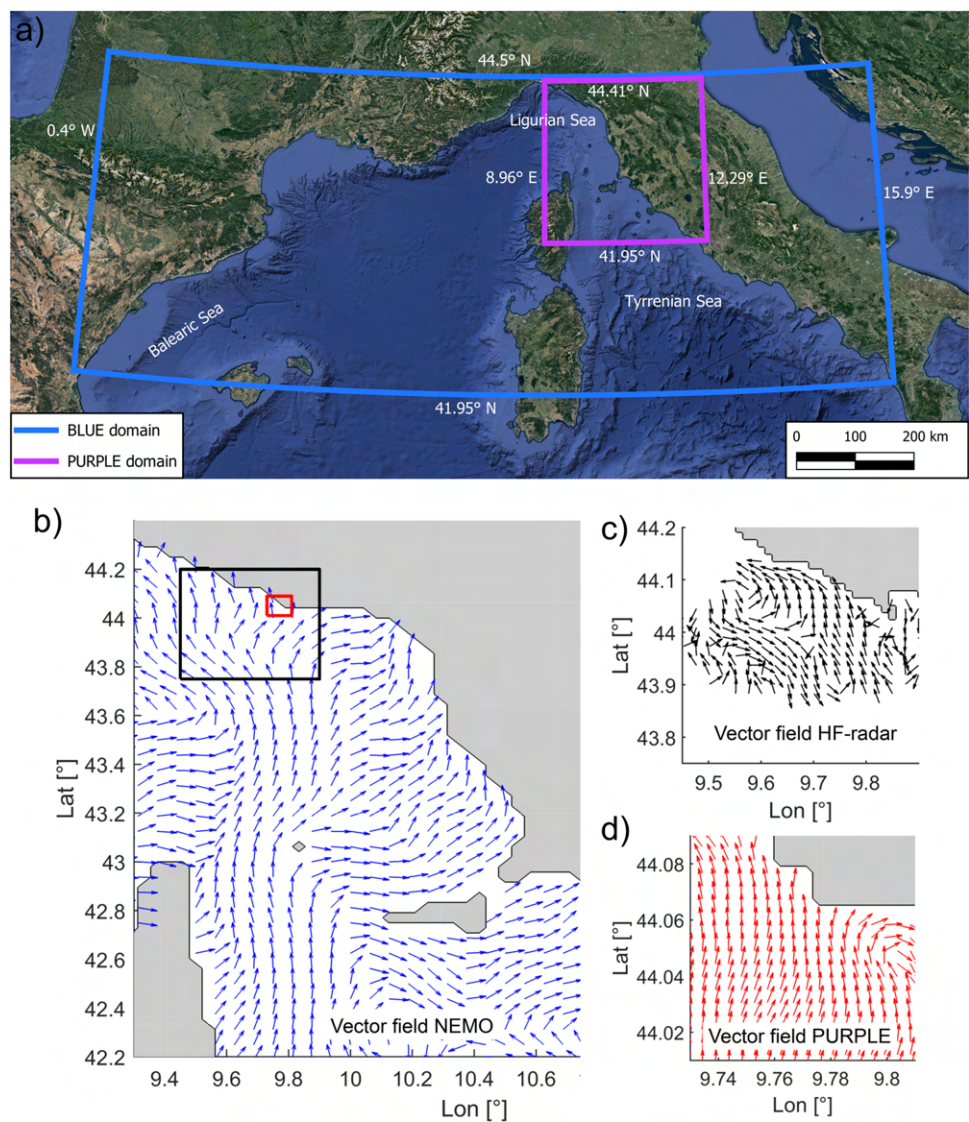
employed, whereas the sea-atmosphere interaction is modelled through the Fairall et al. (1996) parametrization.

Bathymetry has been derived from bathymetric data provided by the EMODnet portal (http://www.emodnet-bathy metry-eu). Raw data have been interpolated at the grid points and smoothed through a sequence of Laplacian filtering with a target rx_0 parameter equal to 0.25 (Haidvogel et al. 2000), using the approach proposed by Sikirić et al. (2009). The model run with $rx_0 = 0.25$ and $rx_1 = 5.2$ (Haney 1991) for the BLUE grid, and $rx_0 = 0.12$ and $rx_1 = 2.4$ for the PURPLE grid. The contact file that allows the grids to interact is created by means of the Matlab package provided by the ROMS developers (https://www.myroms.org/wiki/Matlab_Scripts).

The downscaling covers the whole year 2017 on a weekly basis. Each run lasted 10 days, 3 of which are considered to be for spin-up and are removed, similarly to Trotta et al.

(2016). Such a short spin-up period is chosen also with a view to implement an operational version of the downscaling procedure, for which a longer spin-up would be unfeasible. Baroclinic time step is equal to 45 and 15 s for parent and child models, respectively, and 30 barotropic time steps are set between each baroclinic one.

2.2 Initial and boundary conditions and atmospheric forcing

Initial and boundary conditions for the prognostic variables temperature T, salinity S, sea surface height η, and horizontal velocity u and v are retrieved from the MED-SEA_REANALYSIS_PHY_006_004 CMEMS reanalysis product (Simoncelli et al. 2019), on a daily-averaged basis, and 1/16° (roughly 6 km; see Fig. 1b). Vertical resolution varies linearly from 1 to 3 m at the surface to roughly 200 m

at the bottom. The dataset is produced through the NEMO v3.4 model (Madec 2014), atmospheric forcing from ERA-Interim reanalysis (Dee et al. 2011), and a 3D-Var assimilation scheme for the assimilation of in situ temperature and salinity profiles, and sea level anomaly (SLA) along track data (Dobricic and Pinardi 2008).

Daily-averaged boundary conditions are linearly interpolated by ROMS at internal time steps. A radiating-nudging (Marchesiello et al. 2001) condition is set for T, S, u, and v at the southern boundary of the BLUE domain with inflow and outflow nudging timescales equal to 1 day and 5 days, respectively. At the same boundary, a Flather condition (Flather 1976) is applied to the barotropic velocity components, a Chapman condition (Chapman 1985) to the free surface η, and a zero gradient condition to the total kinetic energy (TKE). Boundary conditions for the PURPLE model are provided every timestep by the BLUE model. Being a one-way configuration, no information flows from the nested grid to the parent one.

Atmospheric forcing is provided by a dynamical downscaling of the ERA5 dataset built upon a nested configuration of the BOLAM and MOLOCH atmosperic models (Buzzi et al. 2014). The former, at 1 h and 7-km resolution, covering the Med-CORDEX domain, the latter, at 1 h and 2.5-km resolution covering part of the North-Western Mediterranean Sea. A detailed description of the implementation of such reanalysis downscaling and the validation of the results is reported in Vannucchi et al. (2021). Outputs from BOLAM and MOLOCH models are employed to force the BLUE and PURPLE domains, respectively. To determine ROMS fluxes of heat, momentum, and freshwater at the sea-air interface, the following physical variables are extracted from the dataset: horizontal velocities at 10 m, temperature and relative humidity at 2 m, atmospheric pressure at mean sea level, total cloud cover and downward short-wave radiation flux.

River discharge is added to the downscaled model by considering only those major rivers included into NEMO, Rhone, and Ebro. Monthly averaged values of river discharge are obtained from the database of the Global Runoff Data Centre, 56068 Koblenz, Germany (GRDC). Salinity values are set equal to 25 PSU for Rhone and 30 PSU for Ebro (Simoncelli et al. 2019); temperature is set equal to sea surface temperature at the mouth.

2.3 Validation against field observations

To assess the capability of the high-resolution PURPLE model to improve the performance of the NEMO model in reproducing the actual sea dynamics, we compare the skills of the two models against two typologies of observations: in situ profiles

of temperature T and salinity S, and surface current components u and v, derived from HF-radar. To measure model skills, we use the following standard statistics: root mean squared error (*RMSE*), centred root mean squared error (*cRMSE*), correlation coefficient (ρ), and mean bias (*MB*). A whole year (2017) of data is employed for the analysis. In the following, subscripts N and P correspond to NEMO and PURPLE models, respectively.

3 Field observations

3.1 In situ temperature and salinity profiles

In situ observations are retrieved from the CMEMS product INSITU_MED_NRT_OBSERVATIONS_013_035, for each month of the year 2017, coming from different instruments and platforms. Each of the downloaded data file contains observations from a specific platform for a specific month, clipped on the spatial domain of the PURPLE model. Each observation file is coded as XX_code, where XX represents the data type, if mentioned, and the code is specific for each platform. Herein, we use data from the following: profiling floats (PF), oceanographic CTD profiles (CT), and gliders (GL). A detailed description of the file-naming convention can be found in Copernicus Marine Team (2020).

Data availability is not uniform during the year; indeed, the months of September, October, and November show a much larger number of vertical profiles with respect to the other months of the year. In particular, in summer months, no observations are available within the analysed area. Data with a quality flag different from "good data" are set to NaN (not a number). As a consequence, not all available profiles are employed for the comparison with model outputs: those having less than 75% of valid data within an upper layer 200 m deep are removed from the analysis.

3.2 Surface currents from HF-radar

The surface velocity components come from the INSITU_GLO_UV_NRT_OBSERVATIONS_013_048 product, provided by CMEMS (Copernicus Marine In Situ TAC 2020). Specifically, we use near surface zonal (u_o) and meridional (v_o) velocities derived from high frequency (HF) radar, part of the European HF-radar network (Mader et al. 2017; Corgnati et al. 2018).

The covered area is located in the Ligurian Sea, adjacent to the Cinque Terre coast (Liguria, Italy), and extends approximately 30 km cross-shore and 30 km alongshore (Fig. 1c). Temporal resolution is 1 hour, and spatial resolution is 2 km in both zonal and meridional directions. For the comparison with modelled values, only those data with quality flag corresponding to "good data" are employed.

4 Results

4.1 Models performance in reproducing temperature and salinity profiles

The first comparison is made for temperature profiles up to a depth of 400 m, independently of the position. For each month, the profiles coming from different platforms are interpolated at common depth values: from 0 to 200 m with 2 m step, and from 200 to 400 m with 20 m step. Then, the bias between observations and model output $\Delta T(z) = T_O(z) - T_M(z)$ is calculated, where $T_O(z)$ represents the measured temperature at a specific time, location, and depth, and $T_M(z)$ the corresponding modelled temperature,

linearly interpolated at the same spatial and temporal coordinates. $\Delta T(z)$ is determined for both NEMO and PURPLE outputs and averaged over multiple profiles, giving the mean biases $MB_{T,N}(z)$ and $MB_{T,P}(z)$ as a function of depth. The calculation of the standard deviation $\sigma_{\Delta T}(z)$ of $\Delta T(z)$ corresponds to the $cRMSE_T(z)$ between observations and model output.

Values of $MB_T(z) \pm \sigma_{\Delta T}(z)$ for the 3 months with the largest amount of observations (September, October, and November), and the number of observation values available for different depth layers (grouped by 25-m intervals), are reported in Fig. 2 for both PURPLE and NEMO models. For each month, the data belonging to different observation platforms are grouped together. The number of observations collected by each observation platforms during the year
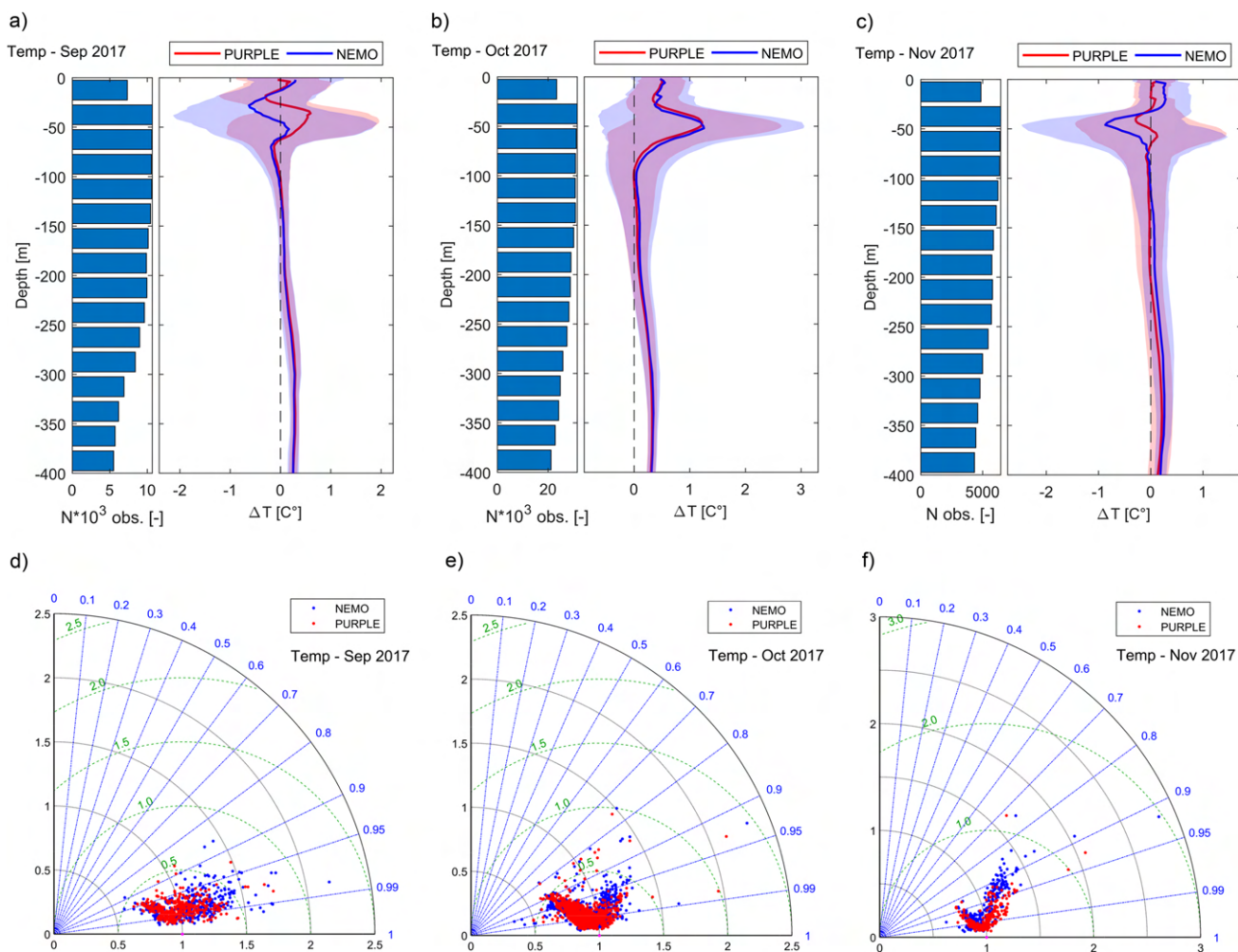

Fig. 2 Mean temperature bias ± standard deviation for PURPLE (red solid line and shaded red area) and NEMO (blue solid line and shaded blue area) models, as a function of depth for different platforms for the months of September, October, and November. The left column of each subfigure reports the distribution of data points with depth, subfigures a, b, and c. Normalized Taylor diagrams of tem-

perature profiles for different platforms for the months of September, October, and November for PURPLE (red dots) and NEMO (blue dots) models. Normalized standard deviation is reported in horizontal and vertical axes (black lines), correlation is read as an angular quantity (blue radials from the origin), and normalized cRMSE is represented by arc of circles (green dotted line), subfigures d, e, and f

are reported in Table 1. Not all months are present, since April, June, July, August, and December have no available measurements.

For all the subfigures (a, b, c), the mean biases $MB_{T,N}(z)$ and $MB_{T,P}(z)$ are larger in the upper layer from 0 to −75 m depth, whereas the remaining portion of the profile tends to show a positive bias (the model underestimates temperature) aroud 0.2–0.3 °C. A similar pattern is found for the variability of the deviation between model and observation, with higher values in the upper layer, with $\sigma_{\Delta T,N}(z)$ and $\sigma_{\Delta T,P}(z)$ ranging from −2 to +3 °C, and lower values below the thermocline given by $\sigma_{\Delta T,N}(z)$ and $\sigma_{\Delta T,P}(z)$ in the order of $O(10^{-1})$ °C.

More specifically, Fig. 2a (September) shows that the PURPLE model is able to reduce the shaded area between −25 and −75 m depth, and therefore the $cRMSE_T$, even if mean bias is improved or worsen based on the depth. The October data (Fig. 2b) show no substantial change in the $MB_T(z)$ between the two models. However, a small reduction of the $cRMSE_T$ for the PURPLE model with respect to NEMO is present for the layer between 0 and −100 m depth. Figure 2c (November) reports a reduction both in the

$MB_{T,P}(z)$ with respect to $MB_{T,N}(z)$, and in the $\sigma_{\Delta T,P}(z)$ with respect to $\sigma_{\Delta T,N}(z)$ for the depths ranging from 0 to −75 m.

In addition, we carried out the analysis considering the behaviour of the single profile, by plotting normalized Taylor diagrams (Taylor 2001) for the three months of September, October, and November (Fig. 2d, e, f). In such a case the different statistics calculated are not referred to specific depths, but to each profile, taken individually. Profiles for September and October (Fig. 2d and e) show the most part of normalized $cRMSE$ lower than 0.5 time the standard deviation of observations and correlation higher than 0.9, for both NEMO and PURPLE models, whereas November data (Fig. 2f) are characterized by normalized $cRMSE$ values up to 1. The improvement of model output due to the downscaling procedure is more clearly detectable in Fig. 2f, where red dots (PURPLE) tend to get closer to the observation than the blue dots (NEMO).

To better understand whether the downscaling procedure leads to an improvement of the solution concerning the T profiles, an estimate of how much the PURPLE root mean square error ($RMSE_{T,P}$) is lower than that of NEMO ($RMSE_{T,N}$), is calculated for each observation file and each

Table 1 Summary of the total number of analysed temperature profiles per month and provenance

Month	Obs. file	Profiles	ER_T [%]	cER_T [%]	$\overline{RMSE}_T$ [°C]		$\overline{cRMSE}_T$ [°C]	
					P	N	P	N
January	PF_6901835	6	2.1	21.5	0.22	0.20	0.17	0.14
	PF_6903199	1	49	57.8	0.10	0.07	0.1	0.06
	XX_IBEX	12	6.9	19	0.30	0.28	0.22	0.18
February	PF_6901835	6	−43.5	5.5	0.25	0.40	0.14	0.12
	PF_6903199	4	11.1	−10.7	0.29	0.24	0.14	0.16
March	PF_6901835	6	−17.4	−17.8	0.19	0.22	0.08	0.1
	PF_6902692	3	25.7	−10.8	0.25	0.20	0.13	0.13
	PF_6903199	6	9.7	−23.1	0.20	0.17	0.09	0.12
	XX_IBIL	10	−15.6	−6.8	0.29	0.34	0.17	0.17
May	PF_FKJB	1	1.5	0.3	0.47	0.46	0.41	0.41
	PF_6901835	5	15.6	13.3	0.25	0.22	0.2	0.17
	PF_6902692	7	7.3	3.2	0.27	0.26	0.21	0.21
	PF_6903202	1	−0.2	−8.1	0.27	0.27	0.1	0.11
	XX_ICGK	12	0.3	0.3	0.44	0.45	0.27	0.28
September	CT_FNCM	8	−74.9	−53.6	0.63	1.50	0.56	1.22
	GL_6801661	164	−21.3	−14.7	0.49	0.61	0.46	0.62
	GL_6801663	251	−16.9	−12.5	0.65	0.76	0.58	0.65
October	GL_6801661	621	−32.3	−22.7	0.52	0.75	0.45	0.57
	GL_6801663	639	−0.5	−0.2	0.75	0.79	0.63	0.63
November	GL_6801661	180	−35.6	−37.6	0.58	0.81	0.47	0.67
	GL_6801663	78	−17.1	−29.1	0.43	0.41	0.28	0.38

ER_T and cER_T measure the reduction in root mean squared error and centred root mean squared error obtained with the downscaling procedure. Averaged root mean squared error $\overline{RMSE}_T$ and centred root mean squared error $\overline{cRMSE}_T$ for both PURPLE (P) and NEMO (N) models are also reported

month, and expressed as an error reduction ER_T. More specifically, ER_T is determined as follows:

$$ER_T = \frac{100}{N_p} \sum \frac{RMSE_{T,P} - RMSE_{T,N}}{\frac{1}{2}(RMSE_{T,P} + RMSE_{T,N})} \qquad (1)$$

where the summation extends to the number of analysed profiles (N_p) for each observation file. In case $ER_T < 0$, we have a general reduction of the temperature $RMSE$ for the PURPLE model with respect to NEMO, the opposite for $ER_T > 0$. In a similar way, cER_T is determined by considering the centred root mean squared error ($cRMSE$).

Such statistics, together with the number of analysed profiles, and the average value of the $RMSE_T$ and $cRMSE_T$ for each observation platform, are reported in Table 1. In the

first months of the year (January to May), characterized by a lower number of available profiles, a majority of positive ER_T values are found; however, this aspect does not hold true for the cER_T, having both positive and negative values. Differently, for the last months of the year (September to November), with a significantly major number of analysed profiles, ER_T and cER_T values are negative, meaning a reduction in model error is obtained by the downscaling procedure. Average $RMSE_T$ and $cRMSE_T$ values for PURPLE range between 0.1 and 0.7 °C, with a median value of 0.27 °C; those for NEMO range between 0.06 and 1.5 °C, with a median value equal to 0.28 °C.

To analyse the skill of the models in reproducing the salinity S profile patterns, we used the same approach as for T profiles. Figure 3 (a, b, c) reports the mean salinity bias
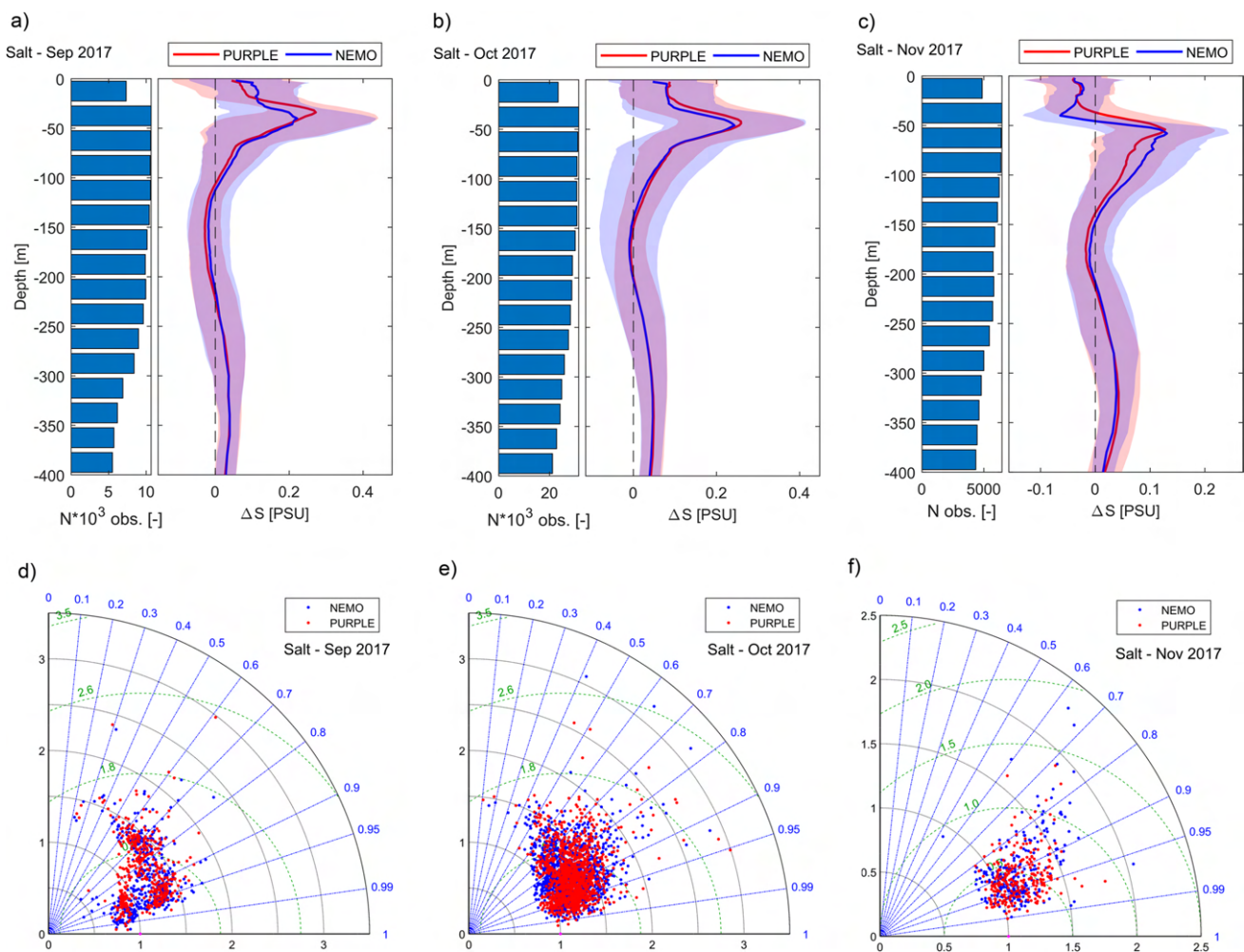

Fig. 3 Mean salinity bias ± standard deviation for PURPLE (red solid line and shaded red area) and NEMO (blue solid line and shaded blue area) models, as a function of depth for different platforms for the months of September, October, and November. The left column of each subfigure reports the distribution of data points with depth, subfigures (**a**, **b**, **c**). Normalized Taylor diagrams of salinity profiles for different platforms for the months of September, October, and November for PURPLE (red dots) and NEMO (blue dots) models. Normalized standard deviation is reported in horizontal and vertical axes (black lines), correlation is read as an angular quantity (blue radials from the origin), and normalized cRMSE is represented by arc of circles (green dotted line), subfigures (**d**, **e**, **f**)

$MB_S(z) \pm$ its standard deviation $\sigma_{\Delta S}(z)$ (i.e. the $cRMSE(z)$), for those months (September, October, and November) having the largest amount of available observations (Table 2).

Principal feature standing out from Fig. 3 is the peak in $MB_{S,N}$ and $MB_{S,P}$, ranging from 0.1 to 0.2 PSU, at a depth between -25 and -100 m, meaning the models tend to underestimate the salinity content. An exception is present for the November dataset (Fig. 3c), where the uppermost layer (0 to -50 m) shows an opposite small bias smaller than 0.1 PSU. These profiles show comparable performances between NEMO and PURPLE models both for MB_S, with differences $O(10^{-2})$ PSU, and $\sigma_{\Delta S}(z)$. The values of $\sigma_{\Delta S}(z)$ are slightly lower for PURPLE for the October dataset (*b*), whereas for the other two analysed months (a and c) non-overlapping shaded areas tend to compensate each other.

Normalized Taylor diagrams for salinity profiles are reported in Fig. 3 (subfigures d, e, f). Salinity profiles appear to be less accurately described by both models. Most part of the data are comprised within 1.5 normalized $cRMSE$, and the correlation ranges from 0.4 to 0.99 (Fig. 3d and e). November data show a lower value for the $cRMSE$ and correlation higher than 0.75 (Fig. 3d). The examination of normalized Taylor diagram does not allow clearly assessing which model performs better since both red and blue dots seem to be similarly spread on the chart for the three datasets.

As for the analysis of temperature profiles, a summary of the statistics used to estimate the potential improvement gained by the downscaling procedure in reproducing salinity profiles, is reported in Table 2.

For salinity profiles, most part of the data for the first months of the year show positive ER_S and cER_S values, meaning the PURPLE model performs worse than NEMO. On the contrary, months from September to November are characterized by both positive and negative values of ER_S and cER_S, with a majority of negative ones. Averaged values of $RMSE_S$ and $cRMSE_S$ have ranges 0.03–0.15 and 0.04–0.23 with equal median values of 0.08 for both PURPLE and NEMO models, respectively.

To merge together the information from temperature and salinity, and to verify the presence of any significant difference in observed and modelled static stability patterns, we determine the Brunt-Väisälä frequency squared $N^2 = -g/\rho \cdot \partial\rho/\partial z$ (Cushman-Roisin and Beckers 2011). The same spatial and temporal interpolation procedure employed for the analysis of temperature and salinity profiles alone ($dz = 2$ m from 0 to -200 m and $dz = 20$ m from -200 to -400 m) is also adopted to calculate N^2, and a 5-point moving average is applied to the profiles derived from observations. Water density ρ is determined using the algorithm provided by Fofonoff and Millard Jr (1983). The values of N^2 derived from observations, PURPLE and NEMO models, for the platform GL_6801661 during the period 1 October–9 November, and for the platform GL_6801663 during the period 21 September–31 October, are reported in Fig. 4a and b, respectively. The first noticeable aspect is that the variability of the peak value of N^2 decreases moving from observation to PURPLE and NEMO models, almost disappearing for the latter. This is more easily detectable if we consider the bottom plot for

Table 2 Summary of the total number of analysed salinity profiles per month and provenance

Month	Obs. file	Profiles	ER_S [%]	cER_S [%]	$\overline{RMSE_S}$ [PSU]		$\overline{cRMSE_S}$ [PSU]	
					P	N	P	N
January	PF_6901835	6	8.6	−2	0.09	0.07	0.06	0.06
February	PF_6901835	6	24.6	4.3	0.15	0.11	0.07	0.07
March	PF_6901835	6	6.1	10.6	0.08	0.07	0.05	0.04
	PF_6902692	3	37.7	45.7	0.11	0.08	0.06	0.03
May	PF_FKJB	1	1.1	−0.8	0.06	0.06	0.03	0.03
	PF_6901835	5	36	19.1	0.07	0.05	0.05	0.04
	PF_6902692	7	3.6	6.1	0.08	0.08	0.07	0.07
September	CT_FNCM	8	−52	−24.5	0.12	0.23	0.1	0.12
	GL_6801661	164	12.4	15.2	0.11	0.1	0.13	0.13
	GL_6801663	251	4	11.4	0.13	0.13	0.12	0.11
October	GL_6801661	621	−10.4	−8.3	0.11	0.12	0.09	0.09
	GL_6801663	639	−6.8	0.9	0.13	0.14	0.11	0.11
November	GL_6801661	180	−20	−11.1	0.07	0.09	0.07	0.08
	GL_6801663	78	16.1	8.8	0.07	0.06	0.06	0.06

ER_S and cER_S measure the reduction in root mean squared error and centred root mean squared error obtained with the downscaling procedure. Averaged root mean squared error $\overline{RMSE_S}$ and centred root mean squared error $\overline{cRMSE_S}$ for both PURPLE (P) and NEMO (N) models are also reported

Fig. 4 Depth-dependent value of the Brunt-Väisälä frequency squared N^2 derived by temperature and salinity profiles from observations (top plot), PURPLE model (second plot from the top), NEMO model (third plot from the top), and depth at which maximum N^2 value is attained (bottom plot) for observations (black dots), PURPLE (red dots), and NEMO (blue dots) models. The black dashed line represents the median value of $z_{max(N^2)}$ for the observation dataset. Subfigures **a** and **b** refer to platform GL_6801661 and GL_6801663, respectively
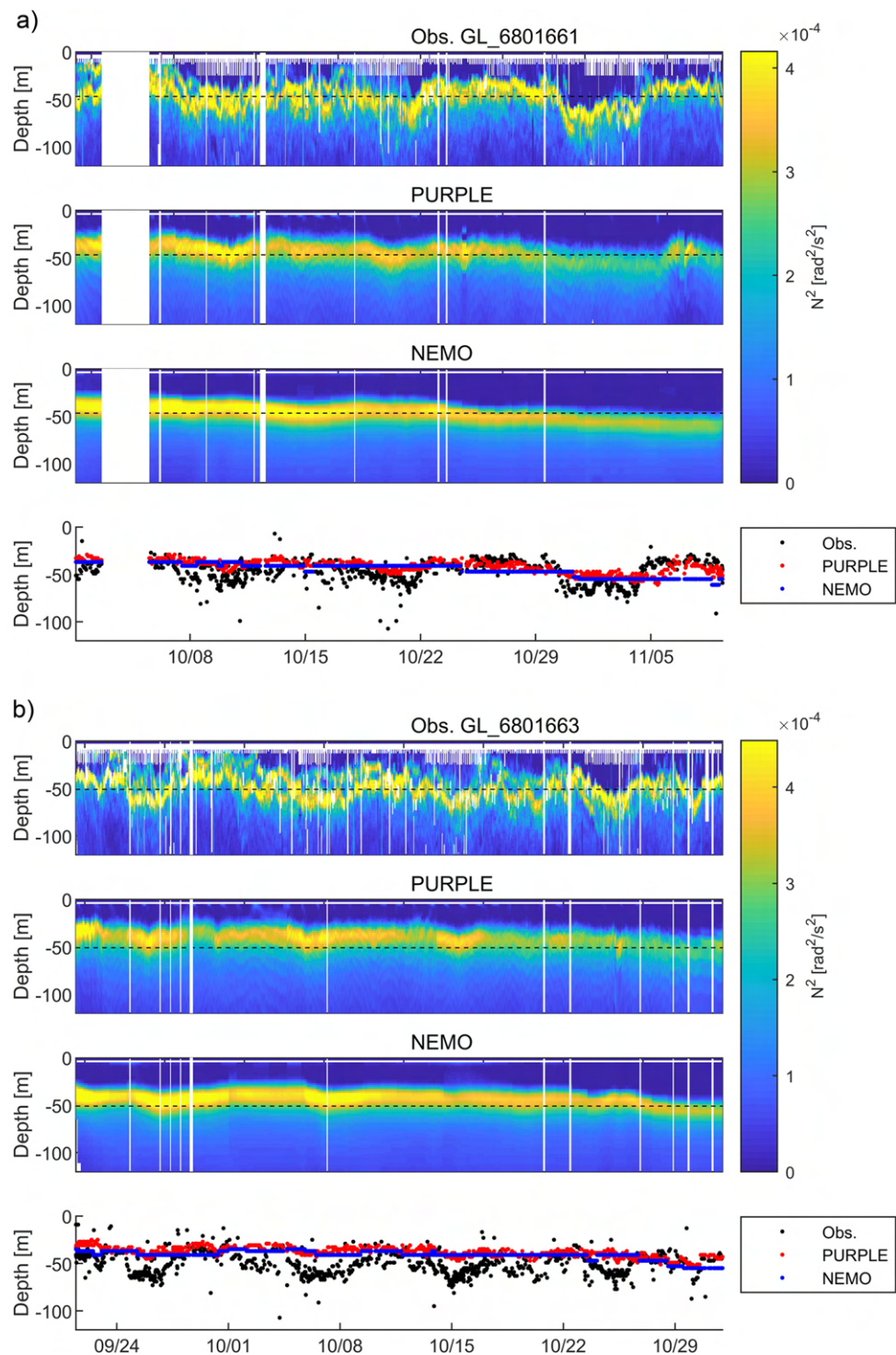
each subfigure, reporting the depth at which the maximum N^2 value is attained, $z_{max(N^2)}$. The black dashed line in the first three plots of each subfigure represents the median value of $z_{max(N^2)}$ for the observation dataset. A slightly better performance of PURPLE model with respect to NEMO

is detectable for the dataset GL_6801661 (Fig. 4a), having the former a higher vertical resolution than the latter within the upper layer. Both models tend to underestimate the depth at which the transition toward a stable stratification occurs, except for the period 4–9 November.

4.2 Models' performance in reproducing surface currents

Surface currents from HF-radar are available for the northern part of PURPLE domain, as shown in Fig. 1c. To assess the improvement obtained by downscaling NEMO surface velocity, several statistics similar to those employed in previous section are calculated. More specifically, we compare observed and modelled values for velocity module $|\mathbf{U}| = \sqrt{u^2 + v^2}$ and direction $\theta = \arctan(v/u)$ by means of $RMSE_{|\mathbf{U}|}(x, y)$ and $RMSE_{\theta}(x, y)$, which account for the spatial distribution of the root mean squared error, given a specific time window, corresponding to each quarter of the year

(JFM, AMJ, JAS, and OND). Furthermore, we determined the correlation coefficient $\rho(x, y)$ between observed and modelled data for the u- and v-component of the velocity, using the same quarterly time windows. To measure the significance of the correlation, we performed a t-test hypothesis with significance level of 0.05.

To make the comparison possible, since the PURPLE model output is provided every 3 h, we interpolate 3-hourly original daily NEMO data, and take HF-radar data every 3 h. Spatial interpolation is carried out linearly and modelled velocities are mapped to HF-radar observation points.

The spatial distribution of the root mean squared error for both $|\mathbf{U}|$ and θ is reported in Fig. 5. It is possible to notice
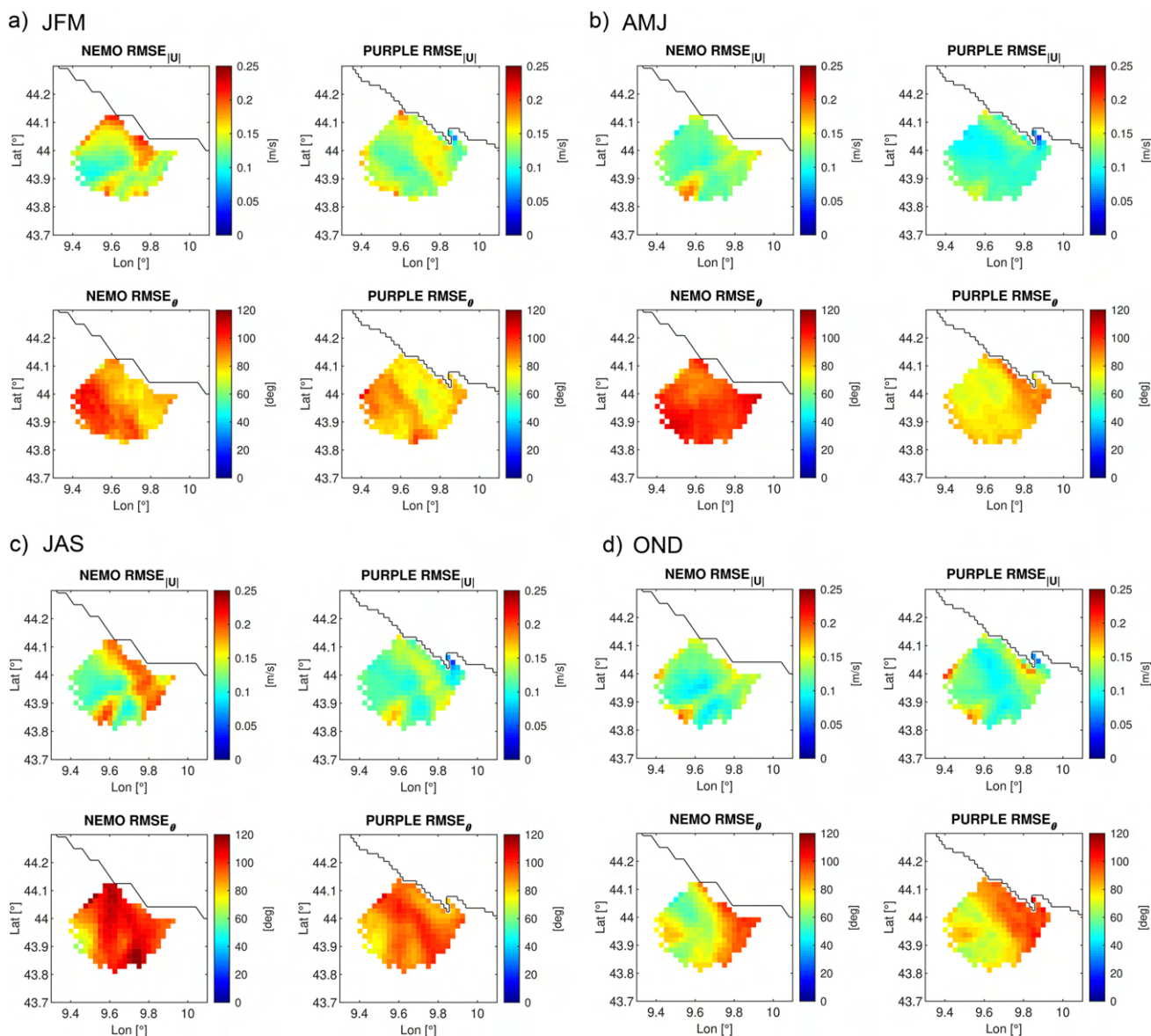

Fig. 5 Spatial distribution of the root mean squared error for the module of velocity $RMSE_{|\mathbf{U}|}(t)$ and its direction $RMSE_{\theta}(t)$, grouped by 3-month intervals

that the two models exhibit similar error patterns for the velocity magnitude ($O(10^{-1})$ m/s), with larger errors toward the coast, whereas errors in velocity direction are rarely lower than 50°for both models and variably located based on the analysed period. For each of the analysed time window, both the $RMSE_{|\mathbf{U}|}(x,y)$ and $RMSE_{\theta}(x,y)$ tend to be lower for the PURPLE model with respect to NEMO, except for the OND quarter (Fig. 5b).

To better quantify the improvement obtained by the downscaling procedure concerning the $RMSE$, the quantity ER_{curr} is defined as follows:

$$ER_{curr} = \frac{100}{N_{xy}} \sum \frac{RMSE_{curr,P} - RMSE_{curr,N}}{\frac{1}{2}\left(RMSE_{curr,P} + RMSE_{curr,N}\right)} \qquad (2)$$

where the summation extends to the number of compuational cells (N_{xy}) over which the comparison between model and radar is evaluated. Results are reported in Table 3. In the same table also the results obtained by daily-averaging PURPLE and HF-radar data are reported ($ER_{curr,d}$). This makes the comparison between the two models more robust being the NEMO data daily-averaged values. For both the quantities ER_{curr} and $ER_{curr,d}$, it is possible to observe a reduction of model error due to the downscaling procedure, except for the velocity direction during the JAS window for the daily-averaged analysis, and during the OND time windows.

In Fig. 6, the correlation coefficients for u and v for those data with a p-value lower than 0.05 are reported. $\rho(x,y)$ values associated to p-values larger than 0.05 are set to NaN (not a number) and not shown in the subfigures. A characteristic pattern for the spatial distribution of the correlation coefficient is not easily identifiable and variations throughout the time windows are present. In general, larger correlation values are noticeable for the PURPLE with respect to the NEMO model. This is also confirmed by the larger amount of data for which a correlation is statistically significant, since some portions of the HF-radar domain for the NEMO model are characterized by $\rho(x,y)$ with p-value > 0.05 (Fig. 6b and c, related to v).

The spatially averaged values of $\rho(x,y)_u$ and $\rho(x,y)_v$ are determined for both NEMO and PURPLE models and reported in Table 4, together with the fraction of data

showing a significant correlation (in round brackets). As for the analysis of the $RMSE$, the values obtained by the analysis carried out on daily-averaged data are reported. For all the analysed time windows, a larger correlation is found for the PURPLE with respect to the NEMO model and a larger proportion of significantly correlated data. Few exceptions are $\overline{\rho}_v$, $\overline{\rho}_{u,d}$, and $\overline{\rho}_{v,d}$ for the OND time window.

5 Discussion

5.1 Models performances and sources of error

Even if the amount of available observation does not uniformly cover the spatial and temporal domain, several aspects regarding the effectiveness of the downscaling procedure in providing more reliable and precise results are emerged and deserve to be analysed.

For temperature profiles, the comparison between modelled and observed data allows us to conclude that the downscaling procedure is able to reduce the error between model and observations for those periods where a huge amount of observation values is available (September to November), whereas this is not clearly detectable in relation to periods poorly covered by field data. Indeed, looking at Table 1, from January to May, we have a predominance of positive values of ER_T (worsening) and a predominance of negative values for cER_T (improvement). Of course, the performance of the downscaling procedure has nothing to do with the number of available observations, but being the uncertainty associated to the periods with a smaller number of data, we argue that a further number of observations in the winter and spring months would be required to more robustly confirm the better performance of the PURPLE model with respect to NEMO.

The results regarding salinity profiles are more uncertain. In such a case, the predominance of negative ER_S and cER_S values, among those months with the higher number of observations (September to November), is not present. Furthermore, for those having at most 7 profiles available, a worsening is detectable (see Table 2). However, mean bias for PURPLE model profiles resembles, for most part of observations, the mean bias from NEMO profiles (Fig. 3), raising the question of how much a downscaled model improves the output, in relation to the degree of inaccuracy of the parent model.

To tackle the issue, the scatter plot of $RMSE_N$ versus $RMSE_P$, for all available temperature and salinity profiles from 0 to 200 m depth, is reported in Fig. 7 (a and c), together with the relative error reduction:

Table 3 Summary of the monthly synthetic values for the error reduction ER_{curr} for surface currents

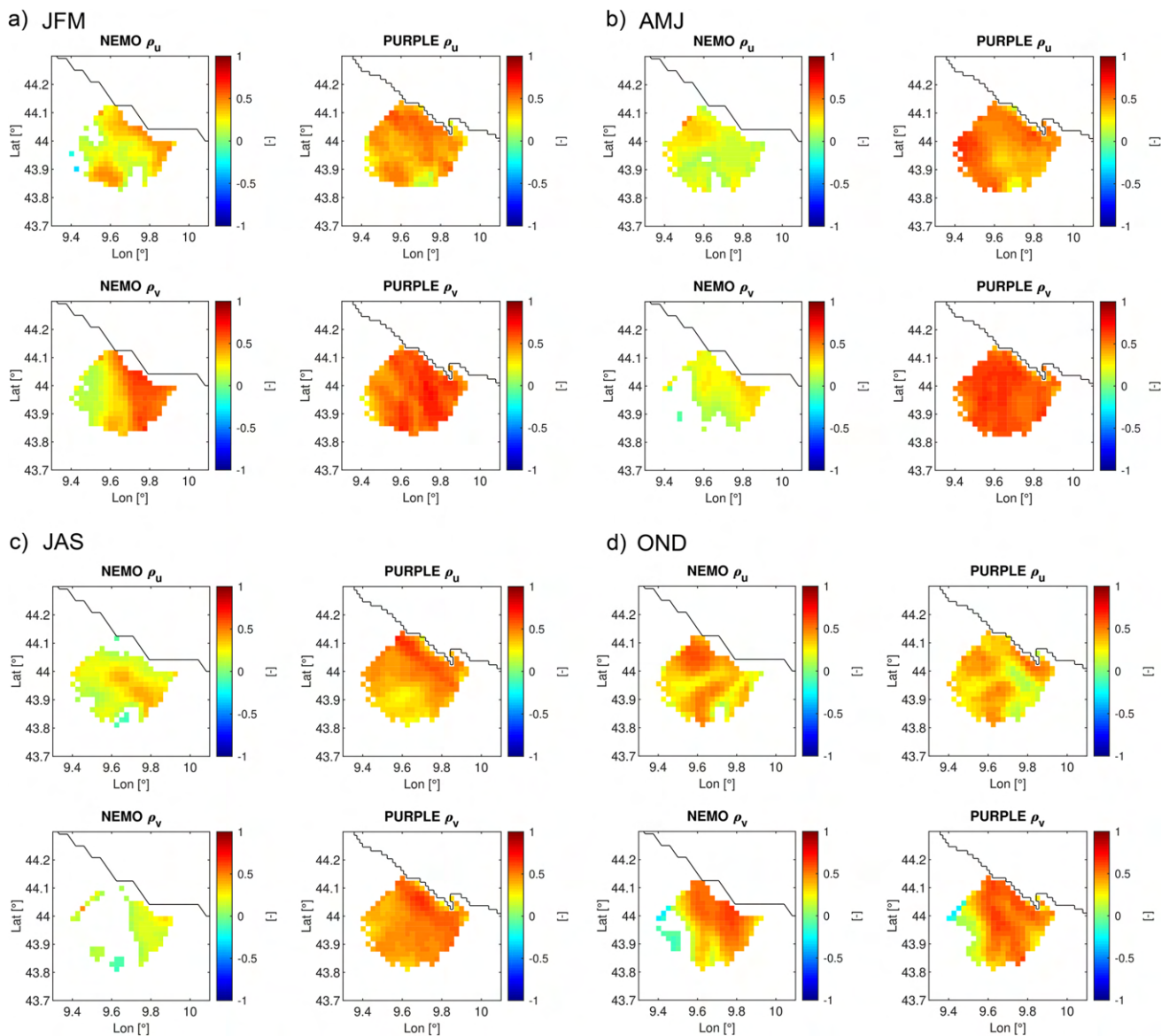| Period	ER_{curr} [%]		$ER_{curr,d}$ [%]	
	$|\mathbf{U}|$	θ	$|\mathbf{U}|$	θ
JFM	−4.1	−9.1	−13.6	−13.4
AMJ	−14.9	−23.3	−21.4	−8.9
JAS	−16.6	−8.6	−16.2	2.7
OND	−2.3	7.2	−4.0	9.5

Fig. 6 Spatial distribution of the correlation coefficient for u and v, grouped by 3-month intervals

Table 4 Summary of the spatially averaged values of the correlation coefficients ρ_u and ρ_v

Period	$\overline{\rho}_u$ (%)		$\overline{\rho}_v$ (%)		$\overline{\rho}_{u,d}$ (%)		$\overline{\rho}_{v,d}$ (%)	
	N	P	N	P	N	P	N	P
JFM	0.26 (84.8)	0.41 (98.3)	0.36 (96.3)	0.54 (99.2)	0.37 (69.6)	0.47 (93.0)	0.50 (72.8)	0.61 (98.7)
AMJ	0.19 (93.7)	0.45 (100)	0.20 (67.0)	0.57 (100)	0.34 (79.6)	0.42 (77.6)	0.28 (40.3)	0.57 (98.4)
JAS	0.21 (88.8)	0.41 (100)	0.14 (40.0)	0.46 (100)	0.34 (74.0)	0.44 (83.1)	0.29 (7.6)	0.46 (97.6)
OND	0.36 (97.8)	0.32 (99.1)	0.37 (86.1)	0.41 (92.7)	0.43 (92.4)	0.40 (81.5)	0.52 (73.1)	0.50 (82.2)

Values obtained with daily-averaged data are indicated with $\rho_{u,d}$ and $\rho_{v,d}$. Percentages in round brackets correspond to the proportion of correlation coefficient values associated to p-values < 0.05. N and P refer to NEMO and PURPLE models, respectively
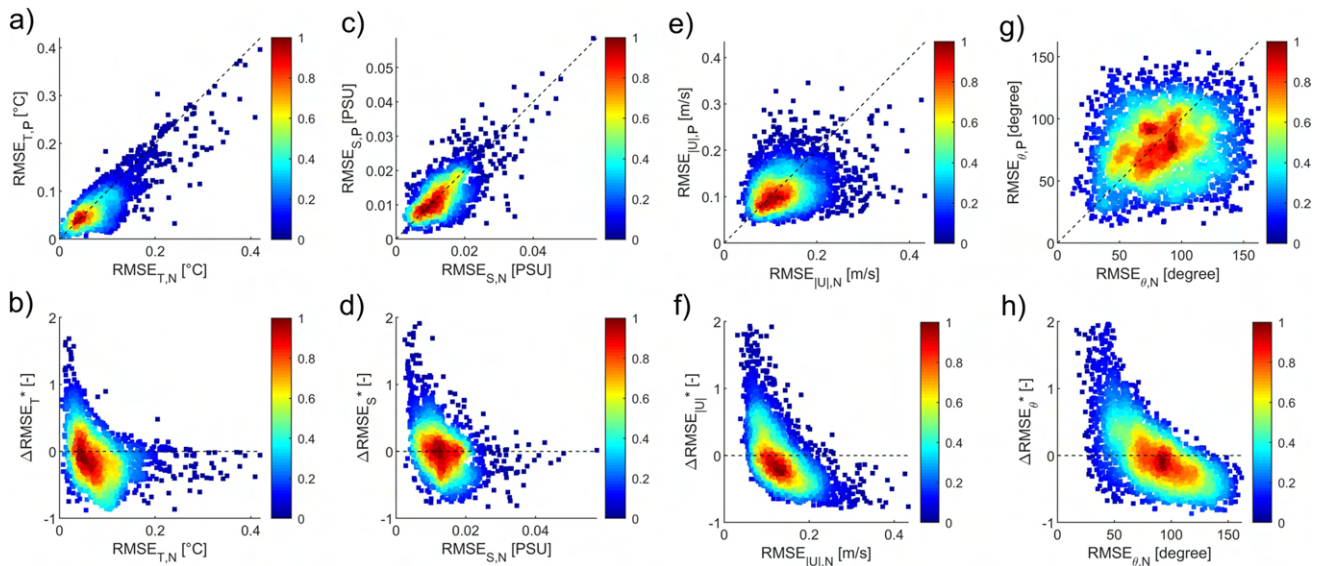
Fig. 7 Scatter plot between NEMO and PURPLE root mean squared error for temperature (**a**) and salinity (**c**); black dashed line indicates 1:1 reference. Scatter plot between relative error reduction $\Delta RMSE^*$ as a function of initial root mean squared error $RMSE_N$ for both temperature (**b**) and salinity (**d**); black dashed line is the horizontal reference between positive and negative values. Scatter plot between NEMO and PURPLE root mean squared error for velocity magnitude (**e**) and direction (**g**); black dashed line indicates 1:1 reference. Scatter plot between relative error reduction $\Delta RMSE^*$ as a function of initial root mean squared error $RMSE_N$ for both velocity magnitude (**f**) and direction (**h**); black dashed line is the horizontal reference between positive and negative values. Colour bars indicate density of data

$$\Delta RMSE^* = \frac{RMSE_P - RMSE_N}{RMSE_N} \qquad (3)$$

as a function of the initial root mean squared error $RMSE_N$ (b and d). Points on the right of 1:1 line (black dashed line) in Fig. 7a and c mean a reduction in the root mean squared error due to downscaling, and indeed, most part of temperature profiles (65%) are improved by PURPLE model (a), whereas if it is less evident (52% profiles improved) for salinity (c). As expected, a correlation seems to exist between $RMSE_N$ and $RMSE_P$, meaning that the performances of the downscaled model are anchored to those of the parent one. Furthermore, looking at Fig. 7b and d, we observe that error reduction (negative $\Delta RMSE^*$ values, Eq. 3) is mostly concentrated within specific ranges of the error coming from the NEMO model (around 0.1 °C for temperature, and between 0.01 and 0.02 PSU for salinity). For those profiles having larger initial errors ($RMSE_{T,N}$ larger than 0.2 °C and $RMSE_{S,N}$ larger than 0.03 PSU), the effect of downscaling is not able to reduce them more than 40% (i.e. only few values of $\Delta RMSE_T^*$ and $\Delta RMSE_S^*$ are lower than −0.4 in Fig. 7b and d).

Figures 2 and 3 (subfigures from a to c) show that most part of inaccuracies concerning temperature and salinity profiles lie within the depth range −25 to −75 m. Looking at Fig. 4, the main source of error appears to be in an underestimation, by the models, of the depth at which the most stable stratification occurs. This was also noticed by

Onken (2017) for a similar application close to the western Sardinian coast (Mediterannean Sea), and might partially be ascribable to a low value of the vertical eddy viscosity, which can be enhanced by surface waves, which are, in fact, not included in the models. Moreover, the underestimation of the stable layer present in the NEMO model is hardly recovered by PURPLE. On the one hand, the daily-averaged output and the vertical resolution of NEMO, which decreases linearly from 4 to 9 m passing from −20 to −100 m depth, might not be sufficient to accurately resolve the stratification pattern. Indeed, the stable layer depth remains almost unchanged even if the NEMO model ingests both temperature and salinity profiles via a 3D-Var assimilation scheme (Simoncelli et al. 2019). On the other hand, the short spin-up period would not allow the nested model to deepen the mixed layer and is just sufficient to slightly increase the variability of the $z_{max(N^2)}$ for PURPLE with respect to NEMO (Fig. 4), whereas the models have almost the same behaviour below −100 m (Figs. 2 and 3).

This gives rise to the issue concerning the implementation strategy of the modelling system: if the interest is on surface processes, a frequent reinitialization avoids the nested model to drift, while allowing the development of superficial variability; if the interest is on deep layer processes, a longer time is required to the nested model to develop a sufficient variability below the surface layer. An assimilation procedure or simple nudging toward measured values in the downscaled model might be useful to reduce

the mismatch between observed and modelled stratification and to avoid the need for frequent reinitializations.

Surface current observations from HF-radar uniformly cover the whole year. To give a measure of the ability of the model to reproduce the main current patterns, the monthly-averaged velocity field is reported in Fig. 8 for the months of January (a) and June 2017 (b). The winter flow field shows a northward transport through the Corsica Channel, together with the westward veering of the water coming from south (Millot 1999), whereas a flow reversal is present during the summer season, as reported by Sciascia et al. (2019). However, these main features at large temporal and spatial scales are strictly linked to the flow pattern coming from the parent model and the downscaling procedure is hardly able to change them significantly.

Overall, the PURPLE model tends to improve the performance of NEMO for most part of the year, but for the last quarter (Tables 3 and 4; Figs. 5 and 6). However, since the wind forcing has a first-order effect on surface currents, even at coastal scale (Lana et al. 2016), we may argue it is a worsening in the atmospheric forcing to be responsible of the increased error in November and December. This aspect deserves to be specifically addressed in future researches through a coupled analysis of the performances of atmospheric and oceanographic models with respect to observed winds and currents, respectively. Furthermore, it is noteworthy to stress the high value of the root mean squared error in the current direction, for both the NEMO and PURPLE models (Fig. 5), ranging from 40 to 120°, and the values of the correlation coefficient which range between 0.36 and 0.61 for the PURPLE model, and between 0.14 and 0.50 for the NEMO model (Table 4). Both ERA-Interim original winds and downscaled ERA5 winds may result imprecise close to the coast, especially where steep topographic

gradients are present in the mainland (Zhang et al. 2013), as for the area monitored by HF-radar. This might be a reason why both NEMO and PURPLE models do not present good skills in reproducing surface current direction and achieve correlation coefficients of no more than 0.6 for u and v.

In addition to the increased spatial resolution, one of the reasons for the improvement obtained by the downscaling procedure, especially for the velocity magnitude, is attributable to the differences in temporal outputs between the two models, daily-averaged values for NEMO and three-hourly intervals for ROMS. The results obtained from three-hourly interpolated and daily-averaged data, reported in Table 3, do not show a clear trend in the differences among results, and a predominance of negative values for the ER_{curr} is present in both approaches. Furthermore, looking at Table 4, it is possible to notice that correlation tends to increase for NEMO when HF-radar data are daily-averaged, whereas it remains almost unchanged for PURPLE data. However, both for three-hourly interpolated and daily-averaged data, the downscaling procedure appears to increase the correlation.

To verify a potential relation between the error from the two models, the scatter plots of $RMSE_{|U|,N}$ versus $RMSE_{|U|,P}$, and $RMSE_{\theta,N}$ versus $RMSE_{\theta,P}$ are reported in Fig. 7 (e and g). They show the lack of clear correlation between the errors for the NEMO and PURPLE models, and a predominance of data values below the reference line (1:1), 60% for velocity magnitude, and 57% for direction.

Furthermore, as for the temperature and salinity profiles, we determine the relative error reduction $\Delta RMSE^*$ for both $|U|$ and θ (Fig. 7f and h). It is possible to observe that the error reduction (negative $\Delta RMSE^*$ values) tends to span the initial $RMSE_N$ values, more uniformly, than observed for temperature and salinity profiles. Indeed, inaccurate initial condition for surface current may be
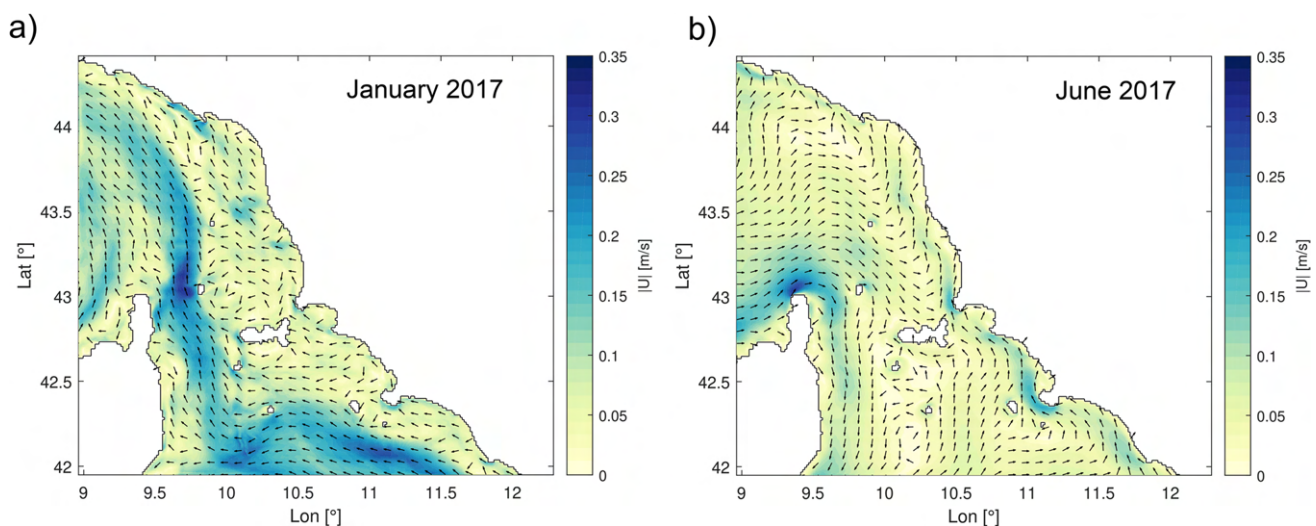

Fig. 8 Monthly-averaged surface currents for January 2017 (**a**) and June 2017 (**b**) for the PURPLE model

corrected by increased resolution or improved wind forcing, being surface dynamics faster than interior in responding to atmospheric forcing.

To consider if the improvement/worsening of model solution due to downscaling is affected by the time after the spin-up period (3 days), we analyse if significant variations in the distribution of the $\Delta RMSE = RMSE_P - RMSE_N$ occur across the three sets composed by all the first, the fourth, and the seventh days after the spin-up period. The check is carried out for temperature, salinity, velocity magnitude and direction. We did not find significant differences in the statistical distribution of the $\Delta RMSE$ across the three sets of days (results not shown).

Another important issue concerns the spatial coverage of the HF-radar observations. They can be used to validate the model on a small portion of the computational domain and it is rather difficult to infer about the skill of the model away from the observation area. It is therefore important to keep in mind that the conclusions about the model's performance, obtained by localized observations which give information on the surface layer, do not necessarily ensure similar results for the remaining part of the domain. This is especially true for the deeper layer, where the potential improvement of the downscaling procedure may not be significant due to the short spin-up period employed. Moreover, the present analysis shows the intrinsic difficulty to unambiguously evaluate the skill of a model to reproduce observed patterns. A large amount of data, even if covering a portion of the analysed domain, requires several statistics and caution to interpret the results.

5.2 A simple metric to compare vector fields

The use of root mean squared error to test the skill of the model to reproduce current direction and magnitude, separately, can be partially misleading. It would therefore be useful to have a metric which combines the performances of the model in reproducing current magnitude and direction. We propose to employ the measure of the error ε, defined as follows:

$$\varepsilon = \frac{\sqrt{\left(\Delta\theta\overline{|\mathbf{U}|}\right)^2 + \Delta|\mathbf{U}|^2}}{\overline{|\mathbf{U}|}} \tag{4}$$

where $\Delta\theta$ is the angular difference in radians between modelled and observed velocity vectors, $\overline{|\mathbf{U}|} = \frac{1}{2}(|\mathbf{U}_o| + |\mathbf{U}_m|)$ and $\Delta|\mathbf{U}| = |\mathbf{U}_o| - |\mathbf{U}_m|$, are the mean and the difference of the velocity magnitudes from observed (subscripts o) and modeled (subscripts m) values, respectively. The first component under the square root accounts for the error in direction, whereas the second for the error in magnitude.

Another possible metric to measure the mismatch between observations and models is simply the ratio of the vectorial difference to the mean magnitude of velocity

$$\varepsilon^* = \frac{\sqrt{(u_o - u_m)^2 + (v_o - v_m)^2}}{\overline{|\mathbf{U}|}} \tag{5}$$

The values assumed by the two measures of error ε and ε^* (Eqs. 4 and 5) for a series of possible mismatches between observed and modelled vectors are reported at the bottom of Fig. 9. Each value of ε and ε^* corresponds to a specific space-time coordinate (x, y, t). The two metrics behave similarly when the velocity vectors have small differences in orientations, whereas the larger the value of $\Delta\theta$, the larger the penalty obtained by ε with respect to ε^*. This is indeed a way to emphasize the effect of mismatch in directions. A certain dose of arbitrariness affects the choice, but we argue it is a good compromise to weigh both the performance of the model in terms of current kinetic energy (i.e. velocity magnitude) and mean direction.

To assess the skill of the model to reproduce the observed current velocity field, the values of ε are time-averaged. Figure 9 reports the spatially distributed $\varepsilon_{\mathrm{avg}}(x, y)$ for the four quarters of the year. All the quarters clearly show a better performance of PURPLE model with respect to NEMO. Moreover, if we determine a measure of the reduction of error ER, analogously to what was done for the other variables $|\mathbf{U}|$ and θ:

$$ER_\varepsilon = \frac{100}{N_{xyt}} \sum \frac{\varepsilon_P - \varepsilon_N}{\frac{1}{2}(\varepsilon_P + \varepsilon_N)} \tag{6}$$

where N_{xyt} represents the total amount of data points in a quarter. ER_ε gets always negative values equal to -40%, -66%, -46%, and -26% for the JFM, AMJ, JAS, and OND quarters, respectively. These values are determined employing velocity field data at 3-hourly time scale.

If the same procedure is carried out using daily-averaged velocity fields, the reduction of error ER is equal to -49%, -50%, -30%, and -23% for the JFM, AMJ, JAS, and OND quarters, respectively. In both cases, the reduction in the error is large and is not significantly affected by the use of 3-hourly or daily-averaged data.

6 Conclusion

In the present work, we try to quantify the improvement gained by downscaling the CMEMS reanalysis product MEDSEA_REANALYSIS_PHY_006_004 (Simoncelli et al. 2019), in reproducing temperature and salinity profiles, and surface currents, in the area between Ligurian and Tyrrhenian basins, for the year 2017. ROMS in
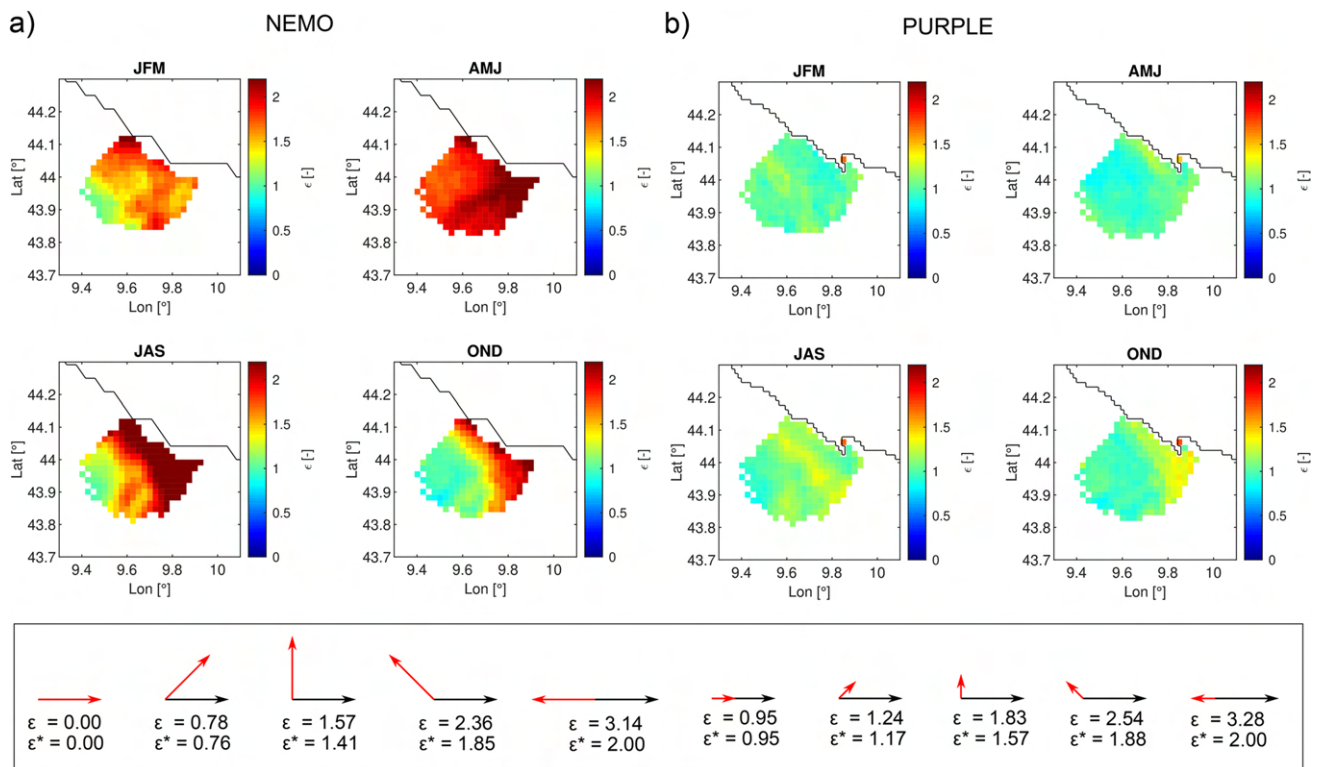
Fig. 9 Spatially distributed time-averaged values $\varepsilon_{avg}(x, y)$ for the quarters of the year from NEMO and PURPLE models, together with sketches showing the values attained by ε and ε^* for different combinations of modelled and observed velocity vector

one-way nested configuration, with atmospheric forcing from ERA5 dynamic downscaling (Vannucchi et al. 2021), is employed to transfer ocean dynamics from 1/16° daily scale (NEMO model) to 1/216° 3-hourly scale (PURPLE configuration).

The comparison of model outputs with respect to temperature profiles shows the downscaled model (PURPLE) has, in general, a lower root mean squared error than the parent one (NEMO). This is supported by a large number of observations $O(10^2{-}10^3)$, which were available during September, October, and November. On the contrary, for the other months where a few observations where available $O(10)$, the two models appear to behave similarly: NEMO having a lower RMSE, PURPLE a lower cRMSE. If we consider salinity profiles, the comparison is less straightforward, being the NEMO model to perform better for most part of the year if we consider observations grouped per months, but it is the PURPLE model to have a lower root mean squared error if we consider field measurements all at once.

The analysis of the static stability profiles shows that both models suffer for an underestimation of mixed layer depth, which is responsible for the largest values in mean bias for temperature and salinity profiles between −25 and −75 m depth. Such an inaccuracy is indeed hardly recovered by the downscaled model.

A simple analysis of the root mean squared error for temperature profiles shows that the downscaling procedure is more effective in reducing those errors from parent model which are in between 0.05 and 0.15 °C, though it is less evident for salinity.

Surface currents from HF-radar cover an area roughly 30×30 km^2 in front of Cinque Terre (Liguria, Italy). The PURPLE model has better performances than NEMO in reproducing surface currents for at least three quarters of the year, being able to reduce the RMSE for both the module and the direction of velocity, and increasing the correlation coefficient for the u and v components. However, such a result is entirely obtained considering the limited area covered by HF-radars and caution is required to infer that the same holds true for the remaining part of the domain.

The analysis of dependencies between errors in surface velocity field from NEMO and PURPLE models shows there is not a preferential error range for which the downscaling procedure helps to improve the representation of surface sea dynamics. Indeed, surface currents are strongly affected by atmospheric forcing, and lose the dependency from the dynamic state provided by initial conditions faster than temperature and salinity profiles.

A simple metric to compare vectorial quantities is proposed to simultaneously account for both differences in

velocity magnitude and direction between observed and modelled values. Using such a metric to assess the performance of the models with respect to surface currents, a more clear improvement is observable as a consequence of the downscaling procedure.

At the present stage, the downscaling procedure of the updated reanalysis product provided by CMEMS, MED-SEA_MULTIYEAR_PHY_006_004 (Escudier et al. 2020) is ongoing, and future works will compare this product with the downscaled model presented herein.

Data availability The datasets generated during and/or analysed during the current study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors declare no competing interests.

References

Aguiar E, Mourre B, Juza M et al (2020) Multi-platform model assessment in the western mediterranean sea: impact of downscaling on the surface circulation and mesoscale activity. Ocean Dynamics 70(2):273–288

Astraldi M, Gasparini G (1992) The seasonal characteristics of the circulation in the north mediterranean basin and their relationship with the atmospheric-climatic conditions. Journal of Geophysical Research: Oceans 97(C6):9531–9540

Astraldi M, Bianchi C, Gasparini G et al (1995) Climatic fluctuations, current variability and marine species distribution-a case-study in the ligurian sea (north-west mediterranean). Oceanologica acta 18(2):139–149

Blayo E, Debreu L (2006) Nesting ocean models. In: Ocean weather forecasting. Springer, p 127–146

Bricheno LM, Wolf JM, Brown JM (2014) Impacts of high resolution model downscaling in coastal regions. Continental Shelf Research 87:7–16

Buzzi A, Davolio S, Malguzzi P et al (2014) Heavy rainfall episodes over liguria in autumn 2011: numerical forecasting experiments. Natural Hazards and Earth System Sciences 14(5):1325

Chapman DC (1985) Numerical treatment of cross-shelf open boundaries in a barotropic coastal ocean model. Journal of Physical Oceanography 15(8):1060–1075

Copernicus Marine In Situ TAC (2020) Copernicus in situ NRT current product user manual (PUM). Ifremer. https://doi.org/10.13155/73192

Copernicus Marine Team (2020) Product User Manual for multiparameter Copernicus In Situ TAC (PUM). Copernicus in situ TAC. https://doi.org/10.13155/43494

Corgnati L, Mantovani C, Rubio A et al (2018) Building strong foundations towards the pan-european high frequency radar network. Boll Geof Suppl 59:246–247

Cucco A, Sinerchia M, Ribotti A et al (2012) A high-resolution real-time forecasting system for predicting the fate of oil spills in the strait of bonifacio (western mediterranean sea). Marine Pollution Bulletin 64(6):1186–1200

Cushman-Roisin B, Beckers JM (2011) Introduction to geophysical fluid dynamics: physical and numerical aspects. Academic Press, New York

Dee DP, Uppala S, Simmons A et al (2011) The era-interim reanalysis: Configuration and performance of the data assimilation system. Quarterly Journal of the Royal Meteorological Society 137(656):553–597

Dobricic S, Pinardi N (2008) An oceanographic three-dimensional variational data assimilation scheme. Ocean Modelling 22(3–4):89–105

Escudier R, Clementi E, Omar M, et al (2020) Mediterranean sea physical reanalysis (cmems med-currents). Copernicus Monitoring Environment Marine Service (CMEMS)

Fairall CW, Bradley EF, Rogers DP et al (1996) Bulk parameterization of air-sea fluxes for tropical ocean-global atmosphere coupled-ocean atmosphere response experiment. Journal of Geophysical Research: Oceans 101(C2):3747–3764

Flather R (1976) A tidal model of the north-west european continental shelf. Memoires de la Societe Royale de Sciences de Liege 6:141–164

Fofonoff N, Millard Jr R (1983) Algorithms for computation of fundamental properties of seawater. endorsed by unesco/scor/ices/iapso joint panel on oceanographic tables and standards and scor working group 51. unesco technical papers in marine science, no. 44. Tech. rep., UNESCO

Fossi MC, Romeo T, Baini M et al (2017) Plastic debris occurrence, convergence areas and fin whales feeding ground in the mediterranean marine protected area pelagos sanctuary: A modeling approach. Frontiers in Marine Science 4:167

Haidvogel DB, Arango HG, Hedstrom K et al (2000) Model evaluation experiments in the north atlantic basin: simulations in nonlinear terrain-following coordinates. Dynamics of Atmospheres and Oceans 32(3–4):239–281

Haney RL (1991) On the pressure gradient force over steep topography in sigma coordinate ocean models. Journal of Physical Oceanography 21(4):610–619

Iacono R, Napolitano E (2020) Aspects of the summer circulation in the eastern ligurian sea. Deep Sea Research Part I: Oceanographic Research Papers 166(103):407

Katavouta A, Thompson KR (2016) Downscaling ocean conditions with application to the gulf of maine, scotian shelf and adjacent deep ocean. Ocean Modelling 104:54–72

Lana A, Marmain J, Fernandez V et al (2016) Wind influence on surface current variability in the ibiza channel from hf radar. Ocean Dynamics 66(4):483–497

López AG, Wilkin JL, Levin JC (2020) Doppio–a roms (v3. 6)-based circulation model for the mid-atlantic bight and gulf of maine: configuration and comparison to integrated coastal observing network observations. Geoscientific Model Development 13(8):3709–3729

Lorente P, García-Sotillo M, Amo-Baladrón A, et al (2019) Skill assessment of global, regional, and coastal circulation forecast models: evaluating the benefits of dynamical downscaling in ibi (iberia–biscay–ireland) surface waters. Ocean Sci 15(4)

Madec G (2014) the nemo team: Nemo ocean engine–version 3.4. Note du Pole de modelisation, Institut Pierre-Simon Laplace (IPSL) 27

Mader J, Rubio A, Asensio J, et al (2017) The European HF radar inventory. publication date 15 Sep 2016 (updated version 30 Jan 2017). Tech. rep., EuroGOOS

Marchesiello P, McWilliams JC, Shchepetkin A (2001) Open boundary conditions for long-term integration of regional oceanic models. Ocean Modelling 3(1–2):1–20

Mason E, Molemaker J, Shchepetkin AF et al (2010) Procedures for offline grid nesting in regional ocean models. Ocean Modelling 35(1–2):1–15

Millot C (1999) Circulation in the western mediterranean sea. Journal of Marine Systems 20(1–4):423–442

Olita A, Ribotti A, Fazioli L et al (2013) Surface circulation and upwelling in the sardinia sea: A numerical study. Continental Shelf Research 71:95–108

Onken R (2017) Validation of an ocean shelf model for the prediction of mixed-layer properties in the mediterranean sea west of sardinia. Ocean Sci 13(2)

Onken R, Robinson AR, Kantha L et al (2005) A rapid response nowcast/forecast system using multiply nested ocean models and distributed data systems. Journal of Marine Systems 56(1–2):45–66

Onken R, Baschek B, Angel-Benavides IM (2020) Very high-resolution modelling of submesoscale turbulent patterns and processes in the baltic sea. Ocean Science 16(3):657–684

Sciascia R, Magaldi MG, Vetrano A (2019) Current reversal and associated variability within the corsica channel: The 2004 case study. Deep Sea Research Part I: Oceanographic Research Papers 144:39–51

Shchepetkin AF, McWilliams JC (2003) A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate. J Geophys Res Oceans 108(C3)

Shchepetkin AF, McWilliams JC (2005) The regional oceanic modeling system (roms): a split-explicit, free-surface, topography-following-coordinate oceanic model. Ocean Modelling 9(4):347–404

Sikirić MD, Janeković I, Kuzmić M (2009) A new approach to bathymetry smoothing in sigma-coordinate ocean models. Ocean Modelling 29(2):128–136

Simoncelli S, Fratianni C, Pinardi N, et al (2019) Mediterranean sea physical reanalysis (cmems med-physics)[data set]. Copernicus Monitoring Environment Marine Service (CMEMS)

Sorgente R, Tedesco C, Pessini F et al (2016) Forecast of drifter trajectories using a rapid environmental assessment based on ctd observations. Deep Sea Research Part II: Topical Studies in Oceanography 133:39–53

Taylor KE (2001) Summarizing multiple aspects of model performance in a single diagram. Journal of Geophysical Research: Atmospheres 106(D7):7183–7192

Trotta F, Fenu E, Pinardi N et al (2016) A structured and unstructured grid relocatable ocean platform for forecasting (surf). Deep Sea Research Part II: Topical Studies in Oceanography 133:54–75

Trotta F, Pinardi N, Fenu E et al (2017) Multi-nest high-resolution model of submesoscale circulation features in the gulf of taranto. Ocean Dynamics 67(12):1609–1625

Umlauf L, Burchard H (2003) A generic length-scale equation for geophysical turbulence models. Journal of Marine Research 61(2):235–265

Vannucchi V, Taddei S, Capecchi V et al (2021) Dynamical downscaling of era5 data on the north-western mediterranean sea: From atmosphere to high-resolution coastal wave climate. Journal of Marine Science and Engineering 9(2):208

Visbeck M (2018) Ocean science research is key for a sustainable future. Nature communications 9(1):1–4

Warner JC, Geyer WR, Arango HG (2010) Using a composite grid approach in a complex coastal domain to estimate estuarine residence time. Computers & Geosciences 36(7):921–935

Zhang H, Pu Z, Zhang X (2013) Examination of errors in near-surface temperature and wind from wrf numerical simulations in regions of complex terrain. Weather and Forecasting 28(3):893–914