Supplementary information for "Intercomparison of Ocean Reanalyses for Interannual Ocean Temperature Variability in the Tropical Atlantic"

Arthur Prigent[1*], Ingo Richter[2], Andrea Storto[3], Riccardo Farneti[1], Rodrigue Anicet Imbol Koungue[4], Yuko M. Okumura[5], Ping Chang[6,7]

[1*]Earth System Physics, The Abdus Salam International Centre for Theoretical Physics, Trieste, Italy.
[2]Application Laboratory, Research Institute for Value-Added-Information Generation, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan.
[3]Institute of Marine Sciences (ISMAR), National Research Council (CNR), Rome, Italy.
[4]Physical Oceanography, GEOMAR Helmholtz Centre for Ocean Research Kiel, Kiel, Germany.
[5]Institute for Geophysics, Jackson School of Geosciences, The University of Texas at Austin, Austin, Texas, USA.
[6]Department of Oceanography, Texas A&M University, College Station, Texas, USA.
[7]Department of Atmospheric Sciences, Texas A&M University, College Station, Texas, USA.

*Corresponding author(s). E-mail(s): aprigent@ictp.it;

The content of this file is:

- Supplementary Texts S1 and S2
- Supplementary Table S1
- Supplementary Figures S1 to S4

Supplementary Text S1: Presentation of the different ocean reanalysis products

GODAS

The monthly mean outputs of the Global Ocean Data Assimilation System (GODAS; **?**) produced by the National Centers for Environmental Prediction (NCEP) are available from January 1980 to present day at a horizontal resolution of 1/3° latitude × 1° longitude and with 40 vertical levels. The ocean model used is the Geophysical Fluid Dynamics Laboratory (GFDL) Modular Ocean Model version 3 (MOM3; **?**). The ocean model in GODAS is forced weekly by heat fluxes, surface buoyancy fluxes and wind stress vectors from the NCEP reanalysis version 2 (NCEP-R2; **?**). The ocean data assimilation system uses the 3D variational technique of **?**. The GODAS SST is relaxed to the Reynolds SST (**?**) with a time scale of 5 days. Similarly, the sea surface salinity (SSS) is relaxed to Levitus monthly climatological SSS fields (**?**) with a time scale of 10 days. The subsurface temperature data that were assimilated were obtained from expendable bathythermographs (XBTs), the tropical atmosphere–ocean (TAO) array of moored buoys, Argo and Argo-like floats.

ORA-S4

The monthly mean outputs of the Ocean Reanalysis System version 4 (ORA-S4; **?**) produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) are available from January 1958 to December 2017 at a horizontal resolution of 1° × 1°, refined to 0.3° × 0.3° at the equator, and with 42 vertical levels each with a thickness of about 10 to 15 meters in the upper 200 meters. The ocean model used is the Nucleus for European Modelling of the Ocean (NEMO) model version 3.0 (**?**). The ocean model in ORA-S4 is forced by daily fluxes of solar radiation, total heat flux, evaporation-minus-precipitation and surface wind stress taken from the ERA-40 reanalysis (**?**) from

September 1957 to December 1989, ERA-Interim reanalysis (**?**) from January 1989 to December 2009, and the ECMWF operational archive from January 2010 onwards. The ocean data assimilation system is NEMOVAR in its 3D-var FGAT mode, using a 10-day assimilation window and a bias correction scheme to address model and forcing errors (**?**). The temperature and salinity profiles assimilated are coming from EN3 v2a (**?**) XBT bias-corrected database (**?**) for the period 1958-2009, incorporating various sources such as XBT, CTD, Argo, and mooring data.

ORA-S5

The monthly mean outputs of the Ocean Reanalysis System version 5 (ORA-S5; **?**) produced by the European Centre for Medium-Range Weather Forecasts (ECMWF) are available from January 1958 to present day at a horizontal resolution of $1/4^\circ \times 1/4^\circ$ with 75 vertical levels. The ocean model used is NEMO version 3.4.1 (**?**) and is forced by ERA-40 before 1979, by ERA-Interim from 1979 to 2015, and by ECMWF NWP from 2015 onwards. The ocean data assimilation system is 3D-VAR FGAT with a 5 day window. The assimilated data are: HadISST2+OSTIA operational for SST and sea-ice concentration, EN4 with XBT and MBT correction from **?** for temperature and salinity; and the AVISO DT2014 for sea level anomaly (**?**).

GECCO3

The monthly mean outputs of the German contribution of the Estimating the Circulation and Climate of the Ocean project version 3 (GECCO3; **?**) are available from January 1948 to December 2018 with a resolution of 0.4° and includes 40 vertical levels. The ocean model used in GECCO3 is the MITgcm ocean model (**?**) and is forced by the NCEP-R1 (**?**). The ocean data assimilation system is the adjoint method (4D-VAR) to adjust initial conditions and atmospheric states every 10 days, ensuring model consistency with various datasets. The assimilated data includes temperature

and salinity from the EN4.2.1 database (**?**) (corrected for XBT following **?**), AVISO along-track sea level anomalies, the DTU10 mean sea surface adjusted by the GOCO5s geoid for mean dynamic topography (**?**), HadISST sea surface temperature (**?**), and WOA18 climatology (**??**).

SODA3.3.2

The monthly mean outputs of the Simple Ocean Data Assimilation version 3.3.2 (SODA3.3.2; **?**) are available at $0.25° \times 0.25°$ horizontal resolution and with 50 vertical levels from January 1980 to December 2019. The ocean model used is the ocean–sea ice component of the GFDL CM2.5 coupled model (**?**) and is forced with the Modern-Era Retrospective Analysis for Research and Applications version 2 (MERRA-2; **?**). The data assimilation system is a linear deterministic sequential filter. The main datasets that SODA ingests are the World Ocean Database of historical hydrographic profiles (**?**) and in situ and remotely sensed SST from the International Comprehensive Ocean-Atmosphere Data set version 5 release 2 SST database (**?**) and from the L3 Pathfinder version 5.2 Advanced Very High Resolution Radiometer SST data (AVHRR; **?**).

CIGAR-CS

The monthly mean outputs of the CNR-ISMAR Global historical reanalysis (CIGAR-CS; **?**) are available at 1° horizontal resolution and 75 vertical levels, with augmented meridional resolution in the Tropics (1/3°), from January 1961 to December 2022. The ocean model used is the NEMO ocean model version 4.0.7 (**?**) and is forced by the ECMWF atmospheric reanalysis version 5 (ERA5; **?**). The assimilation scheme used is 3D-VAR, and the data assimilated are in-situ temperature and salinity profiles from the EN4 database (**?**), along with SST data from the UKMO HadISST (**?**) and the JMA COBE analysis (**?**). In addition to the variational assimilation scheme, a large-scale model bias correction (LSMBC) scheme is applied (**?**).

GLORYS12

The monthly mean outputs of the Global Ocean Physics Reanalysis at 1/12° horizontal resolution (GLORYS12; **?**) from the Copernicus Marine Environment Monitoring Service (CMEMS) are available with 50 vertical levels (22 levels are within the first 100 m) from December 1991 onwards. The ocean model used is NEMO (**?**) and is forced by the ECMWF ERA-interm reanalysis (**?**). The data assimilation system is a reduced-order Kalman filter derived from a singular evolutive extended Kalman filter (**?**) with a three-dimensional multivariate background error covariance matrix and a 7 day assimilation cycle (**?**). The data assimilated are the SLA from CMEMS (**?**), the AVHRR SST from NOAA, Ifremer/CERSAT sea ice concentration, and in situ temperature and salinity vertical profiles from CMEMS quality-controlled CORA database (**?**).

Supplementary Text S2: Definition of the ensemble mean and spread

Following the approach of **?**, the ensemble mean (EM) and spread (ES) of the ORAs are defined by:

$$EM(x,y,z,t) = \frac{1}{N} \times \sum_{k=1}^{k=N} X_k(x,y,z,t) \tag{1}$$

$$ES(x,y,z,t) = \sqrt{\frac{1}{N-1} \times \sum_{k=1}^{k=N} [X_k(x,y,z,t) - EM(x,y,z,t)]^2} \tag{2}$$

where $X_k(x,y,z,t)$ denotes the ocean temperature anomaly for an ocean reanalysis k, and N is the total number of ORA, here $N = 7$. ES is a measurement of the degree of disagreement among the different ORAs. ES can be vertically integrated in the upper

150 m by taking the average ES between 6 m and 156 m (ES_{150}) defined by:

$$ES_{150}(x,y,t) = \sqrt{\frac{1}{NZ} \times \sum_{z=1}^{z=NZ} [ES(x,y,z,t)]^2} \tag{3}$$

where NZ is the total number of vertical levels in the upper 150 m. Here, NZ = 15 since all data were interpolated to a common vertical grid with 10 m increments from 6 m down to 156 m. The spatial and temporal averages of ES(x,y,z,t) in the upper 150 m, ES_{150}(t) and $\overline{ES_{150}}$(x,y), are computed using:

$$ES_{150}(t) = \sqrt{\frac{1}{NX \times NY} \times \sum_{x=1}^{x=NX} \sum_{y=1}^{y=NY} [ES_{150}(x,y,t)]^2} \tag{4}$$

$$\overline{ES_{150}}(x,y) = \sqrt{\frac{1}{NT} \times \sum_{t=1}^{t=NT} [ES_{150}(x,y,t)]^2} \tag{5}$$

where NT is the total number of months over the period and NX and NY are the number of grid points in the x and y coordinates. Finally, the signal-to-noise ratio is defined by:

$$SNR(x,y,z,t) = \frac{|EM(x,y,z,t)|}{ES(x,y,z,t)} \tag{6}$$

Depth (m)	35°W	23°W	10°W	0°E
1	82.1%	89.2%	78.3%	74.2%
5	-	**36.7%**	**8.8%**	-
10	**0.8%**	**57.5%**	**8.8%**	-
13	-	**19.6%**	-	-
20	70.4%	77.1%	75.8%	**67.1%**
23	-	**3.8%**	-	-
40	74.2%	88.3%	78.3%	75%
60	82.5%	79.2%	73.8%	75%
80	81.2%	85%	77.1%	75.4%
100	84.6%	89.2%	77.9%	60.8%
120	87.9%	82.5%	75%	70%
140	93.8%	88.3%	77.9%	74.6%
180	92.9%	91.2%	77.5%	75.4%

Table S1 Percentage of available monthly mean temperature data from January 1998 to December 2017 at the equatorial Atlantic PIRATA mooring sites located at 35°W, 23°W, 10°W, and 0°E. Depths with data coverage below 70% are highlighted in bold.

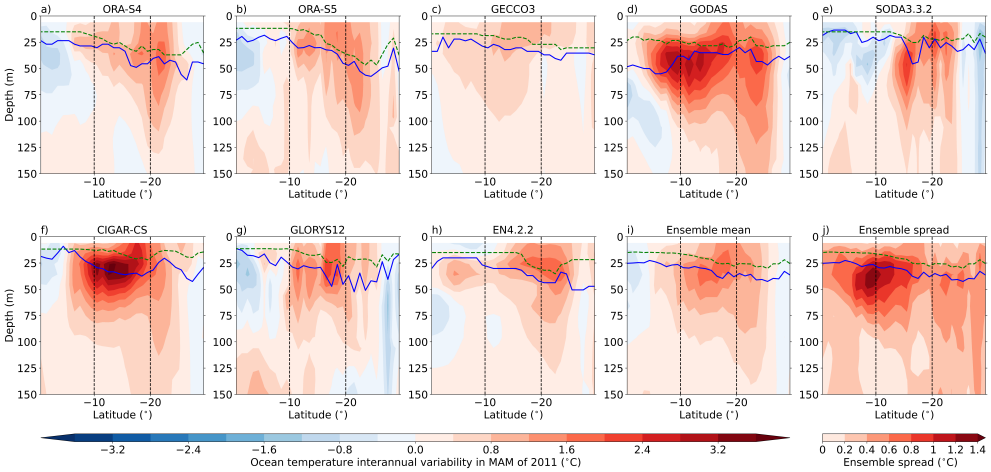

Fig. S1 Southwest African coast (2°S-coast, 30°S-0°N) ocean temperature anomalies in MAM of the year 2011 for (a) ORA-S4, (b) ORA-S5, (c) GECCO3, (d) GODAS, (e) SODA3.3.2, (f) CIGAR-CS, (g) GLORYS12, and (h) EN4.2.2, respectively. (i, j) Ensemble mean and spread of the ocean reanalyses. The ocean temperature anomalies were evaluated over the period from January 1993 to December 2017. Blue (green dashed) lines represent the dT/dz_{max} (MLD) during MAM of the year 2011. The ensemble means of the dT/dz_{max} and MLD are used for (i) and (j).

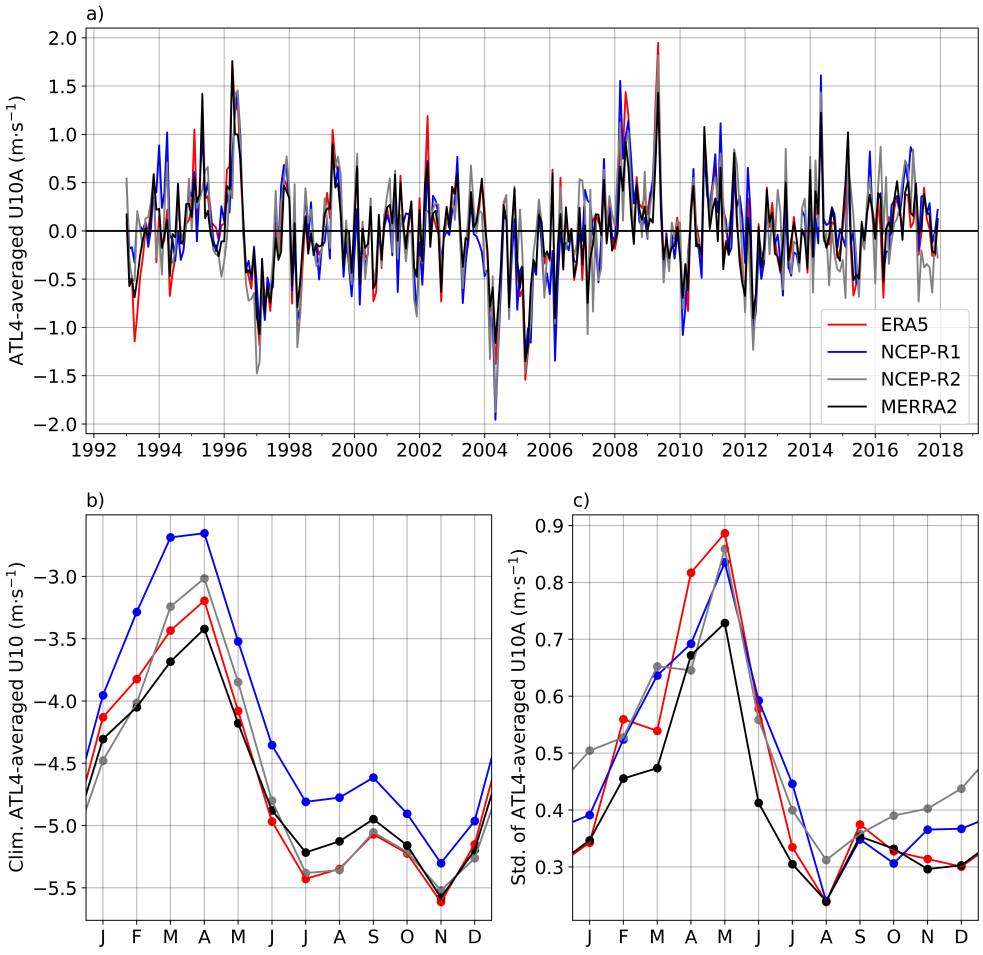

Fig. S2 (a) Timeseries of the ATL4-averaged (40°W-20°W, 3°S-3°N) U10A. (b) Monthly climatology of the ATL4-averaged U10. (c) Monthly climatology of the standard deviation of the ATL4-averaged U10A. Red, blue, grey, and black lines correspond to ERA5, NCEP-R1, NCEP-R2, and MERRA2, respectively.

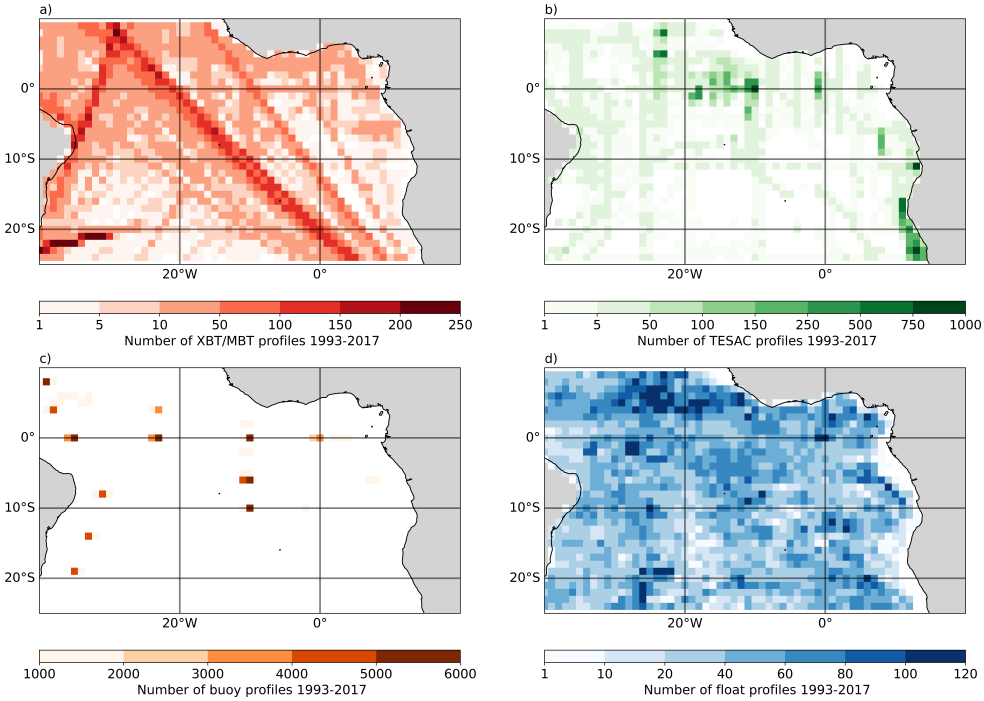

Fig. S3 Spatial distribution of quality-controlled temperature profiles in EN4.2.2. (a) Cumulative number of XBT/MBT temperature profiles from January 1993 to December 2017. (b) Same as (a), but for TESAC profiles. (c) Same as (a), but for buoy temperature profiles. (d) Same as (a), but for profiling floats.

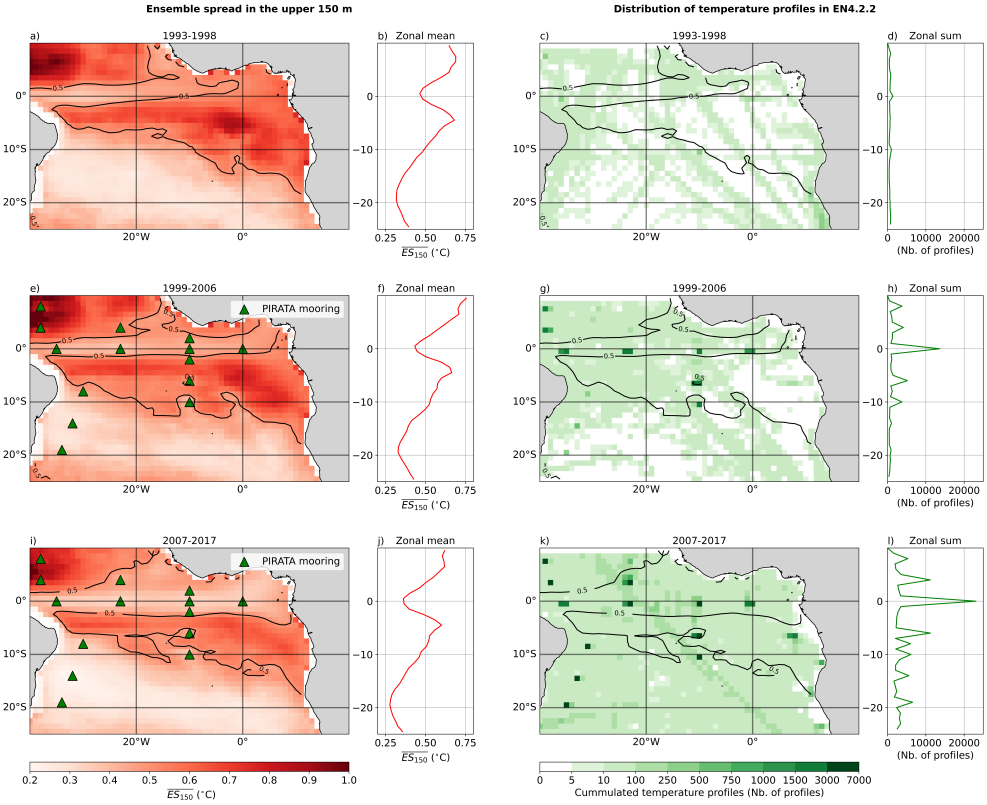

Fig. S4 Same as Figures 11a-d but for different sub-periods: (a-d) from January 1993 to December 1998 dominated by XBT/MBT measurement; (e-h) from January 1999 to December 2006 when the PIRATA array is fully deployed; (i-l) from January 2007 to December 2017 when the proportion of profiling floats becomes large.