CrossMark

Quantifying the effects of observational constraints and uncertainty in atmospheric forcing on historical ocean reanalyses

Chunxue Yang[1] · Andrea Storto[2] · Simona Masina[2,3]

Abstract

Historical ocean reanalyses combine ocean general circulation models with data assimilation schemes that ingest rescued observations. They can be used as a tool to investigate long-term changes in the ocean climate. However, large uncertainties, due to the poorly developed atmospheric and oceanic observing networks in early periods, still remain. Thus, detailed studies to assess the uncertainty and its time dependency are required to quantify the feasibility of historical ocean reanalyses for climate change assessment. In this work, we estimate the ocean heat content variability from a set of ocean reanalyses that cover the period 1900–2010. The ocean reanalyses include realizations forced by two different atmospheric reanalyses (20CRv2 and ERA-20C), combined with different data assimilation strategies, in the attempt to evaluate the relative weight of the atmospheric forcing and observation uncertainties on the resulting ocean heat content estimates. Results suggest that even when observing networks are poor, the observations are able to shape the upper ocean heat content variability, in terms of long-term trends and reproduction of individual warming/cooling events related to volcanic eruptions. The assimilation of in-situ profiles has an effect even on the sea surface temperature variability and is able to constrain the top 700 m heat content since the 1950s with respect to the uncertainty borne by the atmospheric forcing. The vertical propagation of the upper ocean observational information is however slow (with typically decadal time scale). Consequently, the total column heat content is constrained by observations only in the latest two decades. We conclude that upper ocean heat content diagnostics from historical ocean reanalyses bear the climate change signature and may be considered for long-term studies when complemented by proper uncertainty estimation.

1 Introduction

Atmospheric and oceanic reanalyses have been widely used in many research fields, due to the spatial and temporal consistence of their four-dimensional state reconstruction. However, spatial and temporal inhomogeneities may arise mainly because of changes in the observing networks (e.g. Sterl 2004; Masina and Storto 2017) and most of the reanalyses still cover only the last few decades because the lack of observations in early periods limits their ability to constrain the oceanic or atmospheric state. Recently, long-term historical reanalyses became feasible owing to several programs to rescue historical atmospheric and oceanic observations (e.g. International Surface Pressure Databank, ISPD, Cram et al. 2015; International Comprehensive Ocean–Atmosphere Data Set, ICOADS; Freeman et al. 2016). Increasing consensus emerges also on the necessity of sustaining data rescue efforts through permanently established international collaborations (Thorne et al. 2017) because of the potentially enormous value that historical observations represent for all climatic applications (Griffin 2015).

The first historical atmospheric reanalysis with duration over one century was performed at NOAA/CIRES, the 20 Century Reanalysis (20CRv2, Compo et al. 2011). It consists of an ensemble-based data assimilation system that assimilates surface pressure observations from ISPD. The 20CRv2 provides the ocean community with the opportunity to perform historical ocean reanalyses forced by it and ingest rescued oceanic observations. For example, SODAsi.2 (Giese et al. 2016) and the CMCC historical ocean reanalyses (Yang

✉ Chunxue Yang
  c.yang@isac.cnr.it

[1] Institute of Atmospheric Sciences and Climate, National Research Council (CNR-ISAC), via del Fosso del Cavaliere, 100, 00133 Rome, Italy

[2] Fondazione Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Bologna, Italy

[3] Istituto Nazionale di Geofisica e Vulcanologia (INGV), Sezione di Bologna, Bologna, Italy

et al. 2017) are forced by the 20CRv2 reanalysis, although they differ for the use of ocean observations and data assimilation scheme. Penny et al. (2015) also used the 20CRv2 ensemble perturbations to force an ocean reanalysis ensemble that covers recent years. Preliminary investigations had indeed shown that even a relatively poor observing network is able to capture, to some extent, the ocean variability in historical reanalyses (Smith and Murphy 2007; Carton et al. 2012). The latter studies investigated the impact of present-time observations sub-sampled to mimic last century scenarios. In particular, Carton et al. (2012) found that only after the 1960s the observing system was sufficiently developed to resolve oceanic variability.

Long historical reanalyses allow investigating climate signals over centennial time scales, as emerging from a number of studies focusing e.g. on the warming in the western boundary currents (Wu et al. 2012), ENSO oscillations (Yang and Giese 2013) and changes in large-scale current systems (Chen and Wu 2012; Yi et al. 2015). However, the uncertainty of the atmospheric reanalysis, along with poorly observing networks, brings into question the reliability of the decadal ocean variability shown in reanalyses, especially in the early 1900.

Recently, another historical atmospheric reanalysis was produced by ECMWF called ERA-20C (Poli et al. 2016). It is based on the ECMWF atmospheric model and includes 4DVAR data assimilation of surface observations. ERA-20C has a higher horizontal and vertical resolution with respect to 20CRv2, along with several differences in the atmospheric model and data assimilation formulations. Results from the ECMWF 10-member twentieth Century Ocean reanalyses (ORA-20C) forced by the ERA-20C deterministic atmospheric reanalysis have also recently been published (de Boisseson et al. 2017) and stress the importance of surface fluxes and ocean model drift for the ocean heat content variability during the first half of the twentieth Century. The first century long coupled reanalysis conducted in ECMWF is available recently, and the validation of the dataset shows improvements of the representation of several parameters, such as ocean atmosphere heat flux and mean sea level pressure (Laloyaux et al. 2018).

The ERA-20C and 20CRv2 historical reanalyses rely on the approach of assimilating only surface atmospheric observations. In particular, they both assimilate surface pressure observations; ERA-20C additionally assimilates also marine surface wind observations (Poli et al. 2013). This strategy was shown to be a good compromise for realistically reproducing past large-scale weather events and regimes while limiting the influence of the change with time of observational sampling on key climate signals (Compo et al. 2006). On the other hand, neglecting the information coming from upper-air observations and/or remotely sensed data leads to skill scores that are detrimental with respect to the atmospheric reanalyses with full observing network assimilated. See Poli et al. (2013), for a comparison of skill scores between ERA-Interim and ERA-20C and Hersbach et al. (2017) for an evaluation of the impact of upper-air observations in historical reanalyses. Nevertheless, the use of surface observations only does not prevent the reanalyses to suffer from inhomogeneity associated with changes in the observation sampling, the introduction of new stations or suppression of old ones, and intermittent quality. This was identified for instance by Ferguson and Villarini (2014) for 20CRv2, while for ERA-20C de Boisseson et al. (2017) and Laloyaux et al. (2018) showed the spurious signals in the Northern Hemisphere wind speed due to marine wind data assimilation, and the inhomogeneity linked to the observations treatment and pre-processing in poorly sampled periods, respectively. For ocean applications, the fact that air-sea freshwater and heat fluxes are largely determined by unobserved atmospheric parameters that are mostly calculated by model parameterizations—such as precipitation and radiative fluxes—with generally low confidence (Kalnay et al. 1996) complicates further the problem of historical ocean reanalyses because of the uncertainties of the air-sea fluxes (Kato et al. 2013; Valdivieso et al. 2015).

So far, the response of an historical ocean reanalysis to different atmospheric reanalyses forcing for the entire twentieth Century has been little explored: comparing the reanalyzed ocean state forced by different atmospheric historical reanalyses provides a unique possibility to assess the sensitivity of the reconstructed state of the ocean to the uncertainty of the reconstructed state of the atmosphere. Storto et al. (2016) showed that ocean temperature observations are able to constrain the global ocean heat content over the latest three decades with respect to the uncertainty of the atmospheric forcing, quantified by an ensemble of state-of-the-art atmospheric reanalyses. Furthermore, methods for MBT/XBT bias corrections are identified as the largest contributor to the global ocean heat content uncertainty before the deployment of Argo floats (Boyer et al. 2016; Abraham et al. 2013; Ishii and Kimoto 2009).

In this paper, we focus the investigation on the ability of the historical reanalyses to reproduce the long-term ocean heat content (OHC) and the impact of the atmospheric forcing uncertainty on its performance. OHC is a fundamental climate index that sheds light on climate change because the oceans store the large part of the excess Earth's energy (Trenberth et al. 2014). Moreover, as seawater temperature observations are among the longest lasting ocean observing networks, reconstructed OHC has the chance to feature better temporal homogeneity compared to other key climate variables (e.g. transports, freshwater).

As the reanalysis community is rapidly evolving towards historical Earth System reanalyses, such as coupled air-sea-land reconstructions (see e.g. Laloyaux et al. 2016), the

uncertainty assessment is a fundamental task that should be pursued. With this objective, long-term reanalyses differing by the assimilation strategy and atmospheric forcing are presented in this work. Comparisons are mostly conducted in terms of root mean square differences between the experiments that share the same assimilation strategy but are forced by different atmospheric forcing. Even though spread is generally used within an ensemble system, here we denote as spread the dispersion arising from the use of two different atmospheric forcing. In addition, the comparison between control runs without assimilation and reanalyses themselves indicates the impact of the ocean observation networks.

The structure of the paper is as follows: Sect. 2 introduces the atmospheric forcing and briefly compares the two atmospheric reanalyses. Section 3 describes the ocean reanalysis system, observational network and the experimental design. Section 4 presents the main results obtained through the evaluation of the ocean reanalyses. Section 5 concludes and discusses the main achievements.

2 Atmospheric forcing

In this section we present the atmospheric reanalysis datasets used to force the historical ocean reanalyses that are analyzed in this work. In particular, long and short wave radiations, specific humidity and air temperature at 2 m above sea level (q2m, t2m), total and snow precipitation fluxes, and wind vectors at 10 m above sea level (U10m and V10m) from the atmospheric reanalyses are used to calculate heat, water and momentum fluxes based on the CORE bulk formula of Large and Yeager (2004). Freshwater and radiative fluxes are extracted from the atmospheric reanalyses as daily means, while t2m, q2m and the 10 m wind vector are taken as 3-hourly snapshots. Daily mean shortwave radiative fluxes are modulated using the analytical diurnal cycle function from Bernie et al. (2007). The details of the historical atmospheric reanalyses are introduced and compared in the following subsections.

2.1 NOAA 20CRv2

The first atmospheric forcing comes from the ensemble mean of the 20 Century reanalysis version 2 (20CRv2, Compo et al. 2011). The model used in the 20CRv2 is the coupled atmosphere-land model based on the NCEP Global Forecast System (GFS) with a horizontal resolution of T62 (around 210 km) and a vertical resolution of 28 hybrid sigma-pressure levels. The boundary conditions of the 20CRv2 come from the UK Met Office HadISST 1.1 (Rayner et al. 2003) for sea surface temperature and sea ice concentration. The assimilation scheme for the 20CRv2 is the Ensemble Kalman Filter with a 6 h time-window, implemented following the

deterministic approach of Whitaker and Hamill (2002) in order to avoid the spurious covariance sampling that arises when using the perturbed observation method with intermittent observing networks. The ensemble size is composed of 56 ensemble members. 20CRv2 assimilates only surface pressure observations from the International Surface Pressure Databank (ISPD) version 2 (Cram et al. 2015). Quality control procedures are conducted on ISPD observations before being assimilated into the model (Compo et al. 2011).

Before we apply the forcing into the reanalysis system, we compared the 20CRv2 reanalysis data to ERA-Interim (Dee et al. 2011) for the overlap period from 1979 to 2010. ERA-Interim is here considered as a reference for the forcing due to the assimilation of all conventional and satellite observations. Due to large warm biases of the 20CRv2 with respect to ERA-Interim for air temperature at 2 m, long wave and short wave radiations at high latitudes and due to problems in the treatment of sea-ice boundary conditions (Compo et al. 2011; Lindsay et al. 2014), corrections for all the variables (for consistence) used in the bulk formula are applied at high latitudes (50°S to 80°S, and 55°N to 90°N). In particular, the large-scale (low-pass filtered) long-term monthly mean difference between ERA-Interim and 20CRv2, calculated during the overlap period (1979–2010), is added to the inter-annually varying 20CRv2 forcing fields used to force the ocean reanalyses, in an attempt to keep the 20CRv2 inter-annual variability while correcting systematic errors. The inter-annual variability at high latitudes of most parameters seems indeed well captured by 20CRv2 (Brönnimann et al. 2012; Paek and Huang 2012), justifying the recourse to a climatological correction of the dataset.

In the 20CRv2 wind vector fields at 10 m and precipitation fields present stripes (Yang et al. 2017 and; Kent et al. 2013), due to the discontinuities in the spectral transformation of the data and corrections are applied in order to filter them out. The details of the correction methods for all parameters are described in Yang et al. (2017).

2.2 ERA-20C

The second atmospheric forcing comes from ERA-20C, which is conducted at ECMWF (Poli et al. 2016). The atmospheric model of the ERA-20C is based on the ECMWF Integrated Forecast System (IFS), which includes an atmospheric general circulation model with a horizontal resolution of T159 (around 125 km) and 91 vertical levels, and a variational assimilation scheme (a 24 h incremental four-dimensional variational (4D-Var) scheme). ERA-20C is a deterministic reanalysis (i.e. a single-member realization without perturbations, see Poli et al. 2015), although background-error covariances are estimated from a previous ensemble historical reanalysis, and thus include a flow-dependent modulation of the background errors. No ensemble spread

information nor uncertainty estimate is provided with ERA-20C. The boundary conditions of the atmospheric model are from the UK Met Office HadISST version 2 (Titchner and Rayner 2014; Kennedy 2014) including both SST and sea ice. Compared to version 1 used by 20CRv2, HadISST version 2 contains an improved uncertainty budget and slightly higher sea-ice concentrations in both Hemispheres. The atmospheric composition and solar forcing are adapted from the dataset used in the CMIP5 simulations (Taylor et al. 2012) and detailed by Hersbach et al. (2015), who present results from the ERA-20CM integration, i.e. the model simulation counterpart of ERA-20C (without data assimilation). The observations assimilated into the ERA-20C not only include the surface pressure from the ISPD bank version 3.2.6, but also the surface marine wind observations from the ICOADS version 2.5.1 (Woodruff et al. 2011). A quality control procedure is conducted on the observations before being assimilated (Poli et al. 2013).

While 20CRv2 forcing fields were corrected at high latitudes because of known systematic biases over ice-covered regions (Compo et al. 2011; Lindsay et al. 2014), ERA-20C fields were not corrected. The other difference is that we use the ensemble mean of the 20CRv2 and the single member realization of the ERA-20C, implying that 20CRv2 atmospheric forcing fields might be smoother than ERA-20C.

2.3 Comparison between the 20CRv2 and ERA-20C

Differences and similarities between the 20CRv2 and the ERA-20C are listed on the webpage https://climatedataguide.ucar.edu/climate-data/era-20c-ecmwfs-atmospheric-reanalysis-20th-century-and-comparisons-noaas-20cr. Here we give some details of the differences between the corrected 20CRv2 and the ERA-20C only for the variables that we use in the bulk formula to calculate the air-sea fluxes, with the exception of the snow fall (solid precipitation), which impacts the ocean reanalyses only at high latitudes.

Climatological differences for the period from 1900 to 2010 of long and short-wave radiations, t2m and precipitation between corrected 20CRv2 and ERA-20C are presented in Fig. 1, while the seasonal cycle in the Northern and Southern Hemisphere for the atmospheric forcing parameters is shown in Fig. 2.

The maps show that on average the long-wave radiation is larger in the 20CRv2 than ERA-20C at global scale except along the west coast of America and Africa. On the other hand, the shortwave radiation in 20CRv2 is smaller at high latitudes (corrected areas), subtropical Pacific, Atlantic and Indian Oceans but larger in the rest of the global ocean, especially along the west coast of America and Africa. The behavior over the latter areas suggests that 20CRv2 exhibits smaller cloud cover than ERA-20C on the eastern upwelling areas. Note that with the application of the high-latitude

corrections, differences are mitigated but not eliminated; for instance, 20CRv2 is known to overestimate wintertime cloudiness and under-estimate summertime clouds, leading to much larger winter-time longwave and summer-time shortwave radiations at the surface (Zib et al. 2012). This symmetric behavior appears well in the seasonal cycle diagrams (Fig. 2). The two datasets, even after the high-latitude corrections performed over 20CRv2, still present significant differences in the incoming radiative fluxes year-round.

The 2 m temperature is warmer in the 20CRv2. Especially at high latitudes, despite the corrections, the tropospheric temperature bias in 20CRv2 (Lindsay et al. 2014) is clearly visible in the differences with ERA-20C. Temperature and humidity seasonal cycles (Fig. 2) are closely connected and present a rather constant offset between the two datasets, with 20CRv2 showing greater values than ERA-20C.

The main difference between precipitation fields is located in the tropical regions, and in particular 20CRv2 exhibits larger precipitation over the inter-tropical convergence zones (ITCZ). Larger differences occur in spring time in both hemispheres; during these seasons, 20CRv2 is found to overestimate precipitation at high-latitudes (Lindsay et al. 2014).

The differences between 10 m wind speed climatology in the 20CRv2 and ERA-20C are not globally uniform (Fig. 2e). The wind speed is stronger in the 20CRv2 than in the ERA-20C at high latitudes in the Southern Hemisphere and also subpolar region in the Northern Pacific and Atlantic Oceans, and weaker in the tropical oceans. Hemisphere-averaged wind speed differences are generally small (Fig. 2f). The two reanalyses are comparable in term of both intensity and seasonal cycle, except in the Southern Ocean during austral summer, where 20CRv2 has stronger winds in correspondence of the southernmost region in the ACC, uncertainties of the two datasets are large and may lead to poor agreement between the two reanalyses (Wang et al. 2016).

The area-averaged root mean square difference (RMSD) between 20CRv2 and ERA-20C for the global, the Tropical, the Northern and Southern Extra-tropical oceans for all the forcing parameters is given in Fig. 3, as a function of time.

The RMSD for long-wave radiation has a sharp decrease after the 1950s, especially in the Southern Extratropics, associated with the abrupt changes in the observing system. The spread increases during the 1960s and 1970s in the Northern Extra-tropics, consistent with the homogeneity tests performed by Ferguson and Villarini (2014). This latter result, to a lesser extent, also applies to T2M. The shortwave radiative fluxes are characterized by a RMSD peak in the Tropics and are consistent with the large variability; however, it abruptly decreases during the 1950s similarly to the long-wave radiation, suggesting a decrease of cloudiness inconsistency between the two reanalyses.
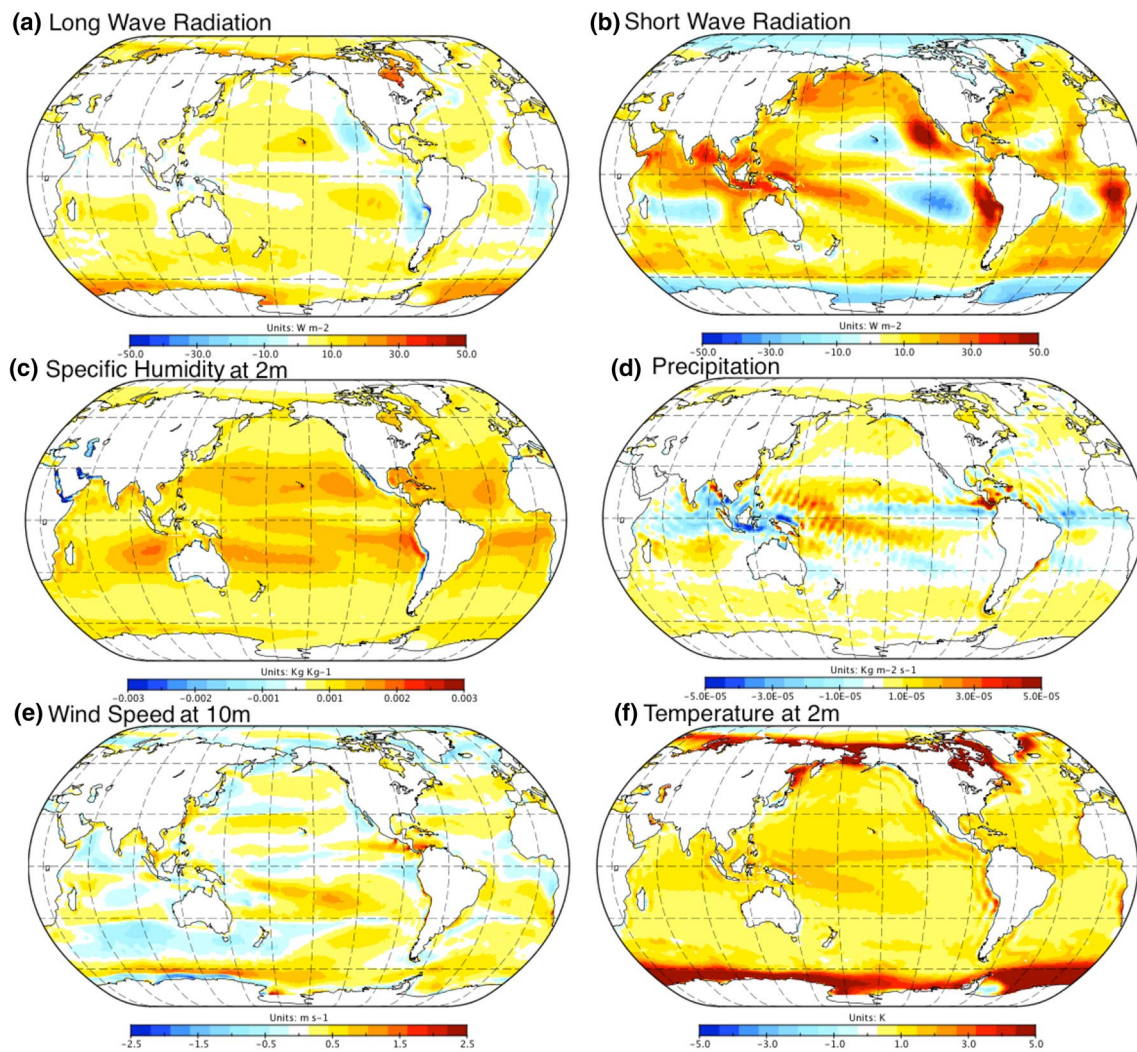
Fig. 1 Long-term mean difference (1900–2010) of the atmospheric forcing parameters from NOAA-20CRv2 minus ERA-20C

The precipitation RMSD remains rather stable along time, and exhibits the largest variability in the Tropical regions, indicating that changes in the observing network do not significantly affect the differences among the two reanalyses for what concerns the rain rate.

For both U10m and V10m the RMSD decreases along time quite constantly, except for spikes in proximity to the World Wars. The increase of observation numbers in the atmospheric reanalyses is able to constrain the marine winds and their accuracy increases along time leading to convergence.

We have also compared the RMSD of ERA20C vs 20CRv2 with the RMSD of any two members of 20CRv2 to verify that the uncertainty associated with 20CRv2 is not larger than that implied by the two atmospheric reanalysis systems (not shown). This analysis indicates that the 20CRv2 RMSD is significantly smaller than that between the two reanalyses (in most cases it counted for the 20% of

the RMSD between ERA-20C and 20CR), implying that we can reasonably focus on the ERA-20C and 20CR forcing data rather than considering also multiple realizations of 20CRv2.

To summarize, the two reanalysis datasets present complex differences that over different regions, seasons, or periods, may largely differ among the forcing parameters, making their effect on the ocean reanalyses difficult to infer. The sensitivity of the ocean state to the different atmospheric forcing on the whole is thus assessed in detail in Sect. 4.

3 Ocean reanalysis system

3.1 Configuration

The ocean model component of the ocean reanalyses is based on the Nucleus for European Modeling of the Ocean
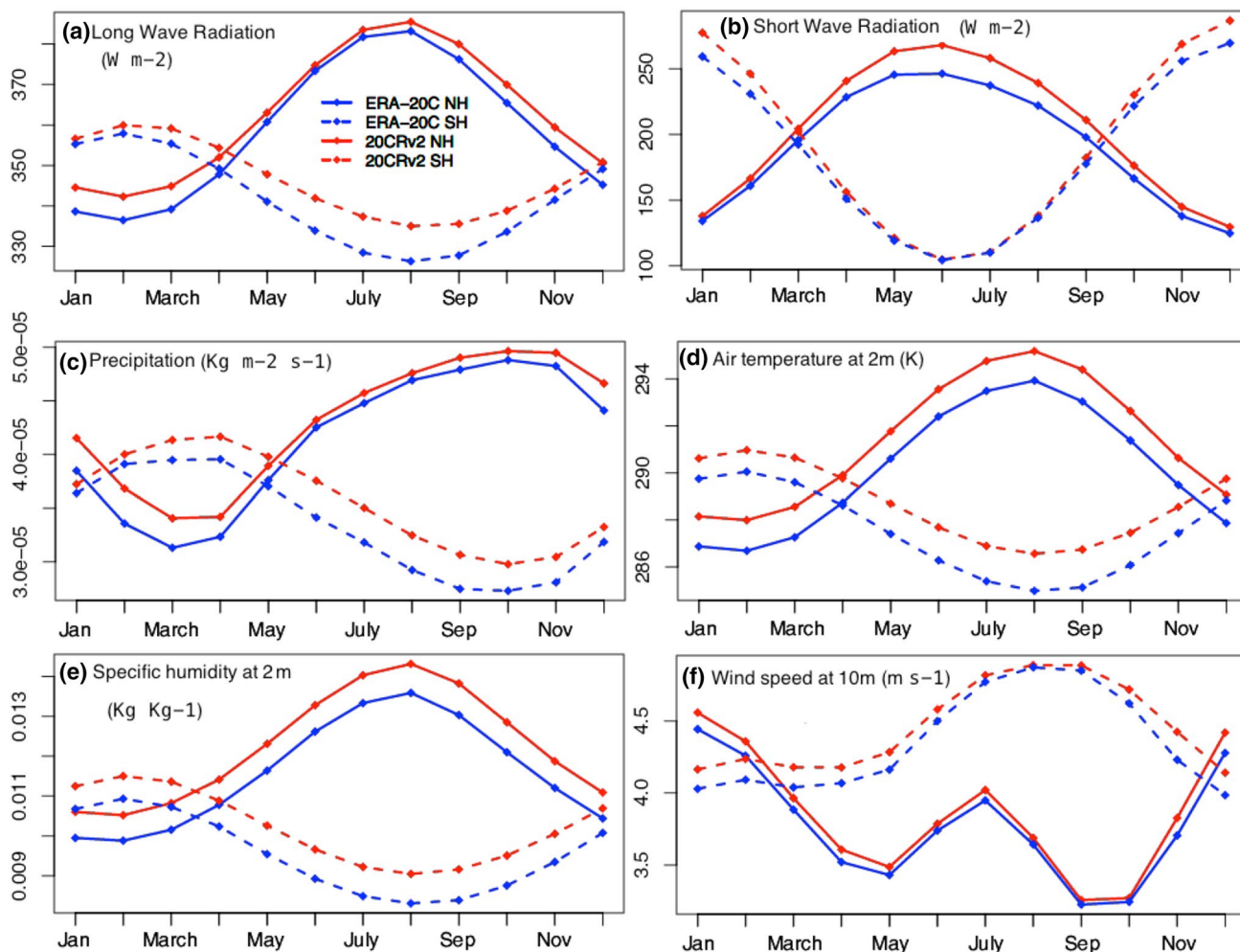
Fig. 2 Seasonal cycle of the atmospheric forcing parameters from NOAA-20C and ERA-20C (period 1900–2010)

(NEMO, Madec and NEMO Team 2012), version 3.4, coupled with the LIM2 sea-ice model (Vancoppenolle et al. 2009), with an approximate horizontal resolution of 0.5° and 75 vertical levels. The runoff is from the monthly climatological discharge provided by Bourdalle-Badie and Treguier (2006), including 99 major rivers and coastal runoffs. The details of the model physics are discussed in Yang et al. (2017). In all the experiments, the sea surface salinity fields are nudged to the monthly climatology of World Ocean Atlas (WOA) 98 (Boyer and Levitus 1998) with a 365-day relaxation time-scale, in order to constrain the surface freshwater at the annual time scale and sustain the deep convection at high latitudes (Behrens et al. 2013). The initial conditions for the experiments with different atmospheric forcing are the same and come from a 110 years spin up control experiment forced by 20CRv2 with relaxation to HadISST1 SST analyses. Details of the initial strategy and the spin up simulation performances are given by Yang et al. (2017). We decided to use the same initial conditions for all

experiments, in order to avoid that different model spin-up compromise the comparison; this is important especially when observations start to be available and inevitably pull the ocean state reproduced by the reanalyses. Furthermore, the adjustment of the experiments forced by ERA-20C is generally small compared to the centennial ocean heat content variability (see next section and Fig. 5 in particular).

We performed pairs of experiments, in which each of the two atmospheric forcing is combined either with a SST nudging scheme to monthly Met Office Hadley Centre's sea ice and sea surface temperature data set (HadISST1) data set (Rayner et al. 2003), or with assimilation of SST observations from an integrated database of ocean temperature and salinity (HadIOD) compiled by UK MetOffice (Atkinson et al. 2014) with the 3D-Var data assimilation scheme. Within SST nudging (or relaxation) experiments, the NEMO model corrects the net heat flux proportionally to the SST gridded innovations of HadISST analyses minus model fields, with a relaxation coefficient equal to

Fig. 3 RMSD between NOAA-20C and ERA-20C for the atmospheric forcing parameters over the period 1900–2010 for four different regions [global ocean, Southern Extratopics (90S–30S), Tropics (30S–30N), Northern Extratropics (30N–90N)]. The orange shaded areas correspond to the two World Wars periods
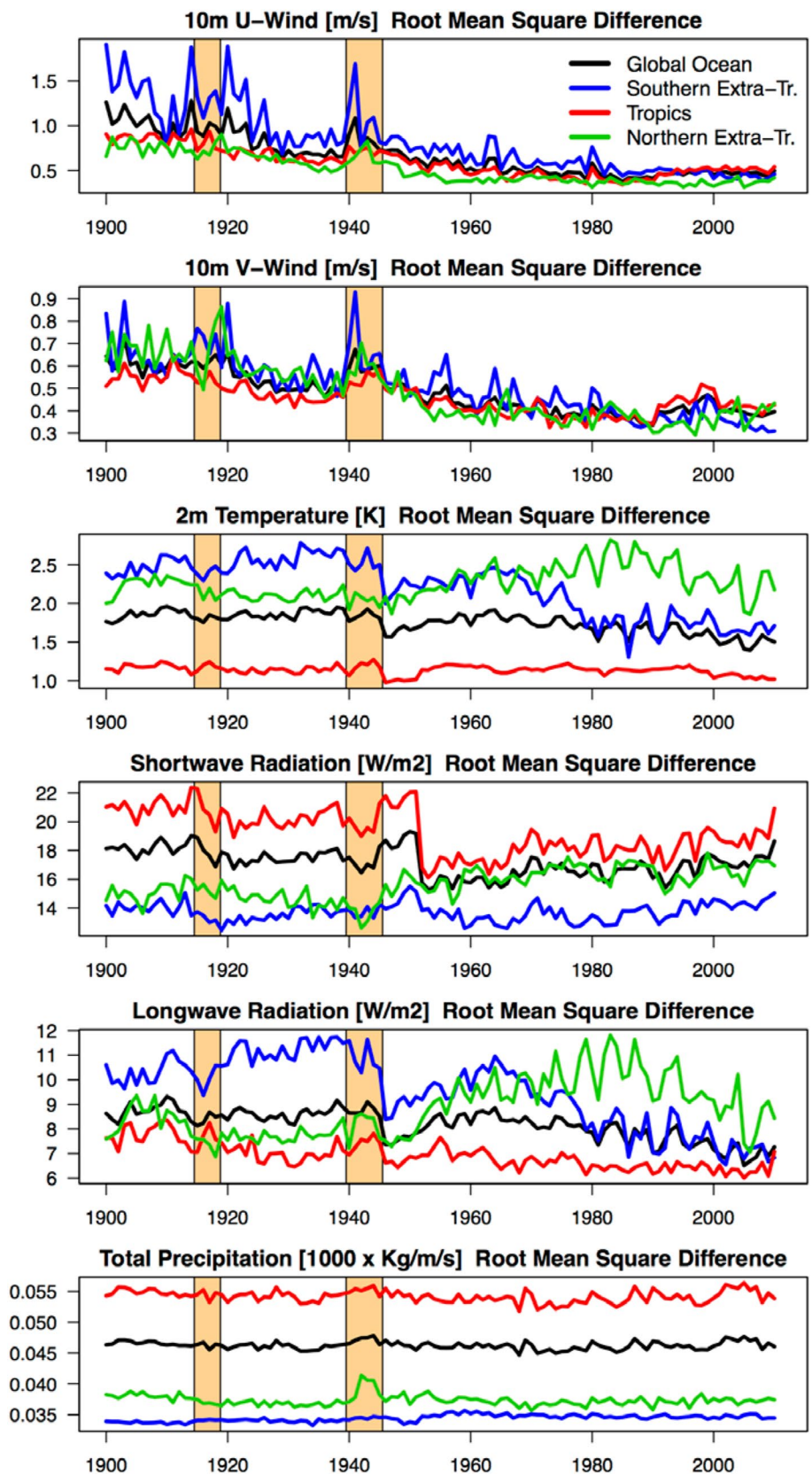
80 W/m^2/K, roughly equivalent to a monthly relaxation time scale (for a 50 m deep mixed layer), consistently with the temporal scale of the HadISST reconstructed fields of SST.

A 3D-Var data assimilation scheme is used in the ocean reanalyses and adapted from Storto et al. (2011, 2014), with a first guess at appropriate time (FGAT) implementation where observations are compared to background fields at observation time for the misfit calculation. The data assimilation step is performed every 7 days, with an assimilation time window of 7 days. A special treatment in the 3D-Var assimilation scheme is devoted to the temporal modulation of the background error covariances. A conventional way of treating background error covariances in variational schemes is to use stationary background errors, which is not suitable for our long-term historical ocean reanalyses. Therefore, in our reanalyses, a global multiplicative factor is used to allow the background error covariances to change with time. The multiplicative factor is spatially uniform, namely it only represents a globally uniform change of background-error covariances due to the changes in the global observing network, and it is estimated using the Desroziers' assimilation output diagnostics (Desroziers et al. 2005) applied to a preliminary version of the historical ocean reanalysis.

Subsurface data, including temperature and salinity profiles come from the Met Office Hadley Centre "EN" series of hydrographic profile collections, version 4 (EN4, version 4.1.0, Good et al. 2013). The main data source of EN4 is the World Ocean Database 2009 (WOD09, Boyer et al. 2009), with in-house quality control and bias correction performed at the UK Met Office. The method for the MBT and XBT bias adjustments is based on Gouretski and Reseghetti (2010). The assimilation of subsurface data is based on 3D-Var assimilation scheme introduced above.

Figure 4 shows the yearly number of assimilated observations from the SST network (top panel) and temperature and salinity hydrographic profiles as a function of depth (middle and bottom panel, respectively), in logarithmic scale. The figure indicates the paucity of profile observations below 200 m before 1915 and during the two World Wars, suggesting that the subsurface ocean is barely constrained by data assimilation before the 1950s. Similar behavior is seen for SST observations, which are however more numerous than the total of temperature depth observations till about the beginning of the 1960s. The number of assimilated observations dramatically drops during the two World Wars and rapidly increases afterwards, with a bottomward propagation of the observation number especially increasing at the beginning of the 1970s and later on with the Argo floats deployment during the mid 2000s. The number of temperature observations is larger with respect to salinity observations at almost all times, except in the last decade where the dominance of Argo floats makes them comparable.

3.2 Experiment design

As previously mentioned, we perform four historical ocean reanalyses that come from the combination of the two atmospheric forcing and the two SST assimilation schemes. Additionally, control experiments without data assimilation forced by ERA-20C or 20CRv2 are conducted in order to assess the impact of the atmospheric forcing on the model simulation, either with or without relaxation to HadISST SST fields. Through comparison with the reanalyses, the control experiments also allow appraising the impact of observation data assimilation.

Table 1 summarizes the 8 experiments analyzed in this work. The set of experiments is conceived as an ensemble of historical ocean reanalysis realizations with different atmospheric forcing and SST assimilation schemes, plus non-assimilative experiments to evaluate the impact of the atmospheric forcing and SST relaxation without in-situ observation assimilation. In Sect. 4.1, the ensemble mean of the reanalyses and that of control runs will be referred to as REA MEAN and CTR MEAN, respectively, including ASSIM_20CRv2, ASSIMALL_20CRv2, ASSIM_ERA-20C and ASSIMALL_ERA-20C the former and CTR_20CRv2, CTRSST_20CRv2, CTR_ERA-20C and CTRSST_ERA-20C the latter.

Table 1 sketches the four pair of experiments (with different atmospheric forcing but same assimilation configuration) that will be used in the next section to evaluate the uncertainty of the atmospheric forcing, in which CTR stands for RMSD between CTR_20CRv2 and CTR_ERA-20C; CTRSST stands for RMSD between CTRSST_20CRv2 and CTRSST_ERA-20C, ASSIMALL stands for RMSD between ASSIMALL_20CRv2 and ASSIMALL_ERA-20C; ASSIM stands for RMSD between ASSIM_20CRv2 and ASSIM_ERA-20C.

4 Results

4.1 Ocean heat content estimates

We first analyze the ocean heat content anomaly during the twentieth century reproduced by the ensemble of reanalyses. The two top panels of Fig. 5 show the top 700 m (left) and the total (right) ocean heat content anomaly for the period 1900–2010. Further to the four reanalyses and four control experiments (see Table 1), also the ensemble mean and standard deviation (shaded areas corresponding to ± one ensemble standard deviation) for the two sets of experiments, with or without in-situ profile data assimilation, are shown. In the top 700 m the ensemble of reanalyses shows a warming larger than that of control experiments, with a significant warming of 0.14 ± 0.04 W/m^2 against 0.07 ±

Fig. 4 Yearly number of assimilated SST (top panel), temperature (middle panel) and salinity (bottom panel) observations. For profile observations, the yearly number is plotted as a function of the depth level. The orange shaded areas correspond to the two World Wars periods in the top panel
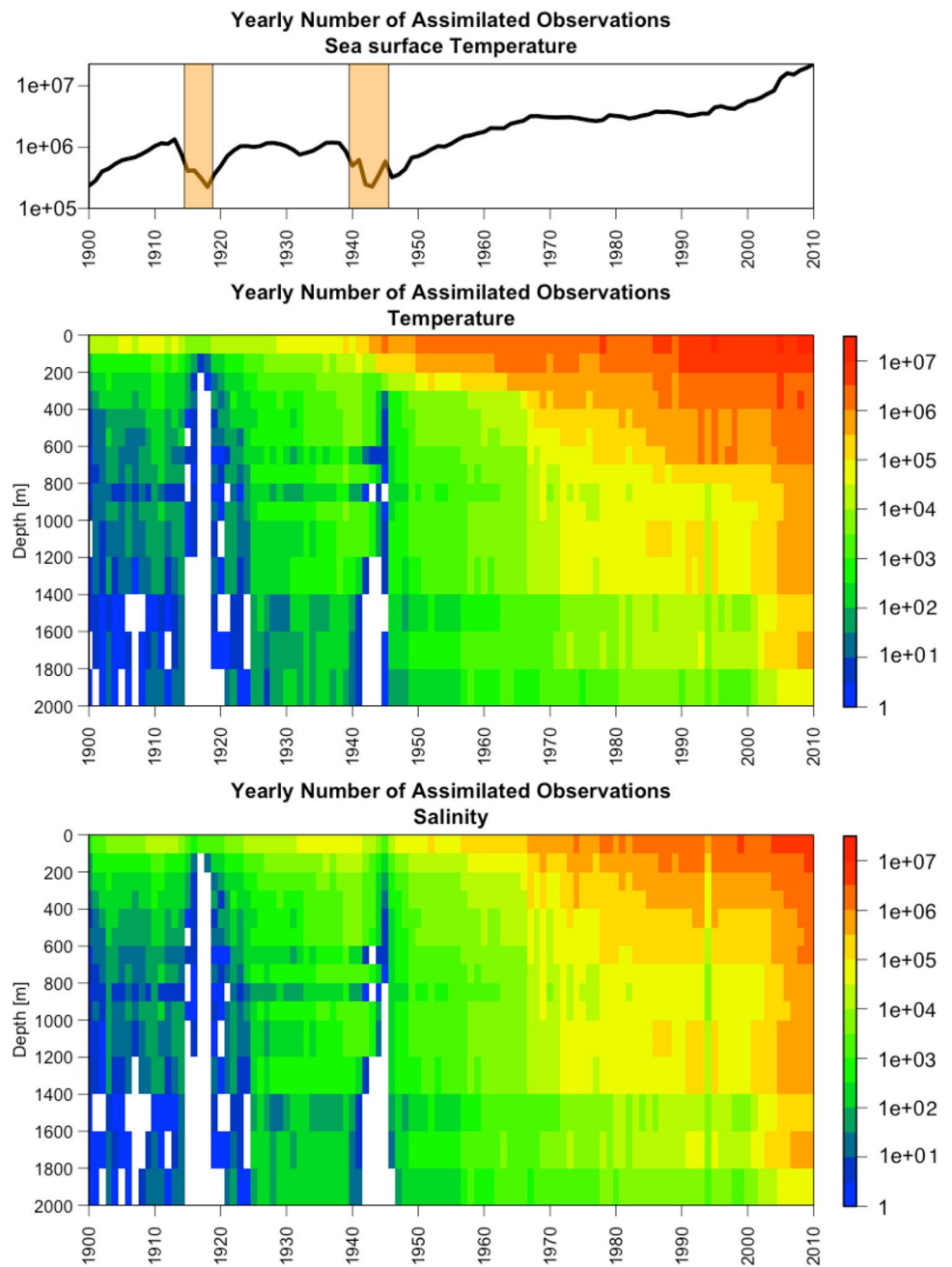

Table 1 List of reanalyses performed, with the name of experiment, period, forcing and assimilation configuration

| Experiment name	Comparison pair	Period	SST Assimilation	Vertical profiles assimilation	Forcing
CTR_20CRv2	CTR	1900–2010	N/A	N/A	20CRv2
CTR_ERA-20C		1900–2010	N/A	N/A	ERA-20C
CTRSST_20CRv2	CTRSST	1900–2010	Nudging	N/A	20CRv2
CTRSST_ERA-20C		1900–2010	Nudging	N/A	ERA-20C
ASSIM_20CRv2	ASSIM	1900–2010	Nudging	3DVAR	20CRv2
ASSIM_ERA-20C		1900–2010	Nudging	3DVAR	ERA-20C
ASSIMALL_20CRv2	ASSIMALL	1900–2010	3DVAR	3DVAR	20CRv2
ASSIMALL_ERA-20C		1900–2010	3DVAR	3DVAR	ERA-20C
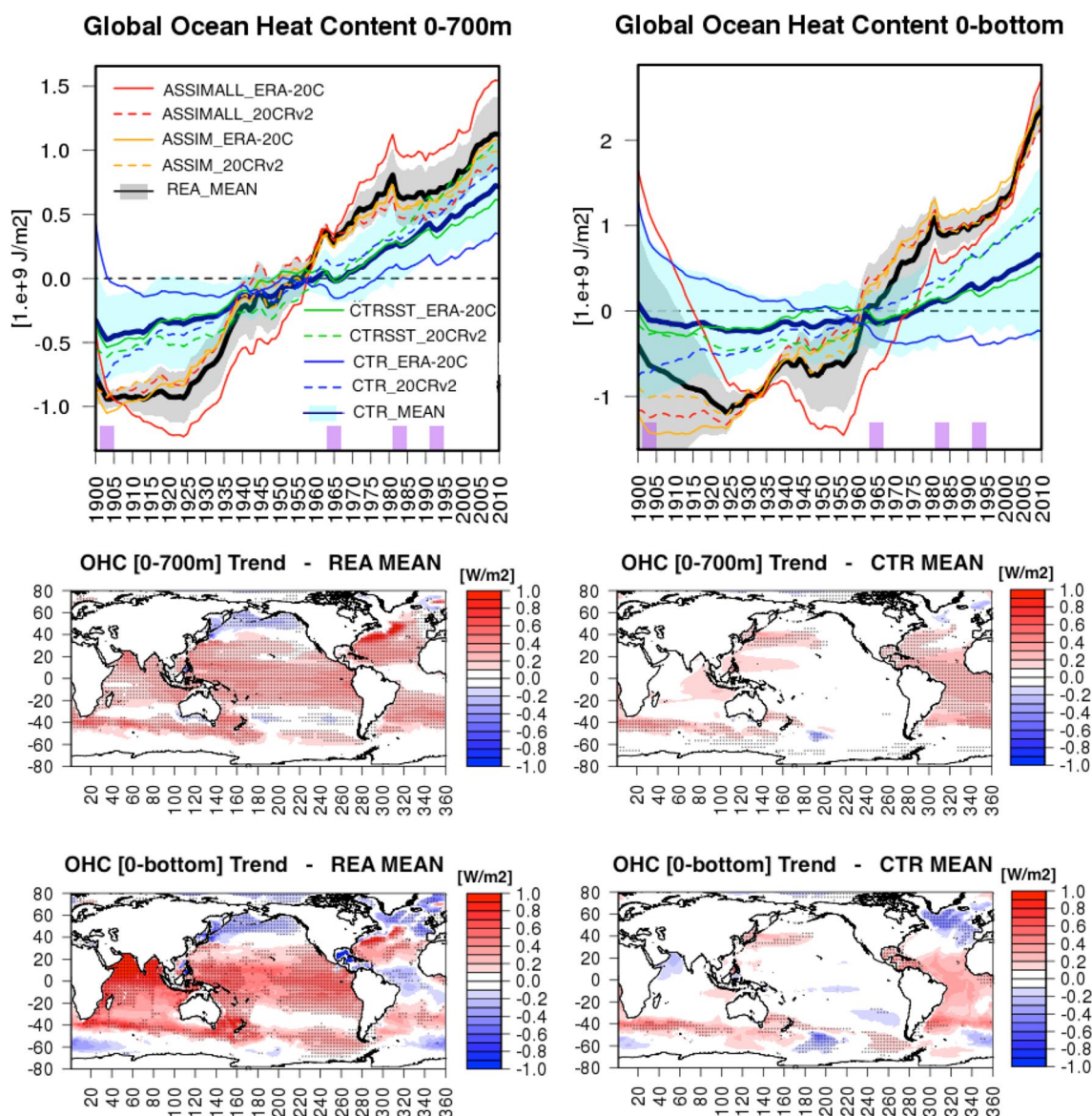
Fig. 5 Top panels: yearly means of global ocean heat content anomaly in the reanalysis period (1900–2010) in the top 700 m (left) and total column (right) from the four reanalyses and their ensemble mean (REA MEAN) and the four control experiments and their ensemble mean (CTR MEAN). Shaded areas correspond to ± 1 ensemble standard deviation. Volcano periods are added indicated with purple bars bottom panels: maps of linear trends (1900–2010) of ocean heat content (W/m^2) from REA MEAN and CTR MEAN in the top 700 m and the total column. Dots super-imposed to contours indicate that trends are significant

0.05 W/m^2 during the entire 111-year period. The temporal variability appears also to reasonably reproduce cooling events linked to volcanic eruptions (1902, 1963, 1982 and 1991, see e.g. Robock 2000) that are generally damped out in control experiments. The reanalyses ensemble exhibits a warming during the period 1925–1940, which does not seem related to the atmospheric forcing, as it consistently appears in all reanalyses and in sea surface temperature data (see also Yang et al. 2017, Fig. 9). In fact control experiments do not show such a quick warming, generally reproducing a fairly constant warming rate during the entire period; this

indicates that the warming is induced by in-situ data assimilation. ASSIMALL_ERA-20C shows the largest warming, which is however linked to the initial cooling (up to 1925). A similar behavior is shown in de Boisseson et al. (2017), probably due to non-optimal initial conditions, which in turn may cause some initial adjustments.

The total column ocean heat content anomaly generally amplifies these features, due to the lack of observations at depths deeper than 1000 m in the first half of the century and owing to the delayed vertical propagation of the upper ocean observations. The ensemble of reanalyses indicates

a warming of 0.19 ± 0.09 W/m^2 against 0.04 ± 0.08 W/m^2 for the ensemble of control experiments, which is a non-significant warming for the latter suite of experiments, visible as a fairly flat timeseries up to 1960s. The control experiments CTR_ERA-20C and CTR_20CRv2 show opposite trends, i.e. ERA-20C implying a negative downward net heat flux while 20CRv2 a positive net heat flux. The uncertainty (ensemble spread) is consequently large. Differences in heat content for the reanalyses experiments that include SST relaxation are generally small as expected, because the SST relaxation corrects the heat fluxes and prevents the heat content from the pair of experiments to diverge. In particular, we have calculated the globally averaged heat flux induced by the SST relaxation, which leads to a warming in the experiments forced by ERA-20C (10 W m^{-2}) and a large cooling (-24 W m^{-2}) in those forced by 20CRv2 during the first half of the century (1900–1950).

The bottom panels of Fig. 5 show maps of 1900–2010 ocean heat content trends for the two vertical regions (top 700 m and total column) and the two ensemble members of experiments, indicating also where the trends are significant. The picture that emerges in the assimilation experiments suggests in general a warming at low latitudes in the Indo-Pacific basin, while a warming at mid latitudes in the Atlantic Ocean, with few cooling areas at high latitudes in the North Hemisphere and a warming in correspondence of the Antarctic Circumpolar Current (ACC). All these features occur for both top 700 m and total column regions, the latter exhibiting greater trends, especially in the Indian Ocean, opposed to a cooling in the Pacific and Atlantic sub-polar gyres. Control experiments show largely different characteristics, except for the warming in the ACC that is still present: significant warming is located only in the Kuroshio extension area and in the Tropical Atlantic Ocean, while cooling in correspondence of the Atlantic sub-polar gyre appears as well. The different distribution of trends in the Atlantic Ocean (left vs right panels) can partially be induced by weaker overturning circulation in the control experiments (8–12 Sv, see also Yang et al. (2017) for a more detailed discussion). This leads in turn to stagnation of meridional heat transports and warming of the Tropical Atlantic, unlike the assimilation experiments where AMOC is sufficiently strong (above 15 Sv during the second half of the century).

## 4.2 Comparison of sea surface temperature long-term spread

In this Section, we analyze the spread and their difference between pairs of experiments performed with different atmospheric forcing but same assimilation configuration. Table 1 recalls the meaning of the acronyms used to identify the pairs of experiments.

Figure 6 shows the SST spread along time for the four pairs of experiments in the case of the global ocean and the three regions we divide the global ocean into Southern (90°S–30°S) and Northern (30°N–90°N) Extra-tropics and Tropics (30°S–30°N). In all regions, differences among the experiments with nudging (ASSIM and CTRSST) are small, around 0.5 °C, and rather stable over time, obviously due to the fact that the use of the relaxation acts as a strong constraint on the sea surface temperature. However, in the second half of the twentieth century in the Tropical and Northern Extra-tropical regions, the spread of ASSIM decreases with respect to CTRSST, meaning that the in-situ observation data assimilation complements the SST relaxation in reducing the sea surface uncertainty. A similar decrease occurs also in the Southern Extra-Tropics, but only after Argo floats deployment.

ASSIMALL has a smaller spread than CTR at all times, even at the beginning of the century and during the two World Wars, suggesting that even a poor observing network is able to reduce the spread linked to the atmospheric forcing. In particular, while during the first half of the twentieth Century the ASSIMALL spread is smaller than that of CTR by rather a constant in-phase offset in all regions, after the 1950s the decrease of the ASSIMALL spread is faster than CTR. Moreover, the spread of ASSIMALL becomes comparable to that of CTRSST since the 1970s for the Tropics and Northern Extra-tropics and since around year 2005 for the Southern Extra-tropics, indicating that the assimilation of in-situ observations provides the same level of uncertainty as the SST relaxation to gridded datasets. The spread decrease of the control experiments indicates that a convergence of the atmospheric forcing is likely due to the increase of atmospheric observation sampling.

The spatial map of the SST spread in CTR during the period 1900–1950 (Fig. 7a) indicates that in general the spread is dominated by the Southern Ocean, as confirmed by the values of the spread in Fig. 6. Peaks of spread are located in the Southern Ocean (along the Antarctic Circumpolar Current and in the South Pacific gyre), in the eastern boundary upwelling systems (off Benguela, Peru and, to a lesser extent, California) and in a few areas of the North Atlantic (subpolar gyre and Canadian archipelago) and North Pacific (Kuroshio area and Bering strait) oceans. Figure 7b shows the percentage reduction of spread in ASSIMALL compared to CTR (calculated as the difference between CTR and ASSIMALL divided by CTR) for the poorly observed period only (1900–1950). Spread decreases everywhere, with large values (more than 50%) in the North Atlantic Ocean and in correspondence of ship routes in the North Pacific and South Atlantic Ocean, and also in the Antarctic region, where possible amplifications occur due to different placements of sea-ice edges.

Fig. 6 Yearly SST RMSD between the four pairs of reanalyses (see Table 1) for four different regions [global ocean, Southern Extratropics (90S–30S), Tropics (30S–30N), Northern Extratropics (30N–90N)]
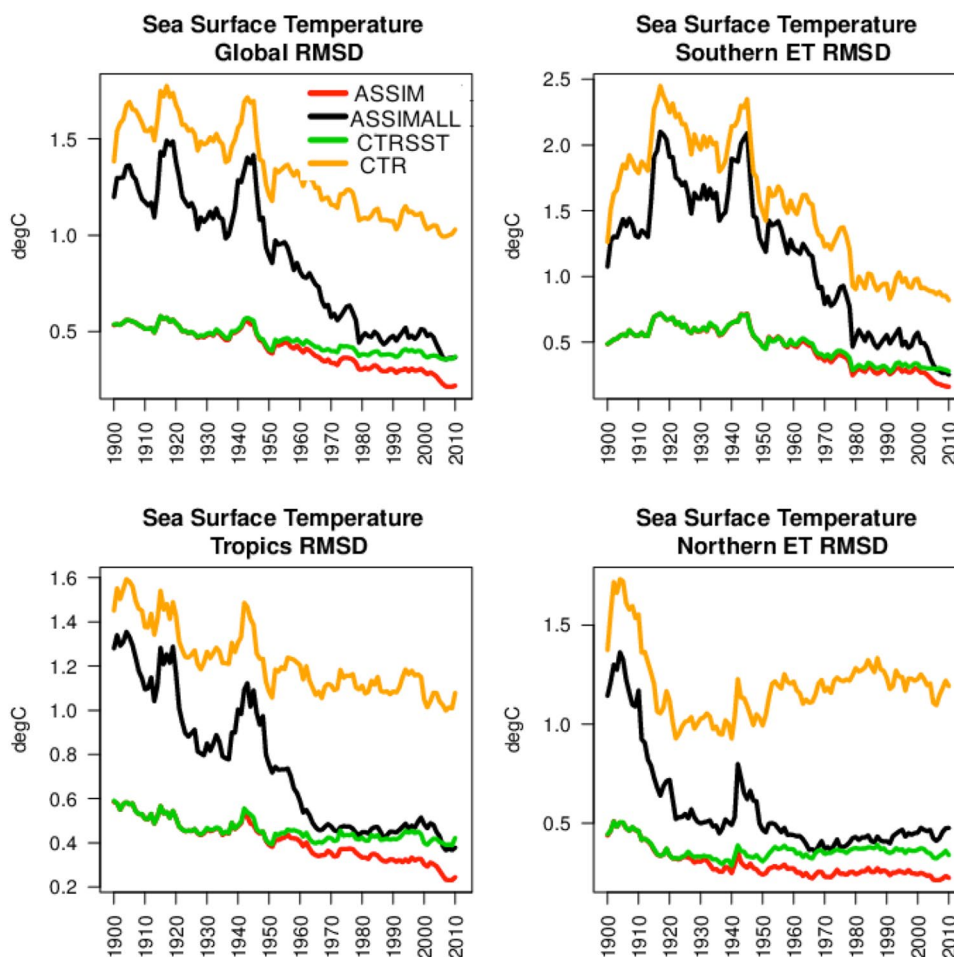
The two bottom panels of Fig. 7 show the percentage spread reduction during 1951–2010 compared to 1900–1950 for CTR (bottom left panel) and ASSIMALL (bottom right panel). The former reproduces the decrease in spread between the two atmospheric forcing, which is visible in the Pacific and Indian oceans, in the South Atlantic Ocean, and particularly large along the ACC. The increase of SST spread for CTR in the Northern Extra-tropics, also seen in Fig. 6, is likely due to the increase of spread in the radiative fluxes (Fig. 3) in this region, where additionally the turbulent fluxes may amplify the net heat flux differences (de Boisseson et al. 2017); indeed, the spread increase is mostly located in the Arctic and North Atlantic Oceans (North-western European shelf) and in the Eastern China and Japan Seas. The spread decrease during 1951–2010 in ASSIMALL is almost everywhere positive, peaking in the Pacific Ocean (affected by larger spread than the Atlantic, as also visible for CTR in the top left panel). Large positive decreases also occur in the Indian Ocean, along the ACC and in the North Atlantic subpolar gyre. The spread increase visible in the Eastern China and Japan Seas, the Indonesian region, in the Arctic Ocean is mostly due to the corresponding increase of

spread visible in CTR and associated with a lack of dense in-situ observing network. To a lesser extent, some spots in the North-western European shelf also exhibit negative values, indicating that the increase of spread in the atmospheric forcing may outnumber the increase of observations in these areas.

4.3 Comparison of ocean heat content long-term spread

A similar analysis on the spread was conducted for the ocean heat content in the top 700 m and in the total column (Figs. 8, 9). The CTR spread in the top 700 m at global scale increases almost linearly along time, apart from an initially rapid increase at the beginning of the time-series due to the initial conditions spinup. This increase appears dominated by the tropical region even after the 1950s, i.e. after the decrease in spread of the radiative fluxes over the tropical regions (see Fig. 3). In the Extra-tropics, after around 1920, the spread remains fairly stable. The spread of ASSIMALL follows that of CTR during the first decades, while later it decreases, indicating that in-situ observation assimilation
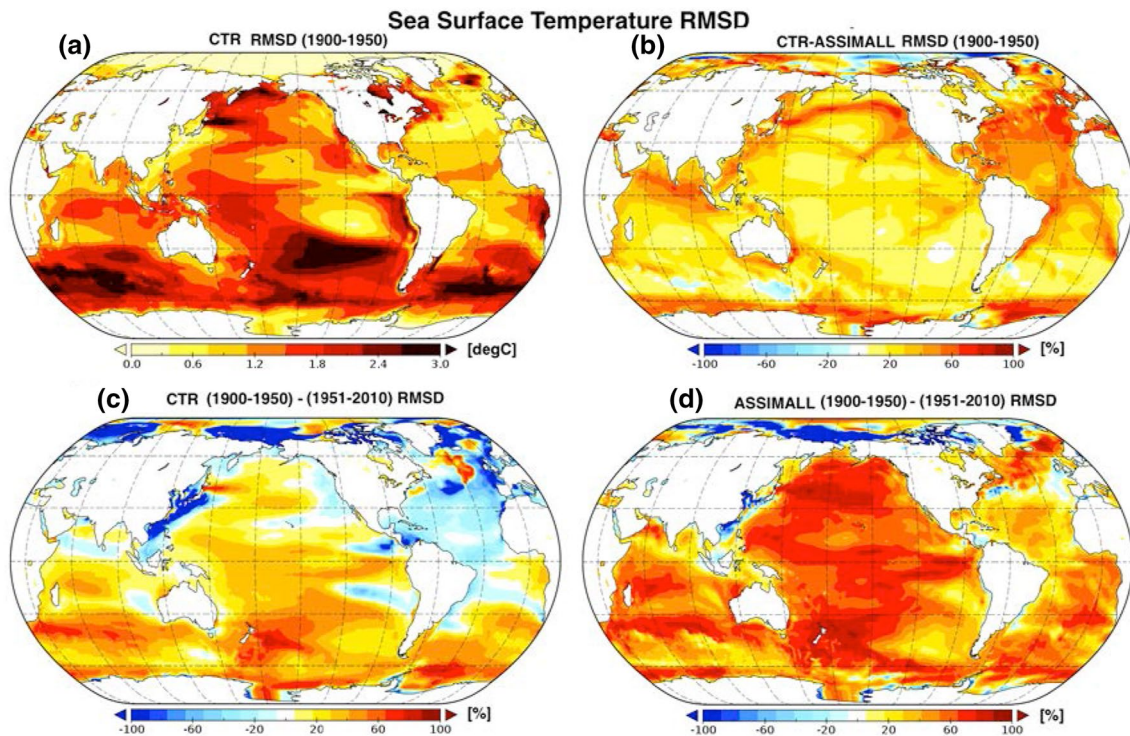
Fig. 7 Sea surface temperature RMSD maps. Top left panel: CTR RMSD over the period 1900–1950; top right panel: percentage difference of CTR minus ASSIMALL (1900–1950); bottom left panel: percentage difference of CTR RMSD over 1900–1950 minus that over 1951–2010; bottom right panel: percentage difference of AS RMSD over 1900–1950 minus that over 1951–2010

constrains the top 700 m heat content. In particular, during the 1950s in the North Extra-tropics and Tropics and during the 1980s in the South Extra-tropics, the spread of ASSI-MALL becomes smaller than CTRSST, and tends to converge towards that of ASSIM at the end of the time-series. Furthermore, there is an increase of the CTRSST spread during the last two decades of the reanalysis period in the North Extra-tropics, linked with the increase of the SST spread in CTR also seen in Fig. 7 in several regions. ASSIM generally shows the smaller spread, resulting from the combination of in-situ observational and SST constraints. The decrease of spread along time of ASSIMALL, especially in the North extra-tropics, is remarkably quick between 1930 and 1960, in correspondence of the rapid increase of the temperature observation sampling therein.

The total column heat content shows qualitatively similar features (Fig. 9). However, important differences can be sketched. For example, at global scale the spread of ASSI-MALL starts to decrease after approximately the mid-1950s but it does never become smaller than that of CTRSST or ASSIM. This suggests that only the heat flux correction implied by the SST relaxation is able to maintain low spread values and directly acts on limiting the propagation of atmospheric forcing uncertainty. Furthermore, the spread of the atmospheric forcing (Fig. 3) in the North Extra-tropics

is so large that it leads to an increase of ASSIMALL spread during the last decades. In the South extra-tropics, during most of the period, the spread of ASSIMALL exceeds that of CTR, indicating that few available observations do not act as a constraint for the total column heat content.

Figure 10 shows the spread from AS during 1900–1950 (top left panel) and 1951–2010 (top right panel) and percentage difference of spread between CTR and ASSIMALL for the two periods (bottom left and bottom right panels, respectively). The spread during the first half of the twentieth century in ASSIMALL is concentrated along the ACC and in the Atlantic Ocean. In the second half of the century, significant spread appears also in the deepest regions of the global ocean, testifying that the uncertainty linked to the atmospheric forcing reached the deep waters. The percentage difference shows that during the first half of the period ASSIMALL has a larger spread than CTR in the South Hemisphere, but generally smaller in the North Hemisphere (except in correspondence of the Mediterranean Sea outflow). The situation largely improves in the second half of the century, where the only regions where CTR shows total heat content spread smaller than ASSIMALL spread are in the Southern Hemisphere areas characterized by very deep bathymetry. These results suggest that when we consider the total column heat content, the vertical propagation of

**Fig. 8** As Fig. 6, but for the ocean heat content in the top 700 m
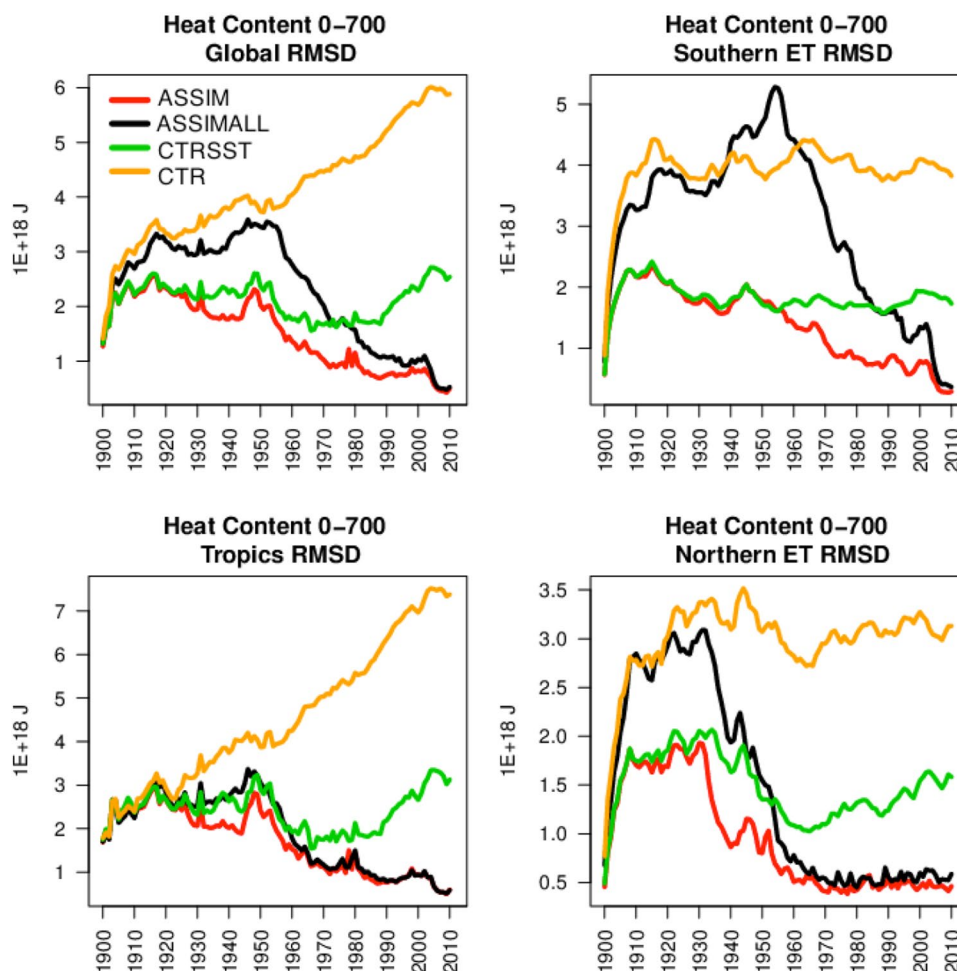
observational constraint following the in-situ observations deployment (especially after World War II) is not sufficient to compensate for the atmospheric forcing uncertainty during the second half of the twentieth century. The opposite happens for the top 700 m heat content.

To better assess the vertical behavior of the temperature spread, Fig. 11 compares the global ocean spread as a function of year and vertical depth from the four pairs of the experiments. The control experiment spread peaks in the top 100 m, with a significant reduction from the 1950s onwards. On the contrary, the pair of experiments that implement SST relaxation exhibits maximum spread in the range 50–200 m, due to the surface observational constraint; however, without in-situ data assimilation (CTRSST, top right panel of Fig. 11), the reduction after the 1950s is not as large as in ASSIM (bottom right panel). The pair of experiments with in-situ and SST data assimilation (ASSIMALL, bottom left panel) shows the largest spread sensitivity to the observational amount, i.e. the spread during the two World Wars increases noticeably, and decreases significantly afterwards. A bottom-wards propagation is however visible, starting from the beginning of the time series and being reduced only

since the 1980s with a reduction located at around 1000 m of depth and propagating both upward and downward. This confirms that with the lack of surface temperature constraint, the uncertainty arising from the atmospheric forcing below the well observed ocean (i.e. below 700 m) is decreasing only with the in-situ observing network developed since the 1960s onwards. We have also found that the top 700 m RMSD is significantly anti-correlated with the number of SST and profile observations per year in ASSIMALL and ASSIM (correlation below −0.7 for these cases). For the total ocean heat content RMSD, the correlation is significant only for ASSIM with respect to the number of SST and in-situ profiles, meaning that the lack of SST relaxation in ASSIMALL leads to a more pronounced effect of atmospheric forcing uncertainty and a non-significant correlation with the observation amount.

The spread derivative with time may give further insights onto the periods and events leading to increase and decrease of the spread, possibly related to the observational sampling, inter-annual ocean variability modes such as ENSO, and external forcing such as volcanic eruptions. However, the spread tendency with time may be

Fig. 9 As Fig. 6, but for the total ocean heat content (surface to bottom)
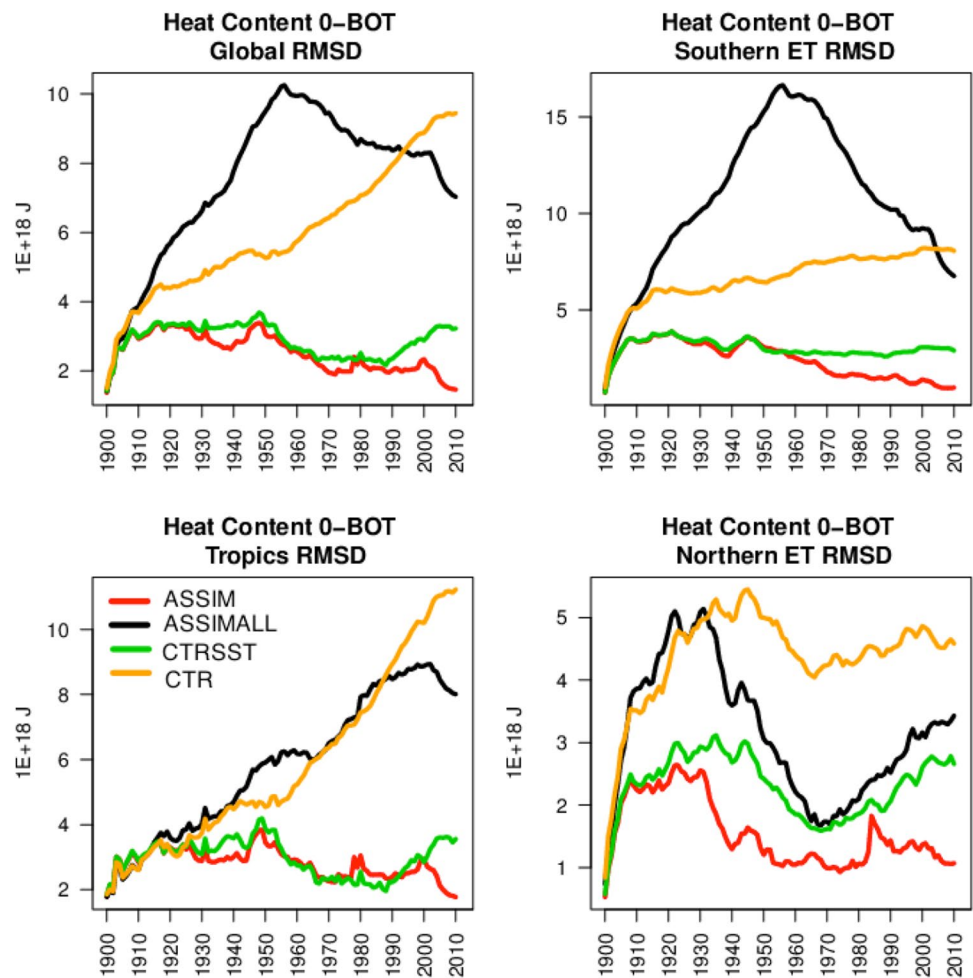
also partly driven by the background-error multiplicative factor included in the data assimilation system, which modulates the decadal error covariances. Figure 12 summarizes these diagnostics, showing the derivative of the global temperature RMSD with time (in degC per decade, calculated year by year as yearly difference of RMSD); SST and profile amount are superimposed, along with the periods of the two World Wars, the strong ENSO events (positive and negative ones) and the volcanic eruptions. Strong spread increases occur at the beginning of the time series, linked to the initial conditions uncertainty and inconsistency between the experiments with different atmospheric forcing, and possibly with the Santa Maria volcanic eruption in 1902. The spread increase is also significant during the World Wars, and the lack of surface constraint leads CTR and ASSIMALL to exhibit stronger spread increase in the top 100 m. These spread increases take approximately about 10 years to penetrate from the top 100 m to below 1000 m of depth, indicating the delay in the deep ocean propagation of the decrease of the upper ocean observational sampling. After the World War II, all pairs of experiments show a quick decrease of spread in

the top 200 m that slowly penetrates below 1000 m (except for the control experiments). The consistent decrease in all pairs indicates that both the oceanic and atmospheric observing networks recover from the lack of observations, thus reducing the spread in all configurations.

Further increases of spread occur during the periods 1971–1974, 1981–1983 and 1990–1995. Although it is apparently difficult to interpret these RMSD increases, the first one (1971–1974) seems to be related to a compensating effect after the post-War decrease jointly with the fact that the ocean observing network does not increase during the 1970s. The other two increase periods do correspond to periods affected by large volcanic eruptions, which likely cause some dispersion in the experiments.

Generally, all pairs of experiments indicate simultaneous RMSD changes, suggesting that the atmospheric forcing is mostly responsible for the positive and negative periods of RMSD change rate. These rates are however amplified in the pair of experiments without SST relaxation, which obviously mitigates the spread between the two atmospheric reanalyses.
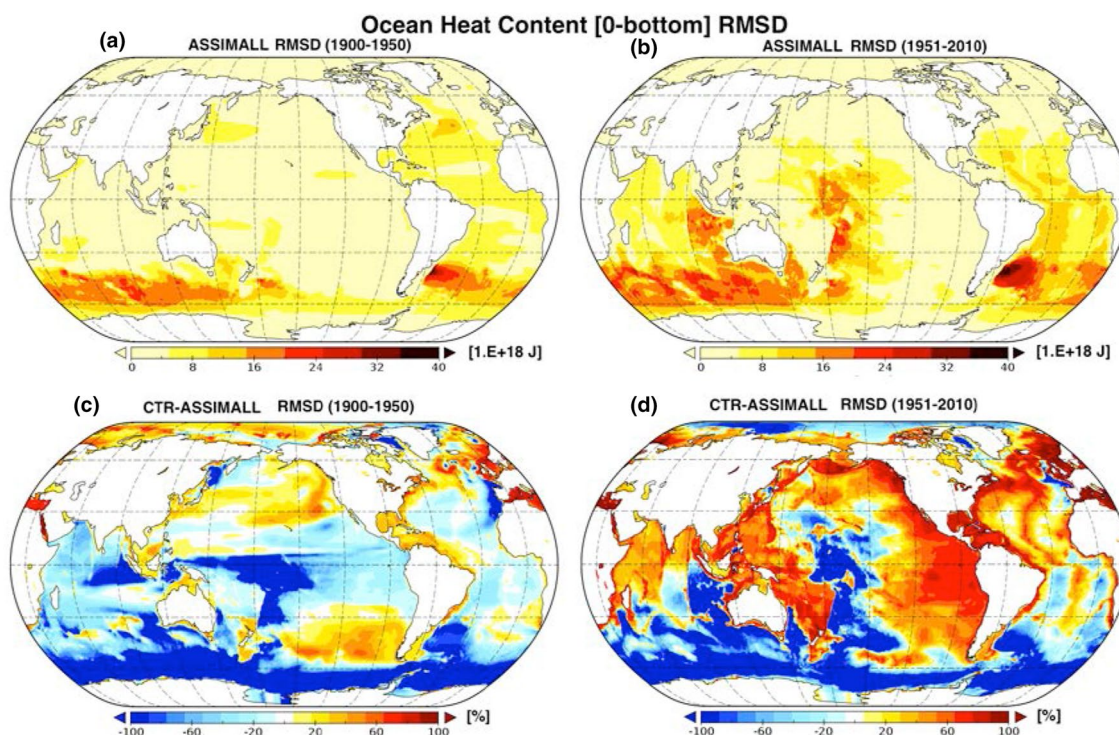
Fig. 10 Total ocean heat content RMSD maps. Top panels: ASSIMALL RMSD over the period 1900–1950 (left) and 1951–2010 (right). Bottom panels: percentage difference of CTR minus ASSIMALL 1900–1950 (left) and 1951–2010 (right)

4.4 Spatial characteristics of the observational constraints

Figures 13 and 14 show the spatial distribution of the first consecutive two years when the RMSD decrease occurs, for the top 700 m and the total column ocean heat content, respectively. The RMSD is calculated separately for each box of size $20° \times 20°$ in the global ocean, with the aim of diagnosing the period when the global ocean heat content starts to be constrained by the observations with respect to the atmospheric forcing uncertainty.

For the top 700 m ocean heat content, experiments without data assimilation (CTR) have a decreasing RMSD in almost all areas within the 1980s, except for two boxes in the Tropical Atlantic and Pacific oceans where the decrease occurs either late (after 2000) or does not occur at all, respectively. For all other pairs of experiments (CTRSST, ASSIM and ASSIMALL), there is in general a RMSD decrease that occurs at latest before the 1960s. The ASSIMALL pair has a slightly delayed decrease of RMSD in the Southern Ocean, i.e. in particular in the Atlantic sector, opposed to the pairs with SST relaxation that exhibit an early occurrence of the decrease in general.

The same diagnostics for the total column ocean heat content (Fig. 14) supply a different picture: the pairs of experiments that implement SST relaxation (CTRSST and

ASSIM) show that the first RMSD decrease generally occur before 1960 in all areas, except in the Tropical Atlantic and Pacific Oceans, where it is delayed up to 1980 and to 1970, respectively. The CTRSST RMSD starts decreasing late also in a few selected areas in the North Pacific Ocean. More heterogeneous is the spatial distribution for the total ocean heat content RMSD in the CTR and ASSIMALL pairs. Here, first RMSD decrease occurs very recently in several areas of the Tropical and Southern oceans. For instance, CTR RMSD shows very recent RMSD decrease in the Tropical and South Atlantic Ocean and in some areas of the Tropical Pacific Ocean. Additionally, RMSD never decreases in the Indonesian region. The situation improves for the ASSIMALL pair, that shows at latest RMSD decrease in the 1990s for a few areas in the Tropical Indian and Pacific oceans.

To summarize, the top 700 m ocean heat content appears well constrained starting from around the 1960s with respect to the atmospheric and oceanic observing networks, i.e. in pairs of experiments either with or without in-situ profile data assimilation, except for localized behavior in the Tropics when no ocean data assimilation is implemented. On the contrary, at the yearly time scales we investigated, the total ocean heat content is constrained only very late or not at all when there is no SST nudging in many areas of the Tropical and Southern oceans. This means that the uncertainty of the atmospheric forcing prevails on the ocean observation
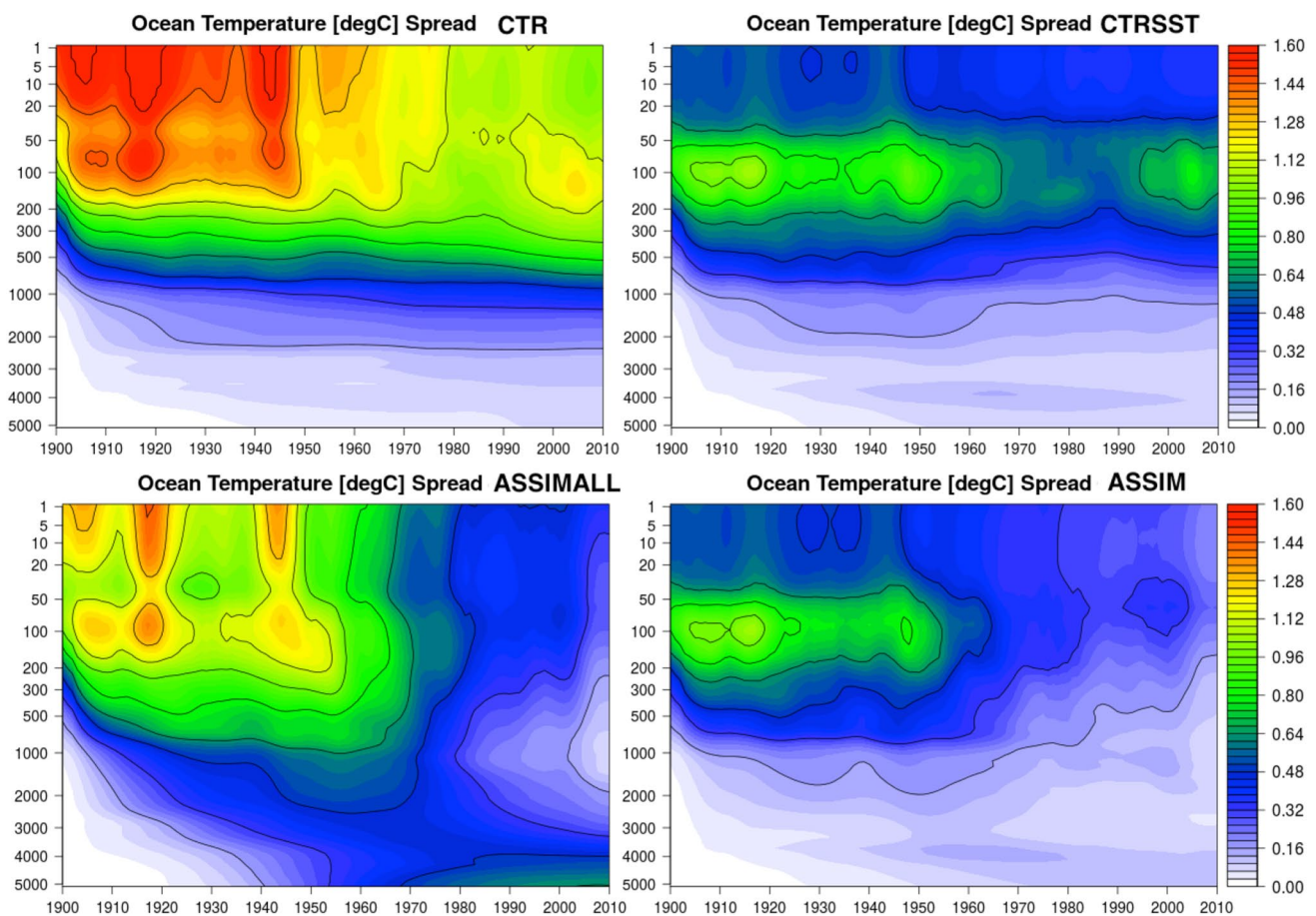
Fig. 11 Yearly temperature RMSD between the four pairs of reanalyses (see Table 1), as a function of the year and vertical depth (in meters)

sampling in these areas, with obvious implications on the reliability of the total ocean heat content estimates during the reanalysis period.

5 Summary and discussion

In this work, we have analyzed the ocean heat content during the period 1900–2010 from a set of ocean reanalyses that are forced by either the 20CRv2 or the ERA-20C atmospheric reanalysis and implement either direct assimilation of SST historical data or nudging to HadISST monthly SST analyses, on top of in-situ profiles data assimilation common to all reanalyses. Complementary experiments without data assimilation were also performed, providing a framework for assessing the impact of the atmospheric forcing and data assimilation of observations on the reconstruction of the heat content variability. The systematic use of the root mean square difference (spread) between pair of experiments with different atmospheric forcing allowed the in-depth analysis of the hierarchy of uncertainties in ocean reanalyses. The assessment contained in this study thus represents an attempt

to estimate the feasibility of reconstructing historical ocean heat content from reanalyses during early periods (first half of the twentieth century) where the ocean observing network is in general poor.

Qualitative assessment of the ocean heat content above 700 m suggests that even poor observational sampling during the first half of the century is able to shape the heat content variability. This promising evidence fosters further activities on historical reanalyses in the ocean community, in a way similar to the efforts devoted by the atmospheric community in designing historical reanalyses during the last decade. Large uncertainties occurring in the first half of the century do not prevent reanalysis systems from detecting global ocean warming during the first half of the twentieth Century, even though large uncertainties and specific challenges (e.g. optimal initialization of the reanalysis at the beginning of the time series, vertical spread of the upper ocean observations) exist.

At the sea surface, the in-situ profile data assimilation complements the SST relaxation in reducing the sea surface temperature spread, and provides a further justification to simultaneously perform both surface and subsurface data
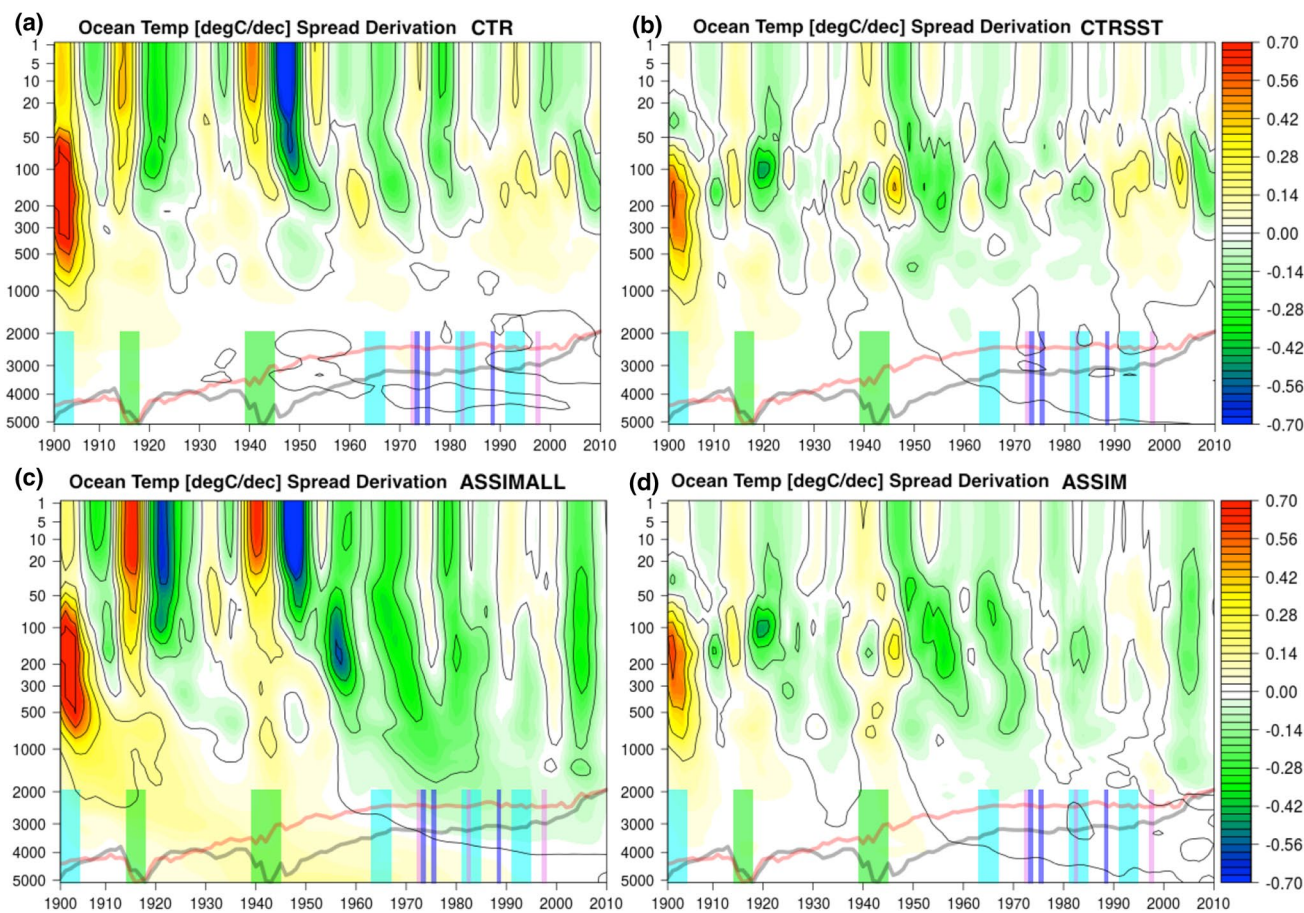
Fig. 12 Yearly temperature RMSD derivative (in degC over decade) between the four pairs of reanalyses (see Table 1), as a function of the year and vertical depth (in meters). In the bottom part of the figure, green bars correspond to the two World Wars periods, light blue bars to the periods of massive volcanic eruptions, dark blue and red to the years of strongly positive and negative ENSO anomalies, respectively, and the black and red line show in logarithmic scale the number of available SST and in-situ profiles, respectively

assimilation within historical ocean reanalyses. This appears particularly crucial in data abundant periods or regions.

Furthermore, in the first half of the century, due to the large uncertainty associated with atmospheric forcing and the small number of available observations, the impact of data assimilation on the top 700 m heat content spread (comparison between ASSIMALL and CTR) is relatively small compared to that in the second half of the century (Fig. 8). During the second half of the twentieth century (since the 1950s in the North Hemisphere and the 1980s in the South Hemisphere), the spread from the pair of experiments with in-situ data assimilation (ASSIMALL) decreases and becomes even smaller than that from the pair of experiments with sea surface relaxation only (CTRSST), meaning that, since then, the observing network is developed enough to tightly constrain the ocean heat content variability. The decrease of spread in the assimilation experiments is indeed remarkably quick between 1930 and 1960, in correspondence of the rapid increase of the temperature observation sampling. This finding is in accordance with the work of

Carton et al. (2012), who also found that the 1960s are the decade since when oceanic observations constrain the ocean variability.

However, the analysis of the top to bottom ocean heat content reveals that the vertical propagation of observational constraint that follows the in-situ observations deployment (especially after World War II) is not yet sufficient to constrain the total heat content during the second half of the twentieth century, unlike the heat content in the top 700 m. This emphasizes the lag between upper ocean observation deployment and deep ocean propagation. Below 700 m, the uncertainty accumulated from the atmospheric forcing is compensated only with the in-situ observing network developed since the 1980s onwards.

The analysis of the spread rate of change also suggests that the atmospheric forcing and large-scale regimes (i.e. the combined information from atmospheric observations and SST analyses used as boundary conditions in atmospheric reanalyses) are mostly responsible for positive and negative phases, because such phases are shared among all
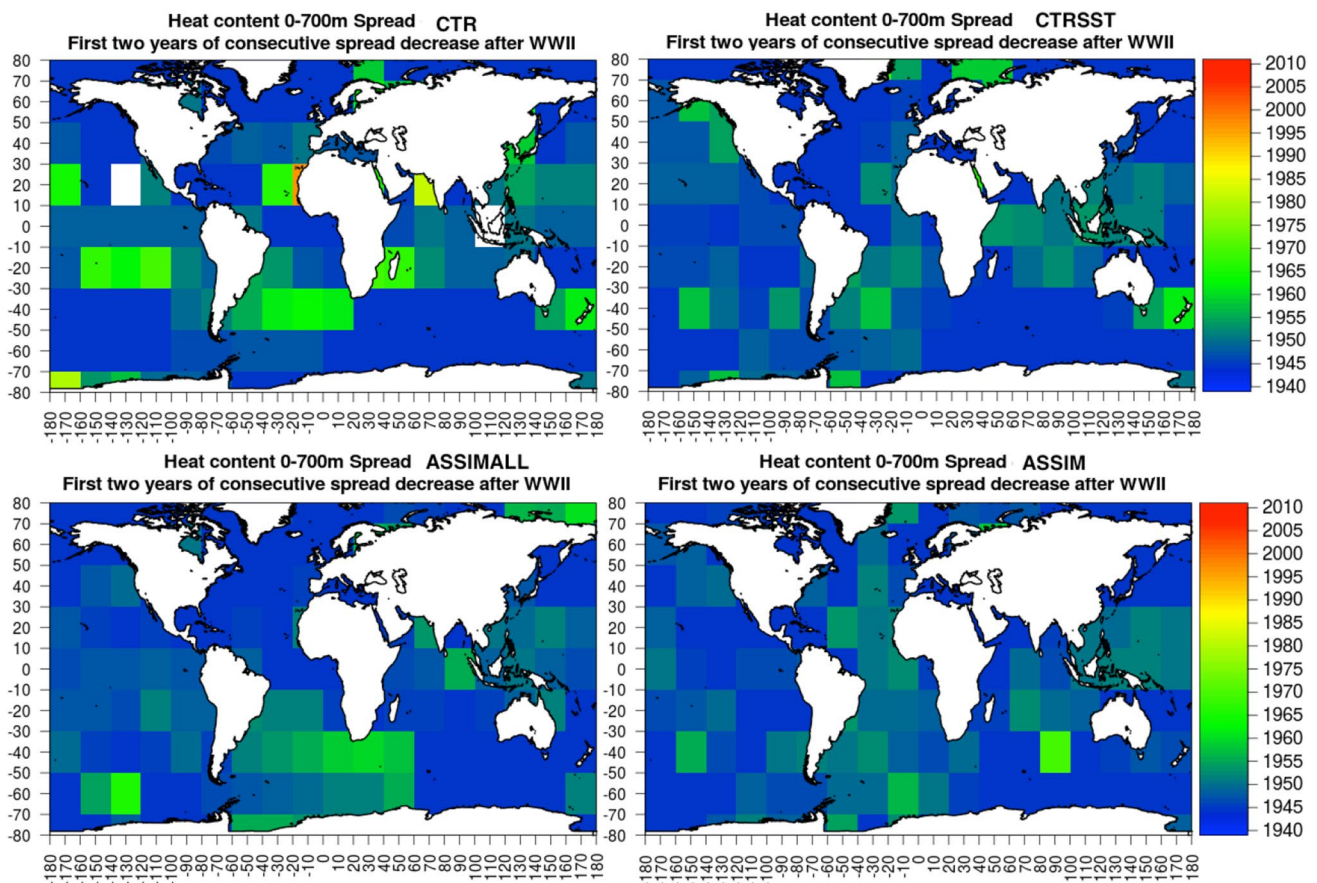
Fig. 13 First occurrence of 0–700 m ocean heat content RMSD decrease for two consecutive years after World War II for the four pairs of reanalyses (see Table 1) as a function of longitude and latitude. The RMSD is calculated for each box of 20° × 20° of size

pairs of experiments. Furthermore, in the top 700 m heat content even experiments without data assimilation have a decreasing spread in almost all areas after the World War II, which stresses how the uncertainty in the atmospheric forcing decreases significantly with time as well, due to the changes in the atmospheric observing network. However, for a few diagnostics it was shown on the contrary an increasing spread with time, often related to an increasing spread with time for some selected atmospheric parameters (e.g. 2 m temperature and shortwave radiation), likely due to the different data streams used by the two atmospheric reanalyses, that are amplified by radiation and microphysics schemes. These considerations are crucial when using historical atmospheric reanalyses for ocean simulations or reconstructions.

Most of results contained in this work suggest that historical ocean reanalyses may be a useful tool for investigating climate variability even in observation poor periods. Uncertainty assessment remains crucially important for defining the reliability of such datasets. It is also suggested that multi-system ensemble approach or, alternatively, multi-forcing ensemble, even in reanalysis systems

implementing variational data assimilation, are fundamental strategies to achieve meaningful uncertainty assessments over long periods. Here we focused on assessing the relative contribution of uncertainty linked with in-situ observational sampling and atmospheric forcing accuracy, which we conceive as the prevailing factors preventing historical ocean reanalyses from stable accuracy with time. However, we recommend that to perform historical ocean reanalyses ensemble we also take into account the uncertainty coming from the in-situ observation bias-correction, the reanalysis initial conditions, and the uncertainty linked with the ocean model parameterizations, in particular the vertical physical schemes responsible for the vertical propagation of the observational information. These sources of uncertainty should all be introduced in the future design of ensemble historical reanalyses.

A possible limitation of this study is that we only show the ensemble historical ocean reanalyses conducted with one single ocean model and our own data assimilation scheme. Indeed, an extension of the present work will be to compare all the existing historical ocean reanalyses (Giese et al. 2016 and; de Boisseson et al. 2017) to detect the uncertainty
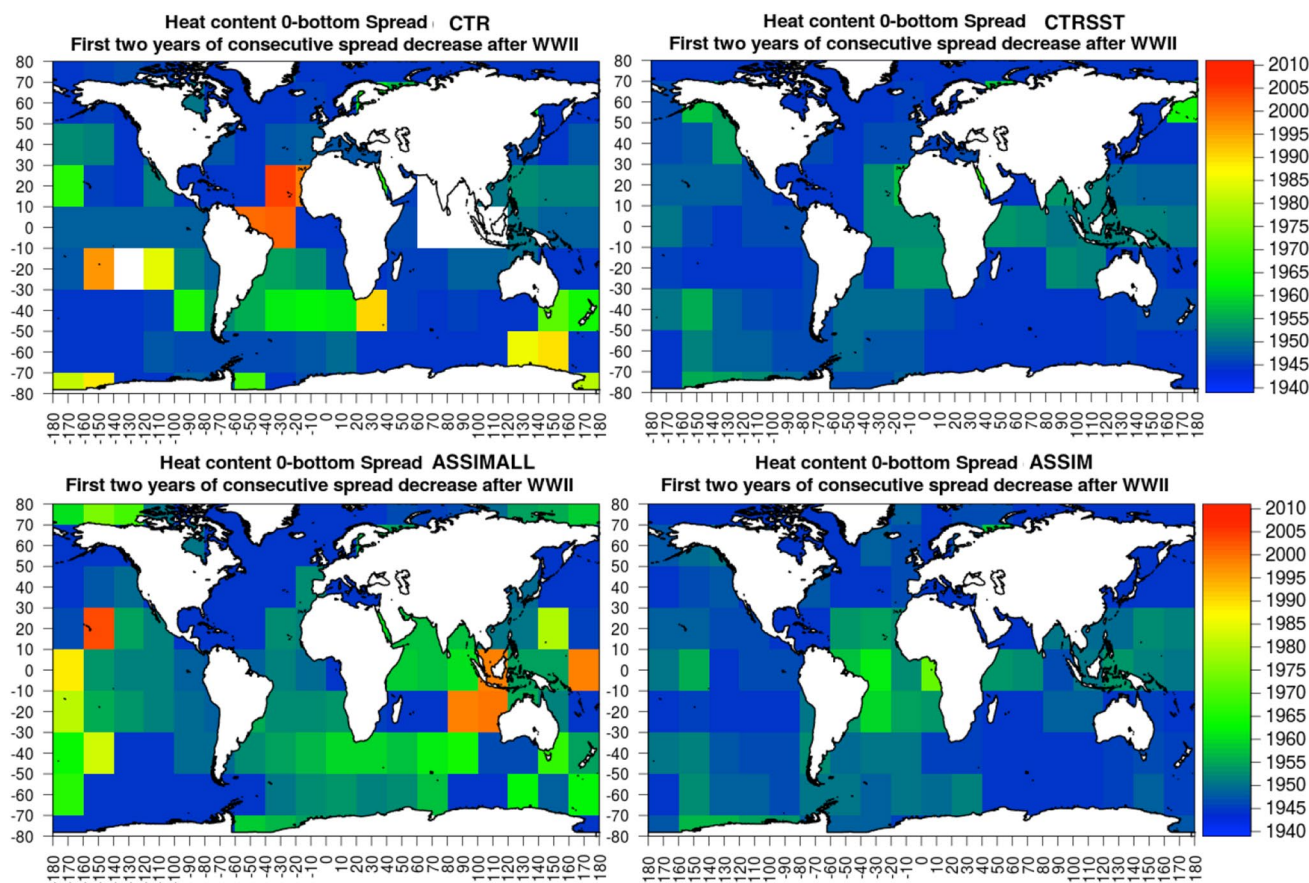
Fig. 14 As Fig. 13, but for the total ocean heat content (surface to bottom)

also coming from different data assimilation schemes, ocean models and atmospheric forcing.

References

Abraham JP, Baringer M, Bindoff NL, Boyer T, Cheng LJ, Church JA, Conroy JL, Domingues CM, Fasullo JT, Gilson J, Goni G, Good SA, Gorman JM, Gouretski V, Isshii M, Jonhson GC, Kizu S, Lyman JM, Macdonald AM, Minkowycz WJ, Moffitt SE, Palmer MD, Piola AR, Reseghetti F, Schukmann K, Trenberth KE, Velicogna I, Willis JK (2013) A review of global ocean temperature observations: Implication for ocean heat content estimates and climate change. Rev Geophys 51:450–483. https://doi.org/10.1002/rog.20022

Atkinson CP, Rayner NA, Kennedy JJ, Good SA (2014) An integrated database of ocean temperature and salinity observations. J Geophys Res Oceans 119:7139–7163. https://doi.org/10.1002/2014JC010053

Behrens E, Biastoch A, Böning CW (2013) Spurious AMOC trends in global ocean sea-ice models related to subarctic freshwater forcing. Ocean Model 69:39–49

Bernie DJ, Guilyardi E, Madec G, Slingo JM, Woolnough SJ (2007) Impact of resolving the diurnal cycle in an ocean–atmosphere GCM. Part 1: A diurnally forced OGCM. Clim Dynam 29:575–590

Bourdalle-Badie R, Treguier AM (2006) A climatology of runoff for the global ocean-ice model ORCA025. MERCATOR report MOO-RP-425-366-MER

Boyer TP, Levitus S (1998) Objective analyses of temperature and salinity for the world ocean on a 1/4 grid. NOAA Atlas NESDIS, Washington, DC

Boyer TP, Antonov JI, Baranova OK, Garcia HE, Johnson DR, Locarnini RA, Mishonov AV, Seidov D, Smolyar IV, Zweng MM (2009) World Ocean Database 2009, Chap. 1: introduction, NOAA Atlas NESDIS 66, Levitus S (ed), U.S. Gov. Printing Office, Wash., DVD, p 216. https://www.nodc.noaa.gov

Boyer T, Domingues CM, Good SA, Johnson GC, Lyman JM, Ishii M, Gouretski V, Willis JK, Antonov J, Wijffels S, Church JA, Cowley R, Bindoff NL (2016) Sensitivity of global upper-ocean heat content estimates to mapping methods, XBT bias corrections and baseline climatologies. J. Clim 29:4817–4842

Brönnimann S, Grant AN, Compo GP et al (2012) A multi-data set comparison of the vertical structure of temperature variability and change over the Arctic during the past 100 years. Clim Dyn 39:1577–1598. https://doi.org/10.1007/s00382-012-1291-6

Carton JA, Seidel HF, Giese BS (2012) Detecting historical ocean climate variability. J Geophys Res 117:C02023. https://doi.org/10.1029/2011JC007401

Chen Z, Wu L (2012) Long-term change of the Pacific North Equatorial Current bifurcation in SODA. J Geophys Res 117:C06016. https://doi.org/10.1029/2011JC007814

Compo GP, Whitaker JS, Sardeshmukh PD (2006) Feasibility of a 100-year reanalysis using only surface pressure data. Bull Am Meteorol Soc 87:175–190. https://doi.org/10.1175/BAMS-87-2-175

Compo GP, Whitaker JS, Sardeshmukh PD, Matsui N, Allan RJ, Yin X, Gleason BE, Vose RS, Rutledge G, Bessemoulin P, Brönnimann S, Brunet M, Crouthamel RI, Grant AN, Groisman PY, Jones PD, Kruk MC, Kruger AC, Marshall GJ, Maugeri M, Mok HY, Nordli Ø, Ross TF, Trigo RM, Wang XL, Woodruff SD, Worley SJ (2011) The Twentieth Century Reanalysis Project. QJR Meteorol Soc 137:1–28. https://doi.org/10.1002/qj.776

Cram TA, Compo GP, Yin X, Allan RJ, McColl C, Vose RS, Whitaker JS, Matsui N, Ashcroft L, Auchmann R, Bessemoulin P, Brandsma T, Brohan P, Brunet M, Comeaux J, Crouthamel R, Gleason BE, Groisman PY, Hersbach H, Jones PD, Jónsson T, Jourdain S, Kelly G, Knapp KR, Kruger A, Kubota H, Lentini G, Lorrey A, Lott N, Lubker SJ, Luterbacher J, Marshall GJ, Maugeri M, Mock CJ, Mok HY, Nordli Ø, Rodwell MJ, Ross TF, Schuster D, Srnec L, Valente MA, Vizi Z, Wang XL, Westcott N, Woollen JS, Worley SJ (2015) The International Surface Pressure Databank version 2. Geosci Data J 2:31–46. https://doi.org/10.1002/gdj3.25

de Boisseson E, Balmaseda MA, Mayer M (2017) Ocean het content variability in an ensemble of twentieth century ocean reanalyses. Clim Dyn. https://doi.org/10.1007/s00382-017-3845-0

Dee DP, Uppala SM, Simmons AJ, Berrisford P, Poli P, Kobayashi S, Andrae U, Balmaseda MA, Balsamo G, Bauer P, Bechtold P, Beljaars ACM, van de Berg L, Bidlot J, Bormann N, Delsol C, Dragani R, Fuentes M, Geer AJ, Haimberger L, Healy SB, Hersbach H, Hólm EV, Isaksen L, Kållberg P, Köhler M, Matricardi M, McNally AP, Monge-Sanz BM, Morcrette J-J, Park B-K, Peubey C, de Rosnay P, Tavolato C, Thépaut J-N, Vitart F (2011) The ERA-Interim reanalysis: configuration and performance of the data assimilation system. QJR Meteorol Soc 137:553–597. https://doi.org/10.1002/qj.828

Desroziers G, Berre L, Chapnik B, Poli P (2005) Diagnosis of observation, background and analysis-error statistics in observation space. QJR Meteorol Soc 131:3385–3396. https://doi.org/10.1256/qj.05.108

Ferguson CR, Villarini G (2014) An evaluation of the statistical homogeneity of the Twentieth Century Reanalysis. Clim Dyn 42(11–12):2841–2866

Freeman E et al (2016) ICOADS Release 3.0: A major update to the historical marine 626 climate record. Int J Climatol. https://doi.org/10.1002/joc.4775

Giese BS, Seidel HF, Compo GP, Sardeshmukh PD (2016) An ensemble of ocean reanalyses for 1815–2013 with sparse observational input. J Geophys Res Oceans 121:6891–6910

Good SA, Martin MJ, Rayner NA (2013) EN4: quality controlled ocean temperature and salinity profiles and monthly objective analyses with uncertainty estimates. J Geophys Res Oceans 118:6704–6716. https://doi.org/10.1002/2013JC009067

Gouretski V, Reseghetti F (2010) On depth and temperature biases in bathythermograph data: development of a new correction scheme based on analysis of a global ocean database. Deep-Sea Res I 57:6. https://doi.org/10.1016/j.dsr.2010.03.011

Griffin RE (2015) When are old data new data? Geo Res J 6:92–97. https://doi.org/10.1016/j.grj.2015.02.004

Hersbach H, Peubey C, Simmons A, Berrisford P, Poli P, Dee D (2015) ERA-20CM: a twentieth-century atmospheric model ensemble. QJR Meteorol Soc 141:2350–2375. https://doi.org/10.1002/qj.2528

Hersbach H, Brönnimann S, Haimberger L, Mayer M, Villiger L, Comeaux J, Simmons A, Dee D, Jourdain S, Peubey C, Poli P, Rayner N, Sterin AM, Stickler A, Valente MA, Worley SJ (2017) The potential value of early (1939–1967) upper-air data in atmospheric climate reanalysis. QJR Meteorol Soc 143:1197–1210. https://doi.org/10.1002/qj.3040

Ishii M, Kimoto M (2009) Reevaluation of historical ocean heat variations with time-varying XBT and MBT depth bias correction. J Oceanogr 65:287–299. https://doi.org/10.1007/s10872-009-0027-7

Kalnay E et al (1996) The NCEP/NCAR 40-Year Reanalysis Project. Bull Am Meteorol Soc 77:437–471

Kato S, Loeb NG, Rose FG, Doelling DR, Rutan DA, Caldwell TE, Yu L, Weller RA (2013) Surface irradiances consistent with CERES-derived top-of-atmosphere shortwave and longwave irradiances. J Clim 26:2719–2740

Kennedy JJ (2014) A review of uncertainty in in situ measurements and data sets of sea surface temperature. Rev Geophys 52:1–32. https://doi.org/10.1002/2013RG000434

Kent EC, Fangohr S, Berry DI (2013) A comparative assessment of monthly mean wind speed products over the global ocean. Int J Climatol 33:2520–2541. https://doi.org/10.1002/joc.3606

Laloyaux P, Balmaseda M, Dee D, Mogensen K, Janssen P (2016) A coupled data assimilation system for climate reanalysis. QJR Meteorol Soc 142:65–78. https://doi.org/10.1002/qj.2629

Laloyaux P, de Boisseson E, Balmasdea M, Bidlot J-R, Broennimann S, Buizza R, Dahlgren P, Dee D, Haiberger L, Hersbach H, Kosaka Y, Martin M, Poli P, Rayner N, Rustemeier E, Schepers D (2018) CERA-20C: a coupled reanalysis of the Twentieth Century. J Adv Modell Earth Syst. https://doi.org/10.1029/2018MS001273

Large W, Yeager S (2004) Diurnal to decadal global forcing for ocean and seaice models: the data sets and climatologies. Technical Report TN-460 + STR, NCAR

Lindsay R, Wensnahan M, Schweiger A, Zhang J (2014) Evaluation of seven different atmospheric reanalysis products in the arctic. J Clim 27:2588–2606. https://doi.org/10.1175/JCLI-D-13-00014.1

Madec G, the NEMO team (2012) "NEMO ocean engine". Note du Pole de modélisation de l'Institut Pierre-Simon Laplace, France, No 27 ISSN No 1288–1619

Masina S, Storto A (2017) Reconstructing the recent past ocean variability: status and perspective. J Mar Sci 75:727–764. https://doi.org/10.1357/002224017823523973

Paek H, Huang H (2012) A comparison of decadal-to-interdecadal variability and trend in reanalysis datasets using atmospheric angular momentum. J Clim 25:4750–4758. https://doi.org/10.1175/JCLI-D-11-00358.1

Penny SG, Behringer DW, Carton JA, Kalnay E (2015) A hybrid global ocean data assimilation system at NCEP. Mon Weather Rev 143:4660–4677. https://doi.org/10.1175/MWR-D-14-00376.1

Poli P et al (2013) The data assimilation system and initial performance evaluation of the ECMWF pilot reanalysis of the 20th-century assimilating surface observations only (ERA-20C). ECMWF ERA Rep 14:59. http://www.ecmwf.int/en/elibrary/11699-data-assimilation-system-and-initial-performance-evaluation-ecmwf-pilot-reanalysis

Poli P, Hersbach H, Berrisford P, Dee D, Simmons A, Laloyaux P (2015) ERA-20C deterministic. ECMWF ERA Rep 20:48. http://www.ecmwf.int/en/elibrary/11700-era-20c-deterministic

Poli P, Hersbach H, Dee DP, Berrisford P, Simmons AJ, Vitart F, Laloyaux P, Tan DH, Peubey C, Thépaut J, Trémolet Y, Hólm EV, Bonavita M, Isaksen L, Fisher M (2016) ERA-20C: an atmospheric reanalysis of the Twentieth Century. J Clim 29:4083–4097. https://doi.org/10.1175/JCLI-D-15-0556.1

Rayner NA, Parker DE, Horton EB, Folland CK, Alexander LV, Rowell DP, Kent EC, Kaplan A (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century J Geophys Res 108(D14):4407. https://doi.org/10.1029/2002JD002670

Robock A (2000) Volcanic eruptions and climate. Rev Geophys 38:191–219. https://doi.org/10.1029/1998RG000054

Smith DM, Murphy JM (2007) An objective ocean temperature and salinity analysis using covariances from a global climate model. J Geophys Res 112:C02022. https://doi.org/10.1029/2005JC003172

Sterl A (2004) On the (In)Homogeneity of reanalysis products. J Clim 17:3866–3873. https://doi.org/10.1175/1520-0442(2004)017%3C3866:OTIORP%3E2.0.CO;2

Storto A, Dobricic S, Masina S, Di Pietro P (2011) Assimilating along-track altimetric observations through local hydrostatic adjustments in a global ocean reanalysis system. Mon Weather Rev 139:738–754

Storto A, Masina S, Dobricic S (2014) Estimation and impact of non-uniform horizontal correlation length-scales for global ocean physical analyses. J Atmos Ocean Tech 31:2330–2349

Storto A, Yang C, Masina S (2016), Sensitivity of global ocean heat content from reanalyses to the atmospheric reanalysis forcing: A comparative study, Geophys Res Lett. https://doi.org/10.1002/2016GL068605

Taylor K, Stouffer R, Meehl G (2012) An overview of CMIP5 and the experiment design. Bull Am Meteorol Soc 93:485–498. https://doi.org/10.1175/BAMS-D-11-00094.1

Thorne PW, Allan RJ, Ashcroft L, Brohan P, Dunn RJ, Menne MJ, Pearce PR, Picas J, Willett KM, Benoy M, Bronnimann S, Canziani PO, Coll J, Crouthamel R, Compo GP, Cuppett D, Curley M, Duffy C, Gillespie I, Guijarro J, Jourdain S, Kent EC, Kubota H, Legg TP, Li Q, Matsumoto J, Murphy C, Rayner NA, Rennie JJ, Rustemeier E, Slivinski LC, Slonosky V, Squintu A, Tinz B, Valente MA, Walsh S, Wang XL, Westcott N, Wood K, Woodruff SD, Worley SJ (2017) Towards an integrated set of surface meteorological observations for climate science and applications. Bull Am Meteorol Soc. https://doi.org/10.1175/BAMS-D-16-0165.1

Titchner HA, Rayner NA (2014) The Met Office Hadley Centre sea ice and sea-surface temperature data set, version 2: 1. Sea ice concentrations. J Geophys Res 119:2864–2889. https://doi.org/10.1002/2013JD020316

Trenberth KE, Fasullo JT, Balmaseda M (2014) Earth's energy imbalance. J Clim 27:3129–3144

Valdivieso M et al (2015) An assessment of air–sea heat fluxes from ocean and coupled reanalyses. Clim Dyn. https://doi.org/10.1007/s00382-015-2843-3

Vancoppenolle M, Fichefet T, Goosse H, Bouillon S, Madec G, Morales Maqueda MA (2009) Simulating the mass balance and salinity of Arctic and Antarctic sea ice. 1. Model description and validation. Ocean Model 27(1–2):33–53

Wang X, Feng Y, Chan R, Isaac V (2016) Inter-comparison of extra-tropical cyclone activity in nine reanalysis datasets. Atmos Res 181:133–153. https://doi.org/10.1016/j.atmosres.2016.06.010. (**ISSN 0169–8095**)

Whitaker JS, Hamill TM (2002) Ensemble data assimilation without perturbed observations. Mon Weather Rev 130:1913–1924

Woodruff SD, Coauthors (2011) ICOADS release 2.5: Extensions and enhancements to the surface marine meteorological archive. Int J Climatol 31:951–967. https://doi.org/10.1002/joc.2103

Wu et al (2012) Enhanced warming over the global subtropical western boundary currents. Nat Clim Chang. https://doi.org/10.1038/NCLIMATE1353

Yang C, Giese BS (2013) El Niño Southern Oscillation in an ensemble ocean reanalysis and coupled climate models. J Geophys Res Oceans 118(9):4052

Yang C, Masina S, Storto A (2017) Historical ocean reanalyses (1900–2010) using different data assimilation strategies. QJR Meteorol Soc 143:479–493. https://doi.org/10.1002/qj.2936

Yi DL, Zhang L, Wu L (2015) On the mechanisms of decadal variability of the North Pacific Gyre Oscillation over the 20th century. J Geophys Res Oceans 120:6114–6129. https://doi.org/10.1002/2014JC010660

Zib B, Dong X, Xi B, Kennedy A (2012) Evaluation and intercomparison of cloud fraction and radiative fluxes in recent reanalyses over the Arctic using BSRN surface observations. J Clim 25:2291–2305. https://doi.org/10.1175/JCLI-D-11-00147.1