Steric sea level variability (1993-2010) in an ensemble of ocean reanalyses and objective analyses

Article

Accepted Version

Storto, A., Masina, S., Balmaseda, M., Guinehut, S., Xue, Y., Szekely, T., Fukumori, I., Forget, G., Chang, Y.-S., Good, S. A., Kohl, A., Vernieres, G., Ferry, N., Peterson, K. A., Behringer, D., Ishii, M., Masuda, S., Fujii, Y., Toyoda, T., Yin, Y., Valdivieso, M., Barnier, B., Boyer, T., Lee, T., Gourrion, J., Wang, O., Heimback, P., Rosati, A., Kovach, R., Hernandez, F., Martin, M. J., Kamachi, M., Kuragano, T., Mogensen, K., Alves, O., Haines, K. and Wang, X. (2017) Steric sea level variability (1993-2010) in an ensemble of ocean reanalyses and objective analyses. Climate Dynamics, 49 (3). pp. 709-729. ISSN 0930-7575 doi: https://doi.org/10.1007/s00382-015-2554-9 Available at https://centaur.reading.ac.uk/52358/

Published version at: http://link.springer.com/article/10.1007/s00382-015-2554-9
To link to this article DOI: http://dx.doi.org/10.1007/s00382-015-2554-9

Publisher: Springer

Steric sea level variability (1993-2010) in an ensemble of ocean reanalyses and objective analyses

Andrea Storto * [1], Simona Masina[1], Magdalena Balmaseda[2], Stéphanie Guinehut[3], Yan Xue[4], Tanguy Szekely[5], Ichiro Fukumori[6], Gael Forget[7], You-Soon Chang[8,24], Simon A. Good[9], Armin Köhl[10], Guillaume Vernieres[11], Nicolas Ferry[12], K. Andrew Peterson[9], David Behringer[4], Masayoshi Ishii[15], Shuhei Masuda[16], Yosuke Fujii[15], Takahiro Toyoda[15], Yonghong Yin[18], Maria Valdivieso[19], Bernard Barnier[23], Tim Boyer[14], Tony Lee[6], Jérome Gourrion[20], Ou Wang[6], Patrick Heimback[7], Anthony Rosati[8], Robin Kovach[11], Fabrice Hernandez[12,25], Matthew J. Martin[13], Masafumi Kamachi[15], Tsurane Kuragano[15], Kristian Mogensen[2], Oscar Alves[18], Keith Haines[19], and Xiaochun Wang[22]

[1]Centro Euro-Mediterraneo sui Cambiamenti Climatici, Bologna, Italy
[2]European Center for Medium-Range Weather Forecast, Reading, UK
[3]Collecte Localisation Satellites, Ramonville Saint-Agne, France
[4]National Center for Environmental Predictions, NOAA, College Park, MD, USA
[5]Institut Universitaire Européen de la Mer, CNRS, Brest, France
[6]Jet Propulsion Laboratory, California Institute of Technology, Pasadena, CA, USA
[7]Massachusetts Institute of Technology, Cambridge, MA, CA, USA
[8]Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA
[9]Met Office Hadley Centre, Exeter, UK
[10]Institute of Oceanography, University of Hamburg, Hamburg, Germany
[11]Global Modeling and Assimilation Office, NASA Goddard Space Flight Center, Greenbelt, MA, USA
[12]Mercator Océan, Ramonville-Saint-Agne, France
[13]Met Office, Exeter, UK
[14]National Oceanographic Data Center, Silver Spring, MA, USA
[15]Meteorological Research Institute, Japan Meteorological Agency, Tsukuba, Japan
[16]Research and Development Center for Global Change, Japan Agency for Marine-Earth Science and Technology, Yokohama, Japan
[18]Centre for Australian Weather and Climate Research, Bureau of Meteorology, Melbourne, Australia
[19]Department of Meteorology, University of Reading, Reading, UK
[20]Centre national de la recherche scientifique, CORIOLIS, Brest, France
[21]Global Environment and Marine Department, Japan Meteorological Agency, Tokio, Japan
[22]Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, CA, USA
[23]Laboratoire des Ecoulements Géophysiques et Industriels, Grenoble, France
[24]Department of Earth Science, Kongju National University, Kongju, South Korea
[25]Institut de Recherche pour le Développement (IRD), Toulouse, France

Abstract

Quantifying the effect of the seawater density changes on sea level variability is of crucial importance for climate change studies, as the sea level cumulative rise can be regarded as both an important climate change indicator and a possible danger for human activities in coastal areas. In this work, as part of the Ocean Reanalysis Intercomparison Project (ORA-IP), the global and regional steric sea level changes are estimated and compared from an ensemble of 16 ocean reanalyses and 4 objective analyses. These estimates are initially compared with a satellite-derived (altimetry minus gravimetry) dataset for a short period (2003-2010). The ensemble mean exhibits a significant high correlation at both global and regional scale, and the ensemble of ocean reanalyses outperforms that of objective analyses, in particular in the Southern Ocean. The reanalysis ensemble mean thus represents a valuable tool for further analyses, although large uncertainties remain for the inter-annual trends. Within the extended intercomparison period that spans the altimetry era (1993-2010), we find that the ensemble of reanalyses and objective analyses are in good agreement, and both detect a trend of the global steric sea level of 1.0 and 1.1 ± 0.05 mm/yr, respectively. However, the spread among the products of the halosteric component trend exceeds the mean trend itself, questioning the reliability of its estimate. This is related to the scarcity of salinity observations before the Argo era. Furthermore, the impact of deep ocean layers is non-negligible on the steric sea level variability (22% and 12% for the layers below 700 and 1500 m of depth, respectively), although the small deep ocean trends are not significant with respect to the products spread.

Keywords: *ocean reanalysis evaluation, sea level variability, altimetry, gravimetry.*

1 Introduction

Sea level change is a key issue in contemporary climate change, as its global and regional variations are both fundamental indicators of climate change itself and may have a strong impact on human activities in coastal areas.

According to recent estimates (e.g. Cazenave and Llovel, 2010), the contribution of thermal expansion to global mean sea level is of the order of 30% ± 12% for the 1993-2007 period, the remaining contribution being mostly given by glaciers and ice sheet melting. Furthermore, future projections of sea level rise indicate that thermal expansion is likely to continue in the XXI century, and may account for the 32 to 36 % of the global mean sea level rise in 2100 on average, depending on the emission scenario, although large uncertainty is associated with the contribution of land-ice melting (IPCC, 2013).

While mean sea level rise may have different causes, regional sea level rise is generally dominated by the steric component (Fukumori and Wang, 2013) and, in particular, the thermosteric one in most areas (Stammer et al., 2013). Sea level projections tend to confirm this tendency for future scenarios (Sukuki and Ishii, 2011). It is also acknowledged that the sea level low-frequency and inter-annual variability is mostly dominated by the steric sea level variability (Piecuch and Ponte, 2011), although there exists several extra-tropical areas where this simplification does not hold (Piecuch et al., 2013). Attention therefore is being devoted to the monitoring of steric sea level variability and the understanding of the mechanisms causing its long-term rise.

* *Corresponding author address:* Andrea Storto, Centro Euro-Mediterraneo sui Cambiamenti Climatici Viale Aldo Moro 44, I-40135 Bologna, Italy.
E-mail: andrea.storto@cmcc.it

Traditional methods to investigate contemporary steric sea level change are the objective analyses (Ishii et al., 2006; Levitus et al., 2012) that perform a statistical interpolation of the available in-situ observations, usually blended with a climatology, or a persistence background field. These methods, which do not make use of a dynamical ocean model, have the advantages of being relatively simple and computationally cheap. However, objective analyses strongly depend on the model used for the correlation function and on the data availability. For data sparse regions (e.g. the Southern Ocean before the full deployment of the Argo floats observing network during the 2000s), the analysis can be unrealistically close to the climatology fields. Furthermore, recent comparisons of the statistical approaches implemented in the objective analyses highlight that the choice of mapping practices and climatological references is non-negligible for the resulting estimates (Abraham et al., 2013; Chang et al., 2014b). This issue has been proven even more important for steric sea level assessment studies that use a combination of in-situ and satellite observations (Chambers, 2006a; Llovel et al., 2010; Leuliette and Willis, 2011).

An alternative methodology is offered by ocean retrospective analyses (or simply reanalyses or ocean syntheses) that typically combine a data assimilation system capable of ingesting most of the available ocean observations (in-situ and remotely sensed) with an ocean general circulation model (OGCM), with surface boundary conditions usually coming from atmospheric reanalysis or similar products. Ocean reanalyses use the same ocean model and analysis configuration (resolution, parametrizations, numerical schemes, data assimilation setup, etc.) throughout the reanalyzed period. Thus, they aim at building a dataset with quality as consistent and coherent over time as possible, in order to provide a robust tool for climate applications. Closely related to ocean reanalyses are ocean state estimates based on smoother, instead of filter methods, for assimilating data (e.g. Wunsch and Heimbach, 2013). For the present purpose we will not make that distinction, and will refer to them as "reanalyses".

Reanalysis approaches contrast with operational oceanographic analysis systems that aim at achieving the day-by-day best analysis and forecast, eventually upgrading the ocean model and analysis system as soon as improvements are available. Analyses for seasonal forecasting – including some of the analyses used here – fit somewhere in between. Consistency throughout the analysis is again paramount, but the need to transition to near real time with suitable continuity is also required in order that the forecast from the real time products can be properly informed by re-forecasts (or hindcasts).

Many efforts have been recently devoted to the assessment of ocean reanalyses, which have now reached some degree of maturity (Lee et al., 2009). This justifies the idea of promoting a coordinated multi-reanalysis intercomparison, the Ocean Reanalyses Intercomparison Project (ORA-IP, Balmaseda et al., 2015), in the framework of which this work was performed. Provided that OGCMs are becoming reliable tools for climate investigations, the use of reanalyses in contrast to objective analyses seems appealing for climate applications because the scarcity of in-situ observations in the pre-Argo ocean observing network may be partly overtaken by satellite data, information from the atmospheric forcing and the ocean model dynamics. In general, model-based products provide a full view of the ocean state, thus allowing process oriented studies. Even without data assimilation, OGCMs are able to capture the bulk of the inter-annual variability of sea level. For instance, Lombard et al. (2009) showed that non-assimilative hindcasts can reproduce steric sea level trends at regional scale, with a reasonable accuracy; similarly, Griffies et al. (2014), in the frame of the CORE-II simulations, found that OGCMs reproduce regional trends over the last two decades in general agreement with satellite estimates, and dominated by the thermal component.

Recent studies (Purkey and Johnson, 2010; Ponte, 2012) pointed out that deep regions may have a non-negligible contribution to the total steric sea level. This makes the recourse to reanalyses appealing, since observations in the upper ocean may partly constrain the deep ocean through the vertical background-error correlations and the physical balances implied by data assimilation systems and OGCMs, respectively, especially in the future when the accuracy and resolution of OGCMs and reanalyses are expected to increase. However, given the paucity of deep ocean observations, validating reanalyses in the deep ocean is an extremely challenging task, especially to evaluate the reliability of reanalyses in capturing the climate change signal in deep waters. The use of an ensemble of ocean reanalyses has also started to be used for key climate indexes (Stammer et al., 2010), as many ocean research groups are continuously producing ocean reanalyses. The underlying assumption is that the averaging operation over the ensemble members is able to reduce the systematic biases of the individual products.

In this paper, we analyze the steric sea level variability from an ensemble of ocean reanalyses and objective analyses, focusing on the performance of the ensemble means. First, data and methods used within the comparison are introduced in Section 2. Second, a comparison against reference steric sea level estimates is presented for a short 8-year period (2003-2010, Section 3). The comparison is then extended to the 1993-2010 period (Section 4). Finally, the main conclusions are given in Section 5.

2 Data and methods

2.1 Strategy for the comparison

The strategy used in this work consists of two separate phases of steric sea level intercomparison, which span different periods: a first validation period (2003-2010), corresponding to the period when ocean syntheses, altimetric and gravimetric data overlap, and a second extended comparison period (1993-2010) that covers the altimetric missions. Within the validation period, reanalyses are compared with reference steric sea level estimates from altimetry minus gravimetry, as detailed later in this Section. For the extended comparison period, the products are compared to estimate their consistency and uncertainty, focusing on the significance of the mean climate signal.

2.2 Estimation of steric sea level

Following Gill and Niiler (1973), Landerer et al. (2007) and Griffies et al. (2014), the sea level anomaly η can be formulated as a function of the anomalies of three terms with respect to the time mean (the space and time dependence is dropped for simplicity):

$$\eta = \eta_a + \frac{p_b}{g\rho_0} + \eta_s \tag{1}$$

where η_a is the contribution from atmospheric pressure at sea level, $\frac{p_b}{g\rho_0}$ accounts for the bottom pressure effects (change in water mass) and η_s is the steric sea level, which is in turn given by

$$\eta_s = -\int_{-H}^{\eta} \frac{\rho}{\rho_0} dz, \tag{2}$$

$z = H$ being the ocean depth and $z = \eta$ the (time and space varying) ocean surface, the integral covering the whole water column. This formulation corresponds to the "local steric effect" defined by Griffies and Greatbatch (2012) (or "global steric effect" in case of global average, see Section 2.4), to whom the reader is referred for a detailed discussion. Note that the "non-Boussinesq steric effect" is not accounted for, because of the Boussinesq approximation made in OGCMs. The integration of Equation (2) is then approximated to extend from the ocean depth to the mean sea surface $z = 0$. The density anomaly can be approximately decomposed into thermal, haline and pressure contributions by considering the variations of density induced by only temperature (T), salinity (S) or pressure anomalies, respectively, keeping constant the other parameters. The pressure-steric sea level contribution to η_s is negligible and is dropped from the equations (Griffies et al., 2014). Thus, Equation (2) becomes:

$$\eta_s \simeq \eta_t + \eta_h = -\int_{-H}^{0} \frac{\rho(T, \bar{S})}{\rho_0} dz - \int_{-H}^{0} \frac{\rho(\bar{T}, S)}{\rho_0} dz, \qquad (3)$$

where the over-bar denotes the time-averaged value. It was shown by Balmaseda et al. (2012) that the inaccuracies introduced by the decomposition in Equation (3) with respect to Equation (2) are negligible.

The global sea level – the spatial average over the whole ocean of the terms of Equation (1) – is therefore given by a barystatic term that accounts for variations on the ocean water mass (from time-varying contributions of evaporation, precipitation, terrestrial runoff and land ice melting) and a steric term that accounts for the expansion and contraction of water. Note that the term "barystatic" denominates the contribution of changes in the ocean mass to the global mean sea level, and is preferred to the term "eustatic", in accordance with Gregory et al. (2013). The steric component, at global scale, is primarily driven by changes in the seawater temperature, the haline contribution being small since, to first approximation, the global water cycle does not change sufficiently to affect the global salinity budget. At basin and local scales, the bottom pressure term corresponds to the redistribution of the seawater masses and the haline contribution of the steric term may be as important as the thermal contribution (e.g. Ivchenko et al., 2008; Lombard et al., 2009), even for long-term basin-scale effects (Durack et al., 2014). However, the barotropic response of the sea level to a local change of ocean water mass occurs on time scales generally of the order of a few weeks to few months because of its fast adjustment (Ponte, 2006); therefore, within studies encompassing monthly to inter-annual time scales the bottom pressure term, which is often dominated by barotropic fluctuations, acts as a globally uniform term, with few notable exceptions (Landerer et al., 2007; Yin et al., 2010; Piecuch et al., 2013; Griffies et al., 2014).

2.3 Reanalyses and objective analyses

The products participating in the comparison are summarized in Table 1. There are 20 products, of which 16 are reanalyses (REA) and 4 are objective analyses (OA). We consider as objective analyses here the products that do not include any dynamical balance through an ocean general circulation model, but make only a statistical use of observations to estimate the three-dimensional state of the ocean. They are ARMOR, CORA, EN3 and IK09. Another fundamental difference is that objective analyses do not use any information about the air-sea fluxes, unlike ocean reanalyses.

The resolution of the products ranges from eddy-permitting (about 1/4 degree) to much coarser resolution, the majority of products being in the range 1/2 degree to 1 degree of horizontal resolution. A few products (CFSR, ECDA and MOVEC) are the ocean components of assimilative coupled atmosphere ocean general circulation models. Furthermore, 13 reanalyses assimilate sea level anomalies, while one objective analysis (ARMOR) projects onto vertical profiles of temperature and salinity the information from satellite altimetry. All the reanalyses except three products assimilate salinity measurements from in-situ observations. The data assimilation methods include optimal interpolation (OI), three- and four- dimensional variational assimilation (3DVAR, 4DVAR), Kalman filter or related sequential smoother (KF, SS), ensemble optimal interpolation and ensemble Kalman filter (EnOI, EnKF). All reanalyses except three (ECCOV4, ECDA and UR025.4) include an additional constraint to avoid model biases and drifts, either implementing a bias correction scheme or restoring to climatological surface or subsurface fields, or a combination of them. Note however that the restoring time scales, when implemented, may vary notably among the products – from a few days to several years . The different characteristics summarized in Table 1 suggest that the collection of products includes a large diversity of model and data assimilation configurations.

The steric sea level anomaly was calculated by means of Equation (3) for all products, using monthly means of temperature and salinity. The steric sea level fields were then interpolated onto a regular grid of 1x1 degree horizontal resolution, regardless of their native resolution. The same computation was also carried out for a number of depth levels: i) 0-100 m; ii) 100-300 m; iii) 300-700 m; iv) 700-1500 m; v) 1500-3000 m; vi) 3000-4000 m; vii) 4000-bottom. This allows us to have an insight into the contributions of different vertical levels. For each of the vertical levels and the total column, not only the steric sea level but also the thermo- and halo- steric components were provided separately. For the sake of comparison with the verifying dataset (see next Section), only the interior ocean (distance from coast greater than 100 km) and the ocean between 70S and 70N contribute to the global mean steric sea level. We introduce also three derived datasets, ALLENS, REAENS and OAENS, which are the ensemble means calculated from all products, from the reanalyses only, and from the objective analyses only, respectively.

2.4 Reference steric sea level estimates

Previous studies have combined gravimetric data with altimetric data in order to estimate the steric sea level. Lombard et al. (2007) and Willis et al. (2008) compared steric sea level inferred from altimetry minus gravimetry with steric sea level objectively analyzed from in-situ observations. They were not able to close the sea level global budget, namely to match the total sea level from altimetry with the the sum of gravimetry and Argo-derived steric sea level, within the error bars of the observational networks, and suggested that the discrepancies may be related to the sampling of the in-situ observing network or inaccuracies in the processing of one or more observing systems. Later, an improved processing of the observational datasets allowed Leuliette and Miller (2009) to close the budget, demonstrating the complementarity of gravimetry, altimetry and Argo network in estimating the sea level budget. Since then, the methodology of combining the different observing networks has largely been explored, also for studies at the basin scale (Garcia-Garcia et al., 2010; Chang et al., 2014a).

Steric sea level estimates for use in the comparison have been calculated from Equation (1) by formulating steric sea level as a difference of total sea level minus bottom pressure and atmospheric pressure. For both terms, we decompose the time-varying sea level anomaly terms into globally averaged and spatially varying sea level variations:

$$\eta(x, y, t) =< \eta(t) >_D +\widetilde{\eta}(x, y, t), \quad (4)$$

where $< ... >_D$ is the spatial averaging operator and $\widetilde{\eta}(x, y, t)$ can be computed as the difference between the previous two terms and has spatial average equal to zero. This allowed us to optimize the processing of data depending on whether the analysis focuses on global mean or regional sea level. For the global mean sea level $< \eta(t) >_D$, we used the dataset of Nerem et al. (2010), which represents a seamless dataset cross-calibrating the TOPEX and JASON altimeter missions. The spatially varying term of Equation (4) for the total component $(\widetilde{\eta}(x, y, t))$ is provided by the AVISO delayed-time monthly gridded altimetric products, with the time-varying global mean removed. Altimetric observations were subject to the usual geophysical removals and multi-satellite cross-correction (Le Traon et al., 1998). For the bottom pressure term (Chambers and Schröter, 2011), the global mean value $< p_b(t) >_D$ was taken from Johnson and Chambers (2013), which uses the Gravity Recovery and Climate Experiment (GRACE) RL05 data only in the ocean interior to avoid possible land and ice contamination. The spatially-varying term $(\widetilde{p_b}(x, y, t))$ was taken from release RL05 of GRACE gravimetric data (Chambers and Bonin, 2012), which is provided with the area-weighted global mean set to zero. The release RL05 disseminates data after the application of a destriping procedure and a 500 km wide Gaussian filter to remove the meridional stripes typical of gravimetric data. The attenuation of systematic biases in the processing chain of GRACE data was attempted using the ensemble mean of the three RL05 releases from CSR (Center for Space Research, University of Texas), GFZ (GeoForschungsZentrum, the German Research Centre for Geosciences) and JPL (Jet Propulsion Laboratory), mostly differing in the data pre-processing. Note that we do not estimate steric sea level errors from the altimetric and gravimetric data errors, as it is beyond the scope of this work.

Results from the construction of the verifying dataset are presented in Figure 1, which reproduces the zonal averages of the 2003-2010 monthly means for the three components (total, bottom pressure, and steric sea level inferred from the previous two). The total sea level seasonality is primarily affected by the heat content seasonal cycle, peaking in September (March) for the Northern (Southern) Hemisphere. Consequently, it is dominated by a strong hemispheric separation during all months. On the other hand, the bottom pressure signal does not exhibit a hemispheric separation. This is due to the fact that at monthly time-scales bottom pressure is rather uniform and driven by the Northern Hemisphere seasonal cycle of water stored inland (snow, ice) and in rivers, reservoirs and underground (Leuliette and Willis, 2011). The resulting steric sea level (bottom panel) resembles the seasonal cycle of the total sea level, but with amplitudes and phases modified according to the bottom pressure component.

Although many studies have suggested the low signal-to-noise ratio of GRACE data over the oceans (e.g. Chambers, 2006b) and the critical role of the data processing in estimating inter-annual trends and variability from gravimetry (Quinn and Ponte, 2010), the use of such data to infer steric sea level allows us to build a validation dataset for the ocean syntheses and further test the consistency between the altimetric and gravimetric datasets and the ocean reanalyses. In other words, the validation allows us to identify the products that better close the sea level budget, given the altimetric and gravimetric missions. The validation dataset is not strictly independent, as most of the products assimilate altimetric data. However, without going into detail, there is a tremendous diversity in the methods that altimetry is assimilated with, ranging from hydrostatic adjustments (e.g. Storto et al., 2011), to simplified barotropic and baroclinic adjustments (e.g. Fukumori et al., 1999), to combined analytic and statistical balances (e.g. Weaver et al., 2005). This suggests that steric sea level estimates from altimetry and gravimetry represent a useful validation dataset, despite the assimilation of altimetric data in most products.

Hereafter, ALT-GRV will denote the steric sea level estimates presented in this Section.

3 Validation period (2003-2010)

3.1 Global steric sea level comparison

In this Section, globally averaged values of the steric sea level of reanalyses are compared to that obtained by altimetry minus gravimetry for the 2003-2010 period. The globally averaged estimates from all products and the validation dataset are shown in Figure 2. For comparison with the verifying dataset (see Section 2.4), global averages of the reanalyses and objective analyses include only the ocean between 70S and 70N. In the figure legend, we report the correlations of the full, seasonal and inter-annual signals, with respect to the verifying dataset. The way through which the signal is decomposed is detailed in the Appendix. Generally, all products except three exhibit a significant correlation (0.21 is the minimum significant correlation according to a two-sided t-test with 95% confidence level). The most skillful product is ARMOR, which shows very high correlations for all the signals.

The performance of the ensemble means are also very satisfactory, with values of (full signal) correlations of 0.79, 0.80 and 0.71 for the ensemble of all products, the ensemble of reanalyses and the ensemble of objective analyses, indicating that the ensemble mean of the reanalyses outperforms the ensemble mean of the objective analyses. The correlation difference between REAENS and OAENS is significant, according to a Steiger's Z test (Steiger, 1980), with 95% confidence level, for two dependent overlapping correlations. Note that not only the correlation of REAENS is greater than OAENS, but is also greater than that of ALLENS, indicating that increased ensemble sizes do not ensure better skill scores. This result is due to the fact that two objective analyses out of four present a negative correlation with the verifying dataset for the inter-annual signal, namely they are not able to capture the year-by-year variations seen by the altimetry minus gravimetry dataset.

The decomposition of the correlation coefficient for the seasonal and inter-annual signal (in the legend of Figure 2) suggests also that the correlation of the seasonal signal is almost always greater than that of the inter-annual signal. Because of the definition of the seasonal signal (corresponding to the detrended signal, see the Appendix), this indicates that linear trends of most of the products are not in agreement with the verifying dataset and are likely affected by biases and drifts. Conversely, the seasonality of the steric sea level is generally well-captured.

To refine the analysis of the seasonal and inter-annual components of the global steric sea level, Figure 3 shows two components (annual and inter-annual), in which we have divided the signal. The semi-annual component is not shown, for sake of simplicity. The first panel (annual component) shows the amplitude and phase in polar coordinates. The majority of the products under-estimate the annual amplitude with respect to ALT-GRV (5.3 mm for ALT-GRV against 3.9 mm for REAENS). OAENS, on the contrary, shows an amplitude of 5.1 mm, very close to the one observed. Almost all products capture the phase of the annual global steric sea level, whose peak occurs in mid-April. The semi-annual amplitude (not shown) shows less consistency among the products, with an amplitude generally under-estimated.

The second panel of Figure 3 shows the linear trends of the global steric sea level for all the products, with the 95% confidence level estimated by means of the bootstrap methodology. There is a large diversity in the trend values, provided that ALT-GRV exhibits a value of 1.0 +/- 0.4 mm yr^{-1}, while the products' trends range from about -2 to 2 mm yr^{-1}, with 9 products (along with OAENS) showing a negative trend. REAENS shows a weakly positive trend (0.2 +/- 0.3 mm yr^{-1}). Although 2003-2010 is a short time period to have a climatologically significant trend, here the purpose is to evaluate the consistency or the diversity of the products in estimating trends in comparison with the verifying dataset. For this well observed period, there is no clear consensus on the inter-annual trend, i.e. the trend of the global mean steric sea level from the individual products has large variability. There may be several reasons for such a result, including the sensitivity of the trend calculation over a short period, the impact of the abrupt change of the observing network (Argo network deployment) in data assimilation systems, or the global mean imbalances in the atmospheric reanalyses used to force the models.

3.2 Regional steric sea level comparison

The regional steric sea level has been evaluated by correlating the point-by-point time-series of the individual products with the verifying dataset. This is summarized in Table 2, which reports for each product and the ensemble means the area averages of the point-by-point correlation. Regional correlations for the ensemble means resemble the correlation coefficients of the global steric sea level (by comparison of Table 2 with Figure 2). However, 11 products exhibit spatially averaged regional correlations greater than that of global steric sea level while 9 products present the opposite behavior, suggesting that there is no clear evidence whether the products capture the regional signal better than the global one. However, all the correlation scores in the Table 2 are significant. For all products the correlation in the tropics is greater than the global average and the one averaged in the southern extra-tropics is smaller, indicating the strong latitudinal dependence of the correlation coefficient.

The top panel of Figure 4 shows the correlation map for REAENS with the verifying dataset, indicating the very high correlation globally (the global mean correlation is 0.8, see Table 2), except at high latitudes. Only south of about 60S, the reanalysis ensemble mean does not show significant correlation (less than 0.21).

The correlation is close to 1 in the tropical band, where reanalyses are most successful at capturing the evolution of density anomalies. Decreases in correlation can be seen in regions around the Kuroshio Extension, the Gulf Stream, the Falkland current, and off the coast of Peru. However, in these regions the ensemble spread among the products is high (not shown), a consequence of the high eddy variability, which most models do not resolve.

The bottom panel shows the map of correlation difference with the verifying dataset between REAENS and OAENS. Where the differences are positive (negative), the ensemble of reanalyses outperforms (underperforms) the ensemble of objective analyses. The picture thus suggests that while in the tropics the differences are negligible due to the high correlations of both datasets, at mid and high latitudes the differences are evident and reveal the better fit of REAENS to the verifying estimates of steric sea level. This is particularly evident for the Southern Ocean, especially in the Indian and Atlantic sectors, where differences reach values up to 0.4. The correlation difference between REAENS and ALLENS (not shown) verifies that ensemble size does not compromise the comparison.

Provided that the correlation coefficients are computed over monthly means and that the annual amplitude of steric sea level is greater than the inter-annual trend and variability (see Figure 3), the general high correlation resides in the high seasonality of the steric signal that may be easily captured by the products. To appreciate the effect of the seasonality on the skill scores previously presented, we show in Figure 5 the same correlation maps of Figure 4 but with the seasonal signal removed. Note that the correlation for the case with inter-annual signal removed behaves closely to that of the full signal, and is not shown. The correlation of REAENS with the ALT-GRV dataset (top panel) shows in this case a generally lower correlation, confirming that the steric sea level seasonality importantly contributes to the correlation scores. However, correlations are still significant all over the global ocean, except south of 60S and over a few other areas (Arabian Sea and Angola Basin) and still quite high, close to 1, in the Tropical Pacific except over the areas corresponding to the Intertropical and South Pacific Convergence Zones. However, within the latter areas the inter-annual variability of the steric sea level is generally small (not shown), contributing to the local decrease of the correlation score. Note also that the dominance of the seasonal signal on the correlation of the full signal may also explain, for some products, the low correlation scores, due to a change of phase in the seasonal signal.

The correlation difference map (bottom panel) exhibits similar patterns to those of Figure 4. However, these differences are slightly larger, confirming that reanalyses capture the steric sea level inter-annual variability better than objective analyses, as for the global mean steric sea level.

There are several interpretations for the ensemble reanalyses to have higher correlations especially in the Southern Ocean. A first interpretation is given by the fact that in this region the in-situ observing network is poor, even for the Argo floats (see e.g. Figure 1d in Storto et al. (2013)). In this case, the background used by objective analyses, which is usually either a climatology or a previous analysis, is not impacted by the analysis step because of the scarcity of observations and does not change, producing no inter-annual variability. On the other hand, reanalyses can still take advantage of space-borne measurements (SLA and SST) over ice-free areas – and SIC over ice-contaminated areas. They also use the information from the atmospheric forcing that, despite its large uncertainty at high latitudes, still provides a time-varying forcing.

Another concurrent explanation is that the Southern Ocean is a current system where deep variations have large contributions to steric sea level variability. A way to show this is given by Figure 6, which reproduces the REAENS explained variance of the 700 m to bottom steric sea level contribution to the total steric sea level. Percentage values of the explained variance are generally low (less than 10%), suggesting that the steric sea level variability can be generally explained by the contribution of the first 700 m. An exception is the Southern Ocean, where percentage values are in the range 20% to 50%, in agreement with a generally deeper mixed layer depth, on the average, in this region (see e.g. de Boyer Montégut et al., 2004). Similar results are found when the linear trend is removed or the explained variance is calculated from OAENS and do not seem affected by possible drifts of individual members (not shown). The sea level variability is characterized by deep steric contributions in the Southern Ocean. This may be better simulated by reanalyses than by objective analyses through the inclusion of the atmospheric forcing, satellite data and the dynamical ocean balances.

4 Extended intercomparison period (1993-2010)

In this Section, we compare the products for the extended period 1993-2010. Within this period, we do not consider any validation dataset and we focus on the statistical significance of the ensemble mean with respect to the ensemble spread (defined hereafter as the standard deviation over the ensemble members). Different regions and vertical levels are considered.

4.1 Global steric sea level comparison

The comparison for global steric sea level is summarized in Figure 7. The figure shows the globally averaged values of steric sea level and its thermo- and halo- steric components for the individual products (gray lines), the ensemble mean of the reanalyses and objective analyses (cyan and blue lines, respectively) and their ensemble spreads (red and orange lines, respectively). Note that y-axes are different for the ensemble means and spreads. The analysis of total steric sea level suggests that although discrepancies and outliers are visible among the individual products, the two ensemble means are reasonably close. This is confirmed by the correlation between the two ensemble means, equal to 0.93. The spread is also generally comparable, although the size of the two ensembles is different. The shape of the spread stems from the definition of the steric sea level data as anomalies with respect to the whole 1993-2010 period. Consequently, different linear trends of the individual products lead to a convex parabola-like shape for the spread.

Qualitatively similar results were found for the thermosteric component, for which the correlation between the two ensemble means is even higher (0.97), and the spread has a very similar magnitude and behavior. We observe a more pronounced seasonality for the reanalysis ensemble spread, while an abrupt increase of spread for the objective analyses in 2003 is likely to be related to the increase of Argo observations. The match of the two ensemble means for halosteric sea level (bottom panel of Figure 7) is lower, the correlation dropping to 0.55. In particular, the seasonal variability is different, with a seasonal amplitude in REAENS of 0.7mm being considerably less than the seasonal amplitude of 2.7mm in OAENS. This may be related to inaccurate representation of the global freshwater budget in many of the reanalyses, and in particular to the fact that some reanalyses unrealistically show an inter-annual variability larger than the seasonal variability.

Although the spread of the halosteric component is smaller than that of the thermosteric one, their values are higher than the variability of the signal. This is illustrated in Table 3, where we report the standard deviation of the ensemble mean of the global time series, along with their time-averaged spread and the normalized spread (the ratio between the spread and the ensemble mean standard deviation), for both reanalyses and objective analyses. While for the steric and thermo-steric sea level the ratios between the spread and the signal variability are smaller than unity, both reanalyses and objective analyses show a value greater than 1 (1.61 and 1.71) for the normalized spread in case of the halosteric component. This means that the uncertainty is greater than the variability of the signal, suggesting the low reliability of the halosteric time-series. The same exercise of computing the normalized spread is repeated for four different reanalysis subgroups, categorized according to possible additional constraints on salinity, i.e. no constraint, bias correction, restoring to climatological sea surface salinity or restoring to subsurface climatological salinity. Results are reported in Table 3, whose caption details the procedure to compute these values accounting for different group sizes. Products with no constraints have small spread also associated with small variability, i.e. the ratio is equal to 1.35. The use of bias correction increases both spread and variability, leading to a ratio comparable with the one from the reanalysis group with no salinity constraints. By using the restoring to sea surface salinity or subsurface salinity fields, reanalyses have an intermediate spread and variability, leading to a smaller normalized spread (1.20 and 1.12, respectively), still greater than 1. This suggests that, based on our ensemble of reanalyses, while the restoring helps in slightly decreasing the normalized spread, namely increasing the signal-to-noise of the halosteric component, all subgroups of reanalyses have a spread exceeding the variability, suggesting that other issues such as the global freshwater budget uncertainty linked to the uncertainty in the modeling of the hydrological cycle together with the scarcity of salinity observations lead to this result, regardless of the bias and drift correction possibly implemented in the reanalyses.

By looking at the geographical patterns of the normalized spread (not shown), the steric and thermosteric components have values less than 1 everywhere except in the ACC region, with minimum values in the Equatorial region (high signal-to-noise ratio). On the contrary, the halosteric component shows values greater than unity everywhere (low signal-to-noise ratio), with high values especially in the Atlantic Ocean, indicating the large uncertainty of the salinity content in this basin.

The reanalyses spread also exhibits local maxima at the beginning and in the middle of the period, with a parabola-like shape (minimum or maximum of the halosteric sea level at the middle of the period) due to a few outlier products.

Global steric sea level trends are reported in Figure 8 in mm/yr. The total steric sea level trends range from 0.1 to 3.1 mm/yr, and are all positive. For the ensemble means, the trends have values of 1.02 $\pm$ 0.05 mm/yr and 1.11 $\pm$ 0.08 mm/yr for the reanalyses and objective analyses, respectively, with a standard deviation of the trends (red bars) of the order of 0.5 mm/yr for both. Similar results apply to the thermosteric sea level, although two products exhibit negative trends. For the ensemble means, the trends equal 1.03 $\pm$ 0.05 mm/yr and 1.17 $\pm$ 0.05 mm/yr, for the reanalyses and objective analyses, respectively, the standard deviation of trends being equal to 0.6 mm/yr for both. The halosteric trend exhibits no clear consistency between the products, ranging from about -0.8 to 0.9 mm/yr, most products showing a negative trend. The reanalyses ensemble mean filters out these discrepancies, exhibiting almost no trend (-0.01 $\pm$ 0.01 mm/yr). OAENS shows a slightly negative trend (-0.07 $\pm$ 0.06 mm/yr).

4.2 Regional steric sea level comparison

4.2.1 Steric sea level trends and their significance

Maps of linear trends (ensemble mean of all products) are presented in Figure 9 for the total steric sea level and the two components separately. The total steric and thermosteric sea level look very similar, i.e. the local trends are in general dominated by the thermosteric component. Well-known maxima of the trends are found in the western tropical Pacific (up to 8 mm/yr). Also shown are signal-to-spread ratios (SSR, see the Appendix), with solid (dashed) contours referring to significant positive (negative) trends, i.e. with absolute SSR values greater than 1. Areas of significant positive trend are found in the western tropical Pacific, the central north Pacific, the Indonesian Archipelago and the southern Indian Ocean, a few areas part of the ACC and some areas in the Atlantic Ocean, in particular within the tropics and in the North Atlantic subpolar gyre and the Labrador Sea. Significant negative trends (up to -4 mm/yr) are found only in the eastern Pacific Ocean and in the Alaskan gyre. The patterns of thermosteric trends are similar, with a more pronounced positive trend in the Labrador and North Atlantic subpolar gyre. (up to 7 mm/yr). The halosteric component trend map shows only few regions with SSR values greater than 1. While there are several regions with positive significant trends within the ACC and in the Western Tropical Pacific, the only significant negative trends are located within the North Pacific subpolar gyre. Negative trends characterize the Atlantic Ocean, showing the typical compensating effect between the halosteric and thermosteric components (Lowe and Gregory, 2006).

To better understand why halosteric trends are non-significant almost everywhere, we report in Figure 10 a graphic showing the percentage of the global ocean area with significant trend, as a function of the starting and ending year for the trend computation. We consider 5 years as minimum period for calculating the trend. Trends are considered significant if their ensemble mean exceeds the spread of the trends. We also report the number of in-situ observations for clarity. The figure provides insights into the significance of the trends, i.e. the capability of capturing the steric trends in the altimetric period. Generally, thermosteric trends are more significant: the percentage of area with significant trends is almost always above 20%, peaking at 37%. For the halosteric component, this percentage is below 7% except for the last 5-year period, i.e. for trends computed starting at least in 2002 and ending at least in 2008, peaking in the 2006-2010 period with a value of about 15%. For periods starting before 2002, there are small differences in the significant area for the halosteric trends, suggesting a close relation with the number of observations, below 2 million per year before 2001. In other words, before the full deployment of the Argo floats, only a small percentage of the ocean show significant trends, questioning the reliability of the halosteric trend estimates. In contrast, the number of temperature observations is larger (between 3 and 6 million before 2002, generally three to four times larger than the amount of salinity observations). Thermosteric trends are also particularly significant in the last period (for trends computed starting from 2003). However, there are periods of increasing (for trends including the 2000 and 2001 years or the 2004 to 2006 period) and decreasing (for trends starting in 1999 or 2000) percentage, which suggests that the change in observation coverage (i.e. Argo floats deployment) is not the only cause for these variations. In particular, these anomalous increases are primarily related to the ENSO variability. For instance, by comparing the maps of significance for thermosteric trends ending in 2010 but starting either in 1998 or in 1999 (not shown), it turns out that the loss of significance is located in the Western Tropical Pacific, related to a La Nina event. On the contrary, the increase of significance when the years 2001 or 2004-2005 fall within the trend computation results from the inclusion of El Niño events. To summarize, halosteric trends are very dependent on the observational coverage and the assimilation method. Thermosteric trends are, however, more robust. This finding complements the previous result about the low signal-to-noise ratio of the global halosteric component due to the freshwater budget uncertainty.

4.2.2 Thermo- and halo- steric contributions

A secondary objective of the intercomparison is the quantification of the thermal and haline steric sea level contributions to the total steric sea level. This is summarized in Figure 11 in terms of linear trends and explained variance of the full and inter-annual signals for the main ocean basins (see also the Appendix). The figure also reports the standard deviation of the contribution among all products in order to evaluate the significance of the results. All basins exhibit positive trends for the thermosteric contribution. This positive trend is significant with respect to the ensemble spread (signal greater than the spread) except in the Southern Ocean. The Atlantic Ocean exhibits the largest trend (1.7 mm/yr). The halosteric trend never appears significant except in the Atlantic Ocean, which presents a salinification corresponding to a negative trend of the steric sea level equal to -0.5 mm/yr. While non-significant, the Southern Ocean exhibits a positive trend in halosteric sea level, consistent with the recent freshening (Böning et al., 2008). In the Southern Ocean the two contributions are comparable (about 0.3 mm/yr), in agreement with Purkey and Johnson (2010, 2013).

All basins report a value for the halosteric sea level explained variance between 15% and 25%, except for the Southern Ocean where the value reaches 36% (Figure 11, middle and bottom panels). Similar values are found for the inter-annual only signal, except for the Southern Ocean. The latter presents a value of about 45% for the halosteric contribution, indicating that the inter-annual variability of halosteric sea level is comparable to the thermosteric component.

4.2.3 Contributions from vertical levels

In this Section we examine the contributions of the different depth levels. Figure 12 reports the same diagnostics of Figure 11 for the 7 vertical levels analyzed. The global steric sea level exhibits a positive and significant trend for the top 700 m of depth, peaking between 100 and 300 m with around 0.39 mm/yr. Depths below 700 m show no significant trends, with abyssal waters (4000 m to bottom) contributing with a slightly negative trend (-0.08 mm/yr). Similar qualitative results apply for all the basins, with some notable exceptions: i) in the Atlantic Ocean, the largest contribution comes from the 300-700 m level (0.37 mm/yr); ii) the 100-300 m Pacific waters exhibit the largest positive contribution (0.58 mm/yr) along with the Indian waters in the 700-1500 m level. The Pacific abyssal waters also showed the largest negative trend (-0.12 mm/yr); iii) the Southern Ocean presents small values for the vertical contributions, all of them non-significant except in the 100-300 m level (0.19 mm/yr $\pm$ 0.12 mm/yr). Note also that the spread of trends in deep ocean layers are of the same magnitude of upper ocean trends, indicating that while there is decrease of contribution to linear trends from top to the bottom of the ocean, the uncertainty remains unchanged, leading to decreased SSR and suggesting that reanalysis trends below 700 m are not robust (except for the Indian Ocean).

The explained variance shows a predominant role of the 0-100 m level, because of the seasonality of the the air-sea fluxes. This is particularly notable in the Indian and Southern Oceans, where the explained variance is equal to 80 and 65%, respectively, against 43% at global scale. Below 100 m, the values decrease and, in particular below 1500 m, the fluctuation among the products is larger than the value itself, indicating the different behavior of the products in representing the deep ocean variability.

When the seasonal signal is removed (bottom panel of Figure 12), the contribution of the top 100 m decreases. The 100-300 m level gives the main contribution at global scale (30%). The Atlantic Ocean shows a similar fraction of explained variance (around 25%) provided by the 0-100, 300-700 and 700-1500 m levels. The Pacific Ocean shows a dominant contribution of the 0-300 m level, due to the dominant role of the near-surface tropical waters. Finally, the Southern Ocean provides higher values for deeper levels. In particular, the contribution of the waters below 1500 m depth sums up to 22% of the inter-annual variability, in agreement with Ponte (2012), reaching 47% for the waters below 700 m.

Trend maps for thermosteric and halosteric contributions to sea level from the 0-700 m, 700-1500 m, 1500-4000 m and the 4000-bottom levels are shown in Figure 13 and 14 respectively, along with the contour lines corresponding to a SSR equal to 1. The 0-700 m thermosteric trend is found very similar to the 0-bottom trend (middle panel of Figure 9), with all significant patterns being located in the same areas. Intermediate waters (between 700 and 1500 m) show significant positive trends (up to 2.5 mm/yr) in the Labrador Sea and in a some areas of the ACC, while non-significant elsewhere. The evident warming of the intermediate waters in the Labrador Sea is in agreement with in-situ measurements (Avsic et al., 2006), and may be related to a reduction of deep convection in this region (Schott and Brandt, 2007).

Deep waters (between 1500 and 4000 m) generally show positive trends in the Atlantic Ocean (up to 4 mm/yr in the Labrador Sea), which is however smaller than the spread of the ensemble trends.

Abyssal waters (bottom right panel) show smaller trends, generally negative in the Indo-Pacific Ocean and positive in the Atlantic Ocean, except in the Gulf Stream region. The latter feature agrees with observation-based studies (e.g. Purkey and Johnson, 2010), which highlight the abyssal warming in the western Atlantic basin coming from the warming of the Antarctic Bottom Water. Only a few significant areas are visible in the Western Pacific Ocean. The Pacific Ocean cooling appears to be related to the outlying behavior of a few individual members that exhibit a large negative in temperature for that area. Our current approach is to treat all products equally, but in the future it may be better to discard outlying members.

For the halosteric trends, the maps confirm small trends everywhere. For the top 700 m, there exist significant positive values in the ACC, in the region of the South Pacific Convergence Zone and in the Western North Pacific zone, while significant negative trends are located only around the Alaskan gyre and in some Atlantic areas. This map agrees well with the total halosteric map, indicating the significance of the Southern Ocean freshening. The maps of the intermediate and deep waters suggest a negative (significant) trend in the Indonesian Throughflow region, up to -4 mm/yr, and in the Alaskan gyre, other areas exhibiting a smaller and non-significant trend. Finally, although non-significant, the ensemble mean of the abyssal water halosteric trend shows negative contributions in the North Atlantic and North and South Pacific gyres, with peaks of -0.6 mm/yr, while the Gulf Stream region is characterized by a positive trend (up to 0.5 mm/yr). In contrast, the Indian Ocean shows a uniform positive trend (0.2 mm/yr).

5 Summary and discussion

We have analyzed steric sea level variability from an ensemble of global ocean reanalyses and objective analyses during the period 1993-2010, in the framework of the ORA-IP project.

The relatively large number of global ocean products included in the comparison allowed us to follow a multi-system ensemble approach and exploit the statistical properties of the ensemble to detect the consistencies among the different products. It should be noted that this is in contrast to the atmospheric reanalysis community, where the number of the state-of-the-art reanalyses is in general of the order of 4 to 6, rendering an ensemble approach more difficult. We believe that the large number and variety of ocean reanalyses should be extensively exploited in the future. For instance, by objectively discarding ocean reanalyses with an outlier behavior or weighting the reanalyses in a super-ensemble context (e.g. Krishnamurti et al., 2000), it will be possible to construct an optimal steric sea level dataset, which would also contain uncertainty estimates. At this stage, however, the goal of this comparison is not to build such a dataset, but rather to evaluate the performance of the reanalyses and their ensemble mean.

As a preliminary step towards the assessment of the products included in the comparison, we have constructed a validation dataset for the period 2003-2010 by combining altimetric and gravimetric satellite data. This allowed us to evaluate the performance of the individual products and that of the ensemble mean, separated into ensemble means of reanalyses and objective analyses. Within the 8-year validation period, the individual products are satisfactorily able to capture the global steric sea level seasonality, while they show large discrepancies in the inter-annual trends. The ensemble means agree well at both the global and regional scale with the altimetry minus gravimetry dataset, and prove a valuable tool for potential use in studies encompassing longer time periods. Furthermore, the use of the ensemble spread to evaluate the significance of important climate signals such as inter-annual trends seems an appealing strategy, and reinforces the importance of sustaining the development and production of multiple reanalyses and adopting a multi-reanalysis approach to ocean variability studies at inter-decadal scale.

The steric sea level variability in the tropics is particularly well represented in the ensemble of products with a correlation close to 1 with the validating data set. In general, we found that the ensemble mean of the reanalyses outperforms that of the objective analyses at both the global and regional scale. This is particularly evident for the Southern Ocean at the inter-annual scale, and we speculate that this may be caused by the small amount of in-situ observations and the strong contributions to the total steric sea level of deep ocean layers in this area, associated to the fact that reanalyses make use of satellite data and bear information about time-varying atmospheric forcing, affected by large uncertainty, though.

The comparison was then extended to the 1993-2010 period, showing close agreement of the ensemble of reanalyses and objective analyses in reproducing the steric and thermosteric sea level, with a high cross-correlation. Consequently, the estimates of global steric sea level trends are similar (1.02 ± 0.05 and 1.11 ± 0.06 mm/yr, respectively) and in agreement with recent estimates (Hanna et al., 2013, and references therein). However, we did not find a significant consensus among the products for the halosteric trends, at both global and regional scale, although we showed that in a few specific regions (e.g. the Southern Ocean) its inter-annual variability has effects comparable to the thermosteric component.

Given the approximate linear relationship between global mean halosteric and barystatic sea level variations due to the land-ice melting affecting both (Munk, 2003), a global mean barystatic sea level rise of 2 mm/yr roughly corresponds to a global mean halosteric sea level rise of about 0.05 mm/yr. Our comparison thus indicates that many ocean syntheses have unrealistically large global mean halosteric changes, further to a large ensemble spread. The disagreement among the global mean halosteric sea level estimates reflects the uncertainty of the freshwater budget and the disparity among its different modeling, intrinsic inaccuracies in the ocean model (e.g. deep ocean variability), as well as data assimilation assumptions, especially for the pre-Argo era when salinity is in fact unconstrained or constrained by temperature profiles or sea surface observations only or climatological assumptions. The global freshwater budget is poorly known, and most reanalyses adopt a climatological representation for the continental runoff, therefore unable to follow the variations of the continental ice melting. Along with the commonly adopted strategy of climatological restoring of the sea surface and/or subsurface salinity, this allows on one hand to mitigate biases arising from atmospheric forcing and ocean model inaccuracies; on the other hand, it forces the products towards a climatological state and neglects climate change signals (Griffies et al., 2014). In the future, more sophisticated methods to correct biases and drifts without compromising the climate change signal should be explored.

Only a very small region of the global ocean has a statistically significant trend in the halosteric component, although this has recently proven to be non-negligible even at basin scale (Durack et al., 2014). Our ability to measure the temporal variability of this component is hampered by the scarcity of salinity observations before the deployment of the Argo float network, with a clear implication regarding the optimal design and maintenance of the in-situ observing network. This questions the reliability of the estimates of the halosteric inter-annual variability before the 2000s. It will be essential in the future to evaluate whether the next generation of ocean reanalyses is able to reduce the relative uncertainty of the halosteric component by making better use of the few observations available (for instance by improving the cross-parameter correlations, the air-sea coupling, the representation of the global freshwater budget, etc.) or this feature intrinsically resides in the characteristics of the pre-Argo observing network. This gives further validation to the idea of promoting and maintaining inter-comparison exercises in the future. However, provided that the halosteric component explains only a small portion of the total steric component at both global and regional scales, its uncertainty does not result in large uncertainties in the total steric sea level.

Qualitatively similar conclusions can be drawn when analyzing the separate impact of different depth levels. While we found a non-negligible effect of the deep waters on inter-annual variability – about 29% and 12% of explained variance for the waters below 700 and 1500 m (± 9% and ± 7%), respectively, with a 1500 m to bottom explained variance over 20% in the Atlantic and Southern Oceans, steric sea level trends contributed by the waters below 700 m are generally non-significant with respect to the ensemble spread. Thus, while the ensemble mean reproduces some notable processes, such as the abyssal warming in the Western Atlantic basin (Purkey and Johnson, 2010), their quantification remains difficult, and the reanalyses are not able to provide robust results.

Acknowledgments.

During the preparation of this article, our co-author Nicolas Ferry passed away. He was an active and supportive member of the ORA-IP and CLIVAR-GSOP activities. This work has received funding from the Italian Ministry of Education, University and Research and the Italian Ministry of Environment, Land and Sea under the GEMINA project and from the European Commission Copernicus program, previously known as GMES program, under the MyOcean and MyOcean2 projects. The authors thank the CLIVAR Project Office for supporting the participation in conferences and workshops for presenting preliminary results about the steric sea level comparison. Simon Good was supported by the Joint UK DECC/Defra Met Office Hadley Centre Climate Programme (GA01101). K.A. Peterson was supported by the UK Public Weather Service research programme. The GLORY2V3 reanalysis has been developed at Mercator Ocèan with the support of the European Commission funded projects MyOcean (FP7-SPACE-2007-1) and MyOcean2 (FP7-SPACE-2011-1). GRACE ocean data were processed by Don P. Chambers, supported by the NASA MEaSUREs Program, and are available at http://grace.jpl.nasa.gov. The authors would like to thank Don Chambers (USF) for his help in the correct use of gravimetry data. Mean sea level data were provided by the Sea Level Research Group, University of Colorado. The altimeter products were produced by Ssalto/Duacs and distributed by AVISO, with support from CNES (http://www.aviso.oceanobs.com/duacs/). The authors are also very grateful to three anonymous reviewers and the editor for their precious suggestions and valuable comments, and for the improvement of the quality of this paper.

Appendix: Mathematical definitions

We briefly introduce in this Appendix some mathematical definitions that are used in the paper.

Signal-to-spread Ratio

In order to evaluate how distinguishable is the climate signal of the reanalysis ensemble with respect to its uncertainty, we define the signal-to-spread ratio (SSR), or more generally, the signal-to-noise ratio of the ensemble for a generic parameter p as

$$SSR = \frac{EM}{ES} = \frac{< p >}{\sqrt{1/N \sum_{i=1}^{i=N} (p_i - < p >)^2}} \tag{5}$$

with EM and ES being the ensemble mean and spread, respectively, and N being the ensemble size, with

$$< p > = 1/N \sum_{i=1}^{i=N} p_i. \tag{6}$$

Note that the ensemble spread is defined as the sample standard deviation. Values of SSR smaller (greater) than 1 indicate that the discrepancy of reanalyses is greater (less) than their mean signal.

Annual and Seasonal Decomposition

It is also useful to decompose the steric sea level signal onto the seasonal (annual and semi-annual) and linear trend (inter-annual) components. To do this, we assume that every time-series of the variable x be of the form:

$$x(t) = mt + c + A_a \cos(\frac{2\pi}{12}t - \varphi_a) + A_s \cos(\frac{2\pi}{6}t - \varphi_s) + \varepsilon(t) \tag{7}$$

where t is the time (in months), m is the linear trend, A_a and φ_a are the annual amplitude and angular phase, respectively, and A_s and φ_s are the semi-annual amplitude and angular phase and ε are the residuals. The decomposition is carried out by a least-squares fitting of Equation (7), i.e. by minimizing the sum of $\varepsilon^2(t)$.

The seasonal signal x_S, introduced in the text in Sections 3 and 4, is defined as the full signal minus the linear trend:

$$x_S(t) = x(t) - mt, \tag{8}$$

namely it corresponds to the detrended signal. Conversely, the inter-annual signal x_I, is defined as the full signal to which the fitted seasonal signal is subtracted:

$$x_I(t) = x(t) - A_a \cos(\frac{2\pi}{12}t - \varphi_a) + A_s \cos(\frac{2\pi}{6}t - \varphi_s). \tag{9}$$

In the above definitions, only the complementary fitted signal is removed, while the residuals are always kept. Note that the time-series corresponding to the inter-annual and seasonal signals have the same length of the time-series of the full signal, implying that the same minimum values for testing the significance of the correlations apply.

Explained Variance

The (percentage) explained variance of a component y with respect to the (total) component z is defined as

$$EV(y) = 100\frac{VAR(z) - VAR(z-y)}{VAR(z)}, \tag{10}$$

with $VAR(...)$ being the variance operator. When the explained variance of the inter-annual signal is introduced (e.g. in Figure 11), it means that the explained variance is calculated on the timeseries, after removal of the seasonal signal, for both components y and z.

References

Abraham, J., et al., 2013: A review of global ocean temperature observations: Implications for ocean heat content estimates and climate change. *Rev. Geophys.*, **51**, 450–483.

Avsic, T., J. Karstensen, U. Send, and J. Fischer, 2006: Interannual variability of newly formed Labrador Sea Water from 1994 to 2005. *Geophys. Res. Lett.*, **33**, L21S02.

Balmaseda, M., et al., 2015: The Ocean Reanalyses Intercomparison Project (ORA-IP). *J. Oper. Oceanogr.*, **7**, 31 pp, accepted for publication.

Balmaseda, M. A., K. Mogensen, and A. Weaver, 2012: Evaluation of the ECMWF ocean reanalysis system ORAS4. *Q. J. R. Meteorol. Soc.*, **139**, 1132–1161.

Behringer, D., 2007: The global ocean data assimilation system at NCEP. *11th Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans and Land Surface*, Amer. Meteor. Soc.

Blockley, E. W., et al., 2014: Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts. *Geoscientific Model Development Discussions*, **6 (4)**, 6219–6278, doi:10.5194/gmdd-6-6219-2013, URL `http://www.geosci-model-dev-discuss.net/6/6219/2013/`.

Böning, C., A. Dispert, M. Visbeck, S. Rintoul, and F. Schwarzkopf, 2008: The response of the Antarctic Circumpolar Current to recent climate change. *Nature Geoscience*, **1**, 864–869.

Cabanes, C., et al., 2013: The CORA dataset: validation and diagnostics of in-situ ocean temperature and salinity measurements. *Ocean Sci.*, **9**, 1–18.

Cazenave, A. and W. Llovel, 2010: Contemporary Sea Level Rise. *Annu. Rev. of Marine Sci.*, **2**, 145–173.

Chambers, D., 2006a: Observing seasonal steric sea level variations with GRACE and satellite altimetry. *J. Geophys. Res.*, **111**, C03 010, doi:10.1029/2005JC004836.

Chambers, D. and J. Bonin, 2012: Evaluation of Release-05 GRACE time-variable gravity coefficients over the ocean. *Ocean Sci.*, **8**, 859–868.

Chambers, D. and B. Schröter, 2011: Measuring ocean mass variability from satellite gravimetry. *J. Geodyn.*, **52**, 333–343.

Chambers, D. P., 2006b: Evaluation of new GRACE timevariable gravity data over the ocean. *Geophys. Res. Lett.*, **33**, LI7603, doi:10.1029/2006GL027296.

Chang, Y., S. Zhang, A. Rosati, T. Delworth, and W. Stern, 2013: An assessment of oceanic variability for 1960-2010 from the GFDL ensemble coupled data assimilation. *Clim. Dynam.*, **40**, 775–803.

Chang, Y.-S., A. Rosati, and G. A. Vecchi, 2014a: Basin patterns of global sea level changes for 2004-2007. *J. Mar. Syst.*, **80**, 115–124.

Chang, Y.-S., G. A. Vecchi, A. Rosati, S. Zhang, and X. Yang, 2014b: Comparison of global objective analyzed T-S fields of the upper ocean for 2008-2011. *J. Mar. Syst.*, **137**, 13–20.

de Boyer Montégut, C., G. Madec, A. S. Fischer, A. Lazar, and D. Iudicone, 2004: Mixed layer depth over the global ocean: an examination of profile data and a profile-based climatology. *J. Geophys. Res.*, **109**, c12003.

Durack, P., S. Wijffels, and P. Gleckler, 2014: Long-term sea-level change revisited: the role of salinity. *Environ. Res. Lett.*, **9**, 114 017–1–11.

Forget, G. and R. Ponte, submitted: The Partition of Regional Sea Level Variability. *Progress in Oceanography.*

Fujii, Y., N. Nakaegawa, S. Matsumoto, T. Yasuda, G. Yamanaka, and M. Kamachi, 2009: Coupled climate simulation by constraining ocean fields in a coupled model with ocean data. *J. Climate*, **22**, 5541–5557.

Fukumori, I., 2002: A partitioned Kalman filter and smoother. *Mon. Wea. Rev.*, **130**, 1370–1383.

Fukumori, I., R. Raghunath, L. Fu, and Y. Chao, 1999: Assimilation of TOPEX/POSEIDON data into a global ocean circulation model: How good are the results? *J. Geophys. Res.*, **104**, 25 647–25 665.

Fukumori, I. and O. Wang, 2013: Origins of heat and freshwater anomalies underlying regional decadal sea level trends. *Geophys. Res. Lett.*, **40**, 563–567.

Garcia-Garcia, D., B. Chao, and J.-P. Boy, 2010: Steric and mass-induced sea level variations in the Mediterranean Sea revisited. *J. Geophys. Res.*, **115**, C12 016, doi:10.1029/2009JC005928.

Gill, A. and P. P. Niiler, 1973: The theory of the seasonal variability in the ocean. *Deep-Sea Res.*, **20**, 141–177.

Gregory, J., et al., 2013: Twentieth-Century Global-Mean Sea Level Rise: Is the Whole Greater than the Sum of the Parts? *J. Climate*, **26**, 4476–4499.

Griffies, S. and R. Greatbatch, 2012: Physical processes that impact the evolution of global mean sea level in ocean climate models. *Ocean Model.*, **51**, 37–72.

Griffies, S., et al., 2014: An assessment of global and regional sea level for years 1993-2007 in a suite of interannual CORE-II simulations. *Ocean Model.*, **78**, 35–89.

Guinehut, S., A. Dhomps, G. Larnicol, and P.-Y. Le Traon, 2012: High resolution 3D temperature and salinity fields derived from in situ and satellite observations. *Ocean Sci.*, **8**, 845–857.

Haines, K., M. Valdivieso, H. Zuo, and V. Stepanov, 2012: Transports and budgets in a 1/4 global ocean reanalysis 1989-2010. *Ocean Sci.*, **8**, 333–344.

Hanna, E., et al., 2013: Ice-sheet mass balance and climate change. *Nature*, **498**, 51–59.

Ingleby, B. and M. Huddleston, 2007: Quality control of ocean temperature and salinity profiles - Historical and real-time data. *J. Mar. Syst.*, **65**, 158–175.

IPCC, 2013: *Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change.* Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 1535 pp.

Ishii, M., M. Kimoto, K. Sakamoto, and S.-I. Iwasaki, 2006: Steric sea level changes estimated from historical ocean subsurface temperature and salinity analyses. *J. Oceanogr.*, **62**, 155–170.

Ivchenko, V., S. Danilov, D. Sidorenko, J. Schröter, M. Wenzel, and L. Aleynik, 2008: Steric height variability in the Northern Atlantic on seasonal and interannual scales. *J. Geophys. Res.*, **113**, C11 007, doi:10.1029/2008JC004836.

Johnson, G. and D. Chambers, 2013: Ocean bottom pressure seasonal cycles and decadal trends from GRACE Release-05: Ocean circulation implications. *J. Geophys. Res.*, **118**, 4228–4240.

Köhl, A., 2014: Evaluation of the GECCO2 Ocean Synthesis: Transports of Volume, Heat and Freshwater in the Atlantic. *Q. J. R. Meteorol. Soc.*, doi:10.1002/qj.2347.

Krishnamurti, T., C. Kishtawal, Z. Zhang, T. Larow, D. Bachiochi, and E. Williford, 2000: Multimodel ensemble forecasts for weather and seasonal climate. *J. Climate*, **13**, 4196–4216.

Landerer, F., J. Jungclaus, and J. Marotze, 2007: Regional Dynamic and Steric Sea Level Change in Response to the IPCC-A1B Scenario. *J. Phys. Oceanogr.*, **37**, 296–312.

Le Traon, P., F. Nadal, and N. Ducet, 1998: An improved mapping method of multi-satellite altimeter data. *J. Atmos. Ocean. Technol.*, **15**, 522–534.

Lee, T., T. Awaji, M. Balmaseda, E. Grenier, and D. Stammer, 2009: Ocean State Estimation for Climate Research. *Oceanography*, **22**, 160–167.

Leuliette, E. and W. Miller, 2009: Closing the sea level rise budget with altimetry, Argo, and GRACE. *Geophys. Res. Lett.*, **36**, L04 608, doi:10.1029/2008GL036010.

Leuliette, E. and J. Willis, 2011: Balancing the sea level budget. *Oceanography*, **24**, 122–129.

Levitus, S., et al., 2012: World ocean heat content and thermosteric sea level change (0-2000 m), 1955-2010. *Geophys. Res. Lett.*, **39**, L10 603.

Llovel, W., S. Guinehut, and A. Cazenave, 2010: Regional and interannual variability in sea level over 20022009 based on satellite altimetry, Argo float data and GRACE ocean mass. *Ocean Dynam.*, **60**, 1193–1204.

Lombard, A., G. Garric, and T. Penduff, 2009: Regional patterns of observed sea level change: insights from a 1/4 global ocean/sea-ice hindcast. *Ocean Dyn.*, **59**, 433–449.

Lombard, A., et al., 2007: Estimation of steric sea level variations from combined GRACE and Jason-1 data. *Earth Planet. Sci. Lett.*, **254**, 194–202.

Lowe, J. and J. Gregory, 2006: Understanding projections of sea level rise in a Hadley Centre coupled climate model. *J. Geophys. Res.*, **111**, C11 014.

Masuda, S., et al., 2010: Simulated Rapid Warming of Abyssal North Pacific Waters. *Science*, **329**, 319–322.

Munk, W., 2003: Ocean Freshening, Sea Level Rising. *Science*, **300**, 2041–2043.

Nerem, R., D. Chambers, C. Choe, and G. Mitchum, 2010: Estimating Mean Sea Level Change from the TOPEX and Jason Altimeter Missions. *Mar. Geodesy*, **33**, 435–446.

Piecuch, C. and R. Ponte, 2011: Mechanisms of interannual steric sea level variability. *Geophys. Res. Lett.*, **38**, L15 605.

Piecuch, C., K. Quinn, and R. Ponte, 2013: Satellite-derived interannual ocean bottom pressure variability and its relation to sea level. *Geophys. Res. Lett.*, **40**, 3106–3110.

Ponte, R. M., 2006: Oceanic response to surface loading effects neglected in volume-conserving models. *J. Phys. Oceanogr.*, **36**, 426–434.

Ponte, R. M., 2012: An assessment of deep steric height variability over the global ocean. *Geophys. Res. Lett.*, **39**, L04 601, doi:10.1029/2011GL050681.

Purkey, S. and G. Johnson, 2010: Warming of Global Abyssal and Deep Southern Ocean Waters between the 1990s and 2000s: Contributions to Global Heat and Sea Level Rise Budgets. *J. Climate*, **23**, 6336–6351.

Purkey, S. and G. Johnson, 2013: Antarctic Bottom Water Warming and Freshening: Contributions to Sea Level Rise, Ocean Freshwater Budgets, and Global Heat Gain. *J. Climate*, **26**, 6105–6122.

Quinn, K. J. and R. M. Ponte, 2010: Uncertainty in ocean mass trends from GRACE. *Geophys. J. Intern.*, **181**, 762–768, doi:10.1111/j.1365-246X.2010.04508.x.

Schott, F. and P. Brandt, 2007: *Ocean Circulation: Mechanisms and Impacts - Past and Future Changes of Meridional Overturning*, chap. Circulation and Deep Water Export of the Subpolar North Atlantic During the 1990's, 91–118. Edited by A. Schmittner, J. C. H. Chiang and S. R. Hemming. American Geophysical Union, Washington, D. C.

Stammer, D., A. Cazenave, R. Ponte, and M. Tamisiea, 2013: Causes for contemporary regional sea level changes. *Annu. Rev. of Marine Sci.*, **5**, 21–46.

Stammer, D., et al., 2010: Ocean Information Provided through Ensemble Ocean Syntheses. *Sustained Ocean Observations and Information for Society*, OceanObs'09, 21-25 September 2009, J. Hall, D. Harrison, D. Stammer Eds., ESA Publication WPP-306.

Steiger, J., 1980: Tests for comparing elements of a correlation matrix. *Psychological Bull.*, **87**, 245–251.

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

Storto, A., S. Masina, and S. Dobricic, 2013: Ensemble spread-based assessment of observation impact: application to a global ocean analysis system. *Q. J. R. Meteorol. Soc.*, **139**, 1842–1862.

Storto, A., S. Masina, and S. Dobricic, 2014: Estimation and Impact of Nonuniform Horizontal Correlation Length Scales for Global Ocean Physical Analyses. *J. Atmos. Ocean. Technol.*, **31**, 2330–2349.

Sukuki, T. and M. Ishii, 2011: Regional distribution of sea level changes resulting from enhanced greenhouse warming in the Model for Interdisciplinary Research on Climate version 3.2 version. *Geophys. Res. Lett.*, **38**, L02 601.

Toyoda, T., Y. Fujii, T. Yasuda, N. Usui, T. Iwao, T. Kuragano, and M. Kamachi, 2013: Improved analysis of the seasonal-interannual fields by a global ocean data assimilation system. *Theoret. and Appl. Mech. Japan*, **61**, 31–48.

Weaver, A. T., C. Deltel, E. Machu, S. Ricci, and N. Daget, 2005: A multivariate balance operator for variational ocean data assimilation. *Q. J. R. Meteorol. Soc.*, **131**, 605–625.

Willis, J., D. Chambers, and R. Nerem, 2008: Assessing the globally averaged sea level budget on seasonal to interannual timescales. *J. Geophys. Res.*, **113**, C06 015.

Wunsch, C. and P. Heimbach, 2013: *Ocean Circulation and Climate - A 21st century perspective, International Geophysics Series, Vol.103*, chap. Chapter 21: Dynamically and Kinematically Consistent Global Ocean Circulation and Ice State Estimates, 553–579. Edited by G. Sielder, J. Church, S. Griffes, J. Gould, and J. Church. Academic Press, Elsevier.

Xue, Y., B. Huang, Z. Hu, A. Kumar, C. Wen, D. Behringer, and S. Nadiga, 2011: An Assessment of Oceanic Variability in the NCEP Climate Forecast System Reanalysis. *Clim. Dynam.*, **37**, 2511–2539.

Yin, J., S. Griffies, and R. Stouffer, 2010: Spatial Variability of Sea Level Rise in Twenty-First Century Projections. *J. Climate*, **23**, 4585–4607.

Yin, Y., O. Alves, and P. Oke, 2011: An ensemble ocean data assimilation system for seasonal prediction. *Mon. Wea. Rev.*, **139**, 786–808.

Table 1: List of products participating the steric sea level comparison, with details about a bibliographic reference, the product type, the spatial resolution and the data assimilation configuration. The symbol + indicates that the product is the ocean component of a coupled ocean-atmosphere system. For the resolution, the symbol * indicates that the computational grid has a resolution refinement in the Equatorial region. The penultimate column reports the data assimilation method used and the observations assimilated (T: in-situ observations of temperature; S: in-situ observations of salinity; SST: satellite observations of sea surface temperature; SLA: altimetric observations). The last column reports additional constraints included in the product (SSS: restoring to climatological sea surface salinity; 3D: three-dimensional restoring to climatological fields ; BC: bias-correction). The product GLORYS2V3 is called GLORYS2 hereafter.

Product Name	Producing Institute	Reference	Type of Product	Resolution	Assimilation Method (Obs)	Additional Constraint
ARMOR	CLS	Guinehut et al. (2012)	OA	1/3 x 1/3	OI(T, S, SST, SLA)	NO
CFSR+	NOAA/NCEP	Xue et al. (2011)	REA	1/2 x 1/2*	3DVAR(T, SST)	SSS
C-GLORS	CMCC	Storto et al. (2014)	REA	1/2 x 1/2	3DVAR(T, S, SST, SLA)	SSS + BC
CORA	Ifremer	Cabanes et al. (2013)	OA	1/3 x 1/3	OI(T, S)	NO
ECCO-NRT	JPL/NASA	Fukumori (2002)	REA	1 x 1*	KF-SS(T, SLA)	SSS
ECCOV4	MIT/AER/JPL	Forget and Ponte (submitted)	REA	1 x 1*	4DVAR(T, S, SST, SLA)	NO
ECDA+	NOAA/GFDL	Chang et al. (2013)	REA	1 x 1*	EnKF(T, S, SST)	NO
EN3	UK MetOffice	Ingleby and Huddleston (2007)	OA	1 x 1	OI(T, S)	3D
GECCO2	Univ. Hamburg	Köhl (2014)	REA	1 x 1*	4DVAR(T, S, SST, SLA)	SSS + 3D
GEOS5	NASA/GMAO		REA	1/2 x 1/2	EnOI(T, S, SST, SLA)	SSS
GLORYS2V3	MERCATOR		REA	1/4 x 1/4	KF(T, S, SST, SLA)	BC
GLOSEA5	UK MetOffice	Blockley et al. (2014)	REA	1/4 x 1/4	3DVAR(T, S, SST, SLA)	SSS + 3D
GODAS	NOAA/NCEP	Behringer (2007)	REA	1 x 1*	3DVAR(T, SST, SLA)	SSS
IK09	JAMSTEC	Ishii et al. (2006)	OA	1 x 1	OI(T, S)	NO
K7OC (ESTOC)	JAMSTEC	Masuda et al. (2010)	REA	1 x 1	4DVAR(T, S, SST, SLA)	BC
MOVEC+	MRI/JMA	Fujii et al. (2009)	REA	1 x 1*	3DVAR(T, S, SST, SLA)	SSS + 3D + BC
MOVEG2	MRI/JMA	Toyoda et al. (2013)	REA	1 x 1/2*	3DVAR(T, S, SST, SLA)	3D
ORAS4	ECMWF	Balmaseda et al. (2012)	REA	1 x 1*	3DVAR(T, S, SST, SLA)	SSS + 3D + BC
PEODAS	BoM/CAWCR	Yin et al. (2011)	REA	2 x 1*	EnKF(T, S, SST)	SSS + 3D
UR025.4	Univ. Reading	Haines et al. (2012)	REA	1/4 x 1/4	OI(T, S, SST, SLA)	NO

Table 2: Spatially area averages of the point-by-point temporal correlation (2003-2010) of the steric sea level product with the verifying dataset (ALT-GRV). The four regions are defined with respect to latitudinal bands: between 60S and 60N for the global average, and between 20N and 60N, 20S and 20N and 60S and 20S for the Northern Extra-Tropics, the Tropics and the Southern Extra-Tropics, respectively.

Product	Globally Averaged Correlation	Northern Extra-Tropics Correlation	Tropics Correlation	Southern Extra-Tropics Correlation
ARMOR	0.762	0.760	0.853	0.673
CFSR	0.527	0.511	0.701	0.364
C-GLORS	0.841	0.852	0.917	0.762
CORA	0.531	0.509	0.730	0.346
ECCO-NRT	0.618	0.564	0.828	0.438
ECCOV4	0.631	0.580	0.835	0.455
ECDA	0.526	0.502	0.759	0.308
EN3	0.512	0.484	0.726	0.315
GECCO2	0.616	0.501	0.801	0.488
GEOS5	0.555	0.490	0.816	0.329
GLORYS2	0.865	0.876	0.906	0.820
GLOSEA5	0.831	0.896	0.902	0.731
GODAS	0.514	0.493	0.729	0.313
IK09	0.541	0.535	0.754	0.335
K7OC	0.449	0.465	0.533	0.359
MOVEC	0.647	0.687	0.784	0.494
MOVEG2	0.707	0.698	0.869	0.552
ORAS4	0.627	0.579	0.861	0.419
PEODAS	0.565	0.522	0.778	0.375
UR025.4	0.757	0.806	0.895	0.599
REAENS	0.799	0.787	0.901	0.704
OAENS	0.647	0.640	0.818	0.482
ALLENS	0.780	0.766	0.894	0.675

Table 3: Standard deviation, spread and normalized spread of the two ensemble means (REA: reanalyses; OA: objective analyses) for the global steric, thermo- and halo- steric sea level during the period 1993-2010. The table reports in mm the standard deviation of the ensemble means, along with the time-averaged ensemble spread and the ratio between the latter and the standard deviation. The same computation is reported also for the halosteric sea level, by catagorizing the reanalyses in four different groups, depending on constraints on salinity (NOREST: no restoring to SSS, nor to subsurface fields; BCORR: bias correction scheme implemented; SREST : restoring to climatological SSS; 3DREST: restoring to climatological subsurface fields). For the latter, products were subsampled by computing the spread and variability of all combinations of two members included in each group, and then averaging over the combinations, in order to avoid the influence of the group size on the results. Note that each product may belong to more than one group.

Products	Steric Component	Monthly Variability	Mean Spread	Normalized Spread
REA	Steric	8.12	6.40	0.79
	Thermosteric	7.17	6.78	0.95
	Halosteric	2.94	5.02	1.71
OA	Steric	6.27	5.22	0.83
	Thermosteric	6.23	5.35	0.86
	Halosteric	0.75	1.21	1.61
NOREST	Halosteric	2.03	2.75	1.35
BCORR	Halosteric	3.85	4.91	1.28
SREST	Halosteric	3.37	4.05	1.20
3DREST	Halosteric	3.79	4.26	1.12

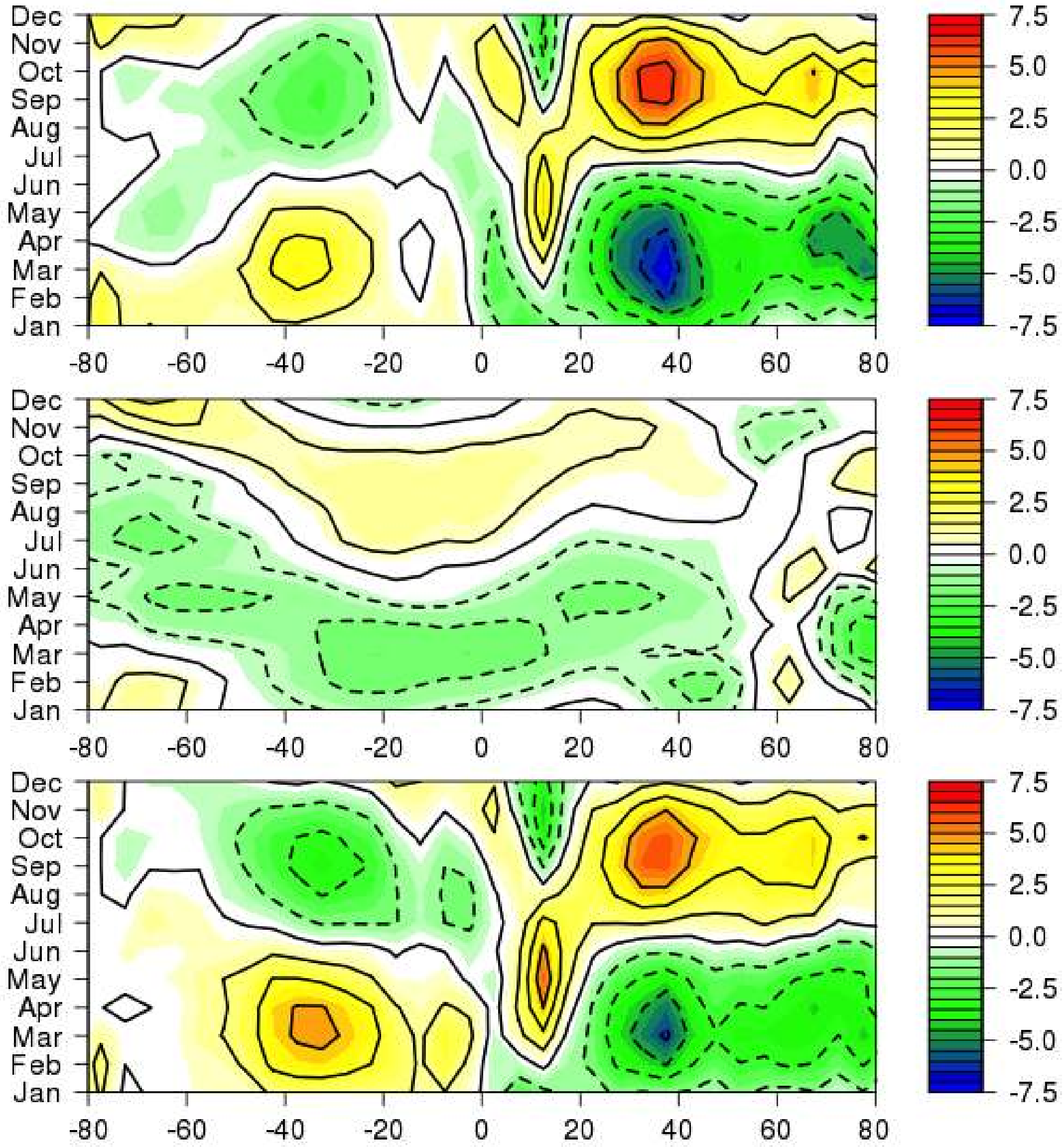

Figure 1: Seasonal cycle (2003-2010 mean) as a function of latitude for the total sea level (top panel) derived by satellite altimetry, mass component (middle panel), derived by satellite gravimetry and steric sea level ALT-GRV (bottom panel), derived by subtracting the barystatic sea level from the total sea level. Data processing is explained in the text. Units are cm.

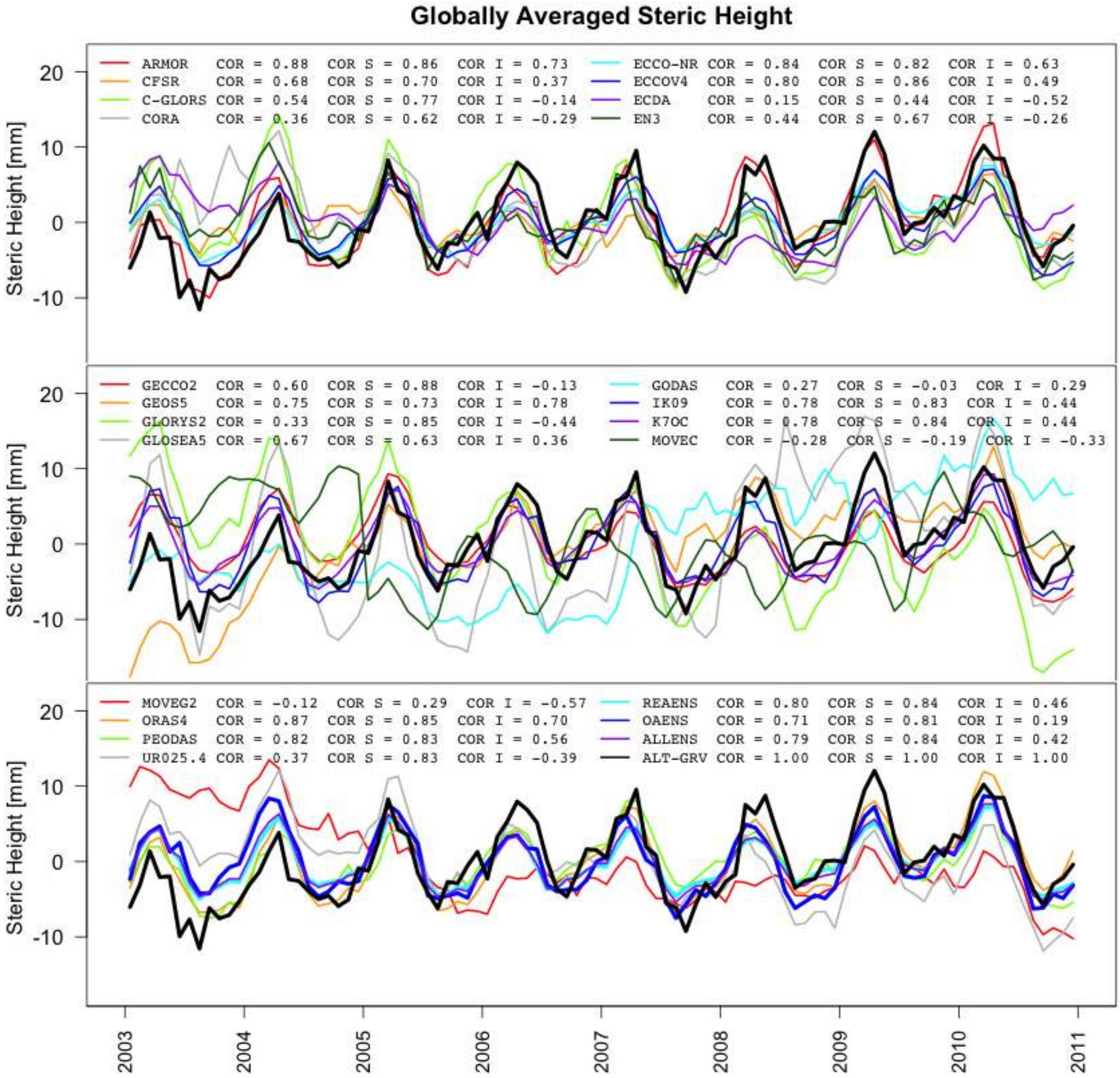

Figure 2: Monthly time-series of global steric sea level for the period 2003-2010 for the different products and the verifying dataset (black lines). The temporal correlation of each product with the verifying dataset for the full signal (COR), the seasonal signal (COR S, i.e. interannual signal removed) and the inter-annual signal (COR I, i.e. seasonal signal removed) is also shown.

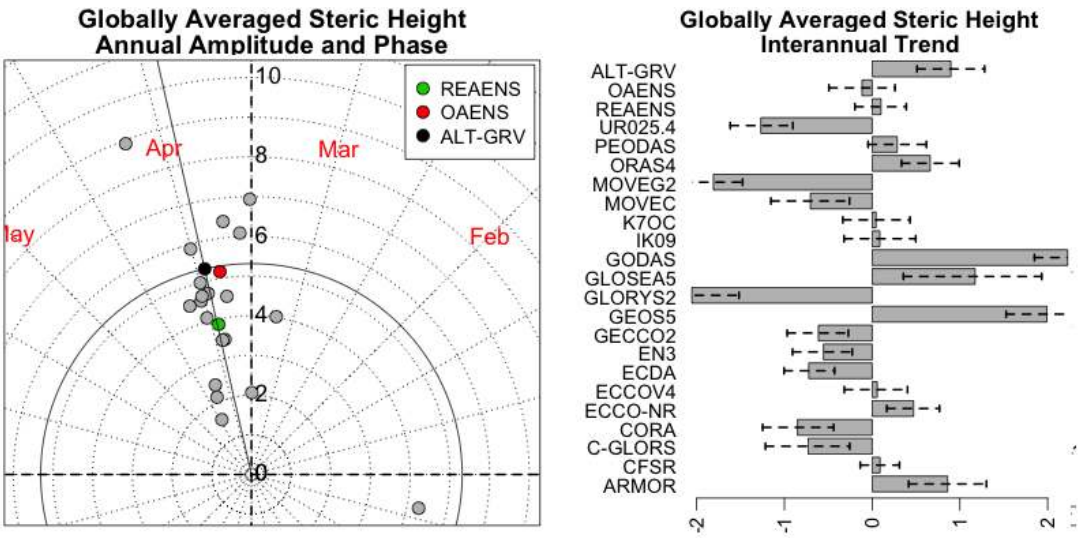

Figure 3: Decomposition of the 2003-2010 steric sea level in annual and inter-annual (linear trend) components for all the products (see Appendix for the definitions). For the annual component, the plot shows the amplitude and phase in polar coordinates, with the radius corresponding to the amplitude and the angle with respect to the x-axis to the phase (corresponding to the maximum reached in the annual cycle), for the individual products (gray circles), the ensemble of the reanalyses (green circle) and the objective analyses (red circle) and the verifying dataset (black circle). The y-axis reports the values for the amplitude in mm, while the phase is given in months from the beginning of the year, respectively, and reported in red. Note that the month labels are located at the middle of the month, and the radial graduation is by half month. For the inter-annual linear trend, units are mm yr^{-1}.

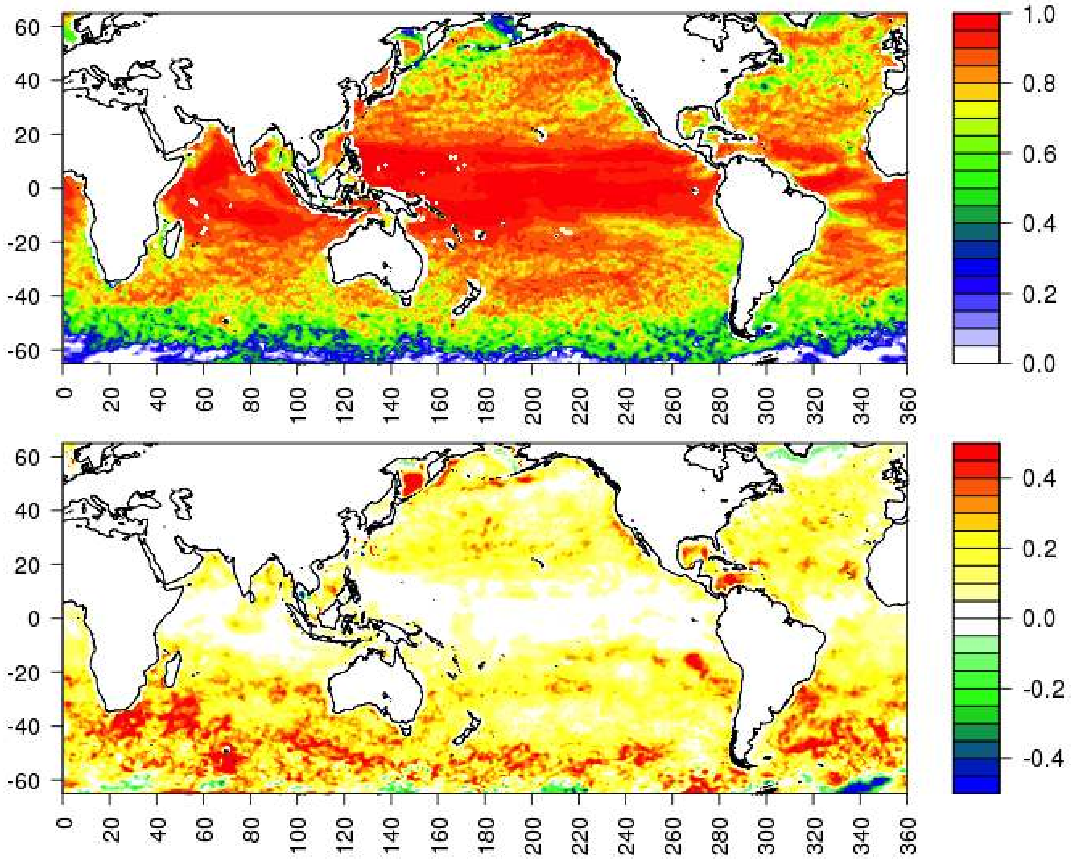

Figure 4: Correlation map of the reanalysis ensemble with the verifying dataset (top panel) and difference of correlation with the verifying dataset between REAENS and OAENS (bottom panel). Note that the color palettes are different.

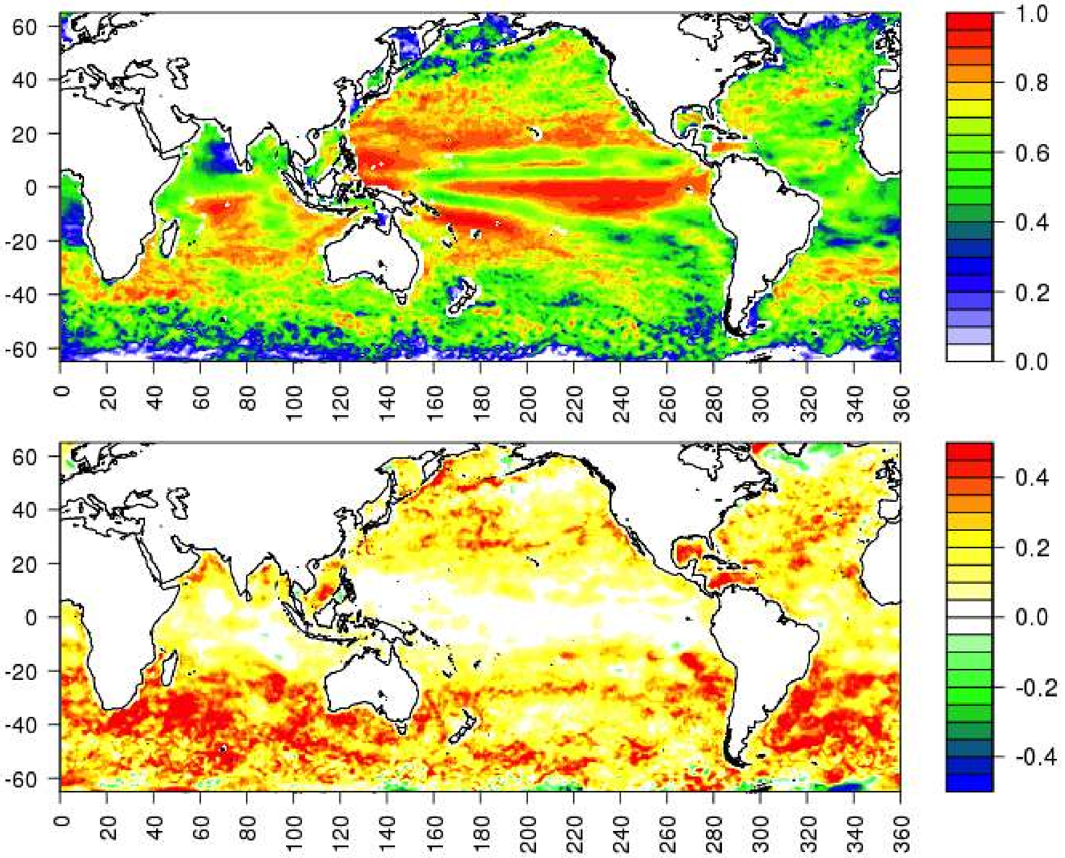

Figure 5: As in Figure 4, but for the inter-annual signal only (seasonal signal removed).

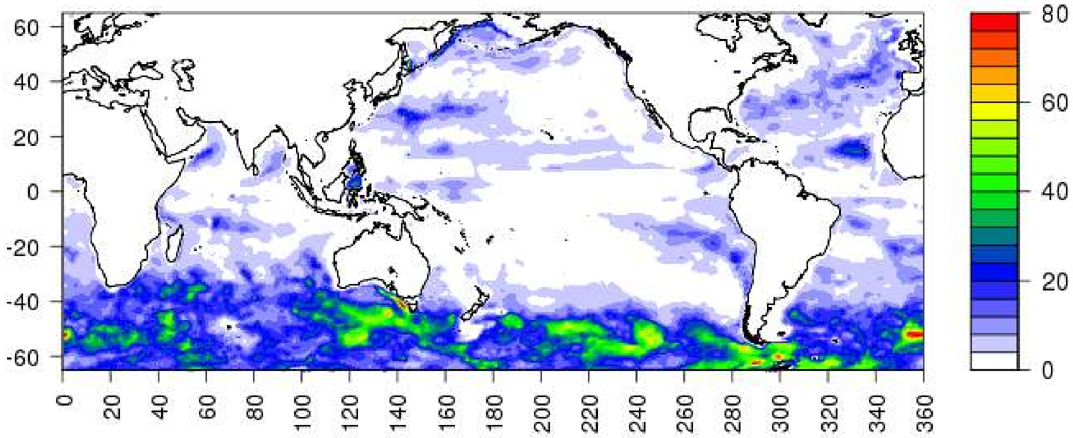

Figure 6: Explained variance of the steric sea level in the layer 700 m to bottom with respect to the surface to bottom variance, calculated on the reanalysis ensemble mean during the period 2003-2010. Units are %. The figure provides a quantitative evaluation of the percentage impact of the relatively deep ocean (below 700 m of depth) on the steric sea level variability.

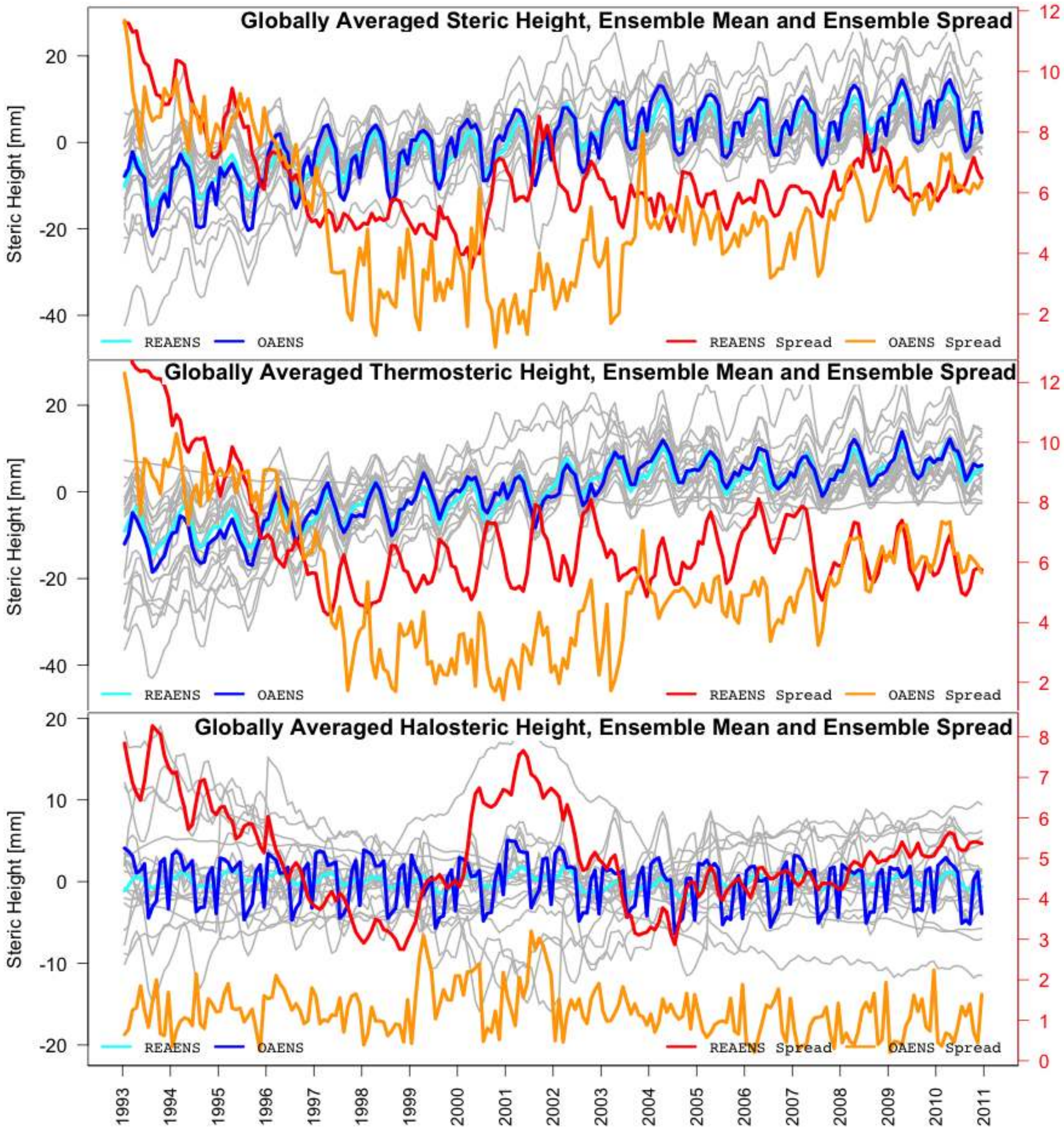

Figure 7: Monthly time-series of global steric (top panel), thermosteric (middle panel) and halosteric (bottom panel) sea level for the period 1993-2010 (Extended Intercomparison Period). Gray lines correspond to the individual products, while the cyan (blue) line corresponds to the ensemble mean of reanalyses (objective analyses). Also shown in red (orange) is the time-series of the ensemble spread from the reanalyses (objective analyses). For the ensemble spreads, the y-axis is in red on the right-side of the panels and the unit is mm.

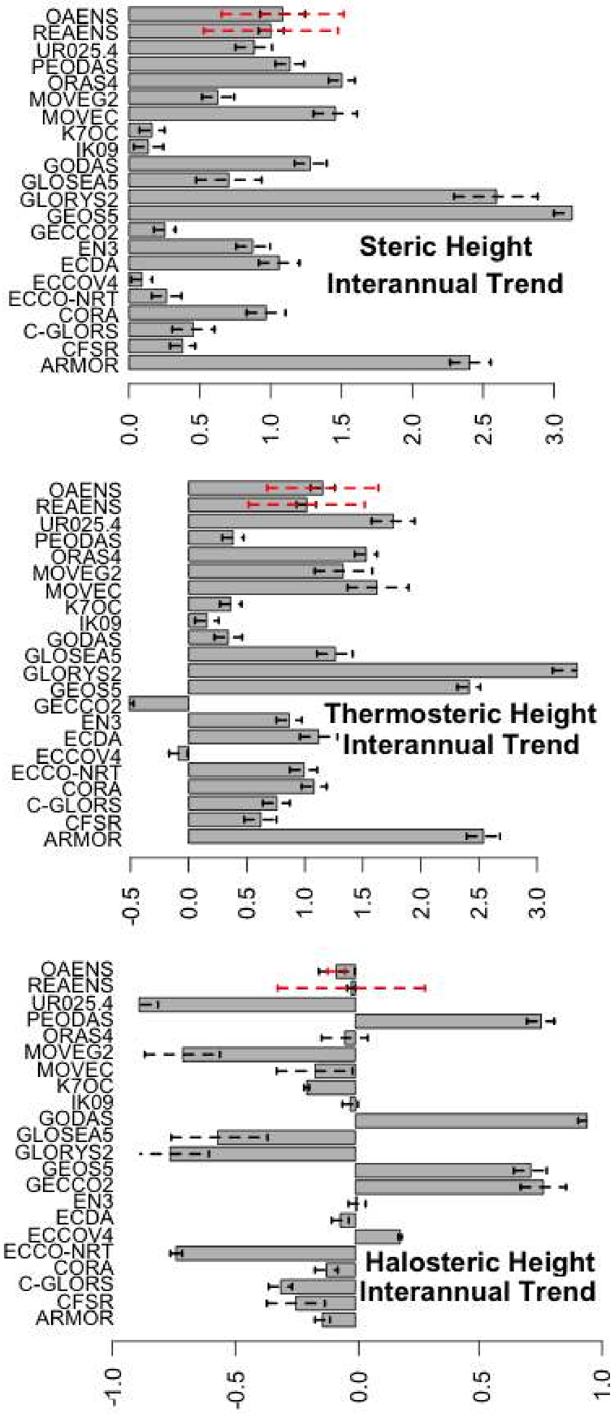

Figure 8: Global sea level linear trends (1993-2010) from all the products, the ensemble mean of reanalyses (REANS) and objective analyses (OAENS) for the steric (top panel), thermosteric (middle panel) and halosteric (bottom panel) sea level, with the 95% confidence level calculated using a bootstrap algorithm. Units are mm yr^{-1}. For REAENS and OAENS, red bars correspond to the spread (standard deviation) of the trends from the individual products.

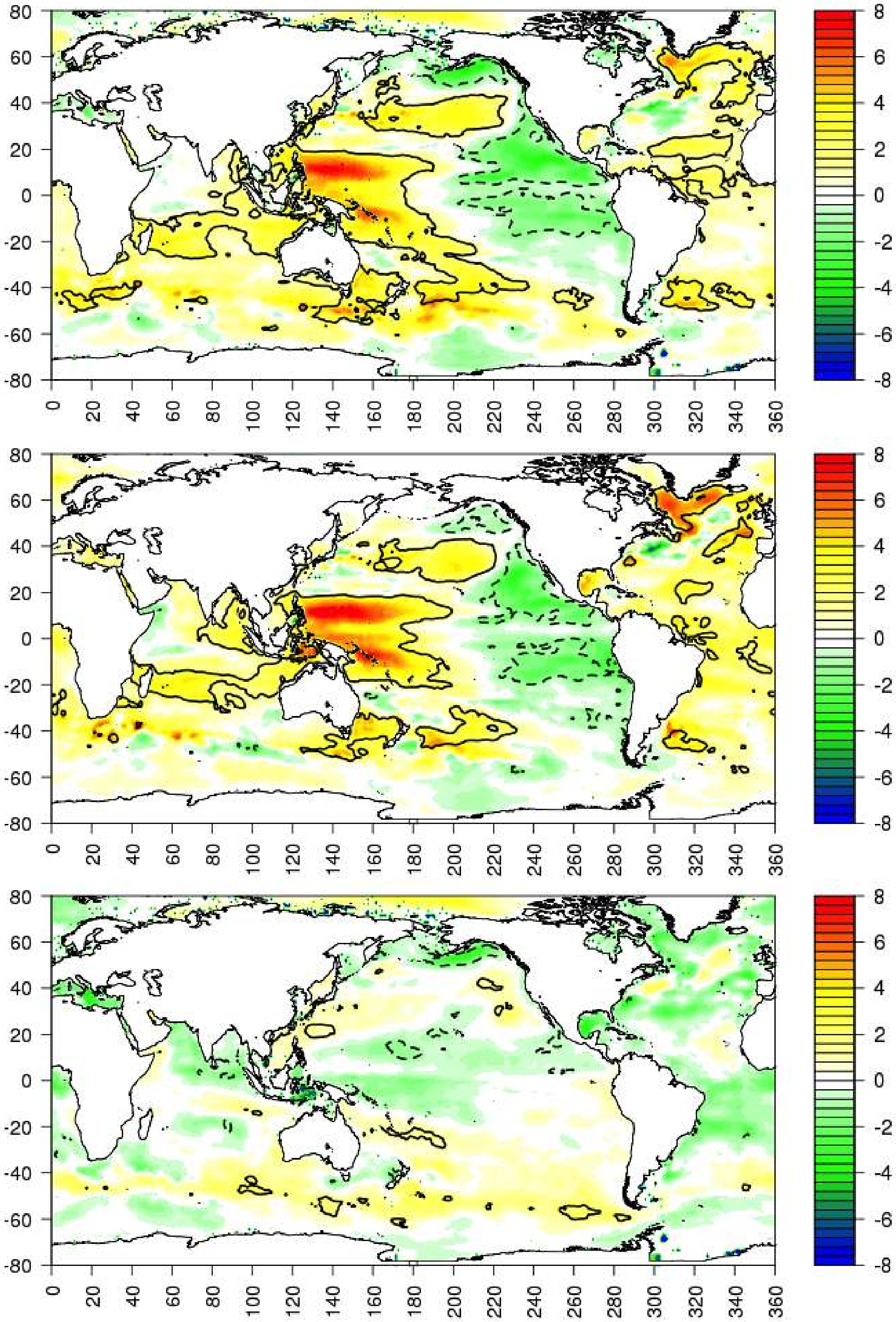

Figure 9: Map of 1993-2010 linear trends ensemble mean (including all the products) for the steric (top panel), thermosteric (middle panel) and halosteric (bottom panel) sea level. Units are mm yr^{-1}. Solid (dashed) contour lines denote regions of signal-to-noise ratio (ensemble mean divided by ensemble spread) equal to 1 (-1).

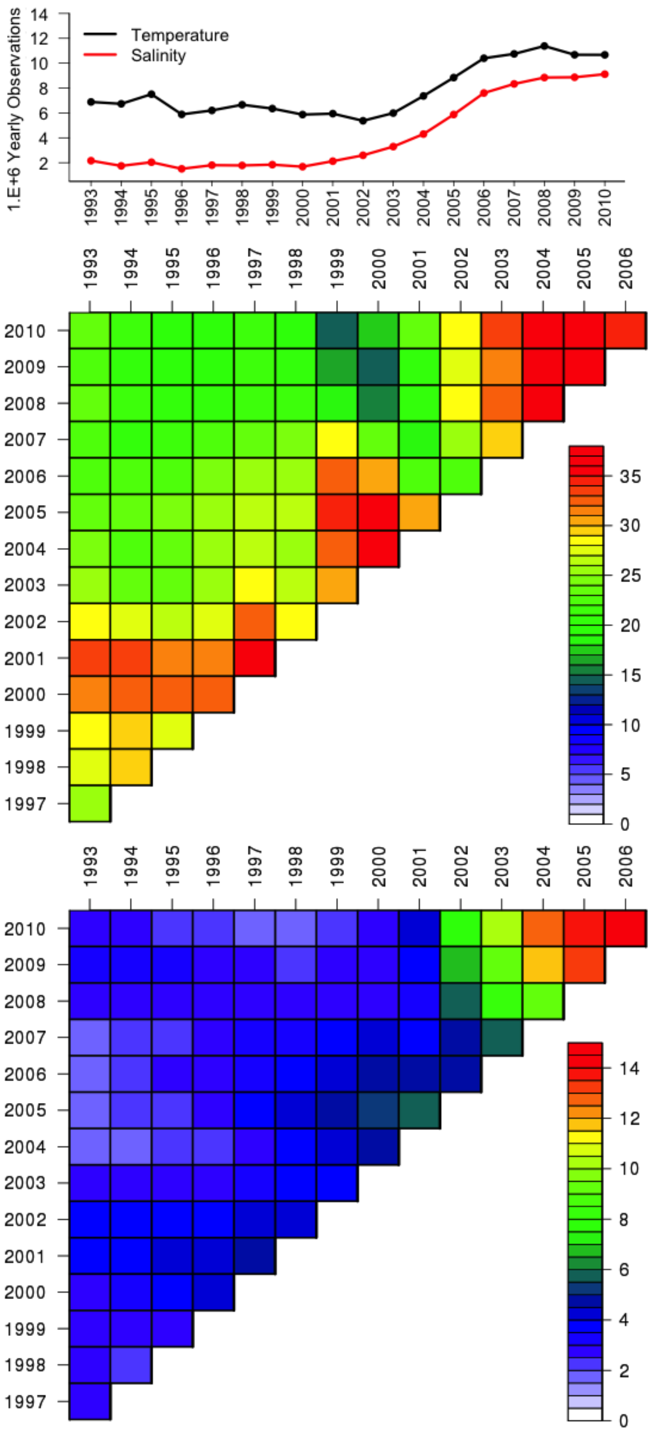

Figure 10: Top: Number of in-situ observations per year from the EN3 dataset (Ingleby and Huddleston, 2007). Units are millions of observations per year. Bottom: Area percentage of the global ocean exhibiting a signal-to-spread ratio greater than the unit for the linear trend as function of starting and ending year, for the thermosteric (middle panel) and the halosteric (bottom panel) sea level. The x-axis above the triangular plots refers to the starting year for the trend computation, while the y-axis refers to the ending year. A minimum period of 5 years is imposed for the linear trend computation. The SSR is computed as ratio between the linear trend ensemble mean and the linear trend ensemble standard deviation, using all the available products, regardless whether they are reanalyses or objective analyses.

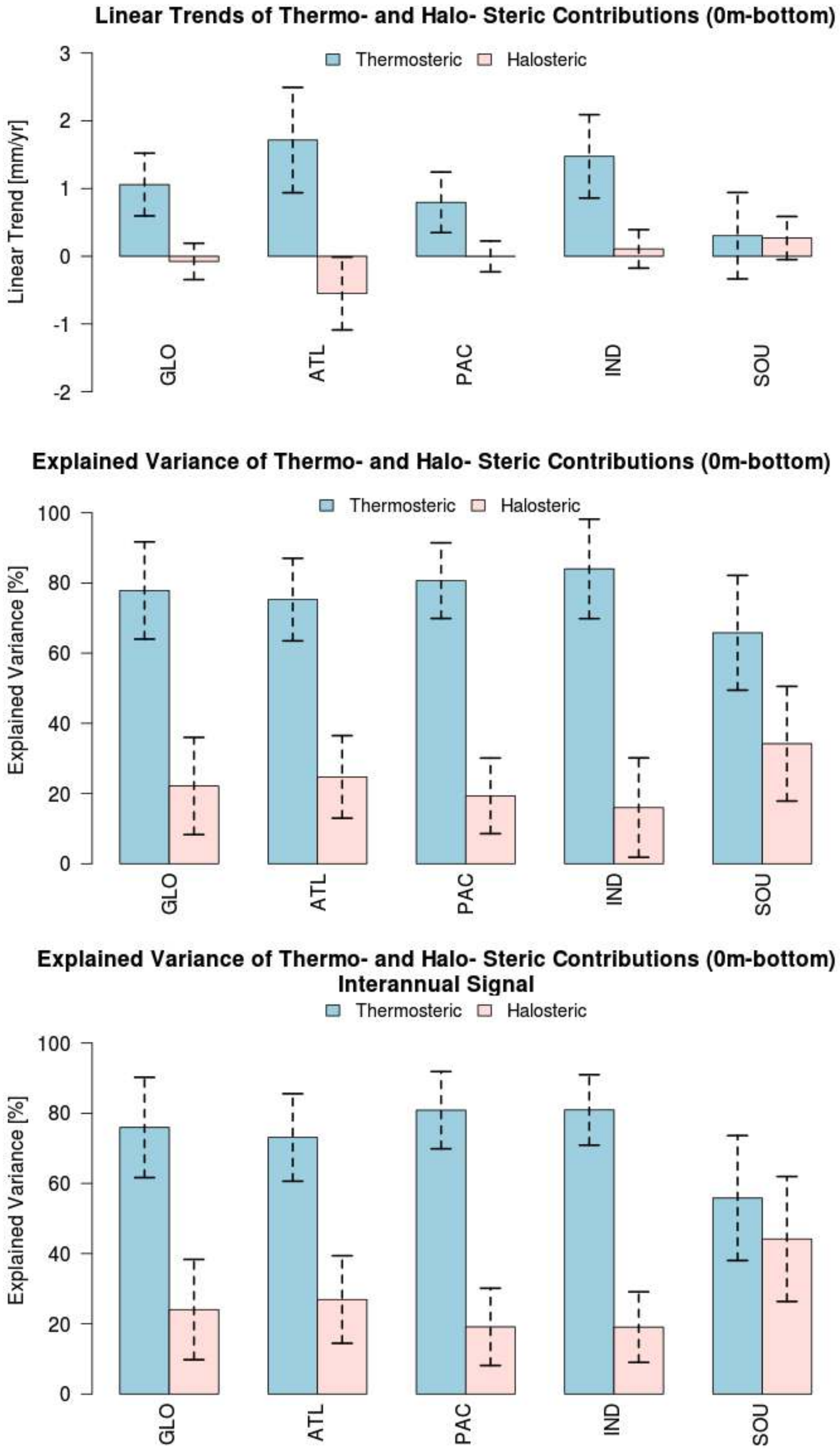

Figure 11: Contribution of the 1993-2010 linear trends from the thermosteric and halosteric component (top panel), and explained variance of the full (middle panel) and inter-annual (bottom panel) signal, for the main Ocean basins (GLO: Global Ocean; ATL: Atlantic Ocean; PAC: Pacific Ocean; IND: Indian Ocean; SOU: Southern Ocean). The Southern Ocean is defined as the part of Ocean south of 50S. The box-plot shows the mean and the standard deviation of the values from the different products.

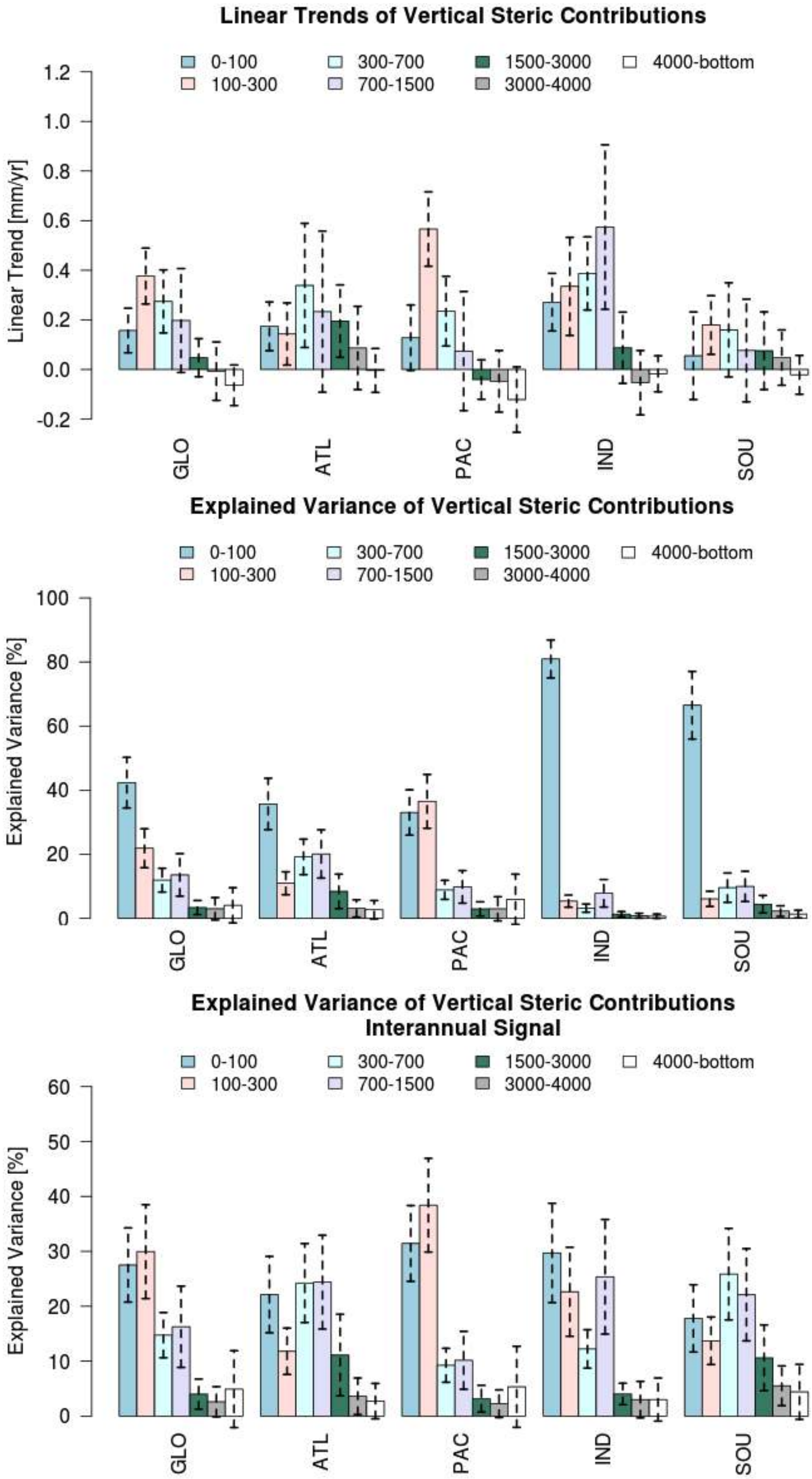

Figure 12: Contribution of the 1993-2010 linear trends from the vertical layers (top panel), and explained variance of the vertical layers for the full (middle panel) and inter-annual (bottom panel) signal, for the main Ocean basins (as in Figure 11).

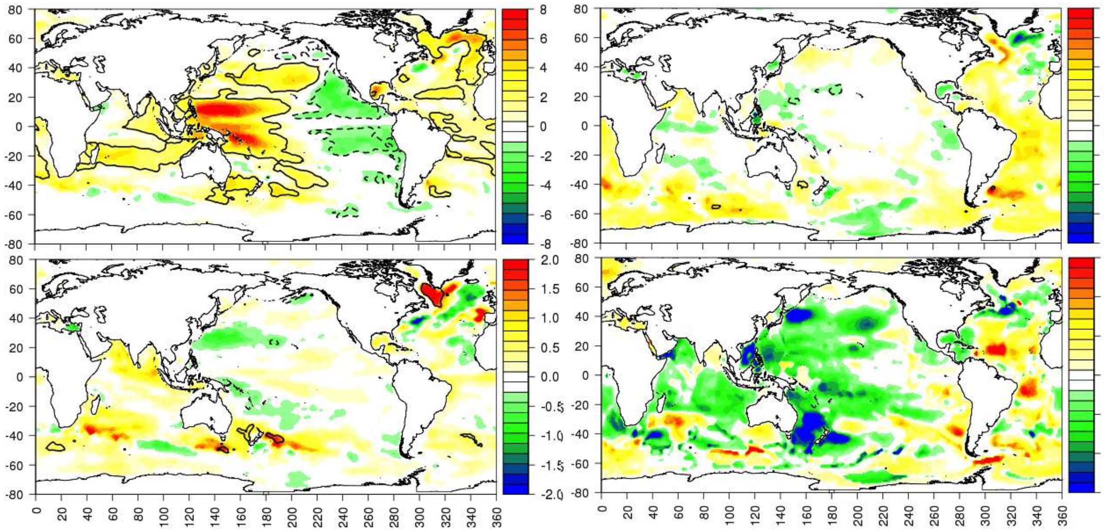

Figure 13: Map of the 1993-2010 ensemble mean of linear trends for the thermosteric sea level in the upper waters (0-700 m of depth, top left panel), intermediate waters (700-1500 m of depth, bottom left panel), deep waters (1500-4000 m of depth, top right panel) and abyssal waters (4000 m to bottom, bottom right panel). Units are mm yr^{-1}. Note that the color palettes are different for the four panels. Solid (dashed) contour lines denote regions of signal-to-noise ratio (ensemble mean divided by ensemble spread) equal to 1 (-1).

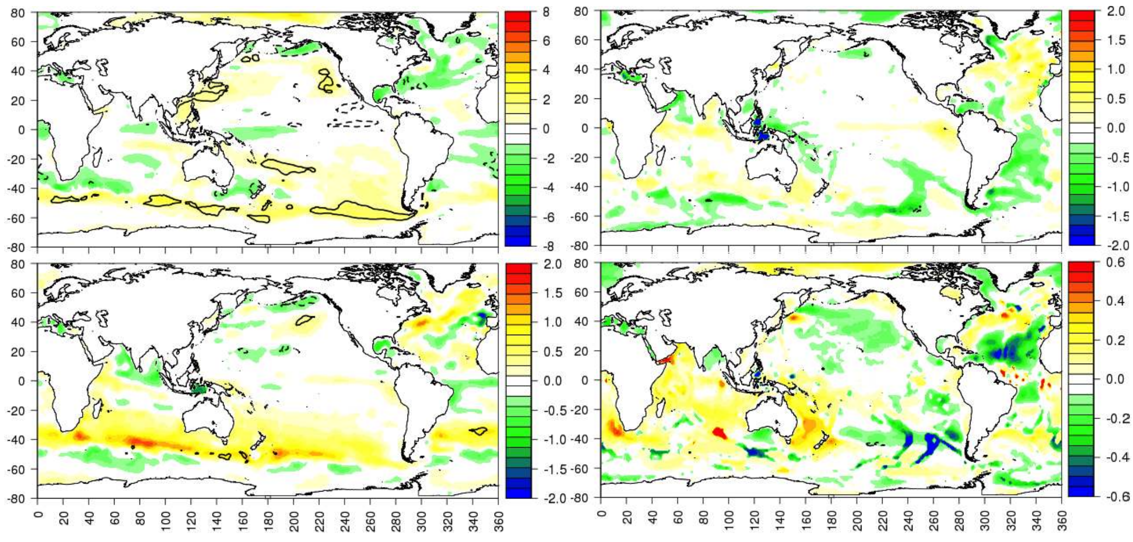

Figure 14: As in Figure 13, but for the haline component of the steric sea level. Note that the color palettes are different for the four panels.