The Ocean Reanalyses Intercomparison Project (ORA-IP)

M. A. Balmaseda[1], F. Hernandez[2,12], A. Storto[3], M. D. Palmer[4], O. Alves[5], L. Shi[5], G. C. Smith[6], T. Toyoda[7], M. Valdivieso[8], B. Barnier[22], D. Behringer[21], T. Boyer[9], Y-S. Chang[10,23], G. A. Chepurin[11], N. Ferry[12], G. Forget[24], Y. Fujii[7], S. Good[4], S. Guinehut[13], K. Haines[8], Y. Ishikawa[14], S. Keeley[1], A. Köhl[15], T. Lee[16], M. Martin[4], S. Masina[3,29], S. Masuda[17], B. Meyssignac[18], K. Mogensen[1], L. Parent[12], K. A. Peterson[4], Y. M. Tang[1,4], Y. Yin[5], G. Vernieres[19], X. Wang[20], J. Waters[4], R. Wedd[5], O. Wang[15], Y. Xue[21], M. Chevallier[28], J-F. Lemieux[6], F. Dupont[6], T. Kuragano[7], M. Kamachi[7], T. Awaji[14], A. Caltabiano[25], K. Wilmer-Becker[26], F. Gaillard[27]

1) European Centre for Medium-Range Weather Forecasts (ECMWF), Reading, United Kingdom
2) Institut de Recherche pour le Développement (IRD), Toulouse, France
3) Centro Euro-Mediterraneo sui Cambiamenti Climatici (CMCC), Bologna, Italy
4) Met Office, Exeter, United Kingdom
5) Centre for Australian Weather and Climate Research, Bureau of Meteorology (BOM), Melbourne, Australia
6) Environment Canada, Québec, Canada
7) Meteorological Research Institute, Japan Meteorological Agency (MRI/JMA), Tsukuba, Japan
8) University of Reading (U-Reading), Reading, United Kingdom
9) NOAA/NODC, College Park, Maryland
10) Geophysical Fluid Dynamics Laboratory, National Oceanic and Atmospheric Administration (GFDL/NOAA), Princeton, New Jersey
11) Department of Atmospheric and Oceanic Science, University of Maryland, College Park, Maryland, USA
12) Mercator Océan, Ramonville Saint-Agne, France
13) Collecte Localisation Satellites (CLS), Ramonville Saint-Agne, France
14) Center for Earth Information Science and Technology, Japan Agency of Marine-Earth Science and Technology (CEIST/JAMSTEC), Yokohama, Japan
15) Universität Hamburg (U-Hamburg), Hamburg, Germany
16) Jet Propulsion Laboratory (JPL), California Institute of Technology, Pasadena, California
17) Research and Development Center for Global Change (RCGC), JAMSTEC, Yokosuka, Japan
18) Laboratoire d'Etudes en Géophysique et Océanographie Spatiale' (LEGOS), Centre National d'Etudes Spatiales (CNES) in Toulouse, France.
19) Goddard Space Flight Center, National Aeronautics and Space Administration (GSFC/NASA), Greenbelt, Maryland
20) Joint Institute for Regional Earth System Science and Engineering, University of California, Los Angeles, California
21) Climate Prediction Center, NOAA/NWS/NCEP, Camp Springs, Maryland, USA
22) Centre National de Recherche Scientifique (CNRS), Laboratoire de Glaciologie et Géophysique de l'Environnement (LGGE), Grenoble, France
23) Department of Earth Science, Kongju National University, Kongju, South Korea
24) Program in Atmosphere, Ocean, and Climate, Massachusetts Institute of Technology
25) International CLIVAR Global Project Office, First Institute of Oceanography, State Oceanic Administration, China.
26) GODAE OceanView Project Office, Met Office, Exeter, United Kingdom
27) Laboratoire de Physique des Océans (LPO/IFREMER), France
28) CNRM-GAME, Météo-France, CNRS UMR3589, Toulouse, France
29) Istituto Nazionale di Geofisica e Vulcanologia, Bologna, Italy

1st September 2014

Synopsis

Uncertainty in ocean analysis methods and deficiencies in the observing system are major obstacles for the reliable reconstruction of the past ocean climate. The variety of existing ocean reanalyses is exploited in a multi-reanalysis ensemble to improve the ocean state estimation and to gauge uncertainty levels. The ensemble-based analysis of signal-to-noise ratio allows the identification of ocean characteristics for which the estimation is robust (such as tropical mixed-layer-depth, upper ocean heat content), and where large uncertainty exists (deep ocean, Southern Ocean, sea-ice thickness, salinity), providing guidance for future enhancement of the observing and data assimilation systems.

Author Biography

Magdalena A.Balmaseda works in the research department of the European Centre for Medium Range Weather Forecasts (ECMWF) as principal scientist in the Marine Prediction Section. The work involves development of the ocean model and ocean initialization of operational forecasting systems (medium range, monthly and seasonal forecasts), as well as ocean reanalyses. She is a member of the GODAE Ocean View Science Team, and member of the CLIVAR Global Synthesis and Observation Panel (GSOP).

Introduction

There is increasing demand for historical records of the ocean climate[1,2]. These are needed as a reference for monitoring the current state of the climate, and also to initialize and validate long-range (e.g. seasonal and decadal) forecasts. Observations alone are often inadequate to generate the required estimate of the ocean variables. Ocean model simulations can provide some insight on the ocean variability, but they are affected by biases due to errors in model formulation, specification of initial states and forcing, and are not directly constrained by observations. Ocean reanalyses are the combination of ocean models, atmospheric forcing fluxes and ocean observations via data assimilation methods and have the potential to provide more accurate information than observation-only or model-only based ocean estimations.

The production of ocean reanalyses (ORAs hereafter) is now an established activity in several research and operational centres. ORAs are revisited every so often, and new 'vintages' are produced at intervals of about five years, as improvements in ocean models, data assimilation methods, forcing fluxes or ocean observations become available. The previous vintage of ORAs (produced around 2006) has already been documented[3,4]. A new vintage has recently been generated, which has come about through the availability of new surface forcing fluxes (from new atmospheric reanalyses), improved quality-controlled ocean datasets, including important corrections to the observations[5,6], as well as the steady improvement in the ocean models and data assimilation methods. There are lower resolution reanalyses (~1 degree horizontal resolution), spanning a long time-period of typically 50 years, as well

90 as higher resolution products (about ¼ of degree), available for shorter records, usually
91 the altimeter period 1993-onwards.

92 Although new reanalysis vintages are produced relatively infrequently, some of the
93 ORAs are continuously updated in quasi-real-time, with the model and data assimilation
94 methodology kept fixed. This is the case for the ORAs produced in operational centres to
95 initialize coupled forecasts. These real-time ORAs have the additional advantage that
96 they allow monitoring of relevant climate variables[7]. The monitoring of the tropical
97 Pacific conditions with a multi ocean reanalysis system (multi-ORA) is now a reality, as
98 can   be   seen   in   the   NCEP   ocean   monitoring   pages   (
99 http://www.cpc.ncep.noaa.gov/products/GODAS/multiora_body.html)

100 In spite of the continuous improvements in methodology, the estimation of the historical
101 ocean state with reliable error estimates is a major challenge. In addition to the
102 estimation of the three-dimensional ocean state at a given time (the analysis problem),
103 an ocean reanalysis also provides an estimation of the time evolution. The time
104 evolution represented by an ORA will be sensitive to the temporal variations of the
105 observing system, to the errors of the ocean model, atmospheric fluxes and assimilation
106 system, which are often flow-dependent, and not easy to estimate[8]. All these factors
107 contribute to the so-called structural uncertainty, i.e. the uncertainty associated with the
108 methodology and that cannot be sampled with a single system. A crude but pragmatic
109 way of estimating the current uncertainty in our ability to measure key ocean variables
110 is to carry out an intercomparison of ORAs within the framework of a multi-reanalysis
111 ensemble approach. For it to work, it is necessary that the individual components are
112 sufficiently distinct while at the same time have similar levels of error (i.e. equally
113 likely). The multi-analysis ensemble approach has already been successfully used to
114 study the ocean heat content[9,10], and to initialize seasonal[11,12] and decadal[13,14] forecasts.
115 The ensemble approach is also used in the framework of the EU funded MyOcean
116 project[15] using eddy-permitting reanalyses over the satellite period (1993-onwards).

117 The operational oceanographic community continuously carries-out coordinated inter-
118 comparison of ocean forecasting systems[16,17,18,19,20,21]. In the same way, there is also
119 need for routine coordinated evaluation of ORAs, which would exploit the existing
120 information for a variety of purposes, namely i) quantifying uncertainty, ii) measuring
121 progress in the quality of the reanalyses and iii) producing indices for ocean monitoring
122 with associated error estimates. These are the motivations for the current Ocean
123 Reanalyses Intercomparison Project (ORA-IP). This paper offers just a first glimpse of
124 the emerging results, with focus on the benefits of the ensemble approach both to
125 improve the estimation of the signals and to provide uncertainty ranges.

126 **The current ORA-IP project**
127
128 The joint GODAE OceanView/CLIVAR-GSOP (Global Synthesis and Observation Panel)
129 workshop in Santa Cruz (13-17 June 2011)[22] called for a community action on
130 exploitation of the latest ORAs for real time climate monitoring and intercomparison.

131 Although the ultimate goal is the near real-time monitoring of the ocean through indices
132 based on an ensemble of reanalyses, the first stage was to complete an ORA-IP. A viable
133 proposal was put forward in Santa Cruz. The reanalyses producers were to provide
134 relevant information (gridded fields of basic primary variables) in agreed formats and
135 grids (where applicable), to enable the agreed intercomparison procedure to be carried
136 out. A "processing centre" would take responsibility for the intercomparison of a
137 particular variable in which they had a strong interest and expertise. The processing
138 centres would analyse ensemble statistics based on the input from the individual
139 reanalyses, and create relevant indices, metrics or graphics that could be directly
140 compared.
141
142 *Table 1: List of ocean variables inter-compared, and responsible processing institution*

143 *Table 2: List of Ocean Reanalysis products entering the inter-comparison.*

144
145 Table 1 provides a list of the variables chosen for intercomparison. Table 2 lists the
146 ORAs included in the study, and provides some details about the product name,
147 associated institution, surface forcing, the ocean model, its resolution[*], assimilation
148 method and observations assimilated. The real-time ORAs are shown in blue. The data
149 assimilation column lists the observation types used for their estimation (T/S for
150 temperature and salinity; SLA: altimeter-derived sea level anomalies; SSH: sea surface
151 height -from tide gauges; SST: sea surface temperature, MDT: mean dynamic
152 topography, SIC: sea-ice concentration), as well as assimilation techniques used for
153 reanalysis: Optimal Interpolation (OI), Ensemble Kalman Filter (EnKF), Kalman Filters
154 and Smoothers (KF-FS), Ensemble OI (EnOI), variational methods (3D-var and 4D-var).
155 Some of the observational products also use statistical techniques such as Empirical
156 Orthogonal Functions (EOFs). In addition to ORAs, the table also lists products named
157 Obs-only (OO in what follows), meaning that they are observation-only products that do
158 not include a dynamical ocean model. The OOs provide sea surface height (SSH) or its
159 anomaly (SLA), and/or temperature and salinity (T/S) estimates, and sometimes 3D
160 velocities (U,V), as in the case of ARMOR3D. The atmospheric surface forcing is usually
161 provided by atmospheric reanalyses, using either direct daily fluxes, or different bulk
162 formulations. Sometimes the atmospheric reanalysis forcing is corrected (suffix *corr* in
163 Table 2), using a variety of methodologies. There are also systems that use fluxes from
164 coupled data assimilation systems (Coupled DA), which come in multiple flavours
165 (parameter estimation, EnKF, weakly coupled). The section on "Surface Heat Fluxes"
166 below provides additional discussion. The detailed description of the analysis systems
167 joining ORA-IP and their differences is beyond the scope of this paper. However, more
168 details about the products can be found in the references given in the table.

---

[*] Even the low resolution models resolve the Equatorial Rossby Radius of deformation by including meridional grid refinement close to the Equator.

169   The production centres provided monthly-mean fields interpolated to the standard 1 ×
170   1 degree latitude-longitude grid used by the World Ocean Atlas 2009[23] (WOA09). Heat
171   and salinity content, their steric contribution, and assimilation increments of
172   temperature were provided as vertically integrated quantities from the surface down to
173   a number of depths: 0-100m; 0-300m; 0-700m; 0-1500m; 0-3000m; and 0-4000m.

174   The ORAs can be exploited, among other purposes, to assess the strengths and
175   weakness of the different systems, to identify of gaps in the observing systems, and to
176   identify robust quantities to use in climate monitoring.. The focus of the results
177   presented is to identify the commonalities and differences among the existing
178   reanalyses. To this end, a multi-system ensemble approach is followed, where the signal
179   and its associated uncertainty are measured by the ensemble mean EM(t) and the
180   ensemble standard deviation (ESD(t)) respectively, defined as:

181     $$\text{EM(t)} = \frac{1}{N}\sum_{k=1}^{N} X_k(t) \quad \text{and} \quad ESD(t) = \sqrt{\frac{1}{N-1}\sum_{k=1}^{N}(X_k(t) - EM(t))^2)} \qquad (1)$$

182   where $X_k(t)$ denotes an individual reanalysis product. Let us denote the different
183   signals in a time series (mean, seasonal cycle, interannual variability, etc,...) as the action
184   of a temporal filter F, and EM^F(s) and ESD^F(s) the ensemble mean and ensemble
185   standard deviation of the filtered signal[†]. We define σ_{EM}^F as the temporal standard
186   deviation of the filtered ensemble mean EM^F(s), and σ_{ESD}^F as the quadratic mean of
187   ensemble spread of the filtered ESD^F(s) as follows:

188     $$\sigma_{EM}^F = \sqrt{\left(\frac{1}{M^F-1}\sum_{s=1}^{s=M^F}\left(EM^F(s) - \overline{EM^F}\right)^2\right)}; \; \sigma_{ESD}^F = \sqrt{\left(\frac{1}{M^F}\sum_{s=1}^{s=M^F}\left(ESD^F(s)\right)^2\right)}, \qquad (2)$$

189   with $\overline{EM^F}$ the time mean of the filtered EM, and M^F is the number of independent
190   temporal samples in the filtered timeseries. The signal-to-noise ratio, defined as the
191   ratio $\sigma_{EM}^F/\sigma_{ESD}^F$, provides guidance on whether the estimation is robust. For instance,
192   estimations with signal-to-noise less than unity are usually not considered robust.

193   In what follows, we will use the term EM-ORA, EM-OO to refer respectively to the
194   ensemble mean of ORAs and OOs. The rest of the article presents a brief overview of the
195   preliminary results of the intercomparison of the variables listed in Table 1.

196 **Heat Content**

197   Monthly-mean depth integrated potential temperatures (K*m) were used in this study.
198   The vertically integrated temperature was converted to ocean heat content (OHC) per
199   unit of area by multiplying by reference values for density (1025 kg m-3) and specific
200   heat capacity (3985 J Kg-1 K-1). This quantity, further integrated in the horizontal
201   global domain, and computed relative to a common reference period of 1993-2007, has
202   been used to estimate changes in the global OHC. Note that when the timeseries are
203   dominated by trends, the choice of reference period impacts the time evolution of the

---

[†]Here *s* is a generic temporal index associated to the temporal filter.

204 spread among the timeseries[‡]. The apparent increase in spread among analyses during
205 the 2000s is substantially reduced if one chooses the 2003-2007 reference period (not
206 shown).

207 Time series of global OHC change (Fig. 1) show best agreement for the upper levels and
208 the products start to diverge as the integration is carried out to deeper levels. The
209 largest rates of 0-4000m OHC rise during the 1990s exceed 3 Wm^{-2} (expressed relative
210 to Earth's surface area) for some products initialized in the early 1990s and cannot be
211 considered physical. They are most likely artefacts of system spin-up or "shocks" related
212 with introduction of the altimeter data. Trends over the period 2000-2009 for 0-4000m
213 OHC give values between about 0.1 and 0.8 Wm^{-2}. The OOs products ARMOR3D and EN3
214 are both near the upper end of this range. Ocean heat uptake below 300m appears to
215 increase markedly in the early 2000s for most products[54], qualitatively supporting the
216 results from the ORAS4 system[55], although there is still a large spread in the amplitude
217 of the OHC, and spatial patterns of change below 300m vary among ocean data
218 assimilation products (not shown). The OO products ARMOR3D and EN3 both show a
219 similar signal of deep ocean heat uptake to ORAS4, illustrating that this signal is
220 inherent to the observations[54].

221 Fig. 1 also shows that the ensemble spread of the multi-ORA is larger than the ensemble
222 spread of the ORAS4 system[55]. Whether this holds for other individual ensemble-based
223 ORA would need to be evaluated. A more difficult question is whether the multi-ORA
224 spread is a good estimator of the existing uncertainty. It appears similar to the spread
225 obtained with observation-only estimations[56] .

226 **Steric Sea Level**

227 Steric sea level (SSL) refers to the change of sea level due to ocean density variations
228 associated with thermal and haline expansion or contraction of sea-water. SSL rise is
229 responsible for about 30% to 40% of the total sea level rise during the last decades,
230 according to recent estimates[57,58]. The ORA-IP is being used to investigate the steric sea
231 level variability, by: i) quantifying the global SSL, its uncertainty and consistency with
232 respect to independent estimates; ii) assessing the regional SSL change and the
233 agreement among ocean reanalyses; iii) quantifying the relative contributions of the
234 thermal and haline components and iv) quantifying the relative contributions of
235 different vertical depth ranges[59]. Some of these aspects are clearly related with the
236 ocean heat content variations and with the attribution of sea-level changes, but are not
237 discussed here. Instead, this section focuses on the performance of the EM-ORA
238 compared to EM-OO.

239 SSL can be diagnosed in two different ways: i) as normalized vertical integration of
240 density anomalies (SSL-density), and ii) as the differences between sea-level and
241 bottom pressure anomalies (SSL-residual). The latter is not easy to infer from models,
242 which are volume-preserving by virtue of the Boussinesq approximation[60]. Instead,
243 temperature and salinity monthly means from the ORA and OO products, containing
244 information from in-situ observations, are used to diagnose SSL-density. Satellite-
245 products are used to derive SSL-residual, thus providing an independent validation data

---

[‡] In the case of linear trends, the spread will increase with the distance to the center of
the reference period.

246 set. Here we use monthly means of altimetric sea-level anomaly (from AVISO) minus
247 gravimetric ocean bottom pressure anomaly (from GRACE RL05[61], available from 2005).

248 The top-left panel of Fig.2 depicts the 2005-2009 map of temporal anomaly correlations
249 between SSL-density from EM-ORA and SSL-residual (altimetry minus gravimetry). The
250 high values of the correlation suggest high consistency in SSL between two independent
251 estimates of SSL over most of the Global Ocean. In the Southern Ocean, south of
252 approximately 60S, where the availability of in situ observations is poor, the correlation
253 is lower. The top-right panel shows the temporal anomaly correlations, calculated after
254 the seasonal signal has been removed (i.e. inter-annual signal retained). Although
255 removing the seasonal cycle decreases the correlation value (especially in the Atlantic
256 Ocean and at high latitudes), EM-ORA still exhibits high correlations for the inter-annual
257 signal in the tropical areas and at mid-latitudes.

258 The correlation between SSL-dens and SSL-residual is higher for EM-ORA (0.84) than
259 for any individual product (0.77 at the maximum), and also higher than for EM-OO
260 (0.74, not shown). The latter is especially evident in areas where the in-situ observing
261 network is poor and/or where there is impact of deep and bottom waters. The bottom
262 panels of Fig. 2 show the difference of the anomaly correlation with respect to the
263 validation dataset between the EM-ORA and the EM-OO for the full (left) and inter-
264 annual (right) signals. The high correlation obtained by EM-ORA emphasizes the added
265 value of the dynamical constraints and atmospheric forcing included in the ORAs. This is
266 evident in the full fields (in the Southern Ocean, in the South Atlantic and just south of
267 the Bering Strait), and especially noticeable for the inter-annual signal.

268 Although the EM-ORA proves to be a good estimator of total steric height, uncertainty
269 still remains regarding the partition into thermal and haline components, and the
270 contribution of different depth ranges. Preliminary results over a longer
271 intercomparison period (1993-2009)[59] show a large spread regarding the contribution
272 of deep layers (below 700 m of depth) to SSL trends, with a low signal-to-noise ratio in
273 the trend estimation of less than 1.

Sea Level

275 The sea level from the ORAs in Table 2 and two OO products (ARMOR3D and LEGOS)
276 have been evaluated. (The sea level ARMOR3D is effectively the delayed gridded
277 AVISO[52] product, also called DUACS). This comparison focuses mainly on the
278 interannual variability and regional distribution of the trend, and it uses globally-
279 detrended monthly means of sea level anomalies. For each product, the seasonal cycle
280 was removed at each location of the ocean domain. The global mean sea level for each
281 month was also removed.

282 Two reference data sets have been used for the evaluation: sea level from tide gauges,
283 and the newly reprocessed altimeter-derived sea level from the ESA Climate Change
284 Initiative (SLCCI)[53]. The latter is a gridded dataset where the original altimeter data has
285 been reprocessed with improved algorithms (orbit, wet tropospheric corrections,
286 among others) and ancillary data (using improved atmospheric fields for instance) in
287 order to produce consistent time series of sea level for climate studies. SLCCI has not
288 been assimilated by any of the products in Table 2, although many of these products
289 assimilate along-track satellite altimetry (usually AVISO). Only two products (ECCO-v4
290 and LEGOS) use information from tide gauges.

291   The tide gauges used for the evaluation are the same as the subset from the Global Sea
292   Level Observing System (GLOSS, see http://www.gloss-sealevel.org/) chosen for
293   evaluation of sea level reconstructions[41]. Monthly means of sea-level anomaly at the tide
294   gauge locations were created after removing the effects of tides and inverse barometer
295   from the original tide gauge data. This allows a relevant comparison with sea level
296   anomalies from the reanalysis products because tides and inverse barometer are not
297   represented in the reanalysis products. The ORAs and OOs were spatially interpolated
298   to the tide gauge locations. All the time series involved in the analyses were detrended
299   at each location by removing the product-specific local linear trend.

300   The comparison with tide gauges appears in the top panels of Fig.3. The statistics are for
301   the period 1993-2009. Fig. 3a shows the scatter diagram for the individual products
302   (top-left), with the temporal correlation (x-axis) and the rms error (cm, in the y-axis). A
303   large scatter in the scores is seen among different products, with the best fit generally
304   obtained by the products assimilating SLA, and in particular by those with higher
305   horizontal resolutions, with scores comparable to those obtained by the altimeter-
306   derived SLCCI and AVISO. This result indicates that not all the ORAs are equally likely,
307   and therefore the grand ensemble mean may not be appropriate to estimate coastal sea-
308   level variations. In this case, the ensemble approach is limited to those products that
309   assimilate altimeter EM-ORAalti. Fig. 3b shows the correlation map between tide
310   gauges and EM-ORAalti. Even with the reduced ensemble, the correlation is higher in
311   the open ocean than in the continental shelves, and it appears higher in the tropics than
312   at higher latitudes.

313   A different application of the multi-ORA ensemble is the definition of climate indices
314   relevant for regional climate monitoring, which is illustrated in the following (although
315   more work is needed to define relevant indices). The sea level variability averaged over
316   the Eastern North Tropical Pacific region (0-12°N, 84-108°W) has been chosen as an
317   example, because although different from the traditional equatorial El Niño index, it
318   reflects the impact El Niño in the Western Coast of Mexico related with the coastal
319   propagation of Kelvin waves. In this case, ESACCI is used as validation data set. All
320   products show a coherent interannual variability (Fig. 3d), even when altimeter data
321   are not assimilated, and there is very small spread around EM-ORA (black). The
322   variability is dominated by the El Niño 1997-98, and a significant negative trend of ~3-4
323   mm/y, consistent with the lack of Eastern Pacific ENSO in the last decade, and with the
324   recently reported strengthening of the Pacific trade winds [55,62,63]. Fig. 3c shows the
325   corresponding Taylor diagram for the different ORAs and OOs averaged over this
326   region. In contrast with the tide-gauge evaluation, here the scores of the ORAs versus
327   SLCCI are quite similar (sixteen of the nineteen products show correlations higher than
328   0.95 and rms differences lower than 1.5 cm). The smallest rms error (around 0.5 cm
329   rms error, y-axis in Taylor diagram) is achieved by the EM-ORA. The EM-ORA score is
330   comparable to that achieved by the best members (which assimilate satellite altimetry)
331   and by AVISO, and is larger than 0.99. EM-ORA has a weaker signal than SLCCI (4.8
332   instead of 5.2 cm of standard deviation, x-axis in Taylor diagram), a natural
333   consequence of the ensemble averaging.

334   In other areas, like the North Atlantic (not shown), there are more discrepancies among
335   the reanalysis products and weaker signal-to-noise ratios. Discrepancies can arise from
336   different choices in the assimilation systems[64]. It has been shown that products
337   assimilating altimeter data can be distinguished from those that do not. The different

338   methods used to assimilate altimeter information can also introduce spread. For
339   instance, the altimeter can be used to constrain only the baroclinic mode, or only the
340   barotropic mode, or to constrain the fresh water budget, or the three aspects
341   simultaneously. The altimeter can be assimilated in anomaly mode (using anomaly
342   values relative to a reference period) or using the absolute values (which implies the
343   use of an external mean dynamic topography (MDT), which differs between systems).
344   The ORA-IP can be used to gain insight into the sensitivity arising from the assimilation
345   methods, but this is beyond the scope of this paper.

Surface Heat Fluxes

347   The purpose of this comparison is to assess the global heat closure in ORAs, the
348   consistency of the seasonal cycle and interannual variability between the products, and
349   to compare with other heat fluxes from a variety of sources (primarily satellite, ships,
350   buoys and atmospheric reanalysis). These other sources are not completely
351   independent (with the exception of satellite based radiative fluxes) because they may
352   also use SST or near surface meteorological data to generate products. Nevertheless,
353   they enable some assessment of the uncertainty introduced by the reanalysis methods
354   themselves. Additional datasets include the OAFlux latent and sensible heat flux
355   product[65] combined with ISCCP satellite based radiation[66], the ship-based NOC2.0
356   product[67], the Large and Yeager[68] hybrid flux dataset CORE.2, and two atmospheric
357   reanalysis products, the ECMWF ERA-Interim reanalysis[69] (referred to as ERAi) and the
358   NCEP/DOE reanalysis R2[70] (referred to as NCEP-R2).

359   Most ORAs are forced with bulk formulae using an atmospheric dataset taken from an
360   atmospheric reanalysis product. In these cases, assimilation of sea surface temperature
361   (SST) observations directly influences the net surface heat flux, as the turbulent latent
362   and heat fluxes, computed from bulk formulae, and the outgoing long wave radiation,
363   computed using the Stefan-Boltzmann Law, depend on the SSTs. The ORAs can also
364   close their heat budget through the temperature assimilation increments, since the
365   vertically integrated temperature assimilation increments, with the appropriate unit
366   transformation, are equivalent to a heat flux[55]. Fig. 4 shows the 17-yr mean globally
367   integrated heat fluxes for 15 individual ORAs and for the ensemble mean, as well as for
368   the other global flux products. The interannual variability over the same period is
369   shown by the error bars. Most ocean reanalyses have a positive surface imbalance
370   (mean net surface heat flux into the ocean), usually considerably smaller than for the
371   observational products, e.g., ISCCP/OAFlux and NOC2.0, and smaller than for
372   atmospheric reanalyses in some cases. The largest interannual variability is seen for the
373   PEODAS product which uses ERA-40[71] forcing fields until 2002, and NCEP-R2 based
374   forcing thereafter. Interannual variations over 1993-2009 are only ~1 Wm^{-2} for the
375   ensemble of 15 flux estimates. The contributions from the assimilation increments are
376   mostly negative (removing heat from the ocean on the global average), resulting in a
377   reduction of the net heat flux. The total net heat flux applied (i.e., surface plus
378   assimilation) is still positive and mostly smaller than ±2 Wm^{-2}, consistent with reported
379   warming in global ocean heat content[42,55,56].

380   The seasonal cycle in surface heat fluxes closely agrees in most regions between the
381   reanalysis products (not shown), with monthly spreads generally being smaller than
382   10Wm^{-2} over most of the global ocean, exceptions being the subpolar gyres, the
383   Southern Ocean and some eastern subtropical basin areas[72]. Interannual signal-to-noise

384   ratios for the surface heat fluxes over the period 1993-2009 show strong signals (2+) in
385   the ENSO affected regions and perhaps some signals at higher latitudes, but with
386   signal/noise ~1, longer analysis periods may be needed to identify this variability more
387   clearly. Regional comparisons are being extended to include individual flux components
388   (representing radiative and turbulent transfers), and also validation against in situ flux
389   measurements at a number of OceanSITES moorings[72], which provide an independent
390   check that is not reliably gained from any other source.

391   **Mixed Layer Depth**

392   Mixed layer depth (MLD) is one of the most important variables for both the dynamical
393   process of climate variation and for biogeochemistry. Intercomparison of the seasonal
394   to interannual variability in the global MLD provides a useful gauge of the value of ORAs
395   for the study of climate variability.

396   The MLD used in this study is defined as the depth where potential density exceeds the
397   10-m depth value by $\Delta\rho$ = 0.03 or 0.125 kg m^{-3} (MLDr003/MLDr0125). Similarly, the
398   isothermal layer depth (ILD) is defined as the depth where potential temperature
399   differs from the 10-m depth value by ΔT = 0.2 or $0.5\,^{\circ}$C (ILDt02/ILDt05). Different
400   criteria are used because it is not easy to find a unique threshold that defines the mixed
401   layer depth at all latitudes. As MLD/ILD verification we use the MILA-GPV[73] and
402   deBoyer[74] datasets, also estimated from the individual TS profiles, following the
403   definitions above.   In particular, MILA-GPV uses only the Argo profiles without
404   interpolation between grid points, although the spatio-temporal coverage of the dataset
405   is limited. deBoyer provides the monthly climatological fields (MLDr003 and ILDt02).

406   The MLD/ILD are calculated from monthly-means of temperature-salinity (TS) fields on
407   the individual native grids of three OO products (EN3v2a, ARMOR3D, WOA09) and 16
408   ORAs; these are then interpolated to the regular global longitude-latitude common grid.
409   EM-ORA is estimated as the ensemble average of individual MLD/ILD on the common
410   grid (this will differ from the MLD/ILD calculated from the ensemble mean of TS). The
411   MLD/ILD from the individual ORAs exhibit various biases in the mean fields depending
412   on the diversity of model configurations and assimilation systems (not shown). Here we
413   focus on the evaluation of EM-ORA rather than on the detailed representations of the
414   individual reanalysis fields, which will be described in future work.

415   Fig. 5 presents the zonal-mean monthly MLD/ILD normalized differences of EN3v2a,
416   ARMOR3D, deBoyer, WOA09 and EM-ORA with respect to the MILA-GPV as reference.
417   Note that values averaged over the Argo-rich 2005-2011 period are plotted for MILA-
418   GPV, EN3v2a, ARMOR3D and EM-ORA, while the climatological fields are provided by
419   deBoyer and WOA. The differences between deBoyer and MILA-GPV (MLDr003 and
420   ILDt02) are generally small, since MILA-GPV and deBoyer are comparable datasets that
421   use individual TS profiles. The larger differences appear in high latitudes, where the
422   availability of ocean observations is limited. MLDs/ILDs for EN3v2a and ARMOR3D are
423   biased-shallow due to the use of gridded and monthly mean TS fields. ILDt02s in WOA
424   are 20 to 40% shallower than MILA-GPV globally, due to the use of the climatological TS
425   field[74]. Using larger values for criterion ($\Delta\rho$ = 0.125 kg m^{-3} and ΔT= $0.5\,^{\circ}$C) reduces
426   the shallow biases. The shallow biases in MLDr0125 and ILDt05 for EN3v2a and
427   ARMOR3D are generally less than 20% except at high latitudes. We found that a large
428   portion of these shallow biases result from the coarser vertical resolution of the OO

429 gridded TS products at relevant depths compared with the model based reanalyses[75].
430 Model biases do cancel in most areas in the EM, although large positive biases remain in
431 regions where common biases are well known from coarse resolution models[76] (e.g., the
432 Kuroshio Extension and Antarctic Circumpolar Current regions). In addition, ILDt05
433 values from WOA, EN3v2a, ARMOR3D and EM-ORA are commonly larger than those
434 from MILA-GPV in the subarctic regions and Southern Ocean around spring. This is
435 likely due to the fact that MILA-GPV is the only product that does not use monthly
436 means of TS when deriving MLD/ILD. This specific topic will be described in future
437 work.

438

Salinity in the top 700m

440 Salinity variability has a significant impact on the density structure and dynamics of the
441 ocean. However, it is only in the past few years that assimilation of salinity has received
442 attention, largely because of the advent of Argo (see http://argo.jcommops. org, which
443 has significantly improved the sampling of the global ocean salinity), and because of its
444 importance in obtaining balanced ocean states. For instance, recent studies on seasonal
445 forecasts[77,78] demonstrate that the assimilation of salinity observations results in
446 improving ocean states density and T/S properties, resulting in better ENSO prediction.

447 This study evaluates the averaged salinity in the top 700m of the ocean (S700) as
448 represented by the EM-ORA and compared it with the EM-OO. As discussed, the ESD-
449 ORA gives an indication of uncertainty, and the signal-to-noise ratio provides guidance
450 on where the signal measured by the ensemble mean dominates over the noise
451 measured by the ensemble spread.

452 Fig. 6a shows the difference of annual mean S700 between EM-ORA and EM-OO in Table
453 2 over the period 1993-2010. The difference is largest (~0.2 psu) in regions of strong
454 frontal variability such as the Gulf Stream, Southern Ocean along the Antarctic
455 Circumpolar Current (ACC) region, and to a lesser extent the Kuroshio region. In the
456 tropics the difference is generally less than 0.05 psu.

457 Fig. 6b shows the ESD-ORA of the S700 1993-2010 mean (or σ_{ESD}^{M}, where M denotes
458 1993-2010 temporal mean). In general the largest spread, up to 0.15 psu, is also
459 associated with the areas of strong variability or greatest mean difference compared to
460 the EM-OO analyses. Around most of the ACC, the ESD-ORA is just large enough to
461 encompass the large differences between EM-ORA and EM-OO. The spread is relatively
462 large in the eastern equatorial Atlantic and the western equatorial Indian Ocean, where
463 the spread reaches up to 0.1 psu

464 Fig. 6c shows the correlation of S700 interannual anomalies between the EM-ORA and
465 EM-OO for the period 1993-2010. Correlations are relatively high, greater than 0.75, in
466 the equatorial and sub-equatorial Pacific, particularly in the centre and west. They are
467 also high in the eastern equatorial Indian Ocean, and throughout parts of the sub-
468 tropical and mid-latitude oceans. Correlations are relatively low, less than 0.5, around
469 the northern edge of the ACC, Western Indian Ocean and parts of the sub-tropical
470 Atlantic, particularly downstream of the Mediterranean outflow. Each individual ORA
471 can be correlated with the EM-OO. Then the spread in this correlation gives an
472 indication of the disagreement in the estimate of variability between the different

473 systems. This is shown in Fig. 6d. There is some correspondence between areas with
474 large spread and low correlation in Fig.6c, e.g., the northern edge of the ACC in the
475 Pacific Sector and the northern part of the tropical Atlantic. Equally, the high correlation
476 in the Tropical Pacific, Eastern Indian Ocean, North East Pacific and North East Atlantic,
477 where the spread is low, is indicative of consistency between the different estimates.
478 The Southern Ocean is an exception, presenting relatively large values of the correlation
479 and large values of spread.

480 Fig.6e shows the standard deviation of the interannual S700 anomalies (seasonal cycle
481 removed) of EM-ORA (σ_{EM}^I, where I stands for "interannual"). This gives an estimate of
482 the amplitude and geographical distribution of the S700 interannual signal, which
483 appears highest in subduction areas close to the edge of strong boundary currents. It is
484 also high in the western equatorial Pacific and central Indian Ocean, probably
485 associated with changes in the fresh-water fluxes. Fig. 6f shows the spread in the S700
486 monthly anomalies of the ORAs (or σ_{ESD}^I). The spread is largest in the sub-tropics and
487 mid-latitudes, particularly associated with western boundary currents and the Southern
488 Ocean. In the western boundary current regions and parts of the Southern Ocean it
489 exceeds 0.1 psu. The signal-to-noise ratio is greater than 1 in the equatorial west Pacific,
490 central Indian Ocean and small regions of the mid-latitude ocean. However, over most of
491 the oceans the signal-to-noise ratio is less than 1.[79]

492 An interesting question arising from this study is why the spread appears largest in the
493 Gulf Stream than in other western boundary currents. One possible explanation is
494 related to the stronger salinity fronts in this region, such that small variations in the
495 Gulf Stream path can produce strong salinity anomalies. But other factors can
496 contribute as well, such as the uncertainty associated with deep-water formation, sea
497 ice, and a larger uncertainty in the representation of the Gulf Stream path itself
498 (compared with other western boundary currents). The uncertainty introduced by the
499 assimilation method cannot be discarded either, and it would be interesting to evaluate
500 the uncertainty pattern of ocean-model simulations, as well as that of individual
501 ensemble-based data assimilation systems.

Depth of 20°C Isotherm

503 Variations in the thermocline depth are associated with major modes of tropical climate
504 variability. The depth of the 20°C isotherm (D20) has been considered as part of this
505 intercomparison project as a proxy for thermocline depth and variability in the tropical
506 oceans. D20 monthly means from the different ORAs in Table 2 and from two OO
507 products (EN3v2a and ARMOR3D) have been used.

508 The absolute value of D20 depends on the vertical discretisation of the model used in
509 each reanalysis. Most of the products have between 16-25 levels in the upper 200-m-
510 depth. There is the small group of eddy-permitting, NEMO based reanalyses,
511 characterized by high vertical resolution of the upper ocean (1-m in the first level, then
512 31 levels for the first 200-m depth). There is also some ambiguity regarding the
513 definition of D20 monthly means included in the evaluations: these can be either
514 "monthly means of D20 from instantaneous values" or "the D20 from the monthly
515 means of the temperature field". In this preliminary diagnostics, different groups have
516 used different methods.[80]

Fig.7 shows the spatial pattern of D20 in the EM-ORA (Fig.7a), the differences between the two OO products (Fig.7b), and the difference between EM-ORA and each of the OO products (Fig.7c and Fig.7d). On average, EM-ORA is shallower than the OO products in the centre of gyres, and deeper on both eastern and western boundaries of the ocean basins. There are also large differences at the western boundaries, especially along the Gulf Stream, which may be related with the misrepresentation of the path of western boundary currents by the models. However, differences along the western boundary currents are also large between the OO products EN3v2a and ARMOR3D (Fig.7b). Compared to the OO products, the D20 EM-ORA is slightly deeper in the Equatorial Indian, Atlantic and Eastern Pacific Oceans, and shallower in the Pacific Warm Pool. The reasons for this unexpected difference will be investigated in future work.

Sea-Ice

Several studies have suggested that sea ice thickness may be a predictor for seasonal sea ice extent[81,82]. This highlights a weakness in almost all ice forecasting systems in that they don't include the explicit assimilation of ice thickness observations. Moreover, it remains to be seen how the predictability of the seasonal ice cover depends on the representation of various physical processes and model details, such as spatial resolution and the inclusion of an ice thickness distribution. By intercomparing various properties of the sea ice cover in existing ice-ocean reanalyses, it may be possible to highlight deficiencies and best practises in these systems toward answering the question: Are current ice-ocean reanalyses suitable for initializing seasonal forecasts of the ice cover? Here we present preliminary results from this intercomparison.

The ice-ocean reanalyses considered here use a variety of model resolutions, physics and analysis methods. Reanalysis details are provided in table 2. For the ECMWF reanalysis system, two additional versions of the system were considered whereby only the method of ice assimilation was varied (ERAL-linear, ERAN-non-linear[83]; note that these products do not appear in Table 2).

Sea ice models used here include two community models, the Los Alamos Community Ice model (CICE[84]) and the Louvain sea Ice Model (LIM[85]), as well as independently developed models. While these models and their particular implementation details may vary widely, an important distinction is the representation of the ice thickness distribution. Some models include a sophisticated multi-category approach, while others use a single ice category. This different treatment of ice thickness impacts both ice dynamics and thermodynamics[86].

Another important distinction is in the application of ice assimilation. Many systems employ a simple nudging of ice concentration toward a gridded ice analysis product (e.g. OSISAF, NSIDC), while a few systems use more sophisticated ice assimilation methods (e.g. 3DVar, SEEK). However, perhaps the most important aspect of ice assimilation is in how the increments to concentration affect ice thickness. Two systems (ECMWF and Mercator) supplied different versions of their reanalyses with/without ice assimilation and the impact on ice thickness is non-negligible, albeit unconstrained (not shown).

Fig. 8 shows an example highlighting the large range of mean ice thicknesses found for the various ice-ocean reanalyses in March 2007. Also shown is a satellite estimate of the ice thickness derived from ICESat[87] for February/March 2007. In general, the reanalysis products all exhibit the basic feature of thicker ice cover north of the Canadian Arctic Archipelago and Greenland as seen in the observations, albeit to a varying degree. However, the thickness of ice in the central Arctic and along the Siberian coast varies widely. In particular, the reanalyses tend to cluster toward either overly thin ice (~1m) or overly thick ice (>3m), with perhaps only one or two showing realistic thicknesses of about 2m. These differences are larger than interannual variations and are on the scale of the decadal thinning of the ice cover (not shown). The relative contribution of the various factors (e.g. model physics and resolution, atmospheric forcing, data assimilation) that may be contributing to these differences is a topic of on-going study. Such large biases may limit the usefulness of these products for seasonal forecasting.

Summary

This paper presents the first results of the ORA-IP, which aims at exploiting the diversity of existing ocean reanalyses to identify those aspects that are robustly represented by the different products and those where there is a large level of discrepancy. The agreement can be exploited to define indices for monitoring or verification, while the discrepancies point towards areas for future enhancement of assimilation and observing systems. The paper also illustrates the use of independent evaluation metrics to measure the quality of the ensemble mean and individual products, thus providing guidance on the adequacy of the ensemble approach.

The intercomparison has focused on a small set of ocean variables, interpolated into a common horizontal grid, and for a limited set of vertical levels (when applicable). The intercomparison period is mainly 1993-2010, although shorter periods are also used. Where relevant (mixed layer, ocean heat content, steric height, sea level, salinity and thermocline depth) the ensemble mean of the ocean reanalyses was compared with observation-only estimates, to assess if the model-derived estimates show any systematic differences from the observation-only estimates. The ensemble spread is also used as a measure of the existing uncertainty.

It is shown that in general the ensemble mean is usually a better estimation than any individual ocean reanalyses. However, in the case of coastal sea level variability, the evaluation with tide-gauge data indicates that ORAs with high-resolution models and assimilation of altimeter are more skilful, and the scores are better when using a sub-ensemble including the subset of best ORAs instead of the grand ensemble.

Systematic differences between OOs and ORAs are largest in the tropics, where model-physics and the wind variability are key assets for the ORAs. These differences are seen in the thermocline and mixed layer depth. In addition, the ensemble of ORAs performs better than the OO products in the estimation of steric height variability at seasonal and interannual variability time scales in the Atlantic and outside the tropics.

The surface heat flux estimates from ocean reanalyses were compared with other products, mostly based on atmospheric reanalyses. Although large uncertainty still exists, the ocean reanalyses global surface heat fluxes appear more balanced than the atmospheric-based products, especially when the contribution of the assimilation

604 increments is taken into account. The results suggest that data assimilation methods
605 and ocean observations can contribute to the estimation of surface heat fluxes.

606 The estimation of interannual variability of salinity continues to be a challenge. Signal-
607 to-noise ratios larger than one are confined to the tropical western Pacific, dominated
608 by the ENSO signal. This was the case in the previous intercomparison of reanalyses
609 (circa 2006)[3,4], and continues to be so now, in spite of the increased salinity
610 observations in recent years. More work is needed to establish the source of uncertainty
611 (changing observing system - e.g. differences before and after Argo, forcing fields,
612 assimilation methods and error specification).

613 The intercomparison of sea-ice showed a large uncertainty in the estimation of sea-ice
614 thickness, which is largely unconstrained by the assimilation methods, highlighting the
615 need for observations of ice thickness for both assimilation and validation.

616 This ORA-IP has also identified areas where the uncertainty is large, thus providing a
617 focus for future developments in the observing system and modelling/data assimilation.
618 The deep ocean (below the top few hundred metres), the Southern Ocean (Antarctic
619 Circumpolar Current region), coastal areas and the path of western boundary currents
620 appear as the areas with largest uncertainty in the density, temperature and salinity
621 fields. Not only are the differences between ORAs and OO products the largest, but there
622 is also a large spread among ORAs (as expected from model error), and among OOs
623 (likely because observation representativeness errors are large). These are also
624 important areas for climate.

625 It is clear that we are still a long way from providing ocean estimations that can answer
626 satisfactorily many fundamental questions, and that continuous development of the
627 assimilation and observing system is needed. In the meantime, the multi-model
628 ensemble strategy is a pragmatic approach to exploit the current resources. It is also
629 clear that the evaluation of successive vintages of ocean reanalyses should be a
630 continuous process, since it is needed to assess progress and to identify gaps, thus
631 contributing to setting the directions for future developments.

632

645

References

1. Balmaseda MA and Coauthors. *2010. Role of the Ocean Observing System in an End-to-End Seasonal Forecasting System.* In Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 1), Venice, Italy, 21-25 September 2009, Hall, J., Harrison, D.E. & Stammer, D., Eds., ESA Publication WPP-306, doi:10.5270/OceanObs09.pp.03.

2. *Dee DP, Balmaseda MA , Balsamo G, Engelen R, Simmons AJ and J.-N. Thépaut. 2014. Toward a consistent reanalysis of the climate system. BAMS. e-View doi: http://dx.doi.org/10.1175/BAMS-D-13-00043.1*

3. Stammer D and Coauthors. 2010. *Ocean Information Provided Through Ensemble Ocean Syntheses.* OceanObs'09: Sustained Ocean Observations and Information for Society. doi:10.5270/OceanObs09.cwp.85.

4. Lee T, Awaji T, Balmaseda MA, Grenier E, and Stammer D. 2009. *Ocean state estimation for climate research.* Oceanography, **22**, 160–167. doi:10.5670/oceanog.2009.74.

5. Lyman JM, Good SA, Gouretski VV, Ishii M, Johnson GC, Palmer MP, Smith DM, Willis JK. 2010. *Robust warming of the global upper ocean.* Nature, **465,** pp. 334-337, doi:10.1038/nature09043.

6. Wijffels S, Willis J, Domingues CM, Barker P, White NJ, Gronell A, Ridgway K, Church JA. 2009. *Changing expendable bathythermograph fall rates and their impact on estimates of thermosteric sea level rise.* J. Climate **21**: 5657–5672.

7. Xue Y and Coauthors. 2010. *Ocean state estimation for global ocean monitoring: ENSO and beyond ENSO.* In OceanObs'09: Conference on Sustained Ocean Observations and Information for Society, vol. 2, Venice, 21–25 September 2009. Hall J, Harrison DE, Stammer D. (eds). ESA publication WPP-306, DOI: 10.5270/OceanObs09.

8. Masina S, Di Pietro P, Storto A and Navarra A. 2011. *Global ocean re-analyses for climate applications.* Dyn. Atmos. Oceans, **52**, (1-2), SI , 341-366, doi:10.1016/j.dynatmoce.2011.03.006

9. Xue Y, and Coauthors, 2012: *A Comparative Analysis of Upper-Ocean Heat Content Variability from an Ensemble of Operational Ocean Reanalyses.* J. Climate, **25**, 6905–6929. doi: http://dx.doi.org/10.1175/JCLI-D-11-00542.1

10. Zhu J, Huang B, Balmaseda MA. 2011. *An ensemble estimation of the variability of upper-ocean heat content over the tropical Atlantic Ocean with multi-ocean reanalysis products.* Clim. Dyn. in press. DOI: 10.1007/s00382–011–1189–8.

11. Zhu J, Huang B, Marx L, Kinter III JL, Balmaseda MA, Zhang R-H, Hu Z-Z. 2012. *Ensemble ENSO hindcasts initialized from multiple ocean analyses.* Geophys. Res. Lett. **39**: L09602, DOI: 10.1029/2012GL051503.

12. Zhu J, Huang B, Balmaseda MA, Kinter III JL, Peng P, Hu ZZ, Marx L. 2013. *Improved reliability of ENSO hindcasts with multi-ocean analyses ensemble initialization.* Clim. Dyn. 11/2013; 41(9-10). DOI:10.1007/s00382-013-1965-8

13. Pohlmann H, Doug S, Balmaseda MA, Keenlyside NS, Masina S, Matei D, Muller WA, Rogel P. 2013. *Predictability of the mid-latitude Atlantic meridional overturning circulation in a multi-model system.* Clim. Dyn., **41**, 10.1007/s00382-013-1663-6.

14. Bellucci A, Gualdi S, Masina S, Storto S, Scoccimarro E, Cagnazzo C, Fogli P, Manzini E and Navarra A. 2013. *Decadal climate predictions with a coupled AOGCM initialized with oceanic reanalyses.* Clim. Dyn., **40**, 1483-1497.

15. Ferry N, Barnier B, Garric G, Haines K, Masina S, Parent L, Storto A , Valdivieso M, Guinehut S and Mulet S. 2012. *NEMO: the modeling engine of global ocean reanalyses.* Mercator Ocean Quarterly Newsletter **46**, 46-59.

695   16. Crosnier L, and Le Provost C. 2006. *Internal metrics definition for operational forecast*
696   *systems inter-comparison: Example in the North Atlantic and Mediterranean Sea.* Ocean
697   Weather Forecasting. Springer Netherlands. 455-465.
698   17. Crosnier L, and Le Provost C. 2007. *Inter-comparing five forecast operational systems in*
699   *the North Atlantic and Mediterranean basins: The MERSEA-strand1 Methodology."*
700   Journal of Marine Systems **65.1**, 354-375.
701   18. Xie J, Zhu J, and Li Y. 2008. *Assessment and inter-comparison of five high-resolution sea*
702   *surface temperature products in the shelf and coastal seas around China.* Continental
703   Shelf Research **28.10**. 1286-1293.
704   19. Hernandez, F, Bertino L, Brassington G, Chassignet E, Cummings J, Davidson F, Drévillon
705   M, Garric G, Kamachi M, Lellouche J-M, Mahdon J, Martin M, Ratsimandresy R, and
706   Regnier C. 2009.*Validation and intercomparison studies within GODAE*. Oceanography
707   22(3):128–143, doi:10.5670/oceanog.2009.71.
708   20. Oke PR, Brassington G, Cummings J, Martin M, and Hernandez F. 2012. *GODAE Inter-*
709   *comparisons in the Tasman and Coral Seas.* Journal of Operational Oceanography **5.2**,
710   11-24.
711   21. Herandez F, and Coauthors 2014. *Performance evaluations, near real-time assessment of*
712   *operational oceanography forecast products*. Journal of Operational Oceanography. This
713   issue.
714   22. Oke P, Martin M, Balmaseda MA, Brassington G and Wilmer-Becker K. 2011. <u>*Report on*</u>
715   <u>*the GODAE Ocean View - CLIVAR GSOP Workshop on Observing System Evaluation*</u>
716   <u>*and Intercomparison*</u>. https://www.godae-oceanview.org/outreach/meetings-
717   workshops/task-team-meetings/godae-oceanview-gsop-clivar-workshop/
718   23. Locarnini RA, Mishonov AV, Antonov JI, Boyer TP, Garcia HE, Baranova OK, Zweng
719   MM, and Johnson DR. 2010. *World Ocean Atlas 2009, Volume 1: Temperature*. S.
720   Levitus, Ed. NOAA Atlas NESDIS 68, U.S. Government Printing Office, Washington,
721   D.C., 184 pp.
722   24. Guinehut S, Dhomps AL, Larnicol G and Le Traon PY. 2012. *High resolution 3D*
723   *temperature and salinity fields derived from in situ and satellite observations*. Ocean Sci.,
724   **8**, 845-857, doi:10.5194/os-8-845-2012.
725   25. Mulet S, Rio MH, Mignot A, Guinehut S and Morrow R. 2012. *A new estimate of the*
726   *global 3D geostrophic ocean circulation based on satellite data and in situ*
727   *measurements*. Deep-Sea Res. II., 77-80, 70-81, doi:10.1016/j.dsr2.2012.04.012.
728   26. Saha S, and Coauthors. 2010. *The NCEP climate forecast system reanalysis*. Bull. Am.
729   Meteorol. Soc. 91: 1015–1057.
730   27. Xue Y, Huang B, Hu ZZ, Kumar A, Wen C, Behringer D, and Nadiga S. 2011. *An*
731   *Assessment of Oceanic Variability in the NCEP Climate Forecast System Reanalysis*.
732   Clim. Dyn., **37**, 2511-2539.
733

734   28. Storto A, Dobricic S, Masina S, and Di Pietro P. 2011. *Assimilating along-track*
735   *altimetric observations through local hydrostatic adjustments in a global ocean*
736   *reanalysis system*. Mon. Wea. Rev., **139**, 738-754.
737   29. Fukumori I. 2002. *A partitioned Kalman filter and smoother*. Mon. Wea. Rev., **130**, 1370-
738   1383.
739   30. Wunsch C and Heimbach P. 2013. *Dynamically and Kinematically Consistent Global*
740   *Ocean Circulation and Ice State Estimates*. Chapter 21 in "Ocean Circulation and Climate
741   - A 21st century perspective", *International Geophysics Series*, Vol.103. Edited by G.

Sielder, J. Church, S. Griffes, J. Gould, and J. Church. Academic Press, Elsevier. ISBN: 978-0-12-391851-2.

31. Speer K and Forget G. 2013. *Global distribution and formation of mode waters*. Chapter 9 in "Ocean Circulation and Climate - A 21st century perspective", *International Geophysics Series*, Vol.103. Edited by G. Sielder, J. Church, S. Griffes, J. Gould, and J. Church. Academic Press, Elsevier. ISBN: 978-0-12-391851-2

32. Ingleby B, and Huddleston M. 2007. *Quality control of ocean temperature and salinity profiles - historical and real-time data*. Journal of Marine Systems. **65**, 158-175 10.1016/j.jmarsys.2005.11.019

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

34. Zhang S., Harrison MJ, Rosati A, and Wittenberg AT. 2007. *System design and evaluation of coupled ensemble data assimilation for global oceanic climate studies*. Mon. Weather Rev., **135**(10), doi:10.1175/MWR3466.1

35. Chang YS, Zhang S, Rosati A, Delworth TL, and Stern WF. 2013. *An assessment of oceanic variability for 1960-2010 from the GFDL ensemble coupled data assimilation*. Clim. Dyn., **40**(3-4), 775-803, doi:10.1007/s00382-012-1412-2.

36. Blockley E, and Coauthors. 2013. *Recent development of the Met Office operational ocean forecasting system: an overview and assessment of the new Global FOAM forecasts,* Geosci. Model Dev. Discuss., **6**, 6219-6278, doi:10.5194/gmdd-6-6219-2013, 2013.

37. Waters J, Martin M, While J, Lea D, Weaver A, and Mirouze I. 2014. Implementing a variational data assimilation system in an operational 1/4 degree global ocean model. Submitted to Q. J. R. Meteorol. Soc.

38. Behringer DW. 2007. *The global ocean data assimilation system at NCEP*. Preprints, 11th Symp. on Integrated Observing and Assimilation Systems for Atmosphere, Oceans, and Land Surface, San Antonio, TX, Amer. Meteor. Soc., 3.3. [Available online at https://ams.confex.com/ams/87ANNUAL/webprogram/Paper119541.html.]

39. Masuda S, and Coauthors. 2010. *Simulated Rapid Warming of Abyssal North Pacific Waters*, Science, **329**, 319-322, DOI, 10.1126/science.1188703.

40. Sugiura N, Awaji T, Masuda S, Mochizuki T, Toyoda T, Miyama T, Igarashi H, and Ishikawa Y. 2008. *Development of a four-dimensional variational coupled data assimilation system for enhanced analysis and prediction of seasonal to interannual climate variations*, J. Geophys. Res., **113**, C10017, doi:10.1029/2008JC004741.

41. Meyssignac B, Becker M, Llovel W, and Cazenave A. 2012. *An assessment of two-dimensional past sea level reconstructions over 1950-2009 based on tide gauge data and different input sea level grids*. Survey in Geophysics, online. doi :10.1007/s10712-011-9171-x

42. Levitus S and Coauthors. 2012. *World Ocean heat content and thermosteric sea level change (0-2000 m) 1955-2010*. Geophys. Res. Lett. , **39**, L10603, doi:10.1029/2012GL051106

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

44. Balmaseda MA, Mogensen K, and Weaver AT. 2013. *Evaluation of the ECMWF ocean reanalysis system ORAS4*. Q.J.R. Meteorol. Soc., **139**: 1132–1161. doi: 10.1002/qj.2063

45. Mogensen K, Balmaseda MA, Weaver AT. 2012. *The NEMOVAR ocean data assimilation system as implemented in the ECMWF ocean analysis for System 4*. Tech. Memo. **668.** ECMWF: Reading, UK.

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

47. Toyoda T, Fujii Y, Yasuda T, Usui N, Iwao T, Kuragano T, and Kamachi M. 2013. *Improved analysis of the seasonal-interannual fields by a global ocean data assimilation system*. Theoretical and Applied Mechanics Japan, **61**, 31-48, doi: 10.11345/nctam.61.31.

48. Tsujino H, Hirabara M, Nakano H, Yasuda T, Motoi T, and Yamanaka G. 2011. *Simulating present climate of the global ocean-ice system using the Meteorological Research Institute Community Ocean Model (MRI.COM): simulation characteristics and variability in the Pacific sector*. Journ. of Oceanogr., **67**, 449-479. doi: 10.1007/s10872-011-0050-3.

49. Danabasoglu G, and Coauthors. 2013. *North Atlantic simulations in Coordinated Ocean-ice Reference Experiments phase II (CORE-II). Part I: Mean states*. Ocean Modell., in press, http://dx.doi.org/10.1016/j.ocemod.2013.10.005.

50. Carton JA and Giese BS. 2008. *A Reanalysis of Ocean Climate Using Simple Ocean Data Assimilation (SODA)*. Mon. Wea. Rev., **136**, 2999–3017. doi: http://dx.doi.org/10.1175/2007MWR1978.1

51. Haines K, Valdivieso M, Zuo H, and Stepanov VN. 2012. *Transports and budgets in a 1/4 ° global ocean reanalysis 1989–2010*. Oce. Sci., **8** (3), 333-344, doi:10.5194/os-8-333-2012.002/qj.2063.

52. SSALTO/DUACS User Handbook: (M)SLA and (M)ADT Near-Real Time and Delayed Time Products. CLS-DOS-NT-06-034, SALP-MU-P-EA-21065-CLS, v4.0, 69pp (http://www.aviso.oceanobs.com/fileadmin/documents/data/tools/hdbk_duacs.pdf).

53. Ablain M, and Coauthors. 2013. *Two Decades of Global and Regional Sea Level Observations from the ESA Climate Change Initiative Sea Level Project*. ESA Living Planet Symposium, Edinburgh, United Kingdom, 9–13 September, 2013

54. Palmer M, and Coauthors. 2014. *CLIVAR-GSOP/GODAE intercomparison of ocean heat content: initial results*. CLIVAR EXCHANGES, **64**. Feb 2014.

55. Balmaseda MA, Trenberth KE, and Kallen E. 2013. *Distinctive climate signals in reanalysis of global ocean heat content*. Geophys. Res. Lett., **40**, 1754–1759, doi:10.1002/grl.50382.

56. Palmer M, and Coauthors. 2010. *Future Observations for Monitoring Global Ocean Heat Content*. In Proceedings of OceanObs'09: Sustained Ocean Observations and Information for Society (Vol. 2), Venice, Italy, 21-25 September 2009, Hall, J., Harrison, D.E. & Stammer, D., Eds., ESA Publication WPP-306, doi:10.5270/OceanObs09.cwp.68

57. Cazenave A, and Llovel W. 2010. *Contemporary Sea Level rise*. Annu. Rev. Mar. Sci., **2**, 145–73.

58. Church JA and Coauthors. 2011. *Revisiting the Earth's sea-level and energy budgets from 1961 to 2008*. Geophys. Res. Lett., **38**, L18601, doi:10.1029/2011GL048794.

59. Storto A, and Coauthors. 2014. *Comparison of Steric Sea Level from Ocean Reanalyses and Objective Analyses*. CLIVAR EXCHANGES, **64**. Feb 2014.

60. Greatbatch R. 1994. *A note on the representation of steric sea level in models that conserve volume rather than mass,* J. Geophys. Res., **99**, 12767–12771.

61. Chambers DP, and Bonin JA. 2012. *Evaluation of Release-05 GRACE time-variable gravity coefficients over the ocean*. Ocean Sci., **8**, 859-868.

62. England MH. and Coauthors. 2014. *Recent intensification of wind-driven circulation in the Pacific and the ongoing warming hiatus*. Nat. Clim. Change, **4**, 222–227, doi:10.1038/nclimate2106.

839 63. de Boisséson E., Balmaseda MA, Abdalla S, Källén E and Janssen PAEM. 2014. *How*
840   *robust is the recent strengthening of the Tropical Pacific trade winds?.* Geophys. Res.
841   Lett., **41**, 4398–4405, doi:10.1002/2014GL060257.
842 64. Hernandez F, and Coauthors. 2014. *Sea Level Inter-Comparison: Initial results.*
843   CLIVAR EXCHANGES, **64**. Feb 2014.
844 65. Yu L, Jin X, and Weller RA. 2008. *Multidecade Global Flux Datasets from the*
845   *Objectively Analyzed Air–sea Fluxes (OAFlux) Project: Latent and Sensible Heat Fluxes,*
846   *Ocean Evaporation, and Related Surface Meteorological Variables.* Technical Report
847   OAFlux Project (OA2008–01), Woods Hole Oceanographic Institution.
848 66. ZhangY, Rossow WB, Lacis AA, Oinas V, Mishchenk MI. 2004. *Calculation of radiative*
849   *fluxes from the surface to top of atmosphere based on ISCCP and other global data sets.*
850   Journal of Geophysical Research: Atmospheres (1984–2012) **109** (D19).
851 67. Berry DI and Kent EC. 2009. *A New Air-Sea Interaction Gridded Dataset from ICOADS*
852   *with Uncertainty Estimates.* Bull. Amer. Meteor. Soc **90**(5), 645-656. doi:
853   10.1175/2008BAMS2639.1.
854 68. Large W and Yeager S. 2009. *The global climatology of an interannually varying air-sea*
855   *flux data set.* Clim. Dynamics, Volume **33**, 341-364.
856 69. Dee DP, and Coauthors. 2011. *The ERA-Interim reanalysis: configuration and*
857   *performance of the data assimilation system.* Q.J.R. Meteorol. Soc., **137**: 553–597. doi:
858   10.1002/qj.828.
859 70. Kanamitsu M, Ebitsuzaki W, Woolen J, Yang SK, Hnilo JJ, Fiorino M, Potter G. 2002.
860   *NCEP-DOE AMIP-II reanalysis (R-2).* Bull. Amer. Meteor. Soc., **83**:1631–1643.
861 71. Uppala SM, and Coauthors. 2005. *The ERA-40 re-analysis.* Q. J. R. Meteorol. Soc. **131**:
862   2961–3012.
863 72. Valdivieso M, and Coauthors. 2014. *Heat fluxes from ocean and coupled reanalyses.*
864   CLIVAR EXCHANGES, **64**. Feb 2014.
865 73. Hosoda S, Ohira T, Sato K, and Suga T. 2010. *Improved description of global mixed-*
866   *layer depth using Argo profiling floats.* J. Oceanogr., **66**, 773-787, doi:10.1007/s10872-
867   010-0063-3.
868 74. de Boyer Montegut C, Madec G, Fischer AS, Lazar A, and Iudicone D. 2004. *Mixed layer*
869   *depth over the global ocean: An examination of profile data and a profile-based*
870   *climatology.* J. Geophys. Res., **109**, C12003, doi:10.1029/2004JC002378.
871 75. Toyoda T, and Coauthors. 2014. *Mixed layer depth intercomparison among global ocean*
872   *syntheses reanalyses.* CLIVAR EXCHANGES, **64**. Feb 2014.
873 76. Hasumi H, Tatebe H, Kawasaki T, Kurogi M, and Sakamoto TT. 2010. *Progress of North*
874   *Pacific modeling over the past decade.* Deep-Sea Res. II, **57**, 1188-1200,
875   doi:10.1016/j.dsr2.2009.12.008.
876 77. Zhao M, Hendon HH, Alves O, Yin Y, and Anderson DLTA. 2013a. *Impact of salinity*
877   *constraints on the simulated mean state and variability in a coupled seasonal forecast*
878   *model.* Mon. Wea. Rev., **141**, 388-402.
879 78. Zhao M, Hendon HH, Alves O, and Yin Y. 2013b. *Impact of improved assimilation of*
880   *temperature and salinity for coupled model seasonal forecasts*, Submitted to Mon. Wea.
881   Rev.
882 79. Alves O, and Coauthors. 2014. *An Assessment of Upper Ocean Salinity Reanalyses from*
883   *CLIVAR GSOP/GODAE Systems.* CLIVAR EXCHANGES, **64.** Feb 2014.
884 80. Hernandez F, and Coauthors.2014. *ORA-IP Depth of the 20°C isotherm: First results.*
885   CLIVAR EXCHANGES, **64**. Feb 2014.
886 81. Kauker F, Kaminski T, Karcher M, Giering R, Gerdes R, Voßbeck M. 2009. *Adjoint*
887   *analysis of the 2007 all time Arctic sea-ice minimum.* Geophys. Res. Let., **36**(3).

888 82. Chevallier M, Salas y Melia D, Voldoire A, Deque M, and Garric G. 2013. *Seasonal*
889 *Forecasts of the Pan-Arctic Sea Ice Extent Using a GCM Based Seasonal Prediction*

890 *System*. J. Clim., **26**, 6092–6104. doi: 10.1175/JCLI-D-12-00612.1

891 83. Tang YM, Balmaseda MA, Mogensen KS, Keeley SPE, and P. A. E. M. Janssen PAEM.
892 2013. *Sensitivity of sea ice thickness to observational constraints on sea ice*
893 *concentration*, ECMWF Tech Memo Number **707**.

894 84. Hunke EC, and Lipscomb WH. 2010. *CICE: the Los Alamos sea ice model*

895 *documentation and software user's manual version 4.1*, Tech. Rep. LA-CC-06-012, Los

896 Alamos National Laboratory.

897 85. Fichefet, T., and Maqueda MAM. 1997. *Sensitivity of a global seaice model to the*
898 *treatment of ice thermodynamics and dynamics*. J. Geophys. Res. **102**:12609-12646.

899 86. Martensson S, Meier HEM, Pemberton P, and Haapala J. 2012: *Ridged sea ice*
900 *characteristics in the Arctic from a coupled multicategory sea ice model*. J. of Geophys.
901 Res., **117**(C8).

902 87. Kwok R, and Rothrock, DA. 2009. *Decline in Arctic sea ice thickness from submarine*
903 *and ICESat records: 1958-2008*, Geophys. Res. Let., **36** (15).

904

905 **Tables**

906 *Table 1: List of ocean variables inter-compared, and responsible processing institution*

Variable	
Ocean Heat Content	MetOffice
Steric Height	CMCC
Sea Level	Mercator Ocean
Surface Heat Fluxes	University Reading
Mixed Layer Depth	MRI/JMA
Salinity	CAWCR
Depth of 20 degree Isotherm	Mercator Ocean
Sea Ice	Env Canada

907

908 *Table 2: List of Ocean Reanalysis products entering the inter-comparison.*

Product	Forcing	Configuration	Data Assim. Method
ARMOR3D[24,25] CLS	N/A	1/3° Obs-Only (T/S/SSH/U/V)	*OI (SLA/MDT/T/S/SST)*
CFSR[26,27] NOAA NCEP	Coupled DA	1/2° MOM4 coupled	*3DVAR (T/SST/SIC)*
C-GLORS05V3[28] CMCC	ERAi corr+ Bulk	1/2° NEMO3.2	*3DVAR (SLA/T/S/SST/SIC)*
ECCO-NRT[29] JPL/NASA	NCEP-R1+ CORE Bulk	1° MITgcm	*KF-FS (SLA/T)*
ECCO-v4[30,31] MIT/AER/JPL	ERAi+CORE Bulk	1° MITgcm	*4DVAR (SLA/SSH/T/S/SST)*
EN3 v2a[32] Hadley Center	N/A	1° Obs-Only (T/S)	*OI (T/S)*
GECCO2[33] U. of Hamburg	NCEP-R1+Bulk	1°x1/3° MITgcm	*4DVAR (SLA/T/S/MDT/SST)*
ECDA[34,35] GFDL/NOAA	Coupled DA	1/3° MOM4 coupled	*EnKF (T/S/SST)*

|			
GloSea5[36,37] **UK MetOffice**	ERAi+CORE Bulk	1/4° NEMO3.2	*3DVAR (SLA/T/S/SST/SIC)*
MERRA Ocean GSFC/NASA/GMAO	Merra + Bulk	1/2° MOM4	*EnOI (SLA/T/S/SST/SIC)*
GODAS[38] NOAA NCEP	NCEP-R2 Flux.	1°x1/3° MOM3	*3DVAR (SST/T)*
GLORYS2V1(G2V1) Mercator Océan	ERAi corr+CORE Bulk	1/4° NEMO3.1	*KF+3DVAR (SLA/T/S/SST/SIC)*
GLORYS2V3(G2V3) Mercator Océan	ERAi corr+ CORE Bulk	1/4° NEMO3.1	*KF+3DVAR (SLA/T/S/SST/SIC)*
K7-ODA(ESTOC)[39] JAMSTEC/RCGC	NCEP-R1 corr. Flux	1° MOM3	*4DVAR (SLA/T/S/SST)*
K7-CDA[40] JAMSTEC/CEIST	Coupled DA	1° MOM3 coupled	*4DVAR (SLA/SST)*
LEGOS[41] LEGOS	N/A	1/4° Obs-Only (SL)	*OI+EOF (SLA/SSH)*
NODC[42] NODC/NOAA	N/A	1° Obs-only (T/S)	*OI (T/S)*
PEODAS[43] CAWCR(BoM)	ERA40 to 2002; NCEP-R2 thereafter. Flux	1°x2° MOM2	*EnKF (T/S/SST)*
ORAS4[44,45] ECMWF	ERA40 to 1988; ERAi thereafter. Flux.	1° NEMO3	*3DVAR (SLA/T/S/SST)*
MOVE-C[46] MRI/JMA	Coupled DA	1° MRI.COM2 coupled	*3DVAR (SLA/T/S/SST)*
MOVE-G2[47] MRI/JMA	JRA-55 corr+ Bulk	0.5°x1° MRI.COM3	*3DVAR (SLA/T/S/SST)*
MOVE-CORE[48,49] MRI/JMA	CORE.2 Bulk	0.5°x1° MRI.COM3	*3DVAR (T/S)*
SODA[50] U. of Maryland and TAMU	ERA40 to 2002; ERAi thereafter. Bulk	1/4° POP2.1	*OI (T/S/SST)*
UR025.4[51] U. of Reading	ERAi + CORE Bulk	1/4° NEMO3.2	*OI (SLA/T/S/SST/SIC)*
AVISO[52] CLS	N/A	1/4° Obs-Only (SSH/SLA)	*OI (SLA)*
SLCCI[53] ESA	N/A	1/4° Obs-Only (SSH)	*OI (SSH)*

909

910

911

912

913

914 **Figures**
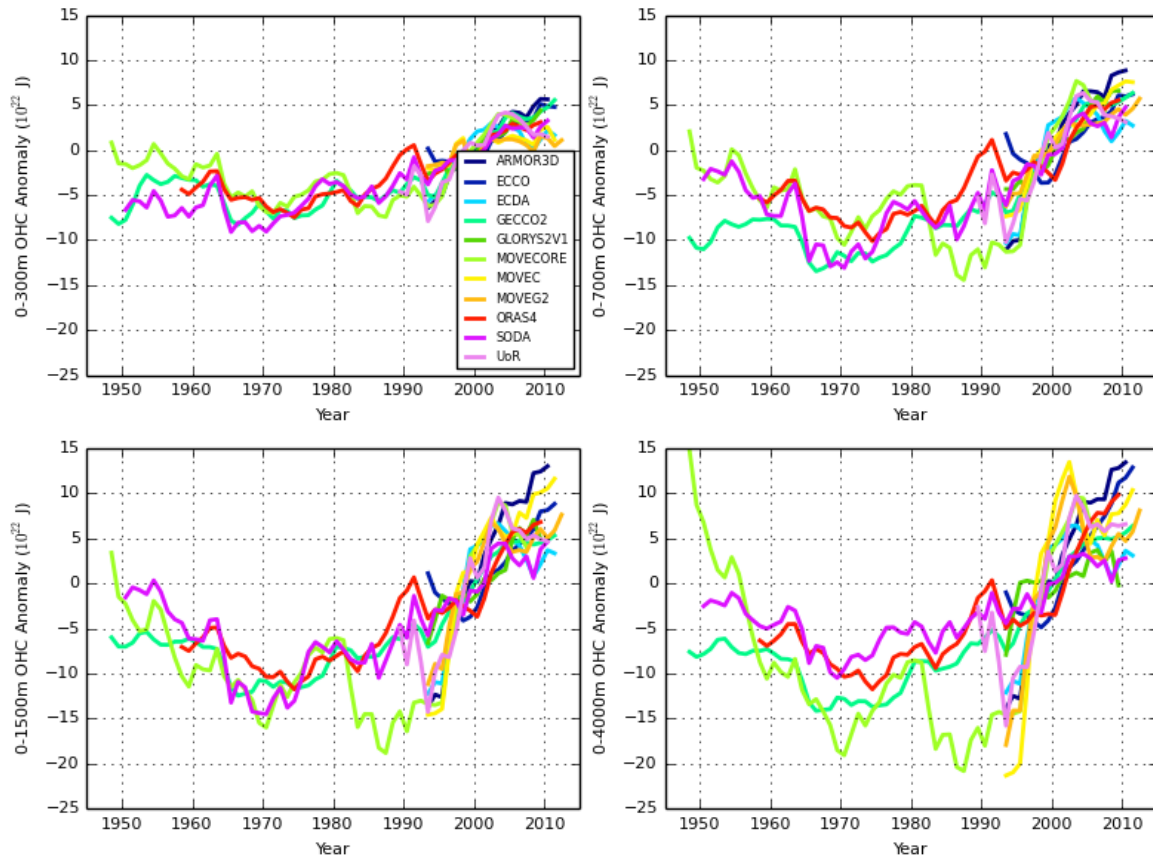
Fig 1: Time series of global ocean heat content anomaly, relative to a baseline period of 1993-2007. Note that SODA only includes grid boxes that span the full column and therefore will tend to underestimate OHC changes as the depth of integration increases. ARMOR3D and EN3 are obs-only analyses and do not include a dynamic model component. [UoR in legend corresponds to the URO25.5 in Table 2].
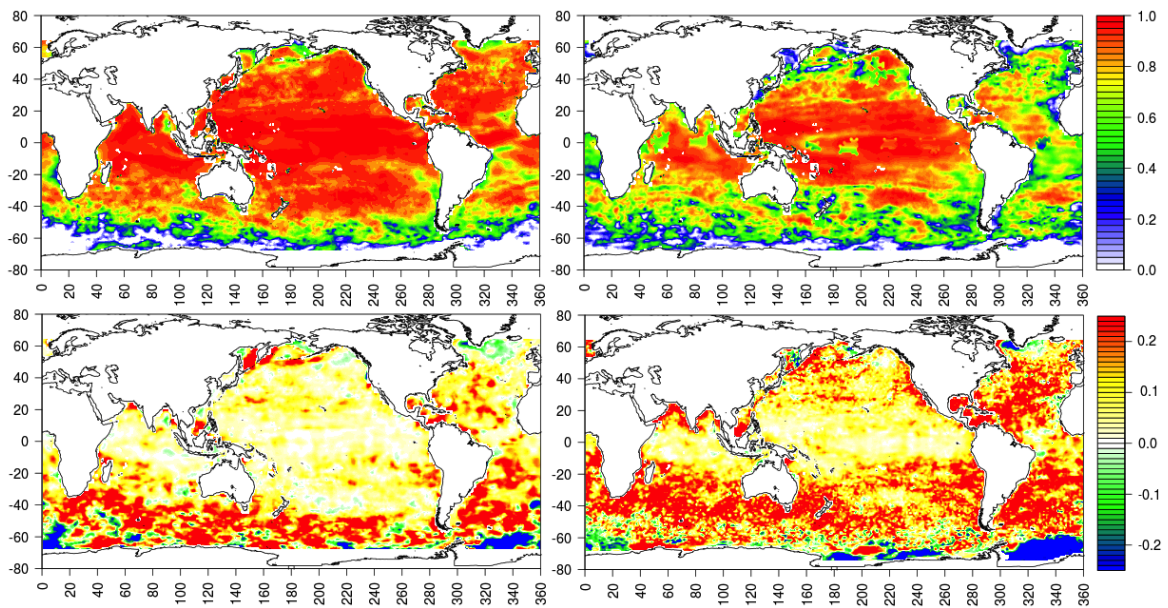
Fig 2: 2005-2009 Steric Sea Level anomaly correlation of EM-ORA with respect to the validation dataset (altimetry minus gravimetry) described in the text, for the full (top-left) and the inter-annual signal (top-right). Correlations higher than 0.25 are significant (at the 95% confidence level). The bottom panels show the map of differences between the EM-ORA anomaly correlation and the EM-OO anomaly correlation for the full (bottom-left) and the inter-annual signal (bottom-right). Positive (negative) values indicate that the correlation is higher (lower) with EM-ORA than with ORA-OO.

915

916

917

918

919

920

921

922

923

924

925

926
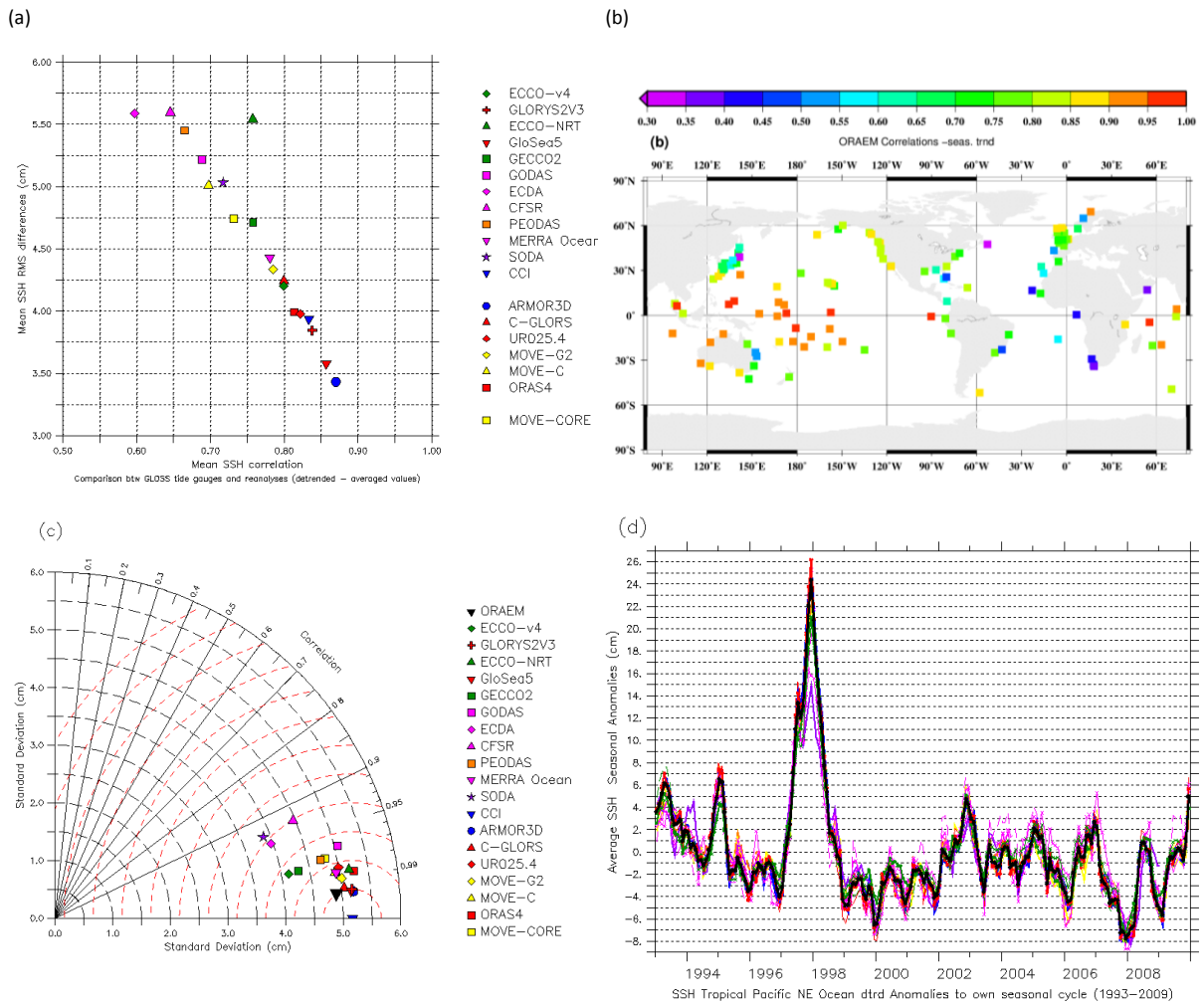
Fig 3: Top: Comparison between tide gauges and ORAs and OOs, after detrending and removing of seasonal cycle: a) RMS/Correlation diagram for the individual products using GLOSS tide gauge data as reference; b) correlations between EM-ORAalti and tide gauges time series, at tide gauge locations. Bottom: Evaluation of a sea level index: c) Taylor diagram using SLCCI as verification; d) Index time series, defined as the area-averaged sea level anomalies over the North-East Tropical Pacific region(0-12°N, 84-108°W). Anomalies and statistics have been computed over the 1993-2009 period.

927

928

929

930
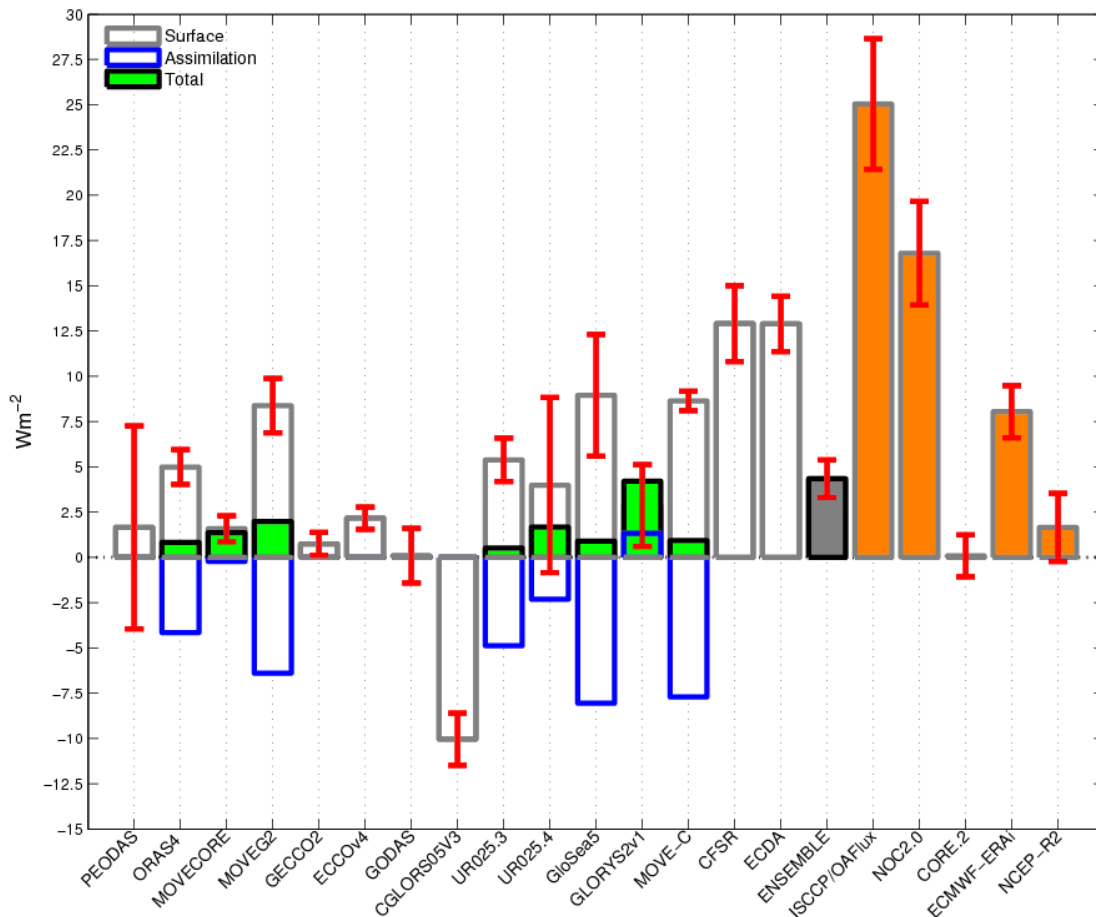
Fig 4: Time-mean global net "Surface" heat fluxes (grey bars) and their interannual standard deviations (red error bars) over the 17 years (1993 ‑ 2009) spanned by all data sets. The 15-member ensemble of "Surface" flux products is also shown (dark grey bar), along with observation based on atmospheric reanalysis products to the right hand side (orange bars). Eight products also have "Assimilation" fluxes (blue bars) computed by integrating the temperature increments from the surface down to the bottom, along with "Total" -fluxes, i.e., "Surface" +"Assimilation" fluxes (green bars). Positive is heat flux into the ocean. Units are in Wm^{-2}.

931

932

933

934

935

936

937

938

939

940
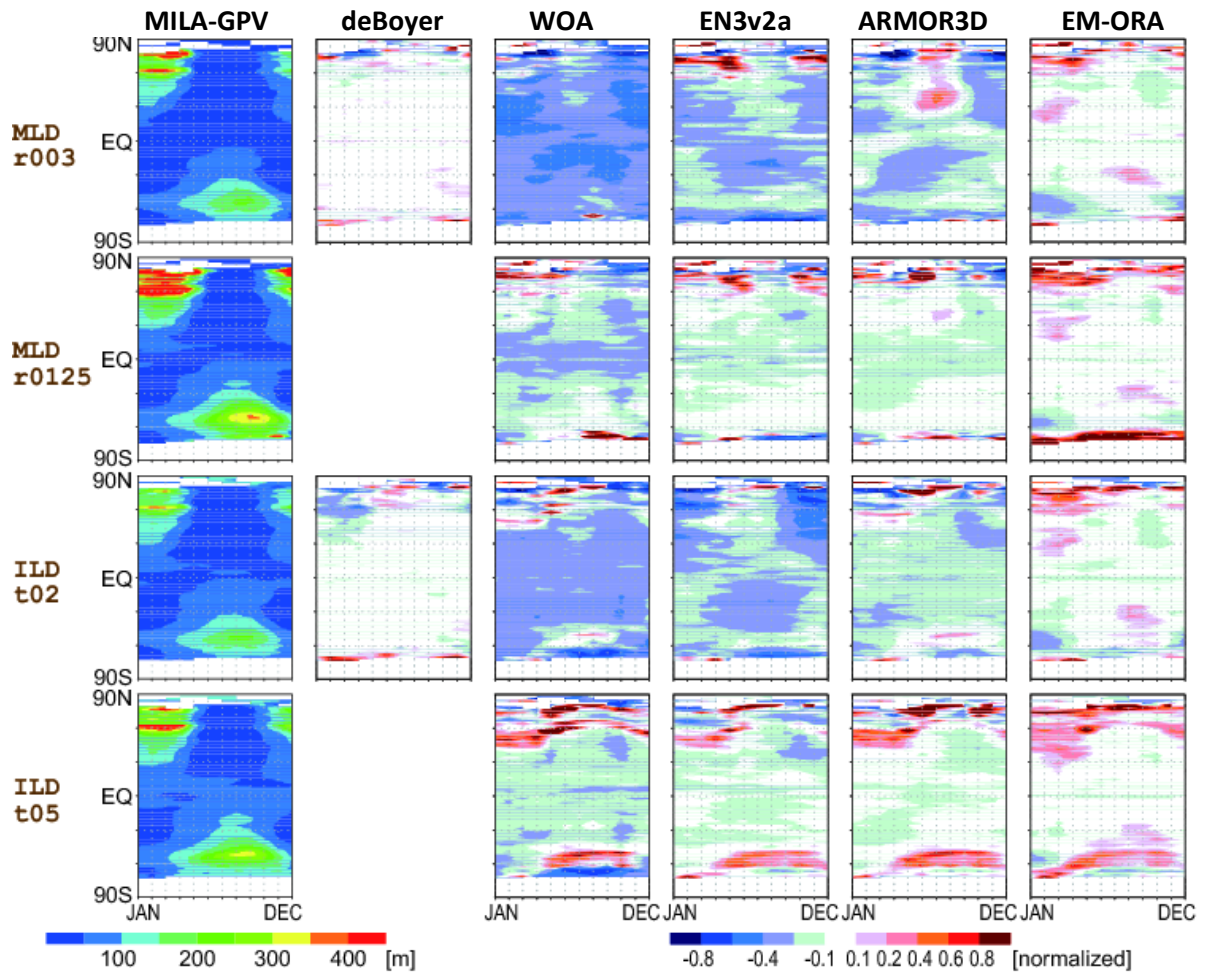
Fig 5: Zonal mean monthly MLDs and ILDs from MILA-GPV averaged over 2005-2011 (left column). Others: Differences from MILA-GPV, normalized by the MILA-GPV values.
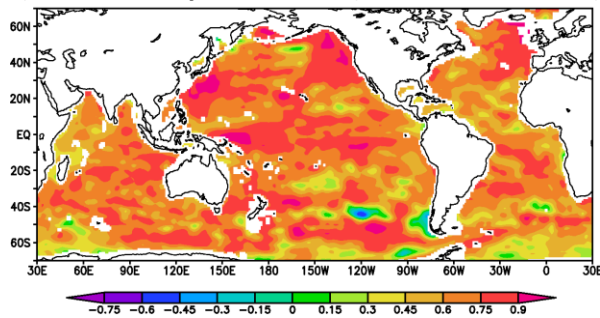
941

942

943

944

945
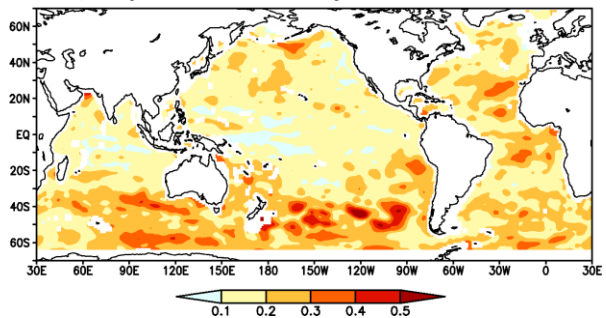
946

947

S700 Ensemble Statistics (1993-2010)

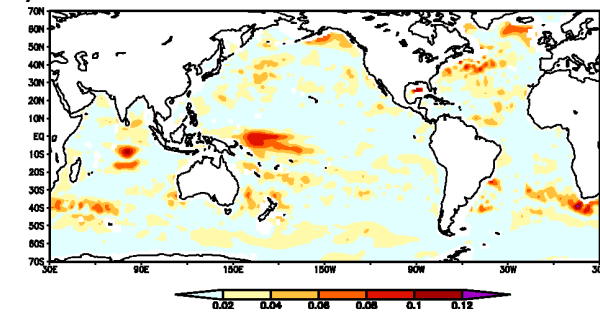a) **1993-2010 Mean Difference EMORA - EMOO**   b) **Ensemble Spread of the 1993-2010 Mean**
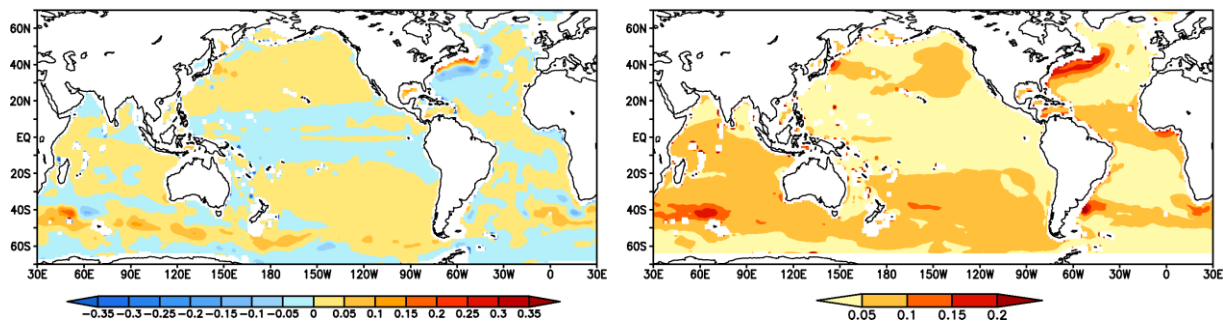
c) **Anomaly Correlation (EMORA and EMOO)**   d) **Spread in Anomaly Correlation**

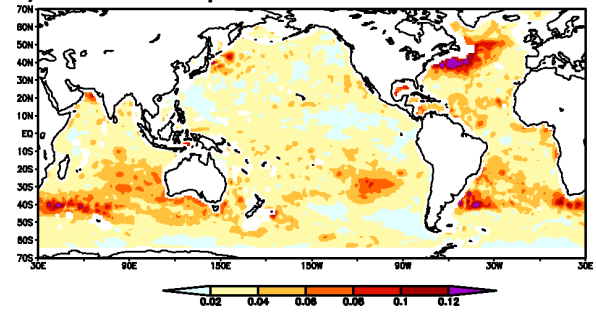e) **Time-Std of the EM Interannual Anom S700**   f) **Ensemble Spread of Interannual Anom**

Fig. 6 a) 1993-2010 mean difference of S700 (Depth-averaged salinity over 0-700m) between EM-ORA and EM-OO. The interval of colour bar is 0.05 psu.

b) The ensemble spread of the mean S700 (ESD-ORA). The interval of colour bar is 0.05 psu.

c) Temporal correlation of S700 monthly interannual anomalies between the EM-ORA and EM-OO, for the period 1993-2010.

d) Spread of correlation coefficients of S700 anomalies from the individual ORAs.

e) The inter-annual standard deviation (1993-2010) of EM-ORA S700, representative of the interannual "signal" (σ_{EM}^{I}). The interval of colour bar is 0.02 psu.

f) The average ensemble spread of the interannual anomalies of S700 (σ_{ESD}^{I} representing the uncertainty or 'noise'). The interval of colour bar is 0.02 psu.
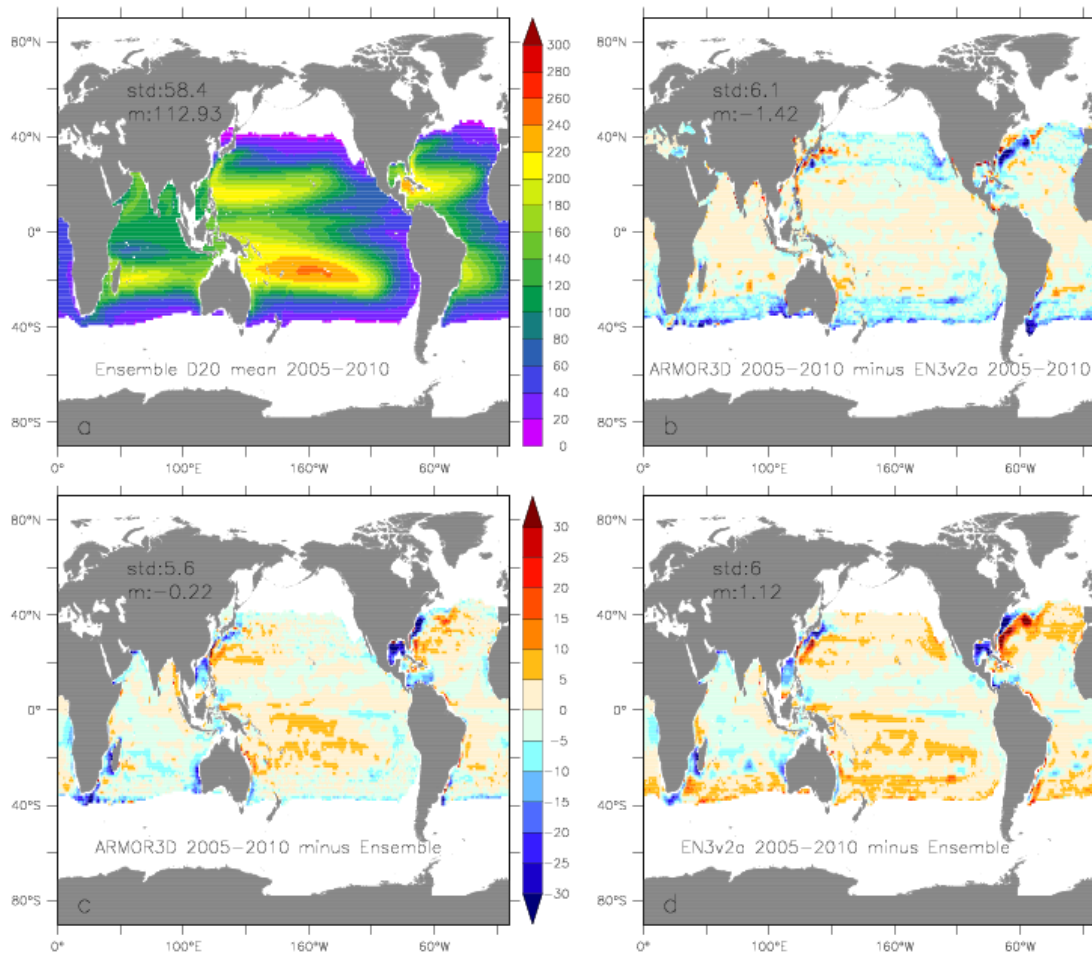
948

949

Figure 7: (a) Global map of mean of D20 from EM-ORA. Differences in mean D20 between (b) the two OO products ARMOR3D and EN3v2, (c) EM-ORA and ARMOR3D and (d) EM-ORA and EN3v2. Units are m. The mean fields have been calculated over the 2005-2010 periods.
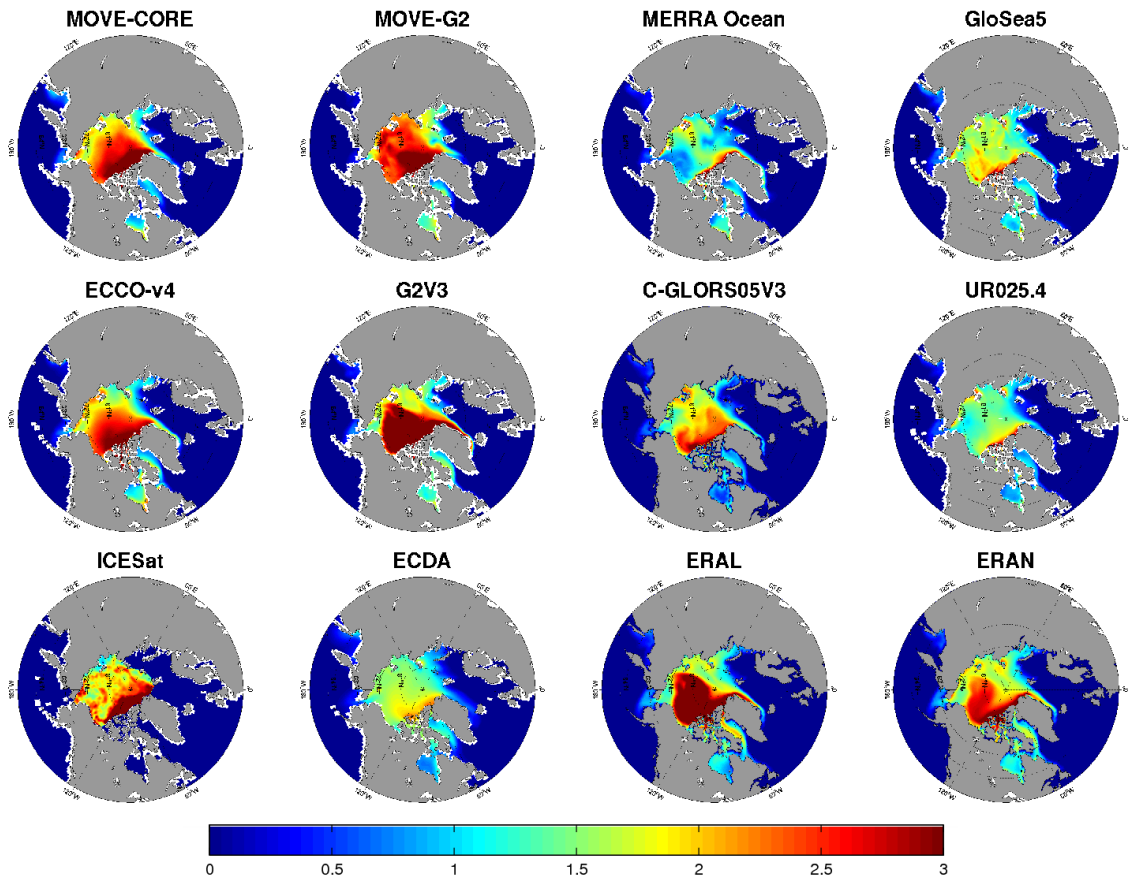
950

951

Fig. 8: Example of mean sea ice thickness for the various ice-ocean reanalyses for March 2007. Also shown is a satellite estimate of sea ice thickness from ICESat (bottom left).

952