Ocean Science

Evaluation of global monitoring and forecasting systems at Mercator Océan

J.-M. Lellouche[1], O. Le Galloudec[1], M. Drévillon[1], C. Régnier[1], E. Greiner[2], G. Garric[1], N. Ferry[1], C. Desportes[1], C.-E. Testut[1], C. Bricaud[1], R. Bourdallé-Badie[1], B. Tranchant[2], M. Benkiran[2], Y. Drillet[1], A. Daudin[1], and C. De Nicola[1]

[1]Mercator Océan, Ramonville Saint Agne, France
[2]CLS, Ramonville Saint Agne, France

Correspondence to: J.-M. Lellouche (jlellouche@mercator-ocean.fr)

Abstract. Since December 2010, the MyOcean global analysis and forecasting system has consisted of the Mercator Océan NEMO global 1/4° configuration with a 1/12° nested model over the Atlantic and the Mediterranean. The open boundary data for the nested configuration come from the global 1/4° configuration at 20° S and 80° N.

The data are assimilated by means of a reduced-order Kalman filter with a 3-D multivariate modal decomposition of the forecast error. It includes an adaptive-error estimate and a localization algorithm. A 3-D-Var scheme provides a correction for the slowly evolving large-scale biases in temperature and salinity. Altimeter data, satellite sea surface temperature and in situ temperature and salinity vertical profiles are jointly assimilated to estimate the initial conditions for numerical ocean forecasting. In addition to the quality control performed by data producers, the system carries out a proper quality control on temperature and salinity vertical profiles in order to minimise the risk of erroneous observed profiles being assimilated in the model.

This paper describes the recent systems used by Mercator Océan and the validation procedure applied to current MyOcean systems as well as systems under development. The paper shows how refinements or adjustments to the system during the validation procedure affect its quality. Additionally, we show that quality checks (in situ, drifters) and data sources (satellite sea surface temperature) have as great an impact as the system design (model physics and assimilation parameters). The results of the scientific assessment are illustrated with diagnostics over the year 2010 mainly, assorted with time series over the 2007–2011 period. The validation procedure demonstrates the accuracy of MyOcean global products, whose quality is stable over time. All monitoring systems are close to altimetric observations with a forecast RMS difference of 7 cm. The update of the mean dynamic topography corrects local biases in the Indonesian Throughflow and in the western tropical Pacific. This improves also the subsurface currents at the Equator. The global systems give an accurate description of water masses almost everywhere. Between 0 and 500 m, departures from in situ observations rarely exceed 1 °C and 0.2 psu. The assimilation of an improved sea surface temperature product aims to better represent the sea ice concentration and the sea ice edge. The systems under development are still suffering from a drift which can only be detected by means of a 5-yr hindcast, preventing us from upgrading them in real time. This emphasizes the need to pursue research while building future systems for MyOcean2 forecasting.

1 Introduction

MyOcean, and now the MyOcean2 project (http://www.myocean.eu), aim to deploy the first integrated pan-European ocean monitoring and forecasting system of the GMES (Global Monitoring for Environment and Security, http://www.gmes.info) Marine Core Service. The target applications include maritime security, oil spill response, marine resources management, climate change, seasonal forecasting, coastal activities, sea ice monitoring and water quality and pollution management. The MyOcean service provides the

best ocean information available for the global and regional scales by combining space and in situ observations and assimilating them into 3-D simulation models.

Scientific quality is one of the criteria for the continuous improvement of MyOcean products. Moreover, metadata on scientific quality greatly help users understand the content of products and their usefulness. It is thus very important for monitoring and forecasting centres (MFCs) to agree on quality standards and to produce homogeneous and accessible information on the scientific quality of their analyses and forecasts. Some of these standards have to be the same as those applied by ocean observation thematic assembly centres (TACs) to ensure that users are given comprehensive and consistent information.

The scientific assessment procedure applied for MyOcean consists of two phases. During the first "calibration phase", new products or developments are checked with a series of metrics before being commissioned. Once the product has been commissioned it then undergoes an "operational validation phase" during which the products are checked against the reference calibration results.

Standards and metrics were defined during the MERSEA (Marine Environment and Security for the European Area, http://www.ifremer.fr/merseaip) integrated project and in the context of GODAE (Global Ocean Data Assimilation Experiment, http://www.godae.org). These standardized diagnostics have enabled comparative exercises between European operational oceanography MFCs (Crosnier and Le Provost, 2007) and others outside of the EU (Hernandez et al., 2009).

Some of these metrics were proposed for calibration and validation purposes. An efficient validation procedure has already been defined by Mercator Océan for its model, including scientific assessment (calibration) and quarterly control bulletins (validation). These documents give a general picture of the normal behaviour of the system in terms of accuracy and realism of the ocean physics. The accuracy is measured by the differences between simulations and observations and the realism by studying particular oceanic processes. In addition, more than a thousand diagnostic checks are routinely performed every day at Mercator Océan.

Since January 2009, Mercator Océan, which has primary responsibility for the global ocean forecasts of the MyOcean project, has developed several versions of its monitoring and forecasting system for the various milestones, V0, V1 and V2, of the MyOcean project (Fig. 1). Other forecasting centres in and outside of Europe also produce global forecasts (Dombrowsky et al., 2009). The Mercator Océan monitoring and forecasting system is based on the NEMO (Nucleus for European Modelling of the Ocean) ocean-modelling platform and on the SAM (*Système d'Assimilation Mercator*) data assimilation system. It is used with several configurations covering different geographical areas with various horizontal and vertical resolutions. The main target configuration is a global high-resolution system with 1/12° horizontal grid spacing and 50 vertical levels. Due to the high computing
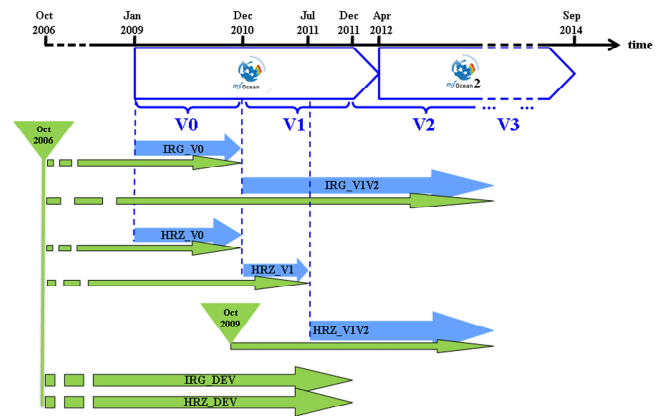

Fig. 1. Timeline of the MyOcean global analysis and forecasting system for the various milestones V0, V1 and V2. Real-time MyOcean production is in blue. Available Mercator Océan simulation is in green, including the catch-up to real time. The latest most advanced versions of systems in terms of scientific developments are named IRG_DEV and HRZ_DEV for intermediate-resolution global and high-resolution zoom systems, respectively.

cost, the major scientific advances are first implemented and tested with a global intermediate-resolution system at 1/4° with 50 vertical levels (hereafter referred to as IRG). High resolution is maintained for the Atlantic and Mediterranean by nesting a high-resolution zoom at 1/12° (hereafter referred to as HRZ).

Global MyOcean products are available for eleven regions (ocean basins) that match the intercomparison zones defined for the international GODAE Oceanview initiative (Hernandez et al., 2009). For each zone the best available product was selected for distribution through MyOcean. The highest resolution available is provided for each zone in order to ensure the highest possible accuracy and consistency.

This article presents the main results of the scientific assessment of MyOcean V2 global MFC at Mercator Océan and shows how refinements or adjustments to the system following validation affect the quality of the system. It focuses on HRZ and IRG, as these are the most scientifically advanced implementations of both the physical model and the data assimilation system. The main characteristics of these monitoring and forecasting systems are described in Sect. 2. The validation methodology is detailed in Sect. 3. The main results of the scientific assessment are given in Sect. 4. For each diagnostic, the results of the MyOcean V2 system (current IRG_V1V2 and HRZ_V1V2) are compared with the results from the latest versions of the systems under development (IRG_DEV and HRZ_DEV) that will be available in MyOcean2 (see Fig. 1). Section 5 contains a summary of the scientific assessment, as well as a discussion of improvements expected for MyOcean2 in terms of performance and validation procedures.

2 Description of the monitoring and forecasting systems

This chapter contains a general description of the systems, particularly of IRG_V1V2 and HRZ_V1V2. The main differences and links between the various versions are summarised in Tables 1 and 2 for all IRG and HRZ systems, respectively.

2.1 A common basis for all forecasting systems

2.1.1 Physical model

Since MyOcean V1, the IRG and HRZ systems have used version 3.1 of NEMO (Madec et al., 2008). Both configurations are based on the tripolar ORCA grid type (Madec and Imbard, 1996) with, for the IRG system, a horizontal resolution of 27 km at the Equator, 21 km at Cape Hatteras (mid-latitudes) and 6 km toward the Ross and Weddell Seas. The HRZ system horizontal grid is three times more refined than the IRG system grid. The 50-level vertical discretization retained for these systems has 1 m resolution at the surface, decreasing to 450 m at the bottom, and 22 levels within the upper 100 m. "Partial cell" parameterization was chosen for a better representation of the topographic floor (Barnier et al., 2006), and the momentum advection term is computed with the energy and enstrophy conserving scheme proposed by Arakawa and Lamb (1981). The advection of the tracers (temperature and salinity) is computed with a total variance diminishing (TVD) advection scheme (Lévy et al., 2001; Cravatte et al., 2007). The high frequency gravity waves are filtered out by a free surface (Roullet and Madec, 2000). A Laplacian lateral isopycnal diffusion on tracers ($300\,\mathrm{m^2\,s^{-1}}$ for IRG, $125\,\mathrm{m^2\,s^{-1}}$ for the HRZ at the Equator and decreasing poleward, proportionally to the grid size) and a horizontal biharmonic viscosity for momentum ($-1\times10^{11}\,\mathrm{m^4\,s^{-1}}$ for IRG and $-1.5\times10^{10}\,\mathrm{m^4\,s^{-1}}$ for the HRZ at the Equator and decreasing poleward as the cube of the grid size) are used. In addition, the vertical mixing is parameterized according to a turbulent closure model (order 1.5) adapted by Blanke and Delecluse (1993), the lateral friction condition is a partial-slip condition with a regionalisation of a no-slip condition (over the Mediterranean Sea) for the HRZ system, and the elastic-viscous-plastic rheology formulation for the LIM2 ice model (hereafter called LIM2_EVP, Fichefet and Maqueda, 1997) has been activated (Hunke and Dukowicz, 1997).

The bathymetry used in the IRG (respectively HRZ) system comes from a combination of interpolated ETOPO1 (respectively ETOPO2) (Amante and Eakins, 2009) and GEBCO8 (Becker et al., 2009) databases. ETOPO datasets are used in regions deeper than 300 m, and GEBCO8 is used in regions shallower than 200 m with a linear interpolation in the 200 m–300 m layer.

The monthly runoff climatology is built with data on coastal runoffs and 100 major rivers from Dai and Trenberth (2002) together with an annual estimate of Antarctica ice sheet melting given by Jacobs et al. (1992). Barotropic mixing due to tidal currents in the semi-enclosed Indonesian Throughflow region has been parameterized in IRG_V1V2 following Koch-Larrouy et al. (2008). The atmospheric fields are taken from the ECMWF (European Centre for Medium-Range Weather Forecasts) Integrated Forecast System. A 3-h sampling is used to reproduce the diurnal cycle with a thickness of 1 m for the uppermost level. According to Bernie et al. (2005), this temporal and vertical resolution is sufficient to capture 90 % of the sea surface temperature (SST) diurnal variability and the maximum heating rates of the diurnal cycle. Momentum and heat turbulent surface fluxes are computed from CORE bulk formulae (Large and Yeager, 2009) using the usual set of atmospheric variables: surface air temperature at a height of 2 m, surface humidity at a height of 2 m, mean sea level pressure and wind at a height of 10 m. Downward longwave and shortwave radiative fluxes and rainfall (solid + liquid) fluxes are also used in the surface heat and freshwater budgets. An analytical formulation (Bernie et al., 2005) is applied to the shortwave flux in order to reproduce an ideal diurnal cycle. Lastly, the system does not include tides and so does not intend to simulate continental shelf areas with large tides.

Considering the physical model, the main improvements of the current systems (IRG_V1V2 and HRZ_V1V2) over the V0 systems concern the use of high frequency (3 h) atmospheric forcing including the diurnal cycle, the use of the CORE bulk formulation instead of the CLIO one (Goosse et al., 2001), and the use of the LIM2_EVP ice model (see Tables 1 and 2). The use of a one-way nesting between the systems is another improvement with respect to V0. Since V1, the IRG system gives boundary conditions for the HRZ system.

2.1.2 Data assimilation scheme

The SAM data assimilation method relies on a reduced-order Kalman filter based on the singular evolutive extended Kalman filter (SEEK) formulation introduced by Pham et al. (1998). This approach has been used for several years at Mercator Océan and has been implemented in different ocean model configurations with a 7-day assimilation window (Tranchant et al., 2008; Cummings et al., 2009). In all Mercator Océan forecasting systems, the forecast error covariance is based on the statistics of a collection of 3-D ocean state anomalies, typically a few hundred. This approach is based on the concept of statistical ensembles in which an ensemble of anomalies is representative of the error covariances. In this way, truncation no longer occurs and all that is needed is to generate the appropriate number of anomalies. This approach is similar to the Ensemble Optimal Interpolation developed by Oke et al. (2008). In our case, the anomalies are computed from a long numerical experiment (free run without assimilation, typically around 10 yr) with respect to a running mean in order to estimate the 7-day scale error on

Table 1. Specifics of the Mercator Océan IRG systems. In bold, the major upgrades with respect to the previous version.
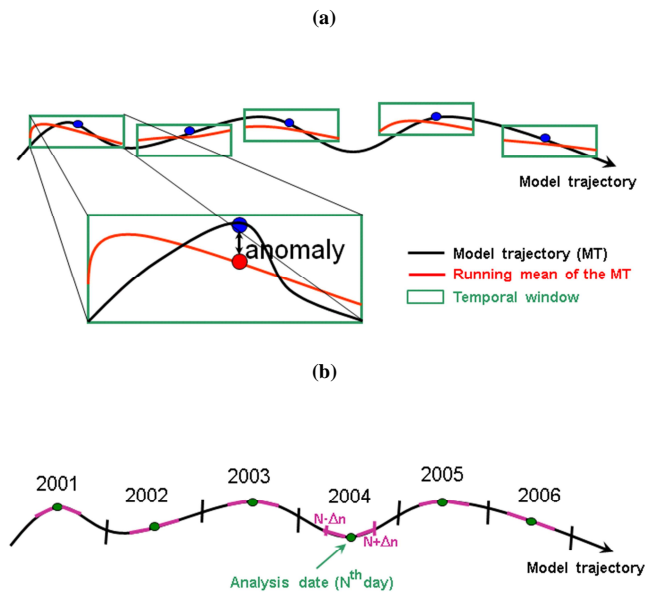
System acronym	Domain	Resolution	Model	Assimilation	Assimilated observations	MyOcean version	Mercator Océan system reference
IRG_V0	Global	Horizontal: 1/4°	ORCA025 NEMO 1.09	SAM (SEEK)	"RTG" SST	V0	PSY3V2R1
		Vertical: 50 levels	LIM2, Bulk CLIO		SLA		
			24-h atmospheric forcing		T/S vertical profiles		
IRG_V1V2	Global	Horizontal: 1/12°	ORCA025 **NEMO 3.1**	SAM (SEEK)	"RTG" SST	V1/V2	PSY3V3R1
		Vertical: 50 levels	**LIM2 EVP, Bulk CORE**	**IAU**	SLA		
			3-h atmospheric forcing	**3-D-Var bias correction**	T/S vertical profiles		
IRG_DEV	Global	Horizontal: 1/4°	ORCA025 NEMO 3.1	SAM (SEEK)	**"AVHRR + AMSRE" SST**		PSY3V3R2
		Vertical: 50 levels	LIM2 EVP, Bulk CORE	IAU	SLA		
			3-h atmospheric forcing	3-D-Var bias correction	T/S vertical profiles		
			New parameterization of vertical mixing	**Obs. errors higher near the coast (for SST and SLA) and on shelves (for SLA)**	**MDT "CNES-CLS09" adjusted**		
			Taking into account ocean colour for depth of light extinction		**Sea Mammals T/S vertical profiles**		
				MDT error adjusted			
			Large-scale correction to the downward radiative and precipitation fluxes	**Increase of Envisat altimeter error**			
				QC on T/S profiles			
			Adding runoff for iceberg melting	**New correlation radii**			
			Adding seasonal cycle for surface mass budget				

the ocean state at a given period of the year for temperature (T), salinity (S), zonal velocity (U), meridional velocity (V) and sea surface height (SSH). More precisely, each temporal anomaly M' corresponds to the difference between the model state M and a running mean $\langle M \rangle_{-\tau}^{+\tau}$ over a fixed time period window ranging from $-\tau$ to τ (see Fig. 2a). Moreover, the signal at a few horizontal grid "Δx" intervals in the model outputs on the native full grid is not physical but only numerical (Grasso, 2000). This signal should not be taken into account when updating an analysis. This is why several passes of a Shapiro filter are applied in order to remove the very short scales that in practice correspond to numerical noise. Consequently, a little subsampling of the model state is applied without aliasing error, and the anomalies are thus calculated on a reduced horizontal grid (1 out of every 2 points in both horizontal directions and all the points along the coast) to limit the storage and the load cost during the analysis stage.

To create the running mean $\langle M \rangle_{-\tau}^{+\tau}$, a Hanning low-pass filter is used:

$$Ha(\nu) = \begin{cases} 0.5 + 0.5\cos\left(\frac{\pi.\nu}{\nu_{\max}}\right) & \text{for } |\nu| \leq \nu_{\max} \\ 0 & \text{for } |\nu| > \nu_{\max} \end{cases}, \quad (1)$$

where ν is the temporal frequency of the model state and $\nu_{\max}$ is the cut-off frequency (equal to $1/36\,\text{days}^{-1}$ in our case). The main characteristic of the anomaly calculation is to filter

Fig. 2. Schematic representation of the anomalies calculation along a model trajectory **(a)** and of the use of these anomalies to build the model forecast covariance **(b)**.

Table 2. Specifics of the Mercator Océan HRZ systems. In bold, the major upgrades with respect to the previous version.

System acronym	Domain	Resolution	Model	Assimilation	Assimilated observations	Inter-dependencies	MyOcean version	Mercator Océan system reference
HRZ_V0	Tropical	Horizontal: 1/12°	NATL12 NEMO 1.09	SAM (SEEK)	"RTG" SST		V0	PSY2V3R1
	North Atlantic	Vertical: 50 levels	LIM2, Bulk CLIO		SLA			
	Mediterranean		24-h atmospheric forcing		T/S vertical profiles			
HRZ_V1	Tropical	Horizontal: 1/12°	NATL12 **NEMO 3.1**	SAM (SEEK)	RTG SST	**OBC from IRG_V1V2**	V1	PSY2V4R1
	North Atlantic	Vertical: 50 levels	**LIM2 EVP, Bulk CORE**	**IAU**	SLA			
	Mediterranean		**3-h atmospheric forcing**	**3-D-Var bias correction**	T/S vertical profiles			
HRZ_V1V2	Tropical	Horizontal: 1/12°	NATL12 NEMO 3.1	SAM (SEEK)	"AVHRR + AMSRE" SST	OBC from IRG_V1V2	V1/V2	PSY2V4R2
	North Atlantic	Vertical: 50 levels	LIM2 EVP, Bulk CORE	IAU	SLA			
	Mediterranean		3-h atmospheric forcing	3-D-Var bias correction	T/S vertical profiles			
				Obs. errors higher near the coast (for SST and SLA) and on shelves (for SLA)	**MDT CNES-CLS09 adjusted**			
				MDT error adjusted				
				New correlation radii				
HRZ_DEV	Tropical	Horizontal: 1/12°	NATL12 NEMO 3.1	SAM (SEEK)	"AVHRR + AMSRE" SST	OBC from IRG_DEV		PSY2V4R3
	North Atlantic	Vertical: 50 levels	LIM2 EVP, Bulk CORE	IAU	SLA			
	Mediterranean		3-h atmospheric forcing	3-D-Var bias correction	T/S vertical profiles	**Spatial mean evaporation minus precipitation from IRG_DEV**		
			New parameterization of vertical mixing	Obs. errors higher near the coast (for SST and SLA) and on shelves (for SLA)	MDT CNES-CLS09 adjusted			
			Taking into account ocean colour for depth of light extinction		**Sea Mammals T/S profiles**			
				MDT error adjusted				
			Large-scale correction to the downward radiative and precipitation fluxes	New correlation radii				
				Increase of Envisat altimeter error				
				QC on T/S profiles				

out temporal scales at low frequencies in order to keep high frequencies for which the period is shorter than two or three assimilation cycles. For an assimilation cycle centred on the N-th day of a given year, ocean state anomalies falling in the window $[N - \Delta n; N + \Delta n]$ of each year of the free run are gathered and define the covariance of the model forecast error (see Fig. 2b). In our case, Δn is equal to 60 days, which means that anomalies are selected over 120-day windows centred on the N-th day of each year of the free run. So in SAM, the forecast error covariances rely on a fixed basis, seasonally variable ensemble of anomalies. This method implies that at each analysis step a sub-set of anomalies is used that improves the dynamic dependency. A significant number of anomalies are kept from one analysis to the other, thus ensuring error covariance continuity. It should also be noted that the analysis increment is a linear combination of these anomalies and depends on the innovation (observation minus model forecast equivalent) and on the specified ob-

servation errors. A particular feature of the SEEK is that the error covariance only gives the direction of the model error and not its intensity. An adaptive scheme for the model error variance calculates an optimal variance of the model error based on a statistical test formulated by Talagrand (1998). The last feature of the model forecast covariance employed is a localization technique which sets the covariances to zero beyond a distance defined as twice the local spatial correlation scale. Because a finite number of ocean state anomalies have been used to build the model forecast covariance, the latter is not significant when further away than this particular distance from the analysis point. This is why it is preferable to not use this information and to set the covariance to zero. Spatial (zonal and meridional directions) and temporal correlation scales (Fig. 3) are then used to define an "influence bubble" around the analysis point in which data are also selected. For the IRG_V1V2 system, spatial and temporal correlation scales were calculated a priori from sea level

anomalies (SLA) observed by satellites (SSALTO/DUACS) from 1993 to 2006. For the IRG_DEV system, the correlation scales (longitude, latitude, time) are deduced a posteriori from the Mercator Océan global 1/4° reanalysis GLORYS2V1 (GLobal Ocean ReanalYsis and Simulation, Ferry et al., 2012). The 2004–2009 Argo (Roemmich et al., 2008) period is used because it strongly constrains the reanalysis. Scales are computed from the GLORYS2V1 2-D hindcast fields of temperature at 100 m and 300 m. At every point, space and time lag correlations are computed with neighbouring points. The distances at which the correlation with the central point falls below 0.4 determine the length scales. Note that all time series are time filtered (no trend, 3–90 days band-pass). To avoid having to change the settings of the SEEK between IRG_V1V2 and IRG_DEV, the radii are increased by a factor of 1.5 in order to have the same order of magnitude on average as the IRG_V1V2 system radii.

To save computing time, the analysis is performed on a reduced grid (1 out of every 4 points in both horizontal directions, all the points along the coast and 1 out of every 2 points in the first 150 km from the coast). As can be seen in Fig. 3, the spatial correlation scales are about 100 km for most of the globe. So, for the IRG system, analyses are computed on a 1° grid, and data influence is set to zero beyond about 200 km and down-weighted at a distance of 100 km. The horizontal spatial structures of the ensemble may be almost completely discarded by the localization. In theory, both the zonal and meridional length scales are not large enough for the 1/4° model (IRG systems), with the potential problem mentioned above at mid latitudes, away from the coast. At high latitudes, the problem is reduced since the distance between two points of analysis varies with the cosine of the latitude. There is no problem in the equatorial band where the radii are larger. However, with this sub-sampled grid, the length scales are large enough for the 1/12° model (HRZ systems). Using more points (1 out of every 2 points in both horizontal directions and all the points along the coast) for the IRG analysis grid does not yield any significant differences (not shown). A one-week test simulation was also performed by imposing radii larger than 200 km (instead of a minimum of 75 km in the IRG_DEV reference run) (see Fig. 3). The increments obtained with both simulations have globally the same structures. Differences concern some extrema linked to the assimilation of temperature and salinity profiles (either bad or isolated) and/or to an incompatibility between these profiles and another type of assimilated data. Part of the increment is rejected because of its inconsistency with the model dynamics. Few unrealistic increments are still taken into account, with a very slight local degradation of the system. The size of the radii is finally imposed by the cost of the system in an operational context.

An important difference of MyOcean V2 systems with respect to more classical forecasting systems is that the analysis is not performed at the end of the assimilation window but at the middle of the 7-day assimilation cycle. The objective is
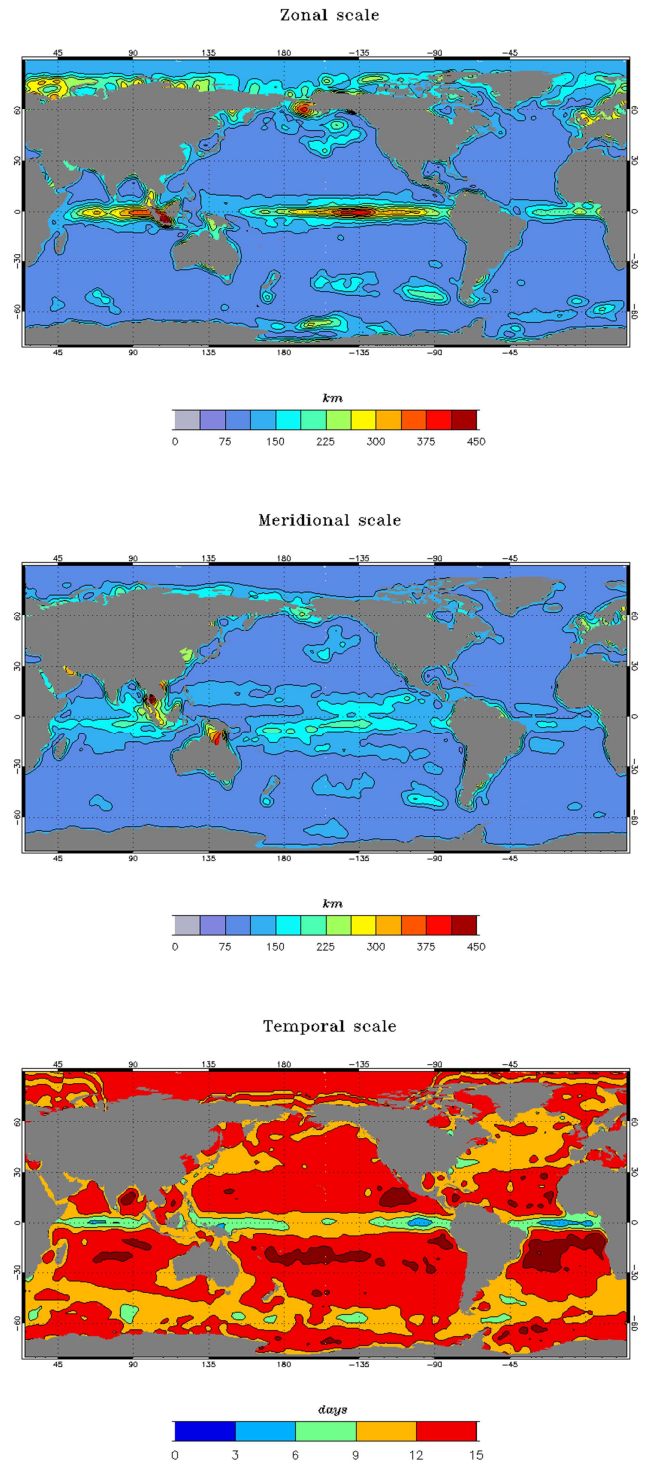
Fig. 3. Zonal, meridional (km) and temporal (days) correlation scales (from top to bottom) used by the IRG_DEV, HRZ_V1V2 and HRZ_DEV systems.

to take into account both past and future information and to provide the best estimate of the ocean centred in time. With such an approach, the analysis, to some extent, acts like a

smoother algorithm. For technical reasons, this could not be done exactly at time = 3.5 days so it has been slightly shifted to time = 4 days.

After each analysis, the data assimilation produces increments of barotropic height, temperature, salinity and zonal and meridional velocity. The SSH increment is the sum of barotropic and dynamic height increments. Dynamic height increment is calculated from the temperature and salinity increments. All these increments are applied progressively using the incremental analysis update (IAU) method (Bloom et al., 1996; Benkiran and Greiner, 2008), which makes it possible to avoid model shock every week due to the imbalance between the analysis increments and the model physics. In this way, the IAU reduces spin-up effects. It is fairly similar to a nudging technique but does not exhibit weaknesses such as frequency aliasing and signal damping. Following the analysis performed at the end of the forecast (or background) model trajectory (referred to as "FORECAST" first trajectory, with analysis time at the 4th day of the cycle), a classical forward scheme would continue straight on from this analysis, integrating from day 7 until day 14. Instead, the IAU scheme rewinds the model and starts again from the beginning of the assimilation cycle, integrating the model for 7 days (referred to as "BEST" second trajectory) with a tendency term added in the model prognostic equations for temperature, salinity, sea surface height and horizontal velocities. The tendency term (which is equal to the increment divided by the length of the cycle) is modulated by an increment distribution function shown in Fig. 4. The time integral of this function equals 1 over the cycle length. In practice, the IAU scheme is more costly than the "classical" model correction (increment applied on one time step) because of the additional model integration ("BEST" trajectory) over the assimilation window.

In addition to the assimilation scheme, a method of bias correction has been developed. This method is based on a 3-D-Var approach which takes into account cumulative innovations over the last 3-month period in order to estimate large-scale temperature and salinity biases when enough observations are available. The aim of the bias correction is to correct the large-scale, slowly evolving error of the model, whereas the SAM assimilation scheme is used to correct the smaller scales of the model forecast error. The bias correction involves several steps. First, temperature and salinity innovations over the last three months are binned and averaged on a coarse resolution ($1° \times 1°$) grid. The two variables are treated separately because temperature and salinity biases are not necessarily correlated. Then, the 3-D-Var method is used to analyse the bias. The bias covariance is constrained by the current patterns and structures of density fronts in the ocean (the bias is large in regions with sharp gradients). There is little bias correction in the mixed layer if the vertical gradient of the thermocline is sharp. The bias correction is fully effective under the thermocline, away from density gradients. The correlations are modelled by means of an anisotropic Gaus-
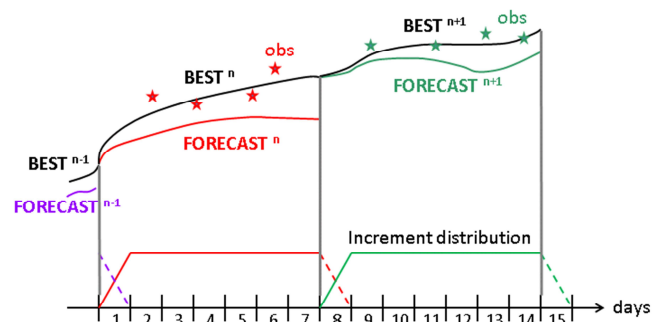
Fig. 4. Schematic representation of the IAU procedure for three consecutive cycles, $n-1$, n and $n+1$. Following the analysis performed at the end of the forecast (or background) model trajectory (referred to as "FORECAST" first trajectory, with analysis time at the 4th day of the cycle), the IAU scheme rewinds the model and starts again from the beginning of the assimilation cycle, integrating the 7-day run (referred to as "BEST" second trajectory) with a tendency term added in the model prognostic equations and modulated by an increment distribution function. The time integral of this function equals 1 over the cycle length.

sian recursive filter. Bias correction of temperature, salinity and dynamic height are then computed and interpolated on the model grid. Lastly, these bias corrections are applied as tendencies in the model prognostic equations, with a 3-month time scale.

The assimilated observations consist of along-track altimeter SLA from AVISO, satellite SST from either NCEP or NOAA, and temperature and salinity in situ vertical profiles from the CORIOLIS data centre, including Argo. Along each track of SLA, only one point each third is conserved to avoid redundant information. Moreover, observations along tracks are smoothed by several altimetric corrections (Le Traon et al., 2001). The first guess at appropriate time (FGAT) method (Huang et al., 2002) is used, which means that the forecast model equivalent of the observation for the innovation computation is taken at the time for which the data is available, even if the analysis is delayed. The concept of "pseudo-observations" or "observed no change" (innovation equal to zero) has also been used to overcome the deficiencies of the background errors, in particular for extrapolated and/or poorly observed variables. We apply this approach to the barotropic height and the 3-D coastal salinity at river mouths and all along the coasts (run off rivers). Pseudo-observations are also used for the 3-D variables T, S, U and V under the ice, between $6°$ S and $6°$ N below a depth of 200 m and near open boundaries of HRZ systems. These observations are geographically positioned on the analysis grid points rather than on a coarser grid in order to avoid generating aliasing on the horizontal. The time of these observations is the same as for the analysis, namely the fourth day of a 7-day assimilation cycle. Given ongoing concern about the need to reduce costs in an operational context, the 3-D variables mentioned

above were sampled on the vertical in order to keep only about ten model levels. Lastly, the mean dynamic topography (MDT) named "CNES-CLS09" derived from observations and described in Rio et al. (2011) is used as a reference for SLA assimilation.

The main data assimilation improvements of the current systems (IRG_V1V2 and HRZ_V1V2) with respect to the previous ones concern the insertion of the zonal and meridional velocity components into the control vector, the use of the IAU procedure, the insertion of new observational operators, the use of the CNES-CLS09 MDT, the introduction of 2-D and 3-D pseudo-observations and the use of a bias correction method (see Tables 1 and 2).

2.2 The HRZ_V1V2 specifics

Unrealistic salinities were diagnosed by several users (of coastal applications) shortly after the real-time HRZ_V1 system went online. This problem occurred in the HRZ_V1 products on the continental shelves, and in particular in tidal areas like the Celtic Seas, the North Sea and the Bay of Biscay. An upgrade of the system, called HRZ_V1V2, was implemented in order to correct these biases and it replaced the previous HRZ_V1 system in July 2011 (see Fig. 1). The updates included a modification of the multivariate data assimilation scheme in order to use an adjusted version of CNES-CLS09 MDT including GOCE observations and bias correction. An intermediate resolution SST at 1/4° including AVHRR and AMSRE observations (Reynolds et al., 2007), referred to below as "AVHRR + AMSRE", was also assimilated instead of NCEP real-time global (RTG) SST (Gemmill et al., 2007), referred to below as "RTG". The observation error variance was increased for the assimilation of SLA near the coast and on the shelves, and for the assimilation of SST near the coast (within 50 km of the coast). The spatial correlation radii were modified everywhere in order to improve the analysis, particularly near the European coast. The system was restarted from October 2009 instead of October 2006, as for all the other systems, because of the computing time required to catch up with real time and the need to correct operational analyses and forecasts quickly. Temperatures and salinities were initialised with climatological conditions from Levitus 2005 (Antonov et al., 2006; Locarnini et al., 2006). The initial condition for sea ice concentration was inferred from the IFREMER/CERSAT products (Ezraty et al., 2007) for October 2009. The sea ice thickness distribution came directly from the Mercator Océan global 1/4° reanalysis GLORYS2V1. A monthly average (October 2009) of the sea ice field was used, the latter having the advantage of being dynamically equilibrated after the 17 yr (1993–2009) of the reanalysis experiment. These changes have solved the problem described above.

2.3 Updates for forthcoming MyOcean systems

The current IRG_V1V2 MyOcean system is built on a physical model configuration (ORCA025) that is extensively used and regularly updated in the ocean modelling community (Barnier et al., 2006; Lombard et al., 2009; Penduff et al., 2007, 2010; Lique et al., 2011). Its operational products feed the open boundaries of the HRZ_V1V2 and give the physical forcings for the Mercator Océan biogeochemical system BIOMER (Elmoussaoui et al., 2011). However, IRG_V1V2 does not benefit from the improvements that were implemented in HRZ_V1V2 (see Sect. 2.2). To ensure consistency between Mercator Océan systems that will be operated at the end of 2012 and to correct for some deficiencies of the latter, trials were carried out with the following additions and changes: (1) Instead of being constant, the depth of light extinction is separated in red-green-blue bands depending on the chlorophyll data distribution from mean monthly SeaWIFS climatology. (2) Satellite-based large-scale corrections have been applied to the downward (shortwave and longwave) radiative and precipitation ECMWF fluxes. For that, ECMWF 2008-2011 mean states have been calculated together with climatology estimations from GPCPV2.1 rainfalls (Huffman et al., 2009) and GEWEX SRB 3.1 (http://eosweb.larc.nasa.gov/PRODOCS/srb/table_srb.html) shortwave and longwave radiative flux satellite data. The ratios between these satellite climatologies and the mean ECMWF fluxes constitute corrective coefficients which have been locally applied to the low-pass filtered ECMWF fluxes. The Shapiro low-pass filter has been tuned to have 80 % amplitude attenuation at the synoptic scales. (3) The estimation of Silva et al. (2006) has been implemented in IRG_DEV to represent the amount and distribution of meltwater which can be attributed to giant and small icebergs calving from Antarctica, in the form of a monthly climatological runoff at the Southern Ocean surface. (4) Despite the previous correction and updates, the freshwater budget is far from balanced. In order to avoid any mean sea-surface-height drift due to the poor water budget closure, the surface freshwater budget is set to zero in IRG_DEV at each time step with a superimposed seasonal cycle (Chen et al., 2005). It helps to reduce errors in SLA assimilation. Because it is difficult to obtain a reliable estimate of the net surface water flux at a regional scale, we constrain the regional average surface flux in HRZ_DEV to be equal to the one in IRG_DEV. However, the concentration/dilution water flux term is not set to zero. (5) As has already been done for the HRZ_V1V2 system, AVHRR + AMSRE SST has been assimilated in IRG_DEV and HRZ_DEV, instead of RTG SST. (6) An error map based on the maximum of sea ice extent was applied to correctly assimilate the Envisat altimeter data for the Arctic. (7) In October 2010, the Envisat altimeter was brought to a lower orbit, which has led to a slight degradation of data quality (Ollivier and Faugere, 2010). This degradation is due to the fact that SLA is computed with respect to a mean sea surface

of lower quality because it falls outside the historical repeat track. This is particularly true at high latitudes where no tracks from other missions are available. For this reason, the Envisat error was increased by 2 cm over the entire domain and by 5 cm above 66° N. (8) New temperature and salinity vertical profiles from the sea mammal (elephant seals) database (Roquet et al., 2011) were assimilated to compensate for the lack of such data at high latitudes. (9) A quality control (QC) of T/S vertical profiles has been implemented to discard suspicious temperature and salinity vertical profiles (see Sect. 3.3). This is done in addition to the quality control procedures performed by the CORIOLIS data centre.

3 Scientific assessment and validation methodology

3.1 Metrics and calibration period

As mentioned above, the scientific assessment performed at Mercator Océan meets the requirements of the MyOcean "scientific calibration phase". During the MyOcean project, scientists from all MFCs and TACs have defined the MyOcean calibration and validation metrics by region and type of product, including observational products. The so-called "Product Quality and Calibration/Validation group" produced a large number of diagnostics and proposed complementary methodologies. All types of metrics used for calibration/validation of the MyOcean global system are shown in Table 3. Many efforts were made to synthesize and homogenise quality information in order to provide quality summaries and accuracy numbers. All these rely on the same basis of metrics that can be divided into four main categories derived from Crosnier and Le Provost (2007).

The consistency between two-system solutions or between a system and observations can be checked by "eyeball" verification. This consists in comparing subjectively two instantaneous or time mean spatial maps of a given parameter. Coherent spatial structures or oceanic processes such as main currents, fronts and eddies are evaluated. This process is referred to as CLASS1 metrics. The consistency over time is checked using CLASS2 metrics, which include comparisons of mooring time series and statistics between time series. Space and/or time integrated values such as volume and heat transports, heat content and eddy kinetic energy are referred to as CLASS3. Their values are generally compared with literature values or values obtained with past time observations such as climatologies or reanalyses. Finally, CLASS4 metrics give a measure of the real-time accuracy of systems, by calculating various statistics of the differences between all available oceanic observations (in situ or satellite) and their model equivalent at the time and location of the observation.

The scientific assessment or calibration procedure thus involves all classes of metrics. It checks improvements between versions of a system, and ensures that a version is robust and its performance stable over time. The assessment must be conducted through a one-year numerical experiment at least, in order to obtain representative results. It is currently very difficult to run real-time systems over many years in the past, for computational reasons, but also due to the recent (and on-going) evolution of the ocean observational network. Different data densities imply different tunings of the data assimilation system. Moreover, homogeneous (reanalysed) atmospheric fluxes are needed to perform experiments that cover several decades. GLORYS2V1 ocean reanalysis spans the 1993–2009 period and is the longest numerical experiment with a system which is similar to the IRG_V1V2 real-time monitoring and forecasting system. The IRG_V1V2 and HRZ_V1V2 numerical experiments only start in October 2006, which is a good trade-off limiting the computational costs needed to catch up with real time and taking advantage of a large enough observation network together with high resolution atmospheric forcing from ECMWF operational forecasts. The results of the scientific assessment are illustrated here with diagnostics from the four classes of metrics, over the year 2010 mainly, assorted with time series over the 2007–2011 period.

3.2 Quality check of real-time analyses and forecasts

Once the scientific assessment has been done, and the system's nominal accuracy values and consistent behaviour have been described, it is possible to apply a regular quality check to the real-time analyses and forecasts. Due to the very large amount of information produced by a global system, real-time quality check is based on a reduced number of metrics, and comparisons with observations are constrained by their availability and timeliness. However, more than a thousand graphics are checked each week (weekly monitoring of the analysis) and each day (consistency check of the daily forecast) by Mercator Océan. The major part of this procedure is currently being automated with indicators based on distribution (percentiles) thresholds computed from the scientific assessment stage.

Numerical weather prediction centres in the world such as Météo-France and ECMWF issue quality reports on a regular basis (quarterly or yearly) which record the strengths and weaknesses of forecasting systems, as well as technical changes to the systems and the spatial and temporal coverage of the input data (see for instance the yearly ECMWF quality report available at http://www.ecmwf.int/publications/newsletters or the quarterly "*Contrôle des produits numériques utilisés pour la prévision météorologique*" published by Météo-France and available on request). Similarly, Mercator Océan has been publishing the Quarterly Ocean Validation Bulletin "*QuO Va Dis?*" since July 2010. It is available at http://www.mercator-ocean.fr/eng/science/qualification. Among the information given, one can find observation minus analysis (called "residual") and observation minus forecast (called "innovation") statistics for T and S vertical profiles, SST and SLA observations that are

Table 3. Types of metrics used for calibration/validation during MyOcean.

MERSEA/GODAE classification	Variable	Region	Type of metric	Reference observational dataset
CLASS1	Monthly T and S (3-D)	Global	Visual inspection of seasonal and interannual signal	Levitus 2005 monthly climatology of T/S
	Sea ice concentration and drift	Antarctic and Arctic	Visual inspection of seasonal and interannual signal	CERSAT sea ice concentration and drift
	SLA and SST	Tropical basins	Visual inspection of Hovmöller diagrams comparisons with satellite observations	AVISO and OSTIA
CLASS2	T, S, U, V, SSH atmospheric forcings	Global at CLASS2 locations	Visual inspection of high frequency comparisons with observed time series	MyOcean: CORIOLIS
CLASS3	Sea ice concentration	Antarctic and Arctic	Time evolution of sea ice extent	NSIDC sea ice extent from SSM/I observations
	U and V (3-D)	Global	Visual inspection of volume transports through sections	Literature
Data assimilation statistics	SLA	Global and regional basins	Error = observation minus model Time evolution of RMS and mean error	MyOcean: on track AVISO SLA observations
	SST	Global and regional basins	Error = observation minus model Time evolution of RMS and mean error	"RTG" SST("AVHRR + AMSRE" SST for HRZ)
CLASS4 and data assimilation statistics	T and S (3-D)	Global and regional basins	Error = model minus observation Time evolution of RMS error on 0–500 m Vertical profile of mean error.	MyOcean: CORIOLIS $T(z)$ and $S(z)$ profiles
CLASS4	SSH	At tide gauges (Global but near coastal regions)	Error = model minus observation Time series correlation and RMS error	GLOSS, BODC, Imedea, WOCE, OPPE, SONEL
	Surface current U	Global and regional basins	Error = model minus observation Mean error and vector correlation	SVP drifting buoys from CORIOLIS
	Surface current V	Global and regional basins	Error = model minus observation Mean error and vector correlation	SVP drifting buoys from CORIOLIS
	Sea ice concentration	Antarctic and Arctic	Error = observation minus model Time evolution of RMS and mean error	CERSAT sea ice concentration

assimilated. Comparisons are also made with independent observations, such as currents at 15 m derived from drifting buoys, sea ice concentration and drift, or tide gauges (the low frequency component of the tide gauges' elevation signals). The systems' SSTs are compared with the high resolution SST OSTIA (Operational Sea Surface Temperature and Sea Ice Analysis, Donlon et al., 2012), which is very close to the assimilated SST from Reynolds et al. (2007). Integrated parameters such as sea ice extent and global mean SST are monitored. Process studies focusing on one process or region, or short research and development validation studies complement the bulletins.

In this article we chose to illustrate the scientific assessment results mostly with the same metrics.

3.3 Quality control on in situ observations and feedback to input data providers

To minimise the risk of erroneous observations being assimilated in the model, the system carries out a quality control (QC), known as "background quality control" in meteorology, on the assimilated T and S vertical profiles. This is done in addition to the quality control procedures performed by data assembly centres.

The basic hypothesis of the data assimilation system is that innovations are normally distributed at each point of the ocean (Gaussian distribution of background errors). Observations for which the innovation is in the tail of the distribution are thus considered to be questionable. Taking advantage of the very large number of temperature and salinity innovations collected in the GLORYS2V1 reanalysis (1993–2009), it was possible to reliably estimate their seasonally and spatially

variable statistics (mean, standard deviation). These parameters were then used to define a space and season dependent threshold value (Cabanes et al., 2013).

The implementation of the QC can be summarised as follows. An observation is considered to be suspect if the two following conditions are satisfied:

$$\begin{cases} | \text{ innovation } | > \text{threshold} \\ | \text{ observation} - \text{climatology} | > 0.5 \ | \text{ innovation } | \end{cases} \quad (2)$$

The first condition is a test on the innovation. It determines whether the innovation is abnormally large which would most likely be due to an erroneous observation. The second condition avoids rejecting "good" observations (i.e. an observation close to the climatology), even if the innovation is high due to the model background being biased. Figure 5 shows an example of a wrong temperature profile detected by the QC in the GLORYS2V1 simulation. Below 400 m in depth, innovations are no longer valid. The two conditions described previously are satisfied and the profile is rejected (Fig. 5a). When this profile is assimilated, an abnormal salinity value appears at the temporal and geographical positions of this profile (Fig. 5b). This is due to the fact that the assimilation algorithm used is multivariate, meaning that an observation of temperature leads to corrections of all the model variables and especially, in this case, the surface salinity.

For each year of the 1993–2009 GLORYS2V1 simulation, all questionable profiles from the CORA3.1 dataset were identified, and percentages of rejection and spatial distribution of questionable observations were produced. Finally, the list of questionable observations was sent to the CORIOLIS data centre, which in turn flagged around 50 % of these observations as bad in the new CORA3.2 dataset (Cabanes et al., 2013), which is used by IRG_DEV and HRZ_DEV systems. MFCs will be routinely transferring an increasing amount of feedback to TACs at the MyOcean level, but also at the international level in the context of GODAE Oceanview.

This QC of oceanic observations based on system innovations has been implemented in IRG_DEV and HRZ_DEV for use both in real time and in delayed time. The parameters (average and standard deviation of the innovations, and therefore threshold value) were calculated from GLORYS2V1, which among other things assimilated the CORA3.1 database. In principle, these parameters are model dependent. However, all our systems suffer from the same kind of defects, mostly related to forcing, or to defects in model parameterizations that are almost the same for all systems. It is therefore considered that the QC parameters built from GLORYS2V1, and in particular the seasonal threshold value, may be applied to other Mercator Océan systems, assuming that the forecast errors or system biases are of the same magnitude or even lower than those of GLORYS2V1.
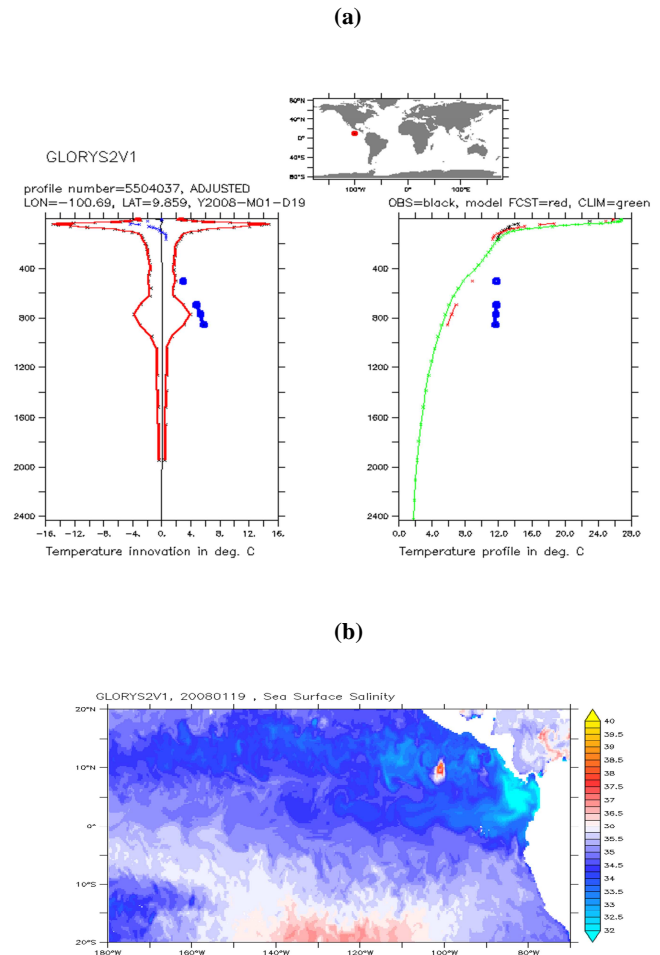
Fig. 5. Example of a suspicious temperature vertical profile at 100.69° W, 9.86° N, highlighted by the QC on the CORA3.1 dataset. **(a)** Left panel shows temperature innovation profile in blue and temperature innovation threshold in red. Right panel shows the absolute vertical temperature profile (observation in black, climatology in green and model in red). Large blue dots correspond to "bad" innovations or "bad" observations. **(b)** When this profile is assimilated, an abnormal value of salinity appears at the position of this profile.

4 Scientific assessment results

This section describes the MyOcean V2 global system's quality assessment, demonstrating the importance of a thorough scientific validation procedure including different points of view. For that, two simulations performed with the IRG_DEV and HRZ_DEV systems were examined.

4.1 Statistics of observation/analysis comparisons

The analyses ("BEST" trajectory, see Sect. 2.1.2) are first assessed by comparing them directly with the assimilated observations, and with observations that have not been

assimilated by the system such as velocity measurements of
drifting buoys or sea ice observations.

4.1.1　Temperature and salinity vertical profiles

The model equivalent at the time and spatial location of
the observation is derived from daily averaged analyses
(CLASS4 metrics). Statistics (mean and RMS differences)
are computed in 2° by 2° bins or in wider (up to basin
scale) regions, and in vertical layers. When the observational
dataset has been assimilated, the resulting scores can be con-
sidered to be residuals of data assimilation. These scores are
given here for in situ temperature and salinity vertical pro-
files which are assimilated in the systems. Note that at the
CLASS4 validation stage, the original observational dataset
is used. However, part of this data is discarded by means of
an external quality check based on regional departures from
climatology. Large differences may thus appear locally in the
CLASS4 scores, with observations discarded by the QC de-
scribed in Sect. 3.3 before the data assimilation takes place.
This process sometimes reveals outliers or erroneous pro-
files. As can be seen in Fig. 6a, IRG_V1V2 yearly mean de-
partures from observed temperature are, on average, not more
than 0.3 °C in many regions of the ocean for the first 500 m
of the ocean. The largest RMS differences take place in
high mesoscale variability regions such as the Gulf Stream,
the Kuroshio, the Agulhas current or the Zapiola eddy. The
thermoclines of the tropical basins also display significant
signatures in the temperature RMS error, especially in the
tropical Pacific where a La Niña event took place through-
out most of the year 2010. The salinity average departures
from observations stay below 0.03 psu in most regions of
the ocean (Fig. 6b). The principal mesoscale activity regions
also display higher salinity RMS values. The IRG_DEV sys-
tem displays similar temperature and salinity RMS differ-
ence patterns (not shown). However, some improvements in
the Mediterranean region, the Bay of Biscay, the Gulf Stream
and the western tropical Pacific can be attributed to the use
of the adjusted MDT and a more adapted specification of ob-
servation errors for SST and SLA. These improvements are
highlighted in time series of basin-scale statistics. Figures 7
and 8 show temperature and salinity statistics performed in
the 5–100 m layer and in the basin-scale zones of the North
Atlantic, Mediterranean Sea and Indian Ocean. In order to
compare all "V1V2" and "DEV" systems, the common pe-
riod of 2007–2011 has been chosen. The HRZ_V1V2 system
(dashed line) is only available from October 2009 (see Fig. 1
and Sect. 2.2). Moreover, the results of this system are not
representative before 1 January 2010 because a few months
of spin up are necessary. Consequently, the time series was
complemented by the HRZ_V1 system (solid line).

We first note that the departures from the observations
are smaller for all the systems than for the climatology. The
RMS residuals in temperature and salinity are significantly
reduced in the North Atlantic and Mediterranean Sea with the
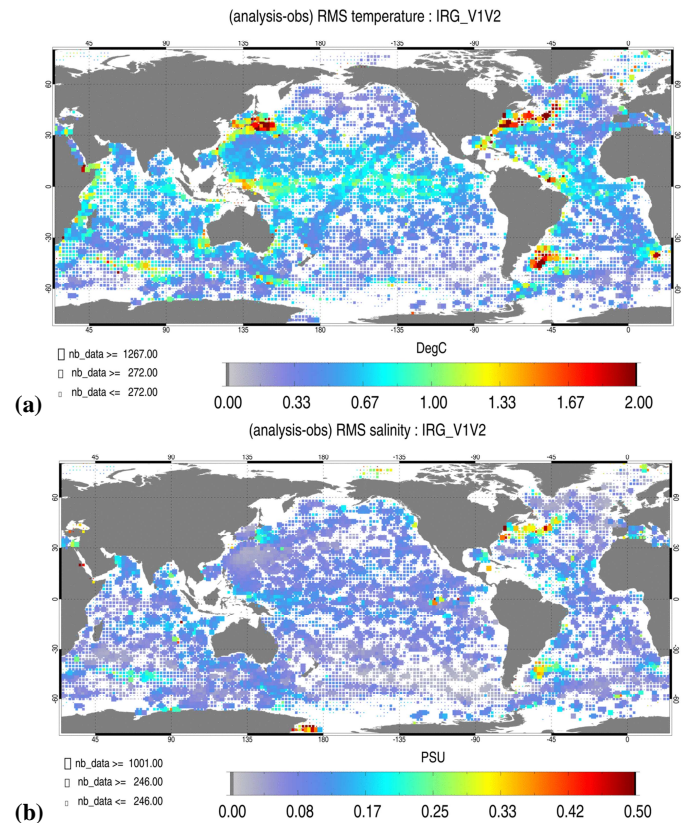
Fig. 6. 2010 temperature **(a)** and salinity **(b)** RMS of the differences
between all available observations from the CORIOLIS database
and daily mean analyses for the IRG_V1V2 system. Averages are
performed in the 0–500 m layer. The size of the pixel is proportional
to the number of observations used to compute the RMS in 2° × 2°
boxes. Observations that differ by more than 8 °C or 1 psu of a
climatological reference are not taken into account when calculating
the diagnostic.

IRG_DEV and HRZ_DEV systems. However, the IRG_DEV
system tends to drift towards a cold and salty subsurface bias
in the Indian Ocean. These biases are of the order of mag-
nitude of 0.2 °C and 0.02 psu on basin average between 5
and 100 m. All regions experience these slight biases (not
shown) except for the North Atlantic, the Mediterranean Sea
and the Arctic Ocean. As a consequence, the HRZ_DEV
system, by its geographical area, does not display biases
except in the tropical Atlantic. Finally, no significant im-
provement has been diagnosed in the Mediterranean for the
HRZ_DEV system compared to the HRZ_V1V2 system be-
cause the latter already benefits from several updates present
in the HRZ_DEV system (see Sects. 2.2 and 2.3).

4.1.2　OSTIA SST analyses

Comparisons between various SST analysis products
(Reynolds and Chelton, 2010) have shown that it was of-
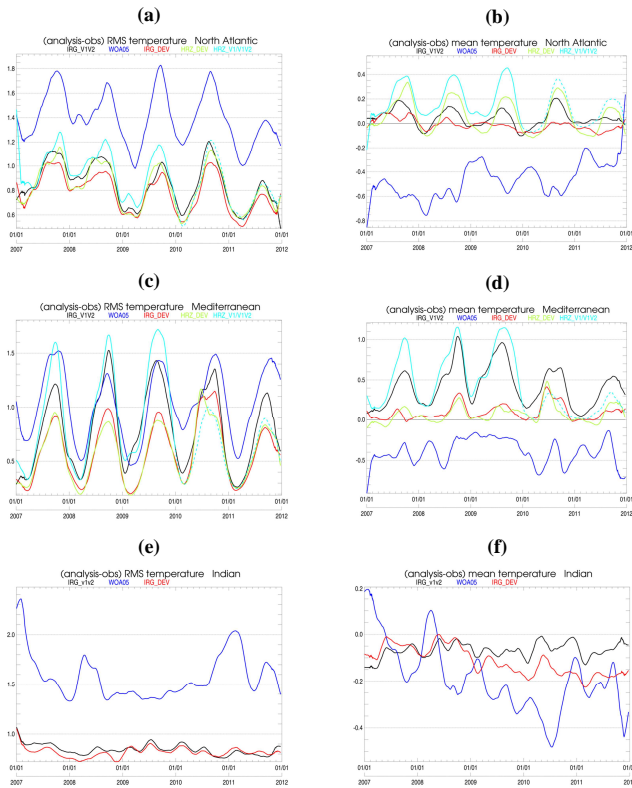ten difficult to choose, from the many that are available,

Fig. 7. Temperature (°C) RMS (**a**, **c** and **e**) and mean (**b**, **d** and **f**) differences (analysis minus observation). For these diagnostics, all available T/S observations from the CORIOLIS database and Mercator Océan daily average analyses, collocated (temporally and spatially) with observations, are used. Statistics are displayed for IRG_V1V2 (black line), IRG_DEV (red line), HRZ_V1 (cyan solid line), HRZ_V1V2 (cyan dashed line), HRZ_DEV (green line) and World Ocean Atlas climatology WOA05 (blue line). Averages are performed in the 5–100 m layer in the basin-scale zones of the North Atlantic (**a** and **b**), Mediterranean Sea (**c** and **d**) and Indian Ocean (**e** and **f**).

Fig. 8. Same as Fig. 7 but for salinity (psu).

the one that is best suited to a particular purpose. Most SST analyses use the same observations such as AVHRR data or in situ SST data from ships and buoys collected via the real-time Global Telecommunication System. However, the OSTIA product shares more observations with the AVHRR + AMSRE product than with RTG, even though there is still some independent information in OSTIA. A recent study ranked them on the basis of a comparison with the Argo array of profiling floats (Guinehut, 2010). This study shows that OSTIA and AVHRR + AMSRE compare well and that RTG has a lower quality level mainly at high latitudes, partly due to the lack of AMSRE.

As SST OSTIA is the MyOcean global SST product, we chose to use it as a reference for validation. Yearly mean model SST differences with OSTIA analysis show that in the Antarctic, Indian and Atlantic basins the model SST is very close to OSTIA, with differences staying below the ob-

servation error of 0.6 °C (Donlon et al., 2012) on average (Fig. 9). However, strong regional biases have been diagnosed in the IRG_V1V2 system, particularly for the high northern latitudes and/or some coastal areas. Part of these coastal biases comes from the use of RTG for data assimilation in IRG_V1V2, as this SST product is known to be too cold in the coastal regions at high latitudes. These biases are no longer found in the new version of the system when AVHRR + AMSRE is assimilated. This SST product has the same quality level as OSTIA and both display better performance than RTG, especially at high latitudes (Reynolds and Chelton, 2010; Guinehut, 2010). Another reason that might explain some of these biases lies in the way the data are assimilated into the system. In the IRG_DEV system, the choice has been made not to trust SST and SLA observations within 50 km of the coast and to prescribe higher observation errors in these coastal areas. In the IRG_V1V2 system the SST and SLA observation errors do not increase near the coast, which partly generate the large coastal biases diagnosed in the Northern Hemisphere.

The cold bias that persists in the eastern part of the Pacific in IRG_DEV is not explained by differences between OSTIA and the assimilated AVHRR + AMSRE. This model bias peaks over 1 °C and reaches its highest amplitude during summer. If we look at the 2010 average SST increment and the part rejected by the system at the end of the assimilation
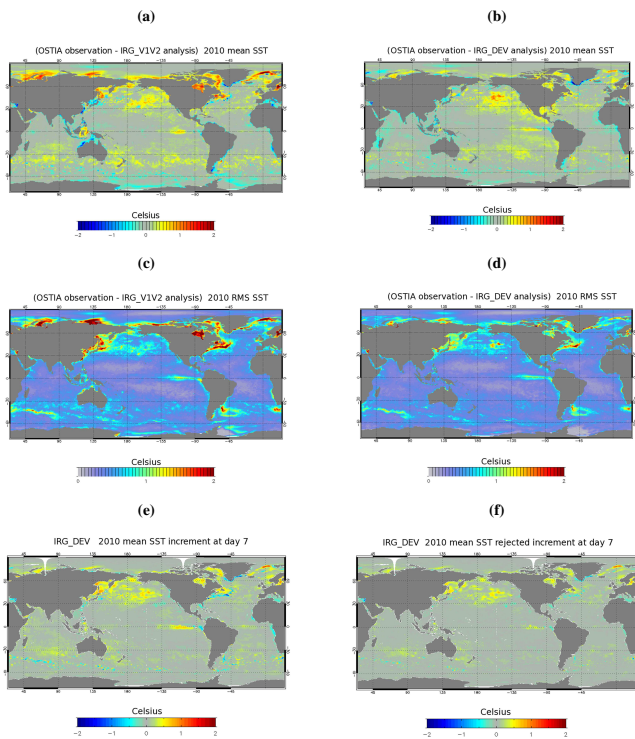
Fig. 9. 2010 SST (°C) mean (**a** and **b**) and RMS (**c** and **d**) differences (observation minus analysis) between OSTIA product and IRG daily average analyses collocated with observations. IRG_V1V2 system (**a** and **c**) assimilates RTG SST and is compared to IRG_DEV system (**b** and **d**) which assimilates AVHRR + AMSRE SST. (**e**) 2010 mean SST increment injected into IRG_DEV system every 7 days. (**f**) 2010 mean SST increment rejected by the system every 7 days. Note that the rejected increment corresponds to the increment weighted by the distribution function (see Fig. 4) minus the difference between the analysis and the background.

window (Fig. 9e, f), we find a significant rejection at the place of the cold bias in the eastern half of the Pacific. This can be explained by the fact that the surface heat fluxes are not included in the estimated state and that the IAU correction of surface temperature does not work in the same way as the bulk forcing function.

4.1.3 Ice observations

The sea ice concentration and drift in the Arctic of IRG_V1V2 analyses were compared with satellite observations in winter (Fig. 10) and in summer (Fig. 11) of the year 2010. The seasonal cycle, the interannual variability and the recent trend of the sea ice extent have already been evaluated and compare well with the satellite estimates (Garric et al., 2008; Lique et al., 2011). The sea ice extent is realistic in all IRG systems. The main spatial patterns of the sea ice drift such as the Beaufort Gyre and the Transpolar Drift Stream are well reproduced by the system. However,
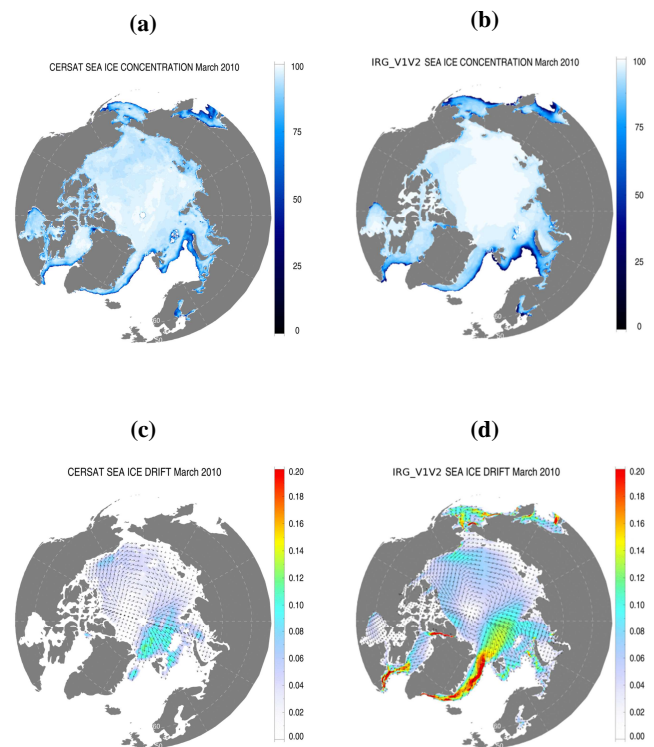
Fig. 10. Sea ice concentration (%) in the Arctic in March 2010: (**a**) from CERSAT satellite measurements, (**b**) from IRG_V1V2. Sea ice drift (m s^{-1}) in the Arctic in March 2010: (**c**) from CERSAT satellite measurements; (**d**) from IRG_V1V2.

the modelled sea ice speed is overestimated whatever the season (Figs. 10 and 11) and gives the highest velocities over the unobserved marginal sea ice open ocean zones. In summer, the sea ice concentration of IRG_DEV is more realistic than IRG_V1V2, especially in the Laptev Sea. The HRZ sea ice concentrations and drifts in the North Atlantic are very close to IRG equivalent fields both in summer and winter (not shown), confirming good performance of the damping of ice condition from IRG to HRZ north open boundaries. The sea ice concentration pattern in the IRG_DEV system is more realistic than in IRG_V1V2, over the Laptev Sea and in the Barents Sea. This is closely linked to the switch of assimilated SST observation from RTG to AVHRR + AMSRE. Indeed, as shown in Fig. 12, a comparison of sea surface temperatures in the two products in the Arctic region shows that AVHRR + AMSRE is substantially warmer, with the misfit with RTG reaching more than 5 °C in summer at a few locations off the Siberian coast. Moreover, during wintertime, these warmer SSTs prevent too great a spread of the modelled sea ice towards the open ocean and lead to a very realistic Arctic sea ice envelope. The time evolution of the sea ice extent in the Arctic and in the Antarctic is shown in Fig. 13. Both the IRG_V1V2 and IRG_DEV systems display a seasonal cycle locked in phase with observations. The
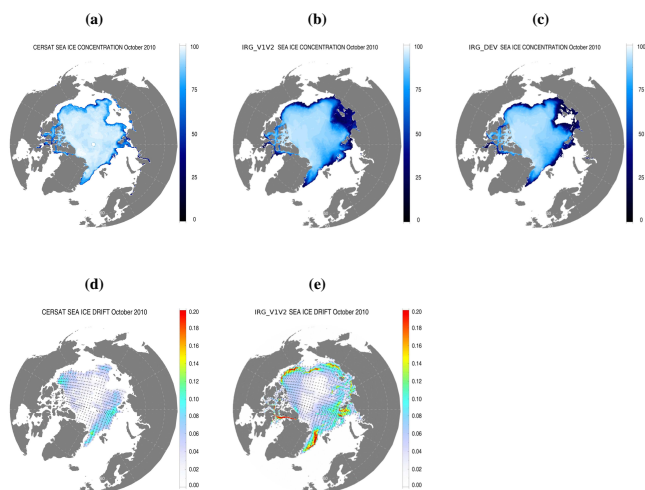
Fig. 11. Sea ice concentration (%) in the Arctic in October 2010: **(a)** from CERSAT satellite measurements, **(b)** from IRG_V1V2, and **(c)** from IRG_DEV. Sea ice drift (m s^{-1}) in the Arctic in October 2010: **(d)** from CERSAT satellite measurements and **(e)** from IRG_V1V2.
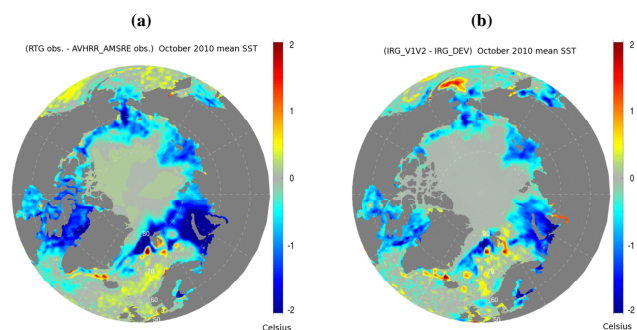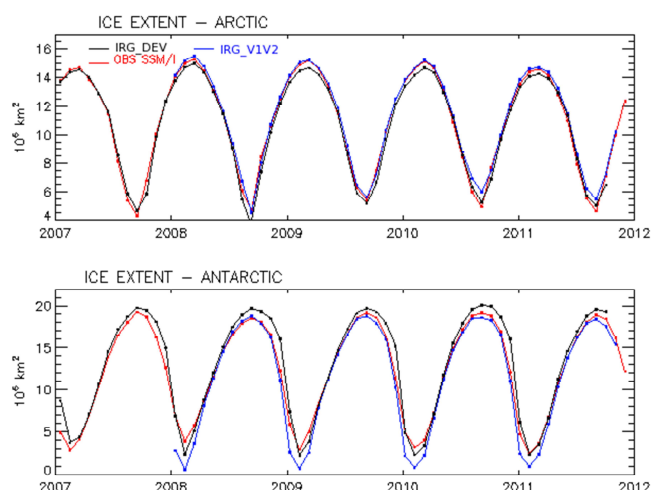

Fig. 13. Sea ice extent (10^6 km^2) in the Arctic (upper panel) and the Antarctic (lower panel) for the 2007–2011 period, from SSM/I satellite measurements (red line), IRG_V1V2 (blue line) and IRG_DEV (black line).

Fig. 12. October 2010 mean difference (°C) between **(a)** assimilated SST products in IRG_V1V2 and IRG_DEV and **(b)** IRG_V1V2 and IRG_DEV SST analyses.

assimilation of a colder AVHRR + AMSRE SST compared with the RTG SST product allows the IRG_DEV system to keep a wider sea ice extent in the Weddell Sea and a presence of sea ice in East Antarctica (not shown). This results in a realistic summer Antarctica sea ice extent. However, these colder SSTs lead to a slight over-estimation of the ice extent during winter.

4.1.4 Drifter velocities

In order to assess the quality of surface currents, IRG_V1V2 ocean velocity analyses at 15 m are compared with measurements from drifting buoys provided by the CORIOLIS data centre. Since Grodsky et al. (2011), it is known that at least 30 % of drifters may have lost their drogues. Hence, a downwind slippage correction for drifting buoy velocities of about 0.07 % of the wind speed at 10 m was computed. Then, an algorithm was applied to detect the presence of un-

drogued drifters in order to add a windage correction (up to 3 %) upon U and V components. This quality control detects about 40 % of the original dataset and "cleans" the high latitude regions (Antarctic Circumpolar Current, North Atlantic Drift). Once this correction (that we will refer to as the Mercator Océan correction) is applied to drifter observations, the zonal and meridional velocities of the model at 15 m in depth are more consistent with the observations (Fig. 14a).

Mean relative speed bias (MRSB) using all drifter observations during the year 2010 can be computed as follows:

$$\text{MRSB}(i,j) = \frac{\text{speed}_{\text{drifter}}(i,j) - \text{speed}_{\text{model}}(i,j)}{\text{speed}_{\text{drifter}}(i,j)}, \qquad (3)$$

where speed$_{\text{drifter}}$ is the drifter horizontal speed, speed$_{\text{model}}$ the model horizontal speed and (i,j) the geographical position of drifter observation.

Figure 14b shows that IRG_V1V2 underestimates the surface speed by 20 to 50 % in the northern and southern mid and high latitude eastward currents such as the Kuroshio, the North Pacific Current, the Gulf Stream, the North Atlantic Drift and the Antarctic Circumpolar Current. On the contrary, the tropical westward currents such as the North and South Equatorial Currents in the Pacific and in the Atlantic are generally overestimated by 20 to 50 %. The direction errors are much smaller than the velocity amplitude errors, and large direction errors are very rare (Fig. 14c). These direction errors generally correspond to poorly positioned strong current structures in high mesoscale variability regions (Gulf Stream, Kuroshio, North Brazil Current, Zapiola Eddy, Agulhas Current, Florida Current, East African Coastal Current, Equatorial Pacific Countercurrent). The IRG_DEV experiment displays similar results (not shown) with a slight improvement

in the western tropical Pacific, which can be attributed to the adjusted MDT used.

4.2 Statistics of observation/forecast comparisons

The forecast skill is assessed in this section. The statistics come from the data assimilation innovations computed from the forecast used as the background model trajectory ("FORECAST" trajectory, see Sect. 2.1.2). This should not be confused with an operational forecast that corresponds to a model trajectory in the future without data assimilation. Even though the quality of real-time forecast is not evaluated here (different atmospheric forcing, different analysis quality), these statistics give an estimate of the skill of the optimal model forecast. These scores are averaged over all seven days of the data assimilation window, which means the results are indicative of the average performance over the seven days, with a lead time equal to 3.5 days.

4.2.1 Sea level anomaly

For SLA, Fig. 15 shows for the IRG_V1V2 and IRG_DEV systems, temporal mean sea level innovation (a and b), temporal mean sea level residual (c and d) and temporal mean sea level RMS residual (e and f) for the year 2010. These diagnostics illustrate the forecast and analysis scores in a geographical context, highlighting the skill of the system. Most discrepancies in the two versions of the IRG system (Fig. 15 top panel), notably near the Hudson Bay, the Indonesian Throughflow, the Caribbean, and the continental shelves (Iceland, United Kingdom), seem to be due to inconsistencies between the MDT and the observed SLA. It may be noted that IRG_DEV innovations within these geographic areas are not reduced by the analysis correction, whereas they are elsewhere (Fig. 15d). This can be attributed to a wider observation error of SLA observation within 50 km of the coast and on the shelves in IRG_DEV (Fig. 15h), meaning that the data in these areas is not trustworthy (same problem with SST). The sea level RMS residuals (Fig. 15e, f) show the skill of the systems in terms of generated error. We find the largest values in areas where the system is biased (continental shelves, Celtic, North and Baltic Seas, Hudson Bay, Indonesian Throughflow, etc.) and in areas of high mesoscale variability. Higher SLA observation errors have been specified for IRG_DEV than for IRG_V1V2 (Fig. 15g, h) in order to give less weight to SLA observations in the multivariate analysis. The specification of a too low SLA observation error resulted in unrealistic increments of ill-controlled variables, such as salinity and/or horizontal velocities (not shown). SLA observation errors in IRG_V1V2 seem to be underestimated for large currents or more specific areas such as Indonesia Throughflow or near Hawaii. SLA errors in IRG_DEV appear to be more consistent with SLA RMS residuals, particularly in these areas.

(a)

(b)

(c)
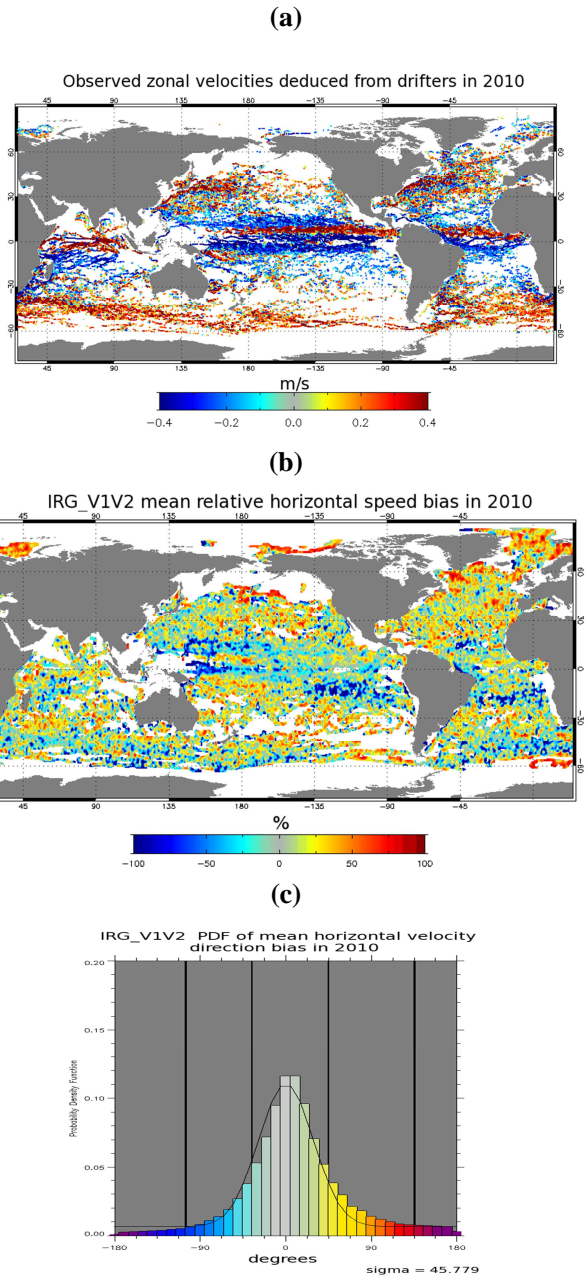
Fig. 14. (a) Zonal velocity observations $(\mathrm{m\,s^{-1}})$ with "Mercator Océan correction" (see text) deduced from all in situ Atlantic Oceanographic and Meteorological Laboratory drifters in 2010. **(b)** Mean relative horizontal speed bias (%) with respect to drifters in IRG_V1V2. **(c)** Probability density function of the mean horizontal velocity direction bias (degrees) with respect to drifters in IRG_V1V2. The Gaussian fit is superimposed, as well as the 1-sigma (thin lines) and 2-sigma interval (thick lines).

Figure 16 shows time series of innovation statistics and illustrates the skill related to the different changes of versions of the HRZ system. For this figure, only Jason1 and Jason2 altimeters, available throughout the period 2007–2011,
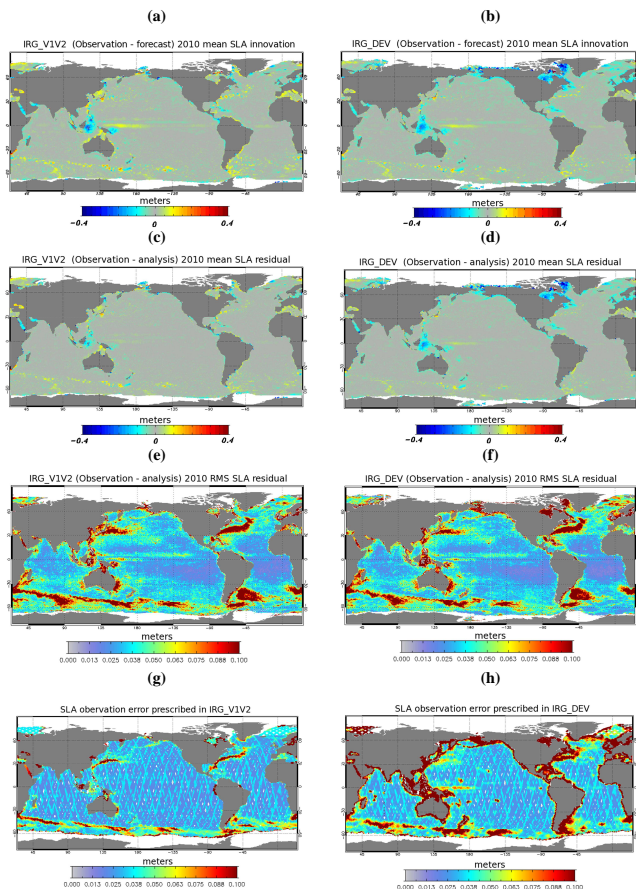
Fig. 15. SLA innovations (**a** and **b**), SLA residuals (**c** and **d**) and SLA RMS residuals (**e** and **f**) for the IRG_V1V2 (**a**, **c** and **e**) and IRG_DEV (**b**, **d** and **f**). Specified SLA observation error for IRG_V1V2 (**g**) and IRG_DEV (**h**). These diagnostics (in meters) illustrate the forecast and analysis scores. These scores are averaged over all seven days of the data assimilation window, with a lead time equal to 3.5 days.

Fig. 16. Time evolution of sea level anomaly (SLA, m) data assimilation statistics averaged over the whole HRZ domain: RMS of innovations (**a**); RMS of innovations divided by quadratic mean of SLA data over the same region (**b**); and mean innovation (**c**). The colours stand for different MyOcean versions of HRZ: HRZ_V0 (black), HRZ_V1 (blue), HRZ_V1V2 (orange) and HRZ_DEV (green). These scores are averaged over all seven days of the data assimilation window, with a lead time equal to 3.5 days.

were considered. The forecasting scores of SLA for the different versions of HRZ are globally similar. The biases are weak as the mean innovation varies around zero with peaks of 1 cm or even 2 cm at times (Fig. 16c). The RMS of this innovation (Fig. 16a) is of the order of 8 cm, on average, over the whole domain but can be smaller over several subdomains. However, this RMS remains higher than the SLA error specified for the systems, which is equal to the sum (in variance) of the SLA instrumental error (about 3 cm on average) and the MDT error (about 5 cm on average, with the largest values being located on shelves, along the coast and mesoscale activity or sharp front areas). Moreover, the model is able to explain the observed signal as shown by the ratio of RMS innovation to RMS data, which decreases with time and converges towards a value less than 1 (Fig. 16b). In particular, the improvement between successive versions can be

seen. The HRZ_V1V2 and HRZ_DEV systems have a weaker RMS innovation than the HRZ_V0 and HRZ_V1 systems.

This paper also includes the Murphy Skill Score (MSS) (Murphy, 1988) for the 2007–2011 period, using all available SLA observations inside the assimilation window. This score is a way of measuring the skill of the model forecast to beat persistence and is computed as follows:

$$MSS = 1 - \frac{\sum (SLA_{obs} - SLA_{forecast})^2}{\sum (SLA_{obs} - SLA_{persistence})^2}. \tag{4}$$

The score is positive (negative) when the accuracy of the forecast is higher (lower) than the accuracy of the persistence. Moreover, MSS = 1 when the forecast is perfect (equal to the observation) and MSS = 0 when the forecast is equivalent to the persistence. Figure 17 shows, for the main GODAE regions, the accuracy of the SLA forecast, which can be translated into a measure of percentage improvement in accuracy simply by multiplying the right-hand side of Eq. (4) by 100. The IRG_DEV and HRZ_DEV systems improve the skill scores for all GODAE regions, compared with previous systems, even though some MSS are still negative, particularly in the North Atlantic and Antarctica.
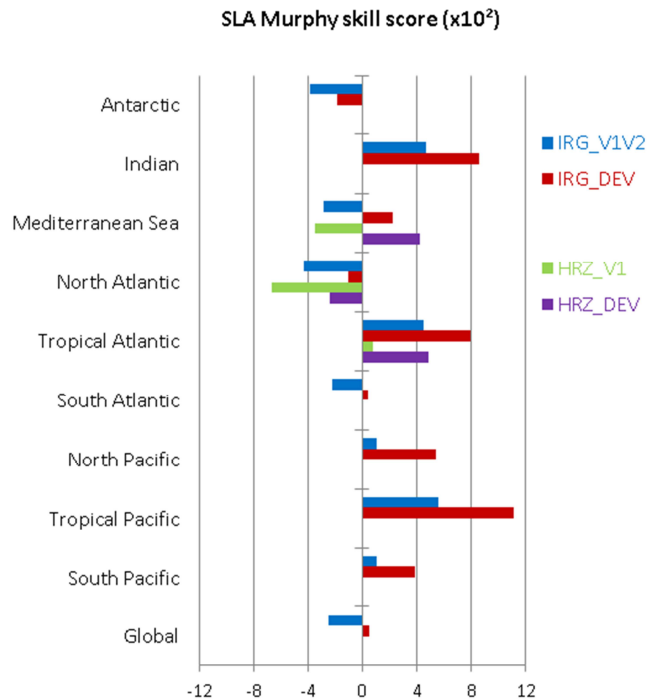
4.2.2 Temperature and salinity vertical profiles

As for the SLA, diagnostics on the assimilated T/S vertical profiles have been performed with the IRG_V1V2 and IRG_DEV systems. Time series of mean innovations of temperature (Fig. 18) and salinity (Fig. 19) in the North Pacific GODAE region reveal a slight drift from the beginning of 2009 with the IRG_DEV system. The model becomes too cold (0.2 °C in the first 300 m) and too salty (0.05 psu in the first 100 m). Some biases have also been observed for free simulation (without data assimilation) and are mainly due to new model parameterizations and atmospheric forcing corrections (see Sect. 2.3). The data assimilation and bias correction stages do not correct this drift but do reduce the errors efficiently. Indeed, the RMS of temperature and salinity innovations is reduced for the IRG_DEV system for all vertical layers.

Let us now examine the Pacific region north of 45° N. This region of low salinity (less than 33 psu) gains about 2 mm day^{-1} of freshwater, not considering the Amur River flowing into the Okhotsk Sea and the American rivers. We saw that the IRG systems suffer from a cold bias in this area (Fig. 9a, b). Figure 20 shows the box-averaged innovations of temperature and salinity as a function of time and depth. The top left panel reveals that the cold bias of IRG_V1V2 only concerns the warm season. There is a lack of stratification
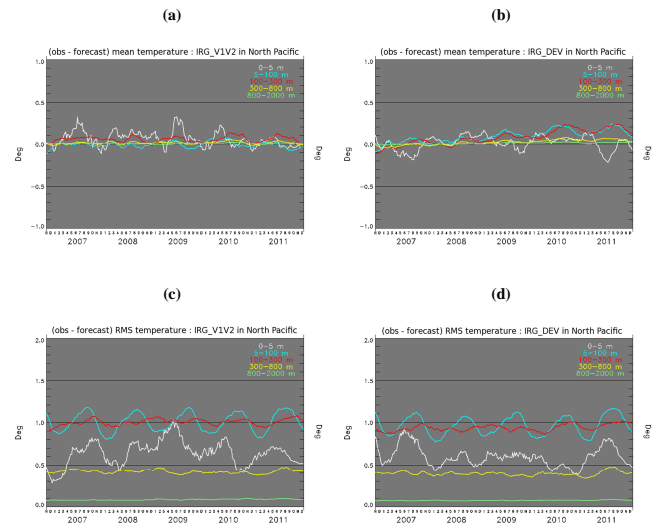
Fig. 18. Temperature data assimilation statistics in the North Pacific GODAE region, and for IRG_V1V2 (**a** and **c**) and IRG_DEV (**b** and **d**): mean (**a** and **b**) and RMS (**c** and **d**) of temperature (°C) innovations (observation minus forecast) computed in layers (0–5, 5–100, 100–300, 300–800, 800–2000) and as a function of time during the 2007–2011 period. For clarity, the time series were smoothed with a 60-day running mean. These scores are averaged over all seven days of the data assimilation window, with a lead time equal to 3.5 days.
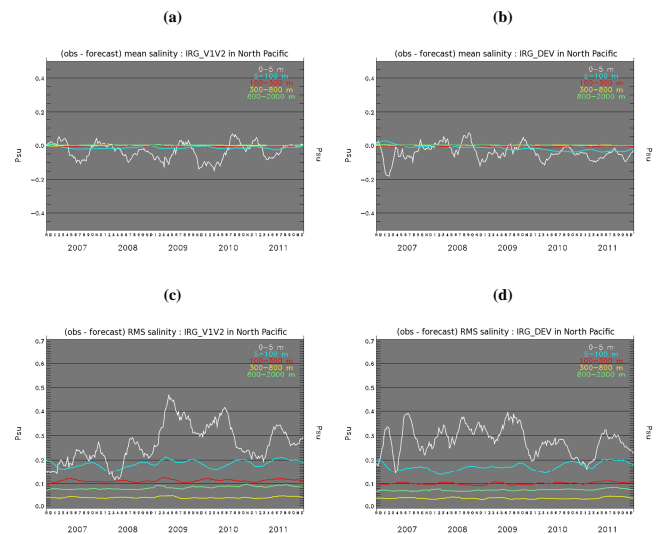

Fig. 19. Salinity data assimilation statistics in the North Pacific GODAE region, and for IRG_V1V2 (**a** and **c**) and IRG_DEV (**d** and **e**): mean (**a** and **b**) and RMS (**c** and **d**) of salinity (psu) innovations (observation minus forecast) computed in layers (0–5, 5–100, 100–300, 300–800, 800–2000) and as a function of time during the 2007–2011 period. For clarity, the time series were smoothed with a 60-day running mean. These scores are averaged over all seven days of the data assimilation window, with a lead time equal to 3.5 days.
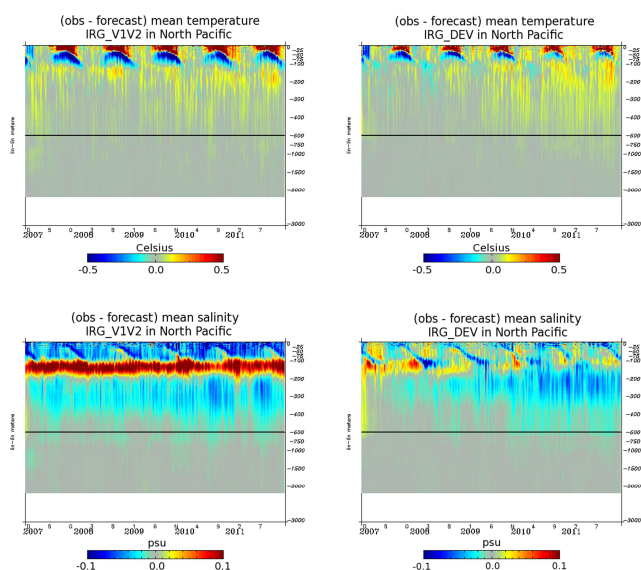
Fig. 20. Assimilation diagnostics with respect to the vertical temperature and salinity profiles over the 2007–2011 period. Mean misfit between observations and model forecast for temperature (top panels) and salinity (low panels), in IRG_V1V2 (left panels) and IRG_DEV (right panels) systems. These scores are averaged over all seven days of the data assimilation window, with a lead time equal to 3.5 days.

above 100 m which disappears in the cold season. The stratification is improved in IRG_DEV (top right), even though the surface cold bias remains. IRG_V1V2 has a salty bias in the mixed layer, and a nearly constant fresh bias just below 100 m (bottom left). Both weaknesses are reduced in IRG_DEV (bottom right). The salinity excess in the warm season is mixed downwards in winter (in blue in Fig. 20). This results in a salty (and cold) bias between 200 m and 300 m, peaking in winter 2010–2011. We checked that the correction of the precipitations (see Sect. 2.3) actually leads to a precipitation deficit in summer. This concerns a broad region with a maximum impact near 150° W, 35° N, and along the coasts from Oregon to British Columbia. It creates a salty bias at the surface and reduces the buoyancy. Another factor is the vertical turbulence closure which uses a vertical profile of energy input from the wind. This profile is fixed in time. The seasonal cycle of surface wave mixing is strongly marked in this region and is poorly represented with the fixed profile. This contributes to excess mixing in summer.

4.3 Physical processes

The accuracy of analysis and forecast does not fully demonstrate that the physics of the model is realistic. We have noticed that the skill of the model's forecast with respect to the persistence is not significant in many regions. In this section, we measure the improvement of the model's analyses from

one version to the other by studying particular physical processes.

4.3.1 Water masses

The OVIDE repetitive section in the North Atlantic has been sampled every other year since 1997 (Lherminier et al., 2007). The North Atlantic is the place of formation of deep water flowing southward, that together with northward surface heat transport via the North Atlantic Drift forms the Atlantic Meridional Overturning Cell. Moreover, this section samples the Mediterranean outflow in the Atlantic near 1000 m, which is an important feature of the circulation and water mass characteristics in the northeast Atlantic near European coasts. A good representation of the northeast Atlantic water masses is crucial for the nesting of MyOcean regional systems such as the Iberian-Biscay-Ireland system (Maraldi et al., 2013) and the North West Shelves system (O'Dea et al., 2010). It is not surprising that both the IRG and HRZ systems compare well with temperature (not shown) and salinity (Fig. 21) measurements along the OVIDE section in 2010, as those observations are assimilated. However, the HRZ_DEV analyses of surface salinity in the vicinity of 11.5° W are closer to OVIDE measurements than any other analysis. The Mediterranean outflow is better represented, especially for salinity. The more accurate salinity representation of HRZ_DEV with respect to HRZ_V1V2 can be partly explained by an earlier initialisation of the system (in October 2006 for HRZ_DEV as opposed to October 2009 for HRZ_V1V2). To that should be added the high quality of the AVHRR + AMSRE SST product and consequently corrections of salinity resulting from the multivariate analysis. Thus, this hydrological section (CLASS2 metric) shows the interest of using the HRZ system as a refinement of the global solution in the North Atlantic and Mediterranean.

4.3.2 Tropical waves

The realism of tropical oceans is crucial for seasonal forecasting applications. The most significant ocean/atmosphere coupling signals are found in the tropical band, for instance the El Niño–Southern Oscillation (ENSO), whose atmospheric teleconnections are global. The Indian Ocean also plays a role in the modulation of ENSO and of the Asian monsoon, and the tropical Atlantic is linked with African monsoons (Redelsperger et al., 2006). While the realism of the heat content is confirmed by examining differences between observations and model results in Sect. 4.1, it is necessary to check that important physical processes taking place in the tropics, such as Kelvin, Rossby and tropical instability waves, are well reproduced. Westward Rossby and eastward Kelvin wave propagations (Delcroix et al., 1991) appear in longitude–time Hovmöller diagrams of SLA at the Equator, as can be seen in Fig. 22. The waves' amplitudes, as well as their propagation speeds, are realistic in the IRG_V1V2
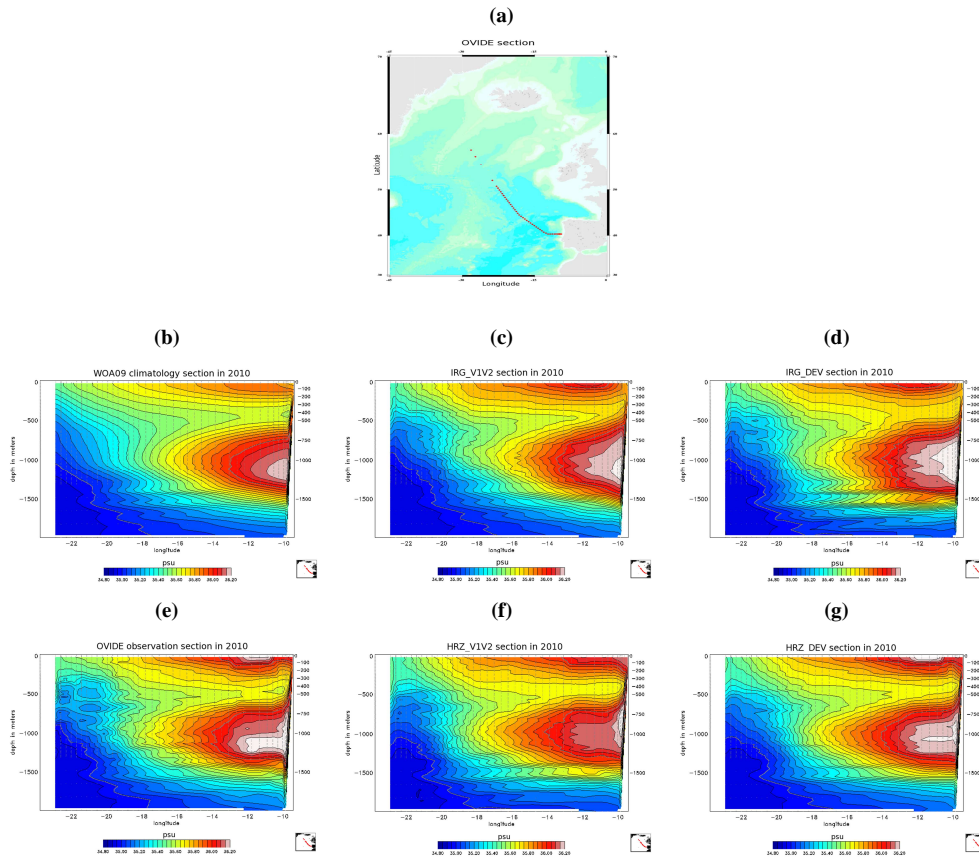
Fig. 21. Salinity (psu) along the OVIDE section **(a)**, in 2010 from WOA09 climatology **(b)**, IRG_V1V2 **(c)**, IRG_DEV **(d)**, in situ observations from CORIOLIS database **(e)**, HRZ_V1V2 **(f)**, and HRZ_DEV **(g)**. The contour interval is 0.05 psu.
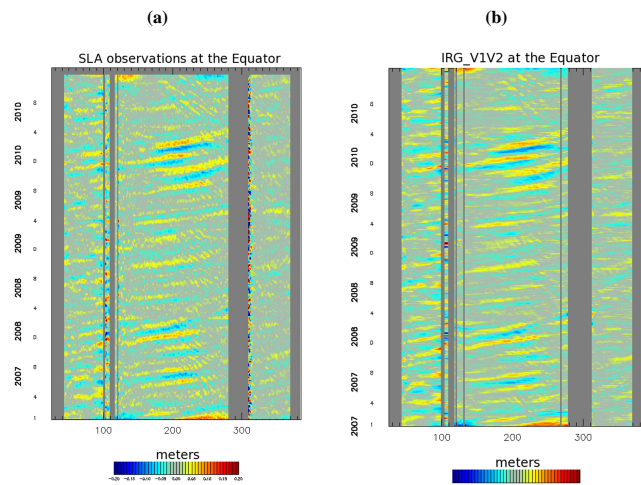

Fig. 22. SLA longitude–time diagrams at the Equator over the world ocean. SLA from AVISO DUACS MyOcean SL TAC **(a)** and from IRG_V1V2 **(b)** are high-pass filtered to keep fluctuations at periods shorter than 128 days.

system with respect to the merged SLA observations. Consistent with Delcroix et al. (1991), the Kelvin waves cross the Pacific Ocean in approximately 2 to 3 months (their phase speed is $O(3\,\mathrm{m\,s^{-1}})$) and their amplitude is $O(15\,\mathrm{cm})$ or more. Rossby waves are three times slower (taking 9 months to cross the Pacific at a speed of $O(1\,\mathrm{m\,s^{-1}})$) and their amplitude is weaker ($O(10\,\mathrm{cm})$). No difference has been observed between IRG_V1V2 and IRG_DEV in this case (not shown). The systems stay close to the observations without inducing shocks or physical inconsistencies thanks to the IAU correction, and the model is allowed to produce smooth propagations. This was an important improvement with respect to previous sequential systems which did suffer initialisation shocks (not shown).

The tropical instability waves (TIW) can be diagnosed from SST (Chelton et al., 2000). These waves initiate at the interface between areas of warm and cold sea surface temperatures near the Equator and form a regular pattern of westward-propagating waves. Longitude–time diagrams at 3° N show TIW mainly in the Pacific and Atlantic Oceans in the eastern parts of the basins (Fig. 23). Similar TIW patterns occur in the IRG_V1V2 system, but their amplitude is slightly underestimated with respect to AVHRR + AMSRE,
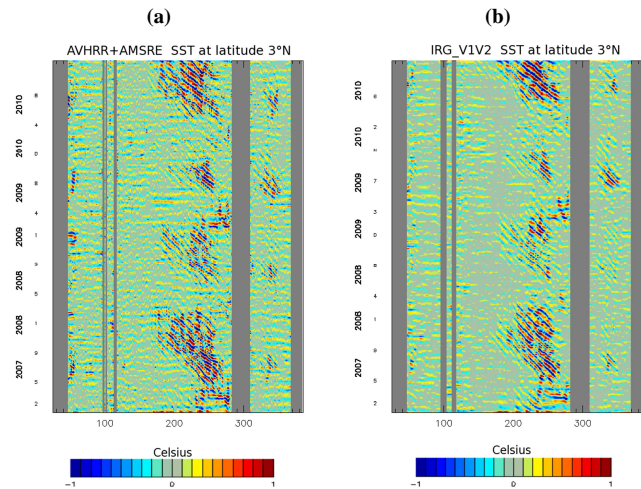
Fig. 23. SST longitude–time diagrams at 3° N over the world ocean. SST from AVHRR + AMSRE (**a**), IRG_V1V2 (**b**) are high-pass filtered to keep fluctuations at periods shorter than 43 days.

as illustrated in Fig. 23. The IRG_DEV system assimilates the AVHRR + AMSRE SST product. Consequently, the IRG_DEV system is even locally closer to this particular SST product than IRG_V1V2 which assimilates RTG (not shown). The HRZ_DEV system, like the HRZ_V1V2 one, displays TIW that have slightly higher amplitude than the IRG system's TIW in the tropical Atlantic (not shown).

4.3.3 Equatorial Undercurrent

We looked more closely at another physical process, the equatorial Pacific current system. Figure 24 shows zonal velocity for equatorial longitude–depth sections and latitude–depth sections at 165° E. We can see that the IRG_V1V2 EUC (Equatorial Undercurrent) does not penetrate west of 165° E and deeper than 100 m. This problem generates an upwelling in the area and changes the water mass properties and transport in the Indonesian Throughflow. The poor representation of the EUC generates errors in the Pacific and Indian Oceans. MDT and its associated error have been adjusted in the IRG_DEV experiment by combining the MDT used by the IRG_V1V2 experiment and GOCE (Gravity field and steady-state Ocean Circulation Explorer) data. An important consequence of this change is that the EUC is continuous over the whole equatorial Pacific in the IRG_DEV experiment, as observed by Johnson et al. (2005) and Marin et al. (2010). Acoustic Doppler Current Profiler (ADCP) observations, from the TAO/TRITON array of moorings (http://www.pmel.noaa.gov/tao), confirm the improvement of subsurface velocities near 165° E at the Equator. The EUC is well marked in the observations from 2007 to 2011, and its amplitude of about 1 m s^{-1} is well represented by IRG_DEV, while it is slightly underestimated by IRG_V1V2.
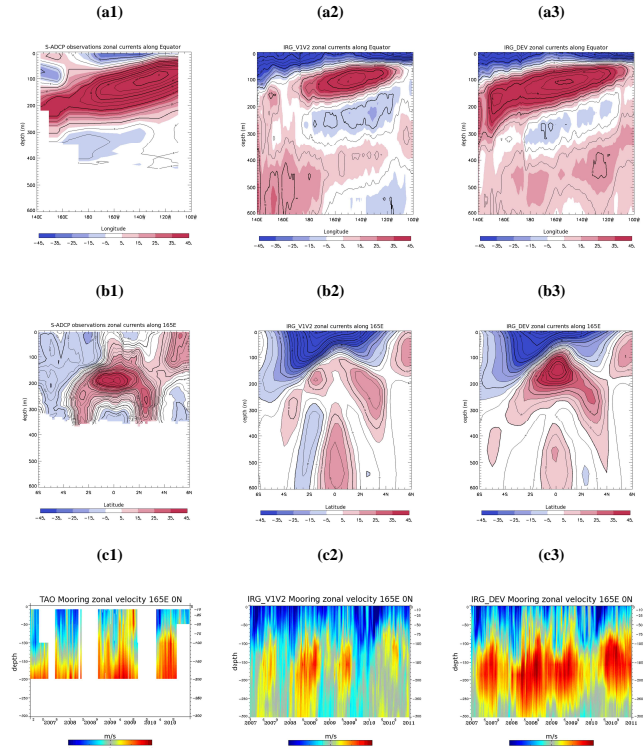

Fig. 24. Zonal velocity (cm s^{-1}) equatorial longitude–depth section (**a1**, **a2**, **a3**) and latitude–depth section at 165° E (**b1**, **b2**, **b3**) in the Pacific Ocean. Zonal velocities over depth and time at the 165° E, 0° N, TAO mooring for the 2007–2010 period (**c1**, **c2**, **c3**). Zonal velocities come from IRG_V1V2 (**a2**, **b2**, **c2**) and IRG_DEV (**a3**, **b3**, **c3**) for the year 2010. Mean zonal sections of ADCP zonal currents come from Johnson et al. (2002) (**a1** and **b1**) and from the TAO ADCP measurements (**c1**).

4.4 Stability in time, trends and biases

Regional biases with respect to assimilated and independent data have been identified in the previous sections, and a set of important physical processes have been validated for 2010. It is now important to verify the long-term behaviour of the system over the whole 2007–2011 period. The linear trend of the temperature at 300 m is estimated with a least squares fit. We call "cumulated trend" the change due to the trend over the period. The global cumulative trend of temperature at 300 m is shown in Fig. 25. There is a noticeable cooling east of the Philippines and two regions of warming west of Australia in IRG_V1V2 (Fig. 25a). These signals are also found in the IRG_DEV (Fig. 25b), but the cooling is generally reinforced. The North and South Atlantic are regions of clear cooling in IRG_DEV. In several regions, the cumulated trend reaches the IRG_DEV bias, as revealed by the 2011 temperature innovations near 300 m (Fig. 26b). This cold bias was not found in the 2007 innovations (Fig. 26a). This confirms the drift shown in Fig. 20 for the North Pacific. There are other regions where the trend from IRG_DEV is not reliable (South
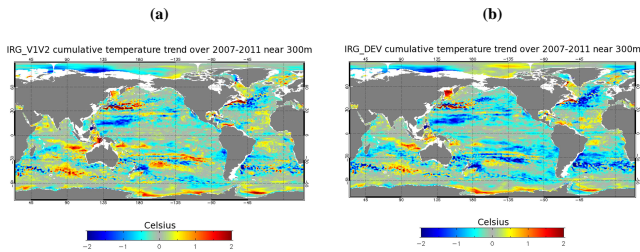
Fig. 25. Cumulative temperature trend (°C) over the 2007–2011 period near 300 m for IRG_V1V2 **(a)** and IRG_DEV **(b)** systems.

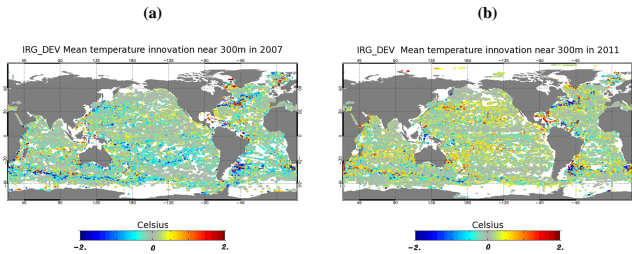**Fig. 26.** Mean temperature innovations (°C) near 300 m in 2007 **(a)** and in 2011 **(b)** for IRG_DEV system.

Pacific east of Australia, South Indian between Madagascar and Australia, etc.).

The large-scale variability of the surface temperature increments can be assessed with an empirical orthogonal function (EOF) analysis. The first EOF of surface temperature increment (Fig. 27) gives the major spatial directions of the surface temperature correction. The associated time series (principal component) shows the time evolution of this dominant correction pattern. The latter confirms the seasonality of the cold bias observed, in particular in the North Pacific (see Fig. 20). The amplitude of this bias is reduced in the IRG_DEV experiment, which indicates that the vertical physics have been improved. The other dominant directions of correction (all variables) are very similar in IRG_V1V2 and IRG_DEV (not shown). These results suggest that the data assimilation system is not responsible for the general cold trend appearing in IRG_DEV and does not accentuate defects of vertical stratification. This confirms the hypothesis of (external) flux correction problems.

To conclude, we have identified drifts that can be and must be corrected if these systems are to be used for climate studies. Even though they may be useful for others studies, this "stability" metric prevents us from releasing the IRG_DEV and HRZ_DEV systems for the moment. These systems are not ready to replace the existing systems in real time.

5 Conclusions

The Mercator Océan global monitoring and forecasting system (MyOcean V2 global MFC) has been evaluated for
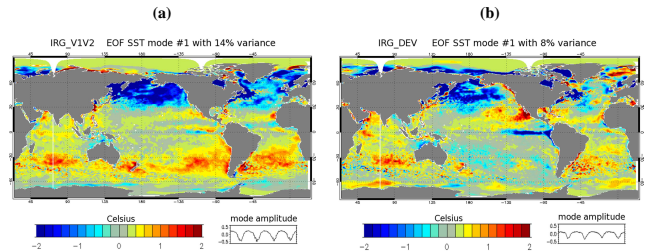

Fig. 27. 1st EOF of surface temperature increment made every 7 days (°C) over the 2007–2010 period for IRG_V1V2 **(a)** and IRG_DEV **(b)** systems. The time series at the bottom of each panel corresponds to the mode amplitude.

the period 2007–2011 by means of a thorough procedure involving statistics of model departures from observations and assessment of physical processes. The accuracy of the V2 system (including the global 1/4° IRG_V1V2 and its 1/12° North Atlantic and Mediterranean zoom HRZ_V1V2) is clearly better than the climatology. Performances are stable over time and should be reliable in the long term.

All monitoring systems are close to SLA observations with a forecast (range 1 to 7 days) RMS difference of 7 cm. This is smaller than the intrinsic variability of the SLA observations. The dominant source of error in sea level comes from the uncertainty in the MDT. The global IRG gives an accurate description of water masses almost everywhere between the ocean bottom and 500 m. Within such layers, departures from in situ observations decrease with the depth and do not exceed 0.2 °C for temperature and 0.03 psu for salinity. Between 0 and 500 m, departures from in situ observations rarely exceed 1 °C and 0.2 psu. Exceptions concern some high variability regions such as the Gulf Stream or the eastern tropical Pacific. Most departures from SST products do not exceed the intrinsic error of these products O(0.6 °C).

During the summer season, the upper ocean is not vertically stratified enough. Excess mixing results in cold biases near the surface and warm biases in subsurface. This bias is particularly marked in IRG_DEV and leads to a large-scale bias of more than 1 °C for the boreal summer in the North Pacific Gyre where the mixed layer is already thin. The bias can be attributed to the atmospheric forcing and/or the bio-optic properties of the seawater. The system is not efficient enough in correcting the SST because heat fluxes are not included in the estimated state. A detailed study reveals that most of the SST correction provided by the analysis is swept away by the bulk forcing function.

A comparison with independent velocity measurements (surface drifters) shows that the surface currents are underestimated in the mid latitudes and overestimated at the Equator with respect to windage and slippage-corrected drifter velocities. This correction developed by Mercator Océan has been applied to cope with the drogue loss of the surface drifter. The underestimation ranges from 20 % in strong currents up

to 60 % in weak currents. The orientation of the current vectors is well represented.

Despite the fact that the IRG systems do not include any assimilation of sea ice data, they reproduce the sea ice seasonal cycle in a very realistic manner. Moreover, thanks to the assimilation of the AVHRR + AMSRE SST product, the IRG_DEV system represents the sea ice edge significantly better, especially during the summer.

The same scientific assessment procedure was conducted on IRG_DEV and HRZ_DEV. Diagnostics were compared with the MyOcean version IRG_V1V2 and HRZ_V1V2. Improvements are the rule. The update of the MDT (and its error) corrects local biases in the Indonesian Throughflow and in the western tropical Pacific. This improves the subsurface currents at the Equator.

Earlier studies have shown that RTG SST data suffer from a cold bias near the coasts. The use of the AVHRR + AMSRE product instead of RTG helps to significantly reduce the surface temperature and salinity biases, as well as RMS differences. The sea ice concentration in both the Arctic and Antarctica also benefits from the assimilation of an improved SST product.

The scientific quality assessment procedure detected a drift in the IRG_DEV system in 2009, two years after the start. A slight cold and salty bias is developing near 100 m. The drift in the HRZ_DEV system is less marked and only concerns water deeper than 2000 m. The drift is a combined effect of the new model parameterizations and surface forcing. Moreover, the data assimilation does not correct the air–sea fluxes, and the IAU of the initial condition is partly cancelled near the surface by the bulk formulae. For instance, a warming increment will be damped by an increased latent heat flux.

IRG_DEV and HRZ_DEV systems have better statistics in spite of this drift. However this drift prevents us from upgrading the system in real time. It demonstrates that several years have to be computed and thoroughly assessed before moving to a new version, especially if new model parameterizations are involved.

We tried to find, for the Mercator Océan systems, the best compromise between getting the "best analysis" and having a sustainable system in an operational context. Using a 1° analysis grid over most of the domain for the 1/4° system with localization radii of 200 km is not satisfactory from a theoretical point of view and constitutes a flaw in the conception of the 1/4° system. This flaw does not alter the mean conclusions of the paper, but reassessment of this aspect of the assimilation is a high priority for future work and will be corrected in further versions of the 1/4° system.

Several scientific and technical choices have been validated such as the use of AVHRR+AMSRE SST for data assimilation, the use of a MDT adjusted with GOCE and with the system innovations, and the parameterization of observation and representativity error covariances. We plan to use the IRG and HRZ systems to test new parameterizations for the future high resolution 1/12° global system (HRG) in order to reduce the computational cost. The current configuration of the free model (without assimilation) of the future HRG system does not exhibit the drifts diagnosed in IRG_DEV. The next version of HRG will include the validated changes mentioned above. A correction of the air temperature will also be introduced in order to avoid the damping of increments via the bulk forcing function.

References

Amante C. and Eakins, B. W.: ETOPO1 1 Arc-minute global relief model: procedures, data sources and analysis, NOAA Technical Memorandum NESDIS NGDC-24, 25 pp., 2009.

Antonov, J. I., Locarnini, R. A., Boyer, T. P., Mishonov, A. V., and Garcia, H. E.: World Ocean Atlas 2005, Vol. 2: Salinity, in: NOAA Atlas NESDIS 62, edited by: Levitus, S., US Government Printing Office, Washington, DC, 182 pp., 2006.

Arakawa, A. and Lamb, V. R.: A potential enstrophy and energy conserving scheme for the shallow water equations, Mon. Weather. Rev., 109, 18–36, 1981.

Barnier, B., Madec, G., Penduff, T., Molines, J. M., Treguier, A. M., Le Sommer, J., Beckmann, A., Biastoch, A., Böning, C., Dengg, J., Derval, C., Durand, E., Gulev, S., Remy, E., Talandier, C., Theetten, S., Maltrud, M., McClean, J., and De Cuevas, B.: Impact of partial steps and momentum advection schemes in a global circulation model at eddy permitting resolution, Ocean Dynam., 56, 543–567, 2006.

Becker, J. J., Sandwell, D. T., Smith, W. H. F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S.H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Trimmer, R., Von Rosenberg, J., Wallace, G., and Weatherall, P.: Global Bathymetry and Elevation Data at 30 Arc Seconds Resolution: SRTM30_PLUS, Mar. Geod., 32, 355–371, doi:10.1080/01490410903297766, 2009.

Benkiran, M. and Greiner, E.: Impact of the Incremental Analysis Updates on a Real-Time System of the North Atlantic Ocean, J. Atmos. Ocean. Tech., 25, 2055–2073, 2008.

Bernie, D. J., Woolnough, S. J., Slingo, J. M., and Guilyardi, E.: Modeling diurnal and intraseasonal variability of the ocean mixed layer, J. Climate, 18, 1190–1202, 2005.

Blanke, B. and Delecluse, P.: Variability of the tropical Atlantic-Ocean simulated by a general-circulation model with 2 different mixed-layer physics, J. Phys. Oceanogr., 23, 1363–1388, 1993.

Bloom, S. C., Takas, L. L., Da Silva, A. M., and Ledvina, D.: Data assimilation using incremental analysis updates. Mon. Weather Rev., 124, 1256–1271, 1996.

Cabanes, C., Grouazel, A., Von Schuckmann, K., Hamon, M., Turpin, V., Coatanoan, C., Guinehut, S., Boone, C., Ferry, N., Reverdin, G., Pouliquen, S., and Le Traon, P. Y.: The CORA dataset: Validation and Diagnostics of ocean temperature and salinity in situ measurement, Ocean Sci., in press, 2013.

Chelton, D. B., Wentz, F. J., Gentemann, C. L., De Szoeke, R. A., and Schlax, M. G.: Satellite microwave SST observations of transequatorial tropical instability waves, Geophys. Res. Lett., 27, 1239–1242, 2000.

Chen, J. L., Wilson, C. R., Tapley, B. D., Famiglietti, J. S., and Rodell, M.: Seasonal global mean sea level change from satellite altimeter, GRACE, and geophysical models, J. Geodesy, 79, 532–539, doi:10.1007/s00190-005-0005-9, 2005.

Cravatte, S., Madec, G., Izumo, T., Menkes, C., and Bozec, A.: Progress in the 3-D circulation of the eastern equatorial Pacific in a climate, Ocean Model., 17, 28–48, 2007.

Crosnier, L. and Le Provost, C.: Inter-comparing five forecast operational systems in the North Atlantic and Mediterranean basins: the MERSEA-strand1 methodology, J. Mar. Syst., 65, 354–375, 2007.

Cummings, J., Bertino, L., Brasseur, P., Fukumori, I., Kamachi, M., Martin, M. J., Mogensen, K., Oke, P., Testut, C. E., Verron, J., and Weaver, A.: Ocean Data Assimilation Systems for GODAE, Oceanography, 22, 96–109, 2009.

Dai, A. and Trenberth, K. E.: Estimates of freshwater discharge from continents: latitudinal and seasonal variations, J. Hydrometeorol., 3, 660–687, 2002.

Delcroix, T., Picaut, J., and Eldin, G.: Equatorial Kelvin and Rossby waves evidenced in the Pacific Ocean through Geosat Sea Level and Surface Current Anomalies, J. Geophys. Res., 96, 3249–3262, 1991.

Dombrowsky, E., Bertino, L., Brassington, G. B., Chassignet, E. P., Davidson, F., Hurlburt, H. E., Kamachi, M., Lee, T., Martin, M. J., Mei, S., and Tonani, M.: GODAE systems in operation, Oceanography, 22, 80–95, doi:10.5670/oceanog.2009.68, 2009.

Donlon, C. J., Martin, M., Stark, J. D., Roberts-Jones, J., Fiedler, E., and Wimmer, W.: The Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system, Remote Sens. Environ., 116, 140–158, 2012.

Elmoussaoui, A., Perruche, C., Greiner, E., Ethé, C., and Gehlen, M.: Integration of biogeochemistry into Mercaor Océan systems, Mercator Océan Newsletter 40: MyOcean Ecosystem Models, 3–14, 2011.

Ezraty, R., Girard-Ardhuin, F., Piolle, J. F., Kaleschke, L., and Heygster, G.: Arctic and Antarctic sea ice concentration and Arctic sea ice drift estimated from Special Sensor Microwave data, User's Manuel Version 2.1, CERSAT, 2007.

Ferry, N., Parent, L., Garric, G., Bricaud, C., Testut, C. E., Le Galloudec, O., Lellouche, J. M., Drevillon, M., Greiner, E., Barnier, B., Molines, J. M., Jourdain, N. C., Guinehut, S., Cabanes, C., and Zawadzki, L.: GLORYS2V1 global ocean reanalysis of the altimetric era (1992–2009) at meso scale, Mercator Newsletter 44, 29–39, 2012.

Fichefet, T. and Maqueda, M. A.: Sensitivity of a global sea ice model to the treatment of ice thermodynamics and dynamics, J. Geophys. Res., 102, 12609–12646, 1997.

Garric, G., Verbrugge, N., Bouillon, S., and Vancoppenolle, M.: Preliminary assessment of sea ice in the global 1/4° Mercator Ocean forecasting system, Mercator Océan Newsletter 28, 2008.

Gemmill, W., Bert K., and Xu L.: Daily Real-Time Global Sea Surface Temperature - High Resolution Analysis at NOAA/NCEP, NOAA/NWS/NCEP/MMAB Office Note Nr. 260, 39 pp., 2007.

Goosse, H., Campin, J. M., Deleersnijder, E., Fichefet, T., Mathieu, P. P., Maqueda, M. A. M., and Tartinville, B: Description of the CLIO model version 3.0, Institut d'Astronomie et de Geophysique Georges Lemaitre, Catholic University of Louvain (Belgium), 2001.

Grasso, L. D.: The differentiation between grid spacing and resolution and their application to numerical modelling, B. Am. Meteor. Soc., 81, 579–580, 2000.

Grodsky, S. A., Lumpkin, R., and Carton, J. A.: Spurious trends in global surface drifter currents, Geophys. Res. Lett., 38, L10606, doi:10.1029/2011GL047393, 2011.

Guinehut, S.: Validation of different SST products using Argo dataset, CLS, Toulouse, Report CLS-DOS-NT-10-264, 42 pp., 2010.

Hernandez, F., Bertino, L., Brassington, G. B., Chassignet, E. P., Cummings, J., Davidson, F., Drévillon, M., Garric, G., Kamachi, M., Lellouche, J. M., Mahdon, R., Martin, M. J., Ratsimandresy, A., and Régnier, C.: Validation and intercomparison studies within GODAE, Oceanography, 22, 128–143, 2009.

Huang, X. Y., Mogensen, K. S., and Yang, X.: First-guess at appropriate time: the HIRLAM implementation and experiments, Proc. HIRLAM Workshop on Variational Data Assimilation and Remote Sensing, Helsinki, Finland, Finnish Meteorological Institute, 28–43, 2002.

Huffman, G. J., Adler, R. F., Bolvin, D. T., and Gu, G.: Improving the Global Precipitation Record: GPCP Version 2.1, Geophys. Res. Lett., 36, L17808, doi:10.1029/2009GL040000, 2009.

Hunke, E. C. and Dukowicz J. K.: An elastic-viscous-plastic model for sea ice dynamics, J. Phys. Oceanogr., 27, 1849–1867, 1997.

Jacobs, S., Hellmer, H., Doake, C. S. M., Jenkins, A., and Frolich, R. M.: Melting of ice shelves and the mass balance of Antarctica, J. Glaciol., 38: 375–387, 1992.

Johnson, G. C., Sloyan, B. M., Kessler, W. S., and McTaggart K. E.: Direct measurements of upper ocean currents and water properties across the tropical Pacific Ocean during the 1990's, Prog. Oceanogr., 52, 31–61, doi:10.1016/S0079-6611(02)00021-6, 2002.

Koch-Larrouy, A., Madec, G., Blanke, B., and Molcard, R.: Water mass transformation along the Indonesian throughflow in an OGCM, Ocean Dynam., 58, 289–309, doi:10.1007/s10236-008-0155-4, 2008.

Large, W. G. and Yeager, S. G.: The global climatology of an interannually varying air–sea flux data set, Clim. Dynam., 33, 341–364, doi:10.1007/s00382-008-0441-3, 2009.

Le Traon, P. Y., Dibarboure, G., and Ducet, N.: Use of a High Resolution Model to Analyse the Mapping Capabilities of Multiple-Altimeter Missions, J. Atmos. Ocean. Tech., 18, 1277–1288, 2001.

Lévy, M., Estublier, A., and Madec, G.: Choice of an advection scheme for biogeochemical models, Geophys. Res. Lett., 28,

3725–3728, doi:10.1029/2001GL012947, 2001.

Lherminier, P., Mercier, H., Gourcuff, C., Alvarez, M., Bacon, S., and Kermabon, C.: Transports across the 2002 Greenland–Portugal OVIDE section and comparison with 1997, J. Geophys. Res., 112, C07003, doi:10.1029/2006JC003716, 2007.

Lique, C., Garric, G., Treguier, A. M., Barnier, B., Ferry, N., Testut, C. E., and Girard-Ardhuin, F.: Evolution of the Arctic Ocean salinity, 2007–2008: contrast between the Canadian and the Eurasian basins, J. Climate, 24, 1705–1717, 2011.

Locarnini, R. A., Mishonov, A. V., Antonov, J. I., Boyer, T. P., and Garcia, H. E.: World Ocean Atlas 2005, Volume 1: Temperature, in: NOAA Atlas NESDIS 61, edited by: Levitus, S., US Government Printing Office, Washington, DC, 182 pp., 2006.

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

Madec, G. and Imbard M.: A global ocean mesh to overcome the North Pole singularity, Clim. Dynam., 12, 381–388, 1996.

Madec, G. and the NEMO team: NEMO ocean engine, Note du Pôle de modélisation, Institut Pierre-Simon Laplace (IPSL), France, No. 27 ISSN, 1288–1619, 2008.

Maraldi, C., Chanut, J., Levier, B., Reffray, G., Ayoub, N., De Mey, P., Lyard, F., Cailleau, S., Drévillon, M., Fanjul, E. A., Sotillo, M. G., Marsaleix, P., and the Mercator team: NEMO on the shelf: assessment of the Iberia-Biscay-Ireland configuration, Ocean Sci., in press, 2013.

Marin, F., Kestenare, E., Delcroix, T., Durand, F., Cravatte, S., Eldin, G., and Bourdallé-Badie, R.: Annual reversal of the equatorial intermediate current in the Pacific: observations and model diagnostics, J. Phys. Oceanogr., 40, 915–933, 2010.

Murphy, A. H.: Skill scores based on the mean square error and their relationships to the correlation coefficient, Mon. Weather Rev., 116, 2417–2424, 1988.

O'Dea, E. J., While, J., Furner, R., Hyder, P., Arnold, A., Storkey, D., John R., Siddorn, J. R., Martin, M., Liu, H., and Holt, J. T.: NEMO-Shelf, towards operational oceanography with SST data assimilation on the North West European Shelf, Mercator Océan Newsletter 39, 25–28, 2010.

Oke, P. R., Brassington, G. B., Griffin, D. A., and Schiller, A.: The Bluelink Ocean Data Assimilation System (BODAS), Ocean Model., 21, 46–70, 2008.

Ollivier, A. and Faugere, Y.: Envisat RA-2/MWR ocean data validation and cross-calibration Activities, Yearly report, Technical Note CLS.DOS/NT/10.018, Contract N SALP-RP-MA EA-21920-CLS, 2010.

Penduff, T., Le Sommer, J., Barnier, B., Treguier, A. M., Molines, J. M., and Madec, G.: Influence of numerical schemes on current-topography interactions in 1/4° global ocean simulations, Ocean Sci., 3, 509–524, doi:10.5194/os-3-509-2007, 2007.

Penduff, T., Juza, M., Brodeau, L., Smith, G. C., Barnier, B., Molines, J. M., Treguier, A. M., and Madec, G.: Impact of global ocean model resolution on sea-level variability with emphasis on interannual time scales, Ocean Sci., 6, 269–284, doi:10.5194/os-6-269-2010, 2010.

Pham, D. T., Verron, J., and Roubaud, M. C.: A singular evolutive extended Kalman filter for data assimilation in oceanography, J. Mar. Syst., 16, 323–340, 1998.

Redelsperger, J. L., Thorncroft, C., Diedhiou, A., Lebel, T., Parker, D. J., and Polcher, J.: African Monsoon Multidisciplinary Analysis (AMMA): an international research project and field campaign, B. Am. Meteorol. Soc., 87, 1739–1746, 2006.

Reynolds, R. W. and Chelton, D. B.: Comparisons of Daily Sea Surface Temperature Analyses for 2007–08, J. Climate, 23, 3545–3562, doi:10.1175/2010JCLI3294.1, 2010.

Reynolds, R. W., Smith, T. M., Liu, C., Chelton, D. B., Casey, K. S., and Schlax, M. G.: Daily high-resolution blended analyses for sea surface temperature, J. Climate, 20, 5473–5496, 2007.

Rio, M. H., Guinehut, S., and Larnicol, G.: New CNES-CLS09 global mean dynamic topography computed from the combination of GRACE data, altimetry, and in situ measurements, J. Geophys. Res., 116, C07018, doi:10.1029/2010JC006505, 2011.

Roemmich, D., Belbeoch, M., Belchi, P. J. V., Freeland, H., Gould, W. J., Grant, F., Ignaszewski, M., Klein, B., Mork, K.A., Owens, W.B., Pouliquen, S., Ravichandran, M., Riser, S., Sterl, A., Suga, T., Suk, M. S., Sutton, P., Thierry, V., Le Traon, P. Y., Wijffels, S., and Xu, J.: Argo: the challenge of continuing 10 years of progress, Oceanography, 22, 46–55, 2008.

Roquet, F., Charrassin, J. B., Marchand, S., Boehme, L., Fedak, M., Reverdin, G., and Guinet, C.: Delayed-mode calibration of hydrographic data obtained from animal-borne satellite relay data loggers, J. Atmos. Ocean. Tech., 28, 787–801, 2011.

Roullet, G. and Madec, G.: Salt conservation, free surface, and varying levels: a new formulation for ocean general circulation models, J. Geophys. Res., 105, 23927–23942, 2000.

Silva, T. A. M., Bigg, G. R., and Nicholls, K. W.: Contribution of giant icebergs to the Southern Ocean freshwater flux, J. Geophys. Res., 111, C03004, doi:10.1029/2004JC002843, 2006.

Talagrand, O.: A posteriori evaluation and verification of analysis and assimilation algorithms, Proc. of ECMWF Workshop on Diagnosis of Data Assimilation System (Reading), 17–28, 1998.

Tranchant, B., Testut, C. E., Renault, L., Ferry, N., Birol, F., and Brasseur, P.: Expected impact of the future SMOS and Aquarius Ocean surface salinity missions in the Mercator Océan operational systems: New perspectives to monitor ocean circulation, Remote Sens. Environ., 112, 1476–1487, 2008.