Ocean Sci. Discuss., 11, 1435–1472, 2014
www.ocean-sci-discuss.net/11/1435/2014/
doi:10.5194/osd-11-1435-2014

Ocean Science
Discussions
Open Access

Forecasting the mixed layer depth in the north east Atlantic: an ensemble approach, with uncertainties based on data from operational oceanic systems

Y. Drillet[1], J. M. Lellouche[1], B. Levier[1], M. Drévillon[1], O. Le Galloudec[1], G. Reffray[1], C. Regnier[1], E. Greiner[2], and M. Clavier[1]

[1]Mercator Océan, Toulouse, France
[2]CLS, Toulouse, France

Received: 28 April 2014 – Accepted: 20 May 2014 – Published: 11 June 2014

Correspondence to: Y. Drillet (yann.drillet@mercator-ocean.fr)

Published by Copernicus Publications on behalf of the European Geosciences Union.

Abstract

Operational systems operated by Mercator Océan provide daily ocean forecasts, and combining these forecasts we can produce ensemble forecast and uncertainty estimates. This study focuses on the mixed layer depth in the North East Atlantic near the Porcupine Abyssal Plain for May 2013. This period is of interest for several reasons: (1) four Mercator Océan operational systems provide daily forecasts at a horizontal resolution of 1/4°, 1/12° and 1/36° with different physics; (2) glider deployment under the OSMOSIS project provides observation of the changes in mixed layer depth; (3) the ocean stratifies in May, but mixing events induced by gale force wind are observed and forecasted by the systems. A statistical approach and forecast error quantification for each system and for the combined products are presented. Skill scores indicate that forecasts are in any case better than persistence, and temporal correlations between forecast and observations are greater than 0.8 even for the 4 day forecast. The impact of atmospheric forecast error, and for the wind field in particular, is also quantified in terms of the forecast time delay and the intensity of mixing or stratification events.

1 Introduction

Operational oceanography has developed since the end of the 90's in several countries with global level partnerships under the GODAE Oceanview initiative (https://www.godae-oceanview.org/) and with European funding through Mersea, MyOcean and then MyOcean2 projects (http://www.myocean.eu/). Mercator Océan is a French institution providing operational ocean forecasts for national requirements and also contributing to international efforts. A suite of global and regional ocean forecast systems has been developed and provides daily forecasts. These forecasts are all available through MyOcean or Mercator Océan services. Built around four operational forecast systems, the aim of this study is to obtain an estimate of the mixed layer depth and the associated uncertainty. The North East Atlantic area was chosen because, since the launch of the V3 MyOcean service at the end of April 2013, four systems with different resolutions are now available on a daily basis. Moreover, glider observations for May 2013 are available in the Coriolis data base (available through MyOcean service) sampling over the whole month in a small $1/2^\circ \times 1/2^\circ$ box centred on 16.25° W and 48.55° N. The physical variable chosen for this study is the mixed layer depth, since the ocean stratifies during the spring months and some mixing events occur which are directly linked to atmospheric forcing. Other studies quantifying the uncertainties in the ocean forecast for several oceanic fields (Lermusiaux et al., 2006), made use of super ensemble techniques (Vandenbulcke et al., 2009; Lenartz et al., 2010; Pistoia, 2012; Scott et al., 2012) or have quantified the impact of medium range atmospheric forecasting on the ocean (Drillet et al., 2009). An ensemble approach is also used in oceanography for estimating variability at a more climatic scale, for example in Zhu et al. (2012), Xue et al. (2012) and more recently in the Clivar Exchange special issue (http://www.clivar.org/sites/default/files/Exchanges/Exchanges_64.pdf). Some fairly complex techniques and diagnostics can be used, but in this study standard statistical techniques are used to compare several estimates of the forecast mixed layer depth. Several systems and forecast lengths, and several initial states and atmospheric forcings were used. The paper is organized as follows: the first section describes the simulations and the observations used in the study. The second section draws on the statistics for quantifying forecast error. The third section describes the mixed layer depth variability during May 2013, and how uncertainties in the observations and forecasts can be estimated. The last section presents the main conclusions of the study.

2 Forecast products and observations

The forecasts used in this study are provided by Mercator Océan using four different operational systems (Table 1). Two global ocean systems, one at $1/4^\circ$ horizontal resolution (Glo4, Lellouche et al., 2013), and the second at $1/12^\circ$ (Glo12) are used. Two regional systems are also used, one covering the North Atlantic and the Mediterranean at 1/12° (Atl12, Lellouche et al., 2013) and the last at 1/36° (Ibi36, Maraldi et al., 2013) covering the North East Atlantic. All these systems are based on the NEMO ocean code (Madec et al., 2008), using the same 50-level vertical grid and forced by ECMWF atmospheric analysis and forecasts. The initial state of each forecast is computed with data assimilation or with re-initialization techniques. The SAM2 method (Tranchant et al., 2008; Lellouche et al., 2013) is used to assimilate in situ and satellite observations. This reduced-order Kalman filter method is based on the singular evolutive extended Kalman filter (SEEK) formulation introduced by Pham et al. (1998). This method is used each week (on Wednesdays as shown in Fig. 1) to produce the initial state of the forecast for the Glo4, Glo12 and Atl12 systems. Two assimilation cycles are performed allowing the assimilation of observations up to two weeks old. The re-initialization method is used for Ibi36 system (also on Wednesdays as shown in Fig. 1); a two week “spin up” is carried out to stabilize the high resolution solution (at 1/36°) which is initialized using the 1/12° analysis produced by the Atl12 system. This method and the effect of the length of spin up time are detailed in Cailleau et al. (2012). The main characteristics of these systems are detailed in Table 1. Figure 1 shows more precisely how the systems are operated on a daily basis. Every day each system provides a hindcast estimate, H, of the ocean state. H is initialized using the “best” ocean state available, and forced with the “best” atmospheric forcing, i.e. the ECMWF analysis. The days following the simulation are F0, F1, F2, F3 and F4, respectively the current day, the one-day forecast and so on. All the ocean forecasts are forced by an atmospheric forecast. Using this scenario, we can build a 4-member ensemble for each forecast length differing mostly in their initial states, and the mean and median of this ensemble can be considered as two other forecasts. We obtain for each date thirty 3-D ocean states which are not independent estimates of the ocean. A reference experiment, hereafter called Atl12 free, was also carried out using the Atl12 system without data assimilation. This simulation was initialised in March 2013 with the analysis provided by Atl12 system, and forced using the atmospheric forecast analysis to the end of May 2013. The results of this experiment will be discussed in Sect. 4.3.4.

This study focuses on May 2013, where the four systems described above were available, and where the Coriolis in situ database contains repetitive in situ profiles obtained from glider observations around 16.5° W and 48.5° N (Fig. 2). These gliders were deployed under the OSMOSIS (Ocean Surface Mixing, Ocean Sub-mesoscale Interaction Study, http://www.bodc.ac.uk/projects/uk/osmosis/introduction/) project, in which care is taken to apply a near real time quality control, disseminating the observations to the Coriolis data centre in real time. The observations used in this study do not represent the full data set but a subsampled one as carried out in the system for all type of in situ observations. For each instrument (in this case gliders) only one profile per day is retained to avoid over sampling (in time and space) of the observations since the global model cannot represent processes at such high resolution. In our subsample database, 74 vertical profiles are available for May 2013 with almost one profile per day. This data set allows a good representation of the day-to-day temporal evolution of the temperature profile, and then of the mixed layer depth during the entire month of May 2013. De Boyer Montégut et al. (2004) defined a 0.2 °C criterion which is used both for the model and the observations in computing the mixed layer depth, and reapplied here in a slightly different context. In their paper, this criterion was used to compute global mixed layer depth climatology based on in situ observations; Fig. 3 shows the availability of the criterion in our area of interest. When the mixed layer is really pronounced as for one profile on 11 May and for the 3 profiles of 28 May, the criterion detects the base of the mixed layer. When the profile is more mixed, as for 18 May, the base of the thermocline is also detected. From a practical point of view, all the in situ profiles are interpolated on the vertical model grid to simplify the comparison between models and observations. The base of the mixed layer, for model and observations, as selected by the criterion is the first level just above the point where the vertical temperature gradient from the surface exceeds 0.2 °C. The precision of this estimate depends on the depth as function of the vertical grid; at the surface this is around 1 m and at 50 m depth around 10 m. This time period is of particular interest as it exhibits the spring re-stratification phase; gusts of wind occur also during this month and their effects on vertical mixing can be quantified. Some meso-scale oceanic structures are also present in this area, associated with strong fronts and currents which induce vertical mixing. In what follows, analyses and statistics computed using the model outputs and observations are based on (i) daily values which are actually daily means for the model but only the mean of all the data available during the day for the observations, and (ii) the spatial mean over the $1/2^\circ \times 1/2^\circ$ box defined previously. This box contains all in the situ profiles available during this month (Fig. 2) and is small enough when compared with the meso-scale structures in this area. This choice, both spatially and temporally, is justified by the fact that the model cannot simulate all the smaller scales available in the in situ observations. To illustrate more precisely the daily variability of the observations, Fig. 2 (right panel) shows a zoomed portion of selected dates. From 11 to 14 May there is a large variability in the observed MLD in a small $1/4^\circ \times 1/4^\circ$ box. All these observations should be in the same model mesh in the $1/4^\circ$ model and in neighbouring meshes in the $1/12^\circ$ models. However, observations show different profiles where the mixed layer depth varies over several tens of metres as for example on 11 and 14 May. This cannot appear in the model because daily average outputs are stored. The best way to compare observations and model outputs at differing horizontal resolutions is to average spatially and temporally, and to consider daily profiles over the month where smaller scales in observation and high resolution models are filtered out.

3 Statistics

The statistics computed are mean bias (not shown), temporal correlation, error standard deviation for the Taylor diagram (Fig. 4), skill score (Fig. 5) (Murphy, 1988) and RMS error (Table 2) for each system and for the ensemble mean and median. The skill scores are computed as follows:

$$\mathrm{SS}_i = 1 - \frac{\sum_{\mathrm{date}}\left(F_i^{\mathrm{date}} - H_i^{\mathrm{date}}\right)^2}{\sum_{\mathrm{date}}\left(P_i^{\mathrm{date}} - H_i^{\mathrm{date}}\right)^2} \tag{1}$$

where H is the hindcast, F the forecast and P the persistence of the initial state or observations, this score being computed independently for each system. The skill score is computed for all the days of May 2013, and index i is the forecast length ranging from 0 for the forecast of the current day to 4 for the 4 day forecast.

The mean bias is small for Atl12 and Ibi36 (less than 2 m up to 3 days of forecast length) and is greater in the 2-global systems, with values greater than 5 m. The 4 day forecast has the same bias amplitude with all systems (around 5 m) but with a negative bias for Atl12 and Ibi36 and a positive bias for Glo4 and Glo12. Generally, there is a positive bias in the Glo4 and Glo12 mixed layer depth, which means that the mixed layer depth is too deep when it is underestimated with Ibi36. These results are consistent with the validation work done regularly for the Mercator Océan real time production (Drevillon et al., 2014). The Taylor diagram (Fig. 4) summarizes the following results: the temporal correlation between forecast and observation is greater than 0.85 for the first forecast day and decreases more or less depending on the system and/or the forecast length. Glo4 system is an exception; it has the lowest correlation for the first forecast lengths (from 0.78 for H to 0.76 for F0 and F1) and then increases to 0.81 for the 4-day-forecast length. The ensemble mean gives the best result even if the Glo4 forecast is far worse than the other systems. The results are very similar (except for Glo4) for H up until the 1 day forecast; the dispersion of the systems (illustrated by the colour) is small in the Taylor diagram (Fig. 4) for all the metrics (correlation, standard deviation or RMS). However, the forecast dispersion increases after the 2 day forecast and in particular there is a significant decrease in correlation to under 0.79 for Glo12, when it remains around 0.85 for Ibi36. The RMS error (Table 2) confirms on previous results with a smaller error for Ibi36 and the ensemble mean (between 15 m and 18 m

RMS) and a larger RMS error for Glo4 (between 27 m and 30 m RMS). The skill score (Fig. 5), which measures improvement of the forecast in comparison to persistence, shows positive values (meaning that the forecast is better than persistence) for all forecasts except F0 in the Glo4 system. As expected, it increases with the forecast length meaning that the 4 day forecast is more efficient than the 1 day forecast in beating persistence. The same kind of diagnostic (Fig. 5 bottom panel) was carried out using persistence of the observations instead of the persistence of the initial state. Observation allows the use of the same "reference" state (in this case the observations) to compare different systems. This diagnostic gives the same information with the same rank among the systems and the same increase in skill score with forecast length. It can be seen that using the persistence of the observations shows more clearly three "classes" of score as in the Taylor diagram (Fig. 4); the best is obtained with Atl12, Ibi36 and the mean and median products, a second with a significant decrease in the score obtained with Glo12, and a third with Glo4. Combining the forecasts in another way, simply by removing one system from the statistics, quantifies the gain (or degradation) obtained with each individual system. Table 2 shows the value of the RMS error for these combinations; the robust result is that the best forecast is obtained for all the forecast length with the mean computed after removing the Glo4 system, and with the Ibi36 system. Removing the Glo4 estimate, it may be noted that the mean of these forecasts is better than all the individual forecasts, showing that each estimate of the ocean state gives pertinent information in terms of statistics for the forecast.

4 Mixed layer depth forecast during May 2013

4.1 Description of the mixing and stratification events

In our area of study, centred on 16.25° W–48.45° N in the North East Atlantic, May 2013 is characterized by mixing and stratification events. Figure 6 illustrates this variability, with three mixing events (referred to as M1, M2 and M3) and three stratification events (referred to as S1, S2 and S3) well marked in both observations and simulations. Figure 6 shows the variability over the same period and in the same area for the main atmospheric forcing parameters, which are respectively wind speed, total heat flux and the fresh water budget. We note the good correspondence between the evolution of the mixed layer depth and atmospheric forcing. The first mixing event (called M1 with a maximum value on 11 May) occurs just after a strong gust of wind ($\sim 13\,\mathrm{m\,s^{-1}}$ on 8 May) and corresponds to an abrupt loss of heat ($\sim 100\,\mathrm{W\,m^{-2}}$) and an evaporation phase. The first stratification phase (called S1) is a short event occurring between 11 and 13 May and corresponds to a sudden change in the heat flux with values decreasing from $+150\,\mathrm{W\,m^{-2}}$ to $-100\,\mathrm{W\,m^{-2}}$ over a few days (3–4 days). The second mixing event (called M2 with a maximum value on 17 May) is longer; it follows a short re-stratification phase before reaching the maximum mixed layer depth and remains around 130 m depth for 3 days. This mixing phase is also associated with strong winds and heat loss but with fresh water fluxes remaining positive. A gradual stratification event (called S2) follows, occurring during a low wind and a warming period which restratifies the entire water column. At the end of the month, a final strong gust of wind, causing heat loss and following excess precipitation, induces the M3 mixing event (28 May). The last rapid re-stratification of the entire water column (S3) occurs when the wind decreases. Several robust conclusions can be drawn from these alternating mixing and stratification events. First, all mixing events are associated with strong winds (exceeding $12\,\mathrm{m\,s^{-1}}$) occurring a few days before the maximum of the mixed layer depth is reached. For M1 and M2, the wind event occurs three days before the mixing maximum, while for M3 the response is faster (only one day). These strong wind events are always associated with a large (less than $-80\,\mathrm{W\,m^{-2}}$) heat loss and follows positive fresh water fluxes. Re-stratification events occur when the wind speed decreases (less than $10\,\mathrm{m\,s^{-1}}$) and when the ocean absorbs heat with total fluxes greater than $100\,\mathrm{W\,m^{-2}}$: one day for the S1 events and over a longer period (6 days) for the S2 event.

The standard deviation of all mixed layer depths available for all systems and all forecast lengths (Fig. 6), is also correlated with the uncertainties in the atmospherics fluxes, estimated as the standard deviation of all atmospheric flux estimates. There is a greater uncertainty for the mixed layer depth during M1, S1, M2 and M3 with a standard deviation around 20 m, also a smaller one around a few metres, during the S2 and S3 events. This is also true for the wind and heat fluxes where uncertainties are greatest during wind events, especially during the maximum of wind speed. For the observations, the uncertainties are represented in Fig. 6 as vertical bars centred on the mean values of the observations in the box for the day. Figure 2 shows the spatial distribution of these observations for each day. The large uncertainty for the M1 and S1 events (from 11 to 14 May) is explained by the fact that we have a large gradient in the mixed layer depth estimate with a value deeper than 80 m in the eastern part, around 60 m in the western part and around 40 m in the middle (Fig. 2). This gradient is smaller for the two other mixing events (M2 and M3). The uncertainty in the observations of the S2 stratification event is quite small and the right-hand panel of Fig. 2 shows a significantly shallow mixed layer of depth less than 20 m for 24, 25 and 26 May as indicated by the small black and purple circles. This uncertainty in the observation is not a robust diagnostic because the number of observations in our case is too small to give a precise estimate of this uncertainty, but nevertheless it gives useful information for evaluating the model. In this particular experiment, at this place, during this month and taking into account the estimate of the uncertainty for the model and the observations, the model is in agreement with the observations.

4.2 Evaluation of the hindcasts

Comparing the hindcasts (hereafter referred to as *H*) in Fig. 7 for the ocean fields and Fig. 8 for the atmospheric fields, all systems describe a stratified period at the beginning of the month with mixed layer depth around 20 m, except for Glo12 where the mixed layer becomes deeper for the same period (around 40 m). On 7 May all the systems simulate the beginning of the M1 mixing event, which reaches its maximum after 4 days but with significantly different amplitudes. Glo4 and Glo12 simulate mixed layer depth greater than 100 m, while Atl12 simulates only 85 m of mixed layer depth and Ibi36 even less so, with only 70 m depth. There is then a re-stratification event (S1), the largest with Glo12 and nothing with Glo4 where the mixed layer remains deeper than 100 m for 8 days. Figures 9 and 10 show the spatial pattern of the mixed layer depth for all systems for 13 and 16 May. In our area of interest (black squares on these figures) there is a strong gradient in the mixed layer depth with a mixed column in the northern part of the area, and a more stratified ocean in the south. In this case the mean profile in this box is not fully representative of the situation and the observation fails to capture this kind of pattern. Statistics computed over a smaller box (taking into account only the northern part of the box from 48.55° N to 48.8° N) are slightly different for the Glo4 system with a deeper M1 mixing event and a more stratified S1 event (not shown). But in this case the number of points in the box is too small for this low resolution system, and the statistical results in terms of bias or RMS values are not as good. In fact, the average applied over the 1/2° × 1/2° box is a small scale filtering which is efficient for the 1/12° or the 1/36° of degree system and consistent with the available observations, but filters no signal for the 1/4° system. Taking into account a larger box for this system could be a solution, but in this case the inconsistency with the available observations which are really concentrated in this small area will induce other biases. The three high resolution systems simulate this re-stratification event followed by a new strong mixing event (M2) for these 8 days (from 11 to 19 May in Fig. 7). The last period of the month is more similar in all systems, with a re-stratification of the entire water column (S2) from 20 to 25 May, a new mixing event (M3) followed by a re-stratification (S3). The temporal evolution of the mixed layer depth agrees well among all the systems with minima and maxima occurring on the same day except for the S1 stratification event in Glo4 between 11 and 13 May. Observations available at this position allow a precise validation of the evolution of the mixed layer during the month. As shown by the statistics, the Ibi36 system is the closest to observations with very good timing of mixing and re-stratification events and a good estimate of the mixed layer depth.

Over the first 2 days of the month the estimate of the mixed layer with Glo12 is the closest to the observations, with a mixed layer depth of around 40 m, while the other systems are more stratified with mixed layer depth around 20 m. The M1 event is too fast and too strong with Glo4 and Glo12 compared with the observations. It is closest to observation with the Ibi36 and Atl12 hindcasts. The S1 event, completely missed with Glo4, is observed and simulated with the other systems. The M2 event observed with a maximum of mixed layer depth on 17 and 18 May is well simulated with the Glo12, Atl12 and Ibi36 systems.

4.3 Discussion on the forecasts

4.3.1 Forecast of the 1st mixing event (M1)

The greatest forecast error is obtained with the Glo4 system during the M1 event. During this first period (between 9 and 12 May) the 1 and 2 day forecasts are consistent with the hindcast (green and blue dots with respect to the black line in Fig. 7) and so deeper than the observations, but the 3 and 4 day forecasts (red and purple dots in Fig. 7) are closer to the observations with a thinner mixed layer. At the beginning of the M1 event, the 4 day forecast misses the mixing. Looking more closely at the forecast for the 9 May (Fig. 7) no 4 day forecast (purple dots) simulates the mixing when the smaller forecast lengths (blue, green, yellow and red dots) capture this event. This is explained by the 4 day wind forecast which is less than the analysis wind (4 $\mathrm{m\,s^{-1}}$ rather than 13 $\mathrm{m\,s^{-1}}$; purple dots for 9 May in Fig. 8, top panels). We also observe the same kind of underestimation on the wind fields used for the Glo4 forecast for other dates and other forecast length (like 6 and 10 May for the 4 day length illustrated by the purple dots on Fig. 8, top left panel) which explains why the 4 day forecast gives too stratified a solution. These differences in the wind field used for the forecast are explained by the fact that in the operational suite all the systems are not launched at the same time. It is then possible to use the different base times of the atmospheric forecast for the ocean forecasts provided by the different systems used in this study (Glo12, Glo4, Atl12 and

Ibi36). As the Glo4 system is the first to be launched in the operational suite, if there is a delay in the atmospheric forcing construction procedure, this system will use the latest atmospheric forecast (using for example the previous analysis cycle). The other systems are able to forecast this mixing of the water column up to 4 days. Glo12 and Atl12 provide an excess of mixing especially for the 3 and 4 day forecast. Ibi36 is in better agreement with observation except for the 11 May where the observed mixed layer is deeper (a depth of 90 m but with high uncertainty) and the forecast, just as with the hindcast, gives too shallow a mixed layer (a depth of between 65 and 75 m).

4.3.2 Forecast of the 1st re-stratification (S1) and 2nd mixing (M2) events

As already discussed for the hindcast in Sect. 4.2, there is no more forecast of S1 re-stratification event with Glo4. Although the 3 and 4 day forecast seem to give good results, it is for the wrong reason; the initial state of these forecasts is too stratified and the strong wind event is not present in the atmospheric forecast. The other systems are able to forecast this re-stratification phase after the 12 May for each forecast length. During the second mixing event (from 12 to 17 May in the observation) the Glo4 forecast (especially from day 2 to day 4) provides a deep mixed layer, deeper than the hindcast and also deeper than the observations. The analysis of the area (Figs. 9 and 10) shows that all systems provide mixing of the water column from 13 to 16 May. This is true for the hindcasts (Figs. 9 and 10) and forecasts (not shown) but at a larger scale than the smaller $1/2^{\circ} \times 1/2^{\circ}$ box which contains the observations, and which is illustrated by the black box in the figures. At this small scale, meso-scale oceanic structures affect the mixed layer and a new source of uncertainty is added to the atmospheric forcing uncertainties. As observed in Figs. 9 and 10, similar large scale mixed layer depth patterns appear in all systems, with a north-south gradient with shallow mixed layer in the south (less than 50 m depth) and a deeper mixed layer in the northern part. Note that the figures show hindcast states and consequently the atmospheric uncertainty is reduced. At smaller scale, the effects of meso-scale, fronts, eddies and associated dynamics are represented by the contour of sea surface height in Figs. 9 and 10. In this case it is noticeable that the horizontal resolution of the system is a key factor in the effect on the mixed layer depth. In Glo4, at 1/4° resolution, there is less consistency between the mixed layer and the sea surface height fields; at 1/12° (in Glo12 and Atl12) and even more so at 1/36° (Ibi36) there are thin structures along fronts, surrounding eddies where the mixed layer is deeper. This influences the statistics when looking at small spatial and temporal scales, as in our case where the spatial scale is less than 50 km and the temporal scale is approximately 1 day. As mentioned in previous sections, this S1 to M2 period contains uncertainties for the mixed layer depth and also for the atmospheric forcing. It is linked to the following phenomena, which all contain uncertainties:

1. Error in the atmospheric forecast (see Fig. 8)
2. Rapid stratification/mixing change occurring over two days. In this case a short delay in the forecast gives a large error
3. M2 event occurs when the mixed layer is still thick; in the case of a shallow mixed layer, the uncertainty is naturally reduced.
4. There are well marked meso-scale structures which affect the mixed layer depth, generating vertical mixing associated with vertical velocities along the front and around eddies.

4.3.3 Forecast of the 2nd and 3rd stratification (S2, S3) and 3rd mixing (M3) events

The S2, M3, S3 time sequence is well forecast in all the systems, with good temporal consistency with observations (Fig. 7). Maximum stratification occurs on 25 May (S2). Then, the water column is mixed until 28 May (M3) and quickly re-stratified until the end of the month (S3). All the forecast lengths are close to the hindcast run except the 4 day forecast for 21 and 28 May. For these dates, all systems give consistent solutions with too rapid a re-stratification for 21 May and a lack of mixing for 28 May. This is explained by the error in the wind forecast (Fig. 8) taking into account a one or two-day lag, which is the typical time taken to mix the water column. For 19 and 20 May the forecast windspeed is too strong with wind speeds exceeding $10\,\mathrm{m\,s^{-1}}$, while analysis give values less than $10\,\mathrm{m\,s^{-1}}$ decreasing to $7\,\mathrm{m\,s^{-1}}$ for 20 May. The opposite occurs for 27 May with a wind forecast of approximately $10\,\mathrm{m\,s^{-1}}$ rather than the $14\,\mathrm{m\,s^{-1}}$ predicted by the analysis.

4.3.4 Atmospheric forcing vs. initial state in the uncertainties

The question of the significance or effect of atmospheric forcing vs. initial state on the mixed layer forecast has to be addressed. One diagnostic computed to quantify these two aspects separately is based on the temporal correlation between several time series. The first step is to compute the temporal correlation between the same forecast lengths with all the available systems. In this case the mean correlation decreases from 0.94 (for the Hindcast time series) to 0.91 (for the 4 day forecast time series). This small decrease in correlation indicates that the initial state has a small effect. In the second step the lag correlation between the Hindcast (H time series) and the Forecast (F0 to F4 time series) is computed independently for each system. In this case the mean correlation decreases from 0.98 (correlation between H and F0 time series with 1 day lag) to 0.83 (correlation between H and F4 with 5 day lag). Even though the correlation is still high, this stronger decrease indicates that atmospheric forcing has a greater effect in comparison with the initial state. A second diagnostic is based on the error growth computed with the standard deviation of the forecast error, normalized with the standard deviation of the observations (Fig. 11). For the atmospheric variables, the main error is due to the fresh water flux which does not drive the variability of the mixed layer depth in our case, as mentioned before. The normalized standard deviation becomes greater than 1, signifying that for the 1 day forecast the error variance is greater than the observation variance. For the wind field, which in this case is the more important, this ratio is smaller in comparison with the other fluxes (heat and fresh water fluxes).

The difference between the forecast over the entire month and that only over the mixing events (illustrated by the dashed line on the top panel in Fig. 11) is small except for the 4 day forecast. For the mixed layer forecast (bottom panel in Fig. 11) considering the entire period there is a small linear increase in the normalized standard deviation which generally remains less than 1 even for the 4 day forecast. The link with the error growth for the wind fields can be made by considering that the largest increase in the error for the 4 day forecast will have an effect on the longer length forecast of the mixed layer (typically for the 5 or 6 days which are not included in this study). Taking only the mixing events into account, the normalised standard deviation is stable for the first 3 days and then increases. The direct link with the increase in standard deviation of the wind field during these mixing events is not obvious. It should be noted that during the stratification events the normalized standard deviation for the mixed layer is greater than one. This is explained by the fact that, in a stratified ocean the error and the mixed layer depth have the same amplitude and a very small variation in the mixed layer gives rise to a large effect for this ratio. As we see in Figs. 9 and 10, there is also a strong spatial variability in the mixed layer which is not driven by atmospheric forcing, especially at small scale. Computing the spatial standard deviation in the small $1/2^{\circ} \times 1/2^{\circ}$ box for all the systems independently, we show that uncertainty at this small scale is as great, or even greater, than the uncertainty estimated as the standard deviation of all systems and all forecast lengths spatially filtered in the $1/2^{\circ} \times 1/2^{\circ}$ box. This standard deviation can reach 50 m to 60 m during the month but the available observations are insufficient to quantify this variability in small spatial scale. To understand the initial state differences, an experiment without data assimilation (Atl12 free) was performed and assimilation statistics between systems were compared. The Atl12 free experiment, driven by the best atmospheric forcing, simulates the mixing and stratification events (not shown); the timing of these events is in good agreement with the Atl12 simulation but the amplitude is quite different. The M1 event is too deep and S1 insufficiently stratified, rather the S2 stratification occurred more quickly and the M3 mixing is insufficiently deep. Statistical results are shown in Fig. 4 where we see that the correlation is still high (0.86), of the same order of magnitude as the 1 day forecast, and the RMS error is comparable with the 2 day forecast. However, the standard deviation is greater than all the Atl12 estimates, showing that data assimilation has a significant effect on the initial state and particularly the stratification which conditions the intensity of the mixing or stratification forecast. Figure 12 shows the SLA increment computed for the three systems (note that there is no data assimilation in the Ibi36 system, which is not presented here). Our area of interest (48.5° N and 16.2° W) is along a well marked front present in all analyses. Positive increments in the northern part and negative in the southern part are deduced from the analysis at 1/4° and 1/12° even though the spatial scales are different with increments containing more meso scale features at 1/12°. This front is more intense in the 1/12° solution and is northern in Glo12 by comparison with Atl12. These centimetre-scale differences affect the circulation and especially the circulation around meso-scale structures as can be seen in the daily mean for 13 and 16 May (Figs. 9 and 10). The temperature increment presented in Fig. 13 illustrates the correction computed on the temperature profile during May 2013 as a result of the data assimilation method. Differences between the three systems are noticeable. As already mentioned, the Atl12 system is the closest to observation with a positive increment around 0.1 °C at the surface and a negative increment of the same order of magnitude at 150 m depth. This correction tends to stratify the ocean (warming in the surface layer and cooling at the base of the mixed layer), as is expected given the previous results (Fig. 7). For the Glo12 system, the temperature increment is negative from the surface down to 150 m depth, but also with greater cooling at the base of the mixed layer than in the surface layer. The effect can be considered equivalent to that for the Atl12 system, neglecting the bias. In Glo4, the increment profile is quite different: in the top first 30 m there is a cooling of the mixed layer and then increments restratify the ocean from 30 m to 150 m just as in the other systems. The dashed line in Fig. 13, computed as the standard deviation of the five increments available for May (we recall that the analysis cycle is one week, and in this case we use the 5 analyses using observations for May 2013), illustrates the large variability in this increment during this month. This might be expected because of the rapid strong mixing and restratification events observed during this month. The conclusion of this part is that evidence of the link between the wind and the mixed layer forecast is clearer than for the initial state in a complex and non-linear operational system. However, with the Atl12 free simulation we have quantified the effect of data assimilation on the initial state including mesoscale processes and ocean stratification. Model physics (vertical mixing scheme) and resolution (from 1/4° to 1/36°) also play a crucial role; they have been discussed and their effects quantified in terms of the statistics generated by the operational system available.

5 Conclusions

This study focuses on a small area in the North East Atlantic during May 2013. Several conditions are met to obtain robust results:

1. A large number of temperature profiles (74) in a small area with a high sampling frequency over the month (more than one per day).
2. Available daily forecasts with four operational oceanic systems containing differences as horizontal resolution from 1/4° to 1/36°, initialization method, vertical mixing scheme, atmospheric forcing, etc.
3. A strong variability in the mixed layer depth during the month with alternating mixing and stratification events.
4. A strong link between atmospheric forcing and ocean response.

As a result of all these conditions, we have shown how operational oceanic systems can provide a mixed layer forecast, and we have quantified the quality of these forecasts with commonly used diagnostics. The mean bias of the mixed layer depth forecast over the month is around a few metres (usually less than 5 m) and is quite stable with the forecast length; the mixed layer depth RMS error increases with the forecast length but remains less than 20 m. The temporal correlation between observation and forecast is usually greater than 0.85 and slowly decreases with forecast length. The skill score shows the benefit of comparing the forecast with the persistence. These statistics are also useful in comparing the performance of the systems from the best to the worst in terms of forecast ability. In our case we have shown that Glo4, which is the system with the lowest resolution, gives the worst results and Ibi36, which has the best resolution, gives the best results close to the Atl12 system. This paper concentrates on temporal variability since with the observations available it is not possible to estimate a spatial distribution of the mixed layer depth. We have shown that temporal variability is mainly driven by atmospheric forcing (especially the wind field) and that the model forecast is often close to the observations with good agreement of the temporal sequence of the mixing and stratification events in the observations and forecasts. Note that a ~ 2 day lag between a strong wind event and the maximum of mixed layer depth is observed, and consequently missing this event on the first day of the wind forecast generates an error in the mixed layer depth forecast.

The availability of four systems providing daily forecasts gives the opportunity to build an ensemble forecast associated with an estimate of the uncertainty of the mixed layer depth. These systems have been developed by Mercator Océan under the MyOcean project, the ocean part of the European Copernicus programme, and have been operated in real time since the end of April 2013 (V3 of MyOcean service). Other ocean forecast products could also have been used to increase the number of members in the ensemble, but for this study we chose to use only these 4-forecasts to separate the effects of atmospheric forcing and initial state. First results show the benefit of the mean or the median of the members as forecast. In our case this ensemble estimate is close to the best forecast, and sometimes this estimate is the best (for example the best correlation for the 1 day forecast is obtained with the median state and with the mean for the 4 day forecast). Computing the same statistics, removing each individual forecast one by one, is a good way to estimate each contribution in the ensemble. We have shown that after removing the worst forecast, which systematically degraded the mixed layer depth estimation, the mean is always better than each individual forecast for every forecast length. Using other operational forecasts, it will be now useful to introduce into the ensemble ocean estimates computed with other atmospheric forecasts, as for example, the product available in MyOcean provided by the UK's Met Office and covering the North West shelf (O'Dea et al., 2012), or other global high resolution forecasts such as that provided by NRL (Cummings, 2005). Uncertainty estimates in the mixed layer in this area based on our 4-forecast systems and 4 day forecast length can reach 50 m during this particular month. The spatial uncertainty for the model in such a small area has the same order of amplitude (~ 50 m). Using the available data an uncertainty of 50 m was also estimated on several dates, though the number of observations might be insufficient to compute a robust level of uncertainty. We have also shown that there is a direct link between the atmospheric uncertainty (especially the wind field) and the mixed layer depth.

Finally we have shown that the temporal variability in the mixed layer depth when changing from the mixing to the stratification phase is driven by the atmospheric focing, but the small and meso ocean scales also have a great local impact. At this smaller scale, resolution, parameterization and assimilation play a role and can impact the forecast score, error or uncertainty. The effect of horizontal circulation, particularly around eddies or along strong fronts, have been illustrated for the model mixed layer. Unfortunately, based on observations this mixing along fronts and around eddies remains difficult to validate properly, The coverage of the in situ observations and the resolution of satellite observations are not sufficient even though the recovery of vertical velocity based on satellite observations is promising (Buongiorno et al., 2012) and though observations of water colour provide high resolution estimates of ocean parameters directly affected by the vertical mixing. Future development of the operational oceanic forecast systems will be crucial in improving forecasts of oceanic parameters or processes such as the mixed layer depth. Within the scientific community, work is in progress to include new types of observation (such as ocean colour and, in the near future, SWOT high resolution sea surface height observations), to increase horizontal and vertical resolution, to improve vertical mixing models and parameterization, to improve ocean and atmosphere interaction due to coupling and to provide better estimates of the uncertainties based on ensemble technics.

Acknowledgements. This research was supported by the MyOcean2 European project and is based on MyOcean products. The authors wish to thank collaborators contributing to the development of the ocean forecasting systems under the MyOcean project, the NEMO consortium and the data centres at CORIOLIS and BODC which disseminate the in situ glider observations collected under the OSMOSIS project.

References

Buongiorno Nardelli, B., Guinehut, S., Pascual, A., Drillet, Y., Ruiz, S., and Mulet, S.: Towards high resolution mapping of 3-D mesoscale dynamics from observations, Ocean Sci., 8, 885–901, doi:10.5194/os-8-885-2012, 2012.

Cailleau, S., Chanut, J., Lellouche, J.-M., Levier, B., Maraldi, C., Reffray, G., and Sotillo, M. G.: Towards a regional ocean forecasting system for the IBI (Iberia-Biscay-Ireland area): developments and improvements within the ECOOP project framework, Ocean Sci., 8, 143–159, doi:10.5194/os-8-143-2012, 2012.

Cummings, J. A.: Operational multivariate ocean data assimilation, Q. J. Roy. Meteor. Soc., 131, 3583–3604, 2005.

De Boyer Montégut, C., Madec, G., Fischer, A. S., Lazar, A., and Iudicone, D.: Mixed layer depth over the global ocean: an examination of profile data and a profile-based climatology, J. Geophys. Res., 109, C12003, doi:10.1029/2004JC002378, 2004.

Drévillon, M., Régnier, C., Levier, B., Desportes, C., and Perruche, C.: Quo va dis? Quarterly Ocean Validation Display #13. Validation bulletin for April-May-June (AMJ), available at: http://www.mercator-ocean.fr/fre/science/qualification/QUOVADIS_num13_AMJ_2013 (last access: 9 June 2014), 2013.

Drillet, Y., Garric, G., Le Vaillant, X., and Benkiran, M.: The dependence of medium range northern Atlantic Ocean predictability on atmospheric forecasts, Journal of Operational Oceanography, 2, 43–55, 2009.

Lellouche, J.-M., Le Galloudec, O., Drévillon, M., Régnier, C., Greiner, E., Garric, G., Ferry, N., Desportes, C., Testut, C.-E., Bricaud, C., Bourdallé-Badie, R., Tranchant, B., Benkiran, M., Drillet, Y., Daudin, A., and De Nicola, C.: Evaluation of global monitoring and forecasting systems at Mercator Océan, Ocean Sci., 9, 57–81, doi:10.5194/os-9-57-2013, 2013.

Lenartz, F., Mourre, B., Barth, A., Beckers, J. M., Vandenbulcke, L., and Rixen, M.: Enhanced ocean temperature forecast skills through 3-D super-ensemble multimodel fusion, Geophys. Res. Lett., 37, L19606, doi:10.1029/2010GL044591, 2010.

Lengaigne, M., Menkes, C., Aumont, O., Gorgues, T., Bopp, L., and Madec, J.-M. A. G.: Biophysical feedbacks on the tropical pacific climate in a coupled general circulation model, Clim. Dynam., 28, 503–516, 2007.

Lermusiaux, P. F. J., Chiu, C.-S., Gawarkiewicz, G. G., Abbot, P., Robinson, A. R., Miller, R. N., Haley, P. J., Leslie, W. G., Majumdar, S. J., Pang, A., and Lekien, F.: Quantifying uncertainties in ocean predictions, Oceanography, 19, 92–105, 2006.

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., Ayoub, N., De Mey, P., Reffray, G., Lyard, F., Cailleau, S., Drévillon, M., Fanjul, E. A., Sotillo, M. G., Marsaleix, P., and the Mercator Research and Development Team: NEMO on the shelf: assessment of the Iberia–Biscay–Ireland configuration, Ocean Sci., 9, 745–771, doi:10.5194/os-9-745-2013, 2013.

Morel, A., Huot, Y., Gentili, B., Werdell, P. J., Hooker, S. B., and Franz, B. A.: Examining the consistency of products derived from various ocean color sensors in open ocean (Case 1) waters in the perspective of a multi-sensor approach, Remote Sens. Environ., 111, 69–88, 2007.

Murphy, A. H.: Skill Scores Based on the Mean Square Error and their relationships to the correlation coefficient, Mon. Weather Rev., 116, 2417–2424, doi:10.1175/1520-0493(1988)116<2417:SSBOTM>2.0.CO;2, 1998.

O'Dea, E. J., Arnold, A. K., Edwards, K. P., Furner, R., Hyder, P., Martin, M. J., Siddorn, J. R., Storkey, D., While, J., Holt, J. T., and Liu, H.: An operational ocean forecast system incorporating NEMO and SST data assimilation for the tidally driven European Northwest shelf, Journal of Operational Oceanography, 5, 3–17(15), 2012.

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

Pistoia, J.: Development of SuperEnsemble Techniques for the Mediterranean ocean Forecasting System. Alma Mater Studiorum Università di Bologna, Dottorato di Ricerca, in: GEOFISICA Ciclo XXIII Settore Scientifico-Disciplinare di afferenza: GEO/10, 2012.

Scott, R., Ferry, N., Drevillon, M., Barron, C. N., Jourdain, N. C., Lellouche, J.-M., Metzger, E. J., Rio, M.-H., and Smedstad, O. M.: Estimates of surface drifter trajectories in the Equatorial Atlantic: a multi-model ensemble approach, Ocean Dynam., 62, 1091–1109, doi:10.1007/s10236-012-0548-2, 2012.

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.

Vandenbulcke, L., Beckers, J.-M., Lenartz, F., Barth, A., Poulain, P.-M., Aidonidis, M., Meyrat, J., Ardhuin, F., Tonani, M., Fratianni, C., Torris, L., Chiggiato, J., Tudor, M., Book, J. W., Martin, P., Peggion, G., and Rixen, M.: Super ensemble techniques: application to surface drift prediction, Prog. Oceanogr., 82, 149–167, doi:10.1016/j.pocean.2009.06.002, 2009.

Xue, Y., Balmaseda, M. A., Boyer, T., Ferry, N., Good, S., Ishikawa, I., Kumar, A., Rienecker, M., Rosati, A. J., and Yin, Y.: A comparative analysis of upper-ocean heat content variability from an ensemble of operational ocean re-analyses, J. Climate, 6905–6929, doi:10.1175/JCLI-D-11-00542.1, 2012.

Zhu, J., Huang, B., and Balmaseda, M. A.: An ensemble estimation of the variability of upper-ocean heat content over the tropical Atlantic Ocean with multi-ocean reanalysis products, Clim. Dynam., 39, 1001–1020, 2012.

Table 1. Main characteristics of the ocean forecasting systems.

System Reference	Glo4 PSY3QV3R3	Glo12 PSY4QV2R2	Atl12 PSY2QV4R4	Ibi36 IBI36QV2R1
Nemo	NEMO3.1			NEMO2.3 including specific development for regional/coastal application
Horizontal resolution	1/4° (∼ 20 km)	1/12° (∼ 6.5 km)	1/12° (∼ 6.5 km)	1/36° (∼ 2.2 km)
Vertical resolution	50 z vertical levels with partial step. 1 m at the surface. 22 levels in the upper 100 m.			
Atmospheric forcing	ECMWF operational analysis and forecast, spatial resolution ∼ 12 km and 3 h temporal frequency. CORE Bulk formulation is used to compute atmospheric stress and fluxes.			
Atmospheric grid	Interpolated on 1/4° grid			Interpolated on 1/12° grid
Solar flux penetration	3-band parameterization for short-wave radiation (Lengaigne et al., 2007)			2-band parameterization for short-wave radiation (Morel et al., 2007)
Vertical mixing	TKE vertical mixing			GLS vertical mixing
Free surface	Filtered free surface			Explicit free surface with time splitting and tide
Initialization	SAM2 assimilation scheme (based on SEEK filter) assimilating SLA along track, SST and in situ temperature and salinity profiles			Initialization with Atl12 analysis and 2-week spin-up.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◄ ►
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

Table 2. RMS error in metres for the mixed layer depth computed with the systems, the mean value and the mean after removing one system. The F0, F2 and F4 forecast lengths are shown. For each forecast length the best forecast is bold underlined, and the other forecast with error not greater than 1 m compared with the best is shown in bold.

System	Ibi36	Atl12	Glo4	Glo12	Mean	M-Ibi36	M-Atl12	M-Glo4	M-Glo12
F0 RMS error	**15.5**	**16.0**	27.4	19.8	17.0	18.7	17.8	**15.3**	17.6
F2 RMS error	**16.5**	18.1	29.4	21.6	18.2	20.2	19.1	**16.7**	18.7
F4 RMS error	**18.8**	19.1	29.8	23.6	**18.3**	19.7	19.3	**17.8**	**18.4**

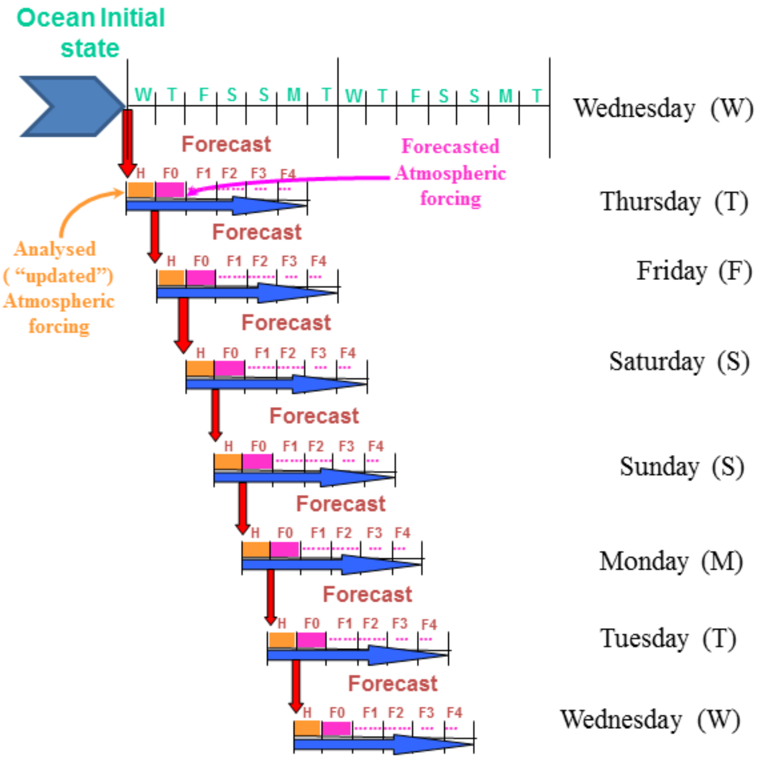

Figure 1. Operational scheme for producing daily forecasts with all the Mercator Océan systems. The ocean initial state is produced once a week on Wednesdays. Then, starting from this state, a hindcast (*H*) is produced each day using analysed atmospheric forcing. Then the forecast for the current day (F0) up to 4 day forecasts (F4) are performed daily, forced by the atmospheric forecasts.

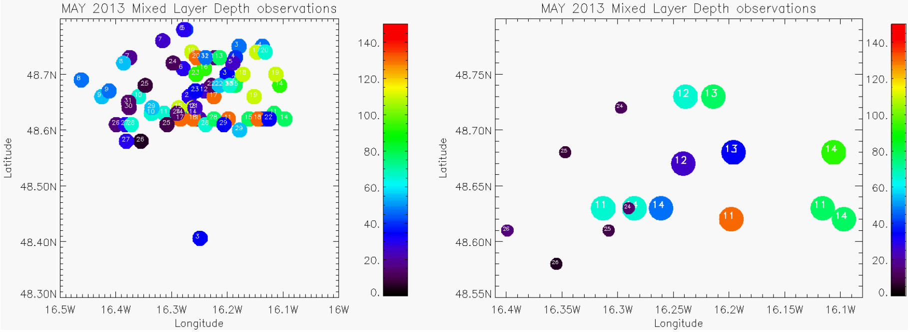

Figure 2. Left panel, position of the 74 profiles available in the area during May 2013. Right panel, selection of profiles from 11 to 14 May during the M1 mixing event and the first S1 re-stratification phase (large circles) and from 24 to 26 May during the S2 stratification phase (small circles). Colours show the mixed layer depth computed for each profile with the 0.2 °C criterion. The number inside of the circles gives the day of the measurement.

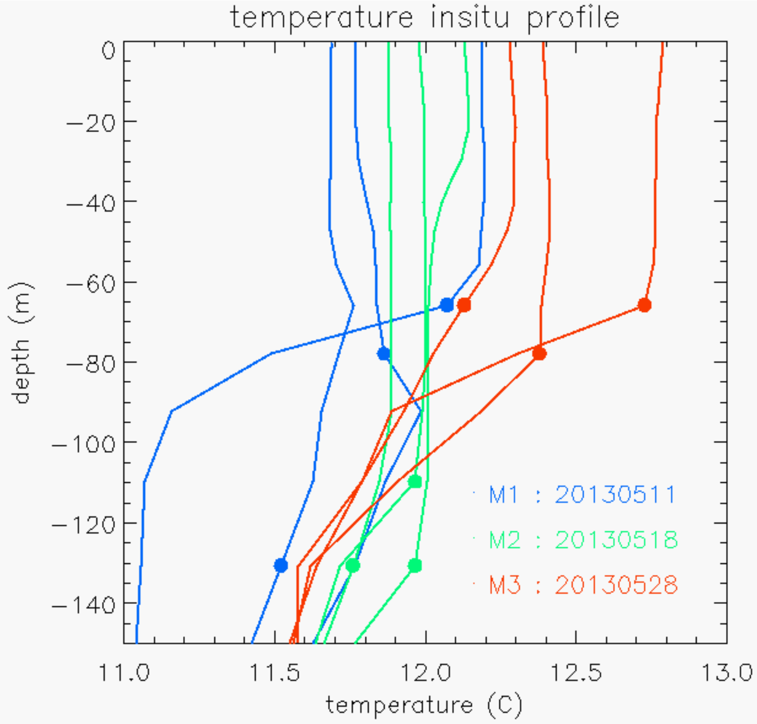

Figure 3. Available in situ profiles for 3 dates corresponding to 3 mixing events (M1, M2 and M3) during May 2013. Note that for these three dates 3 temperature profiles are available, with their geographical positions shown in Fig. 2. The circles indicate the mixed layer depth computed using the 0.2 °C criterion.

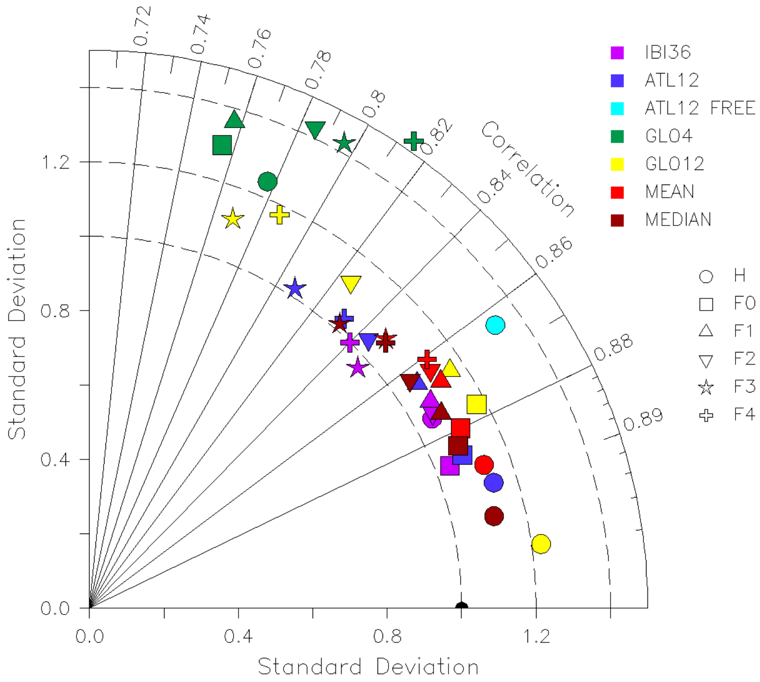

Figure 4. Taylor diagram comparing all available systems (in colour) and forecast lengths (symbol). The black dot with a standard deviation equal to 1 and a correlation of 1 indicates observations.

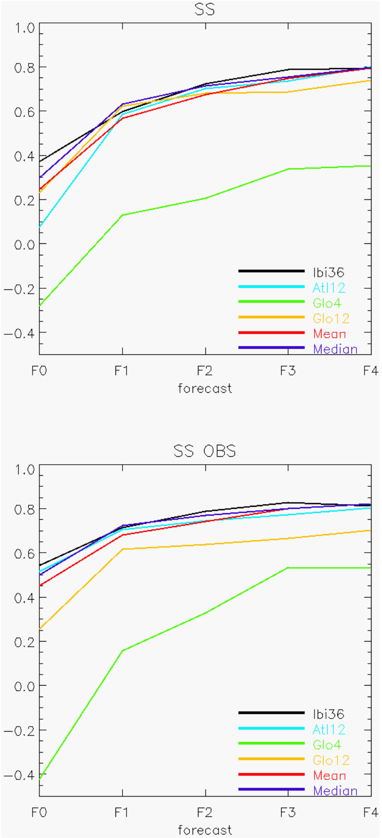

Figure 5. Skill score for the mixed layer depth computed for all the systems and the ensemble mean and median during May 2013. In the top panel the skill score is computed with the persistence of the analysis, and in the bottom panel with the persistence of the observation.

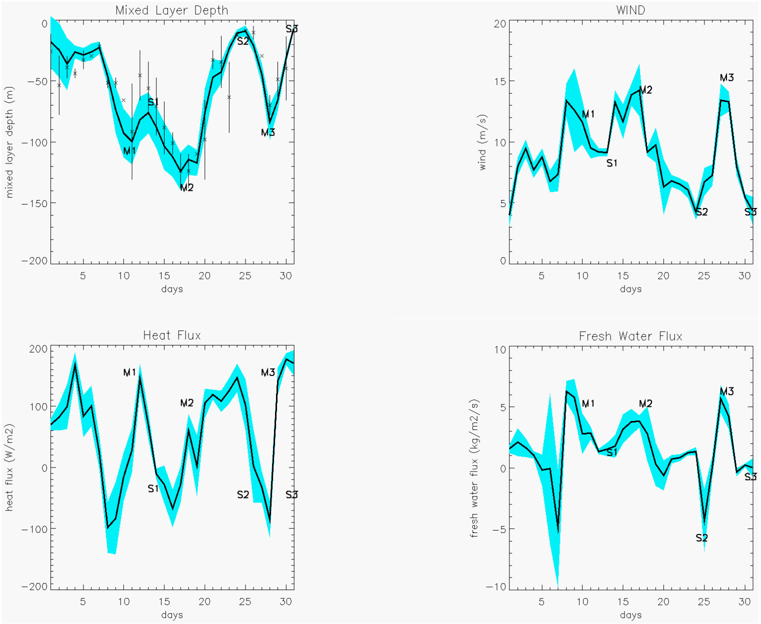

Figure 6. Top left: temporal evolution of the mixed layer simulated by the ensemble with the standard deviation in blue, and observations with associated uncertainties. Top right: wind speed time-series: analysis in black and ±1 standard deviation in blue computed with all the forecast lengths; note that all systems are assumed to be using the same wind speed field, though an exception can occur if a forecast using one system is launched before atmospheric forcing is updated in the real time production. Bottom left: total heat flux time-series, analysis in black and ±1 standard deviation in blue computed with all forecast lengths and with all systems negative flux means that ocean gets heat. Bottom right: fresh water flux time-series, analysis in black and ±1 standard deviation in blue computed with all forecast lengths and with all systems. The fresh water flux includes evaporation minus precipitation and runoff, a negative flux means that ocean gets fresh water.

Title Page
Abstract Introduction
Conclusions References
Tables Figures
◀◀ ▶▶
◀ ▶
Back Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

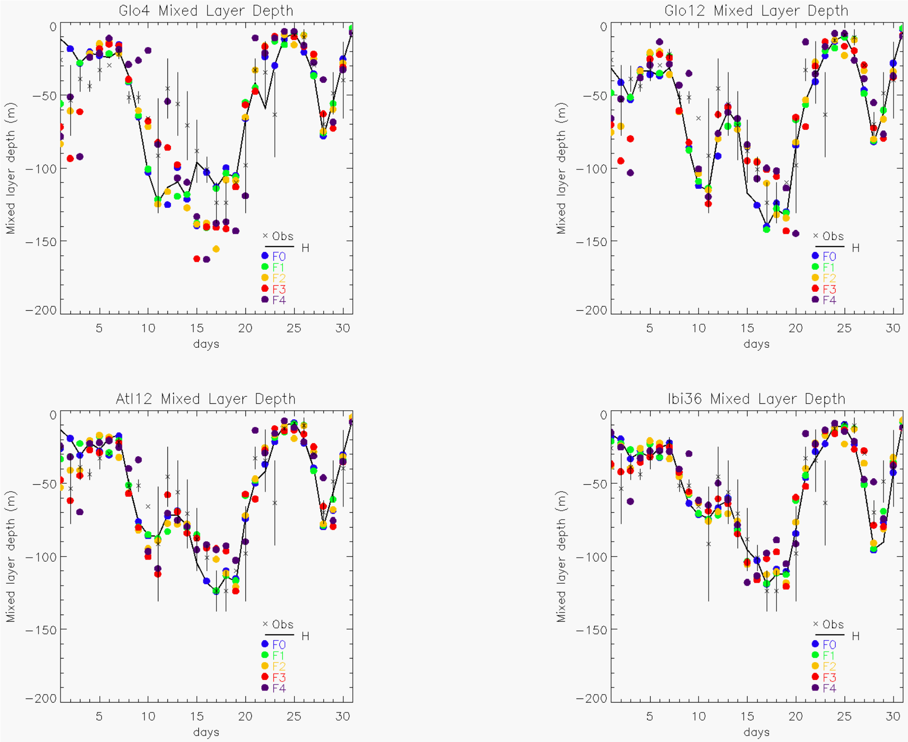

Figure 7. Mixed layer depth evolution during May 2013. The black line is the hindcast and the coloured dots are the forecasts for several forecast lengths. The crosses are the means of the observations and the vertical black lines are error bars computed with the min and max values of the MLD estimated by the profiles during the day.

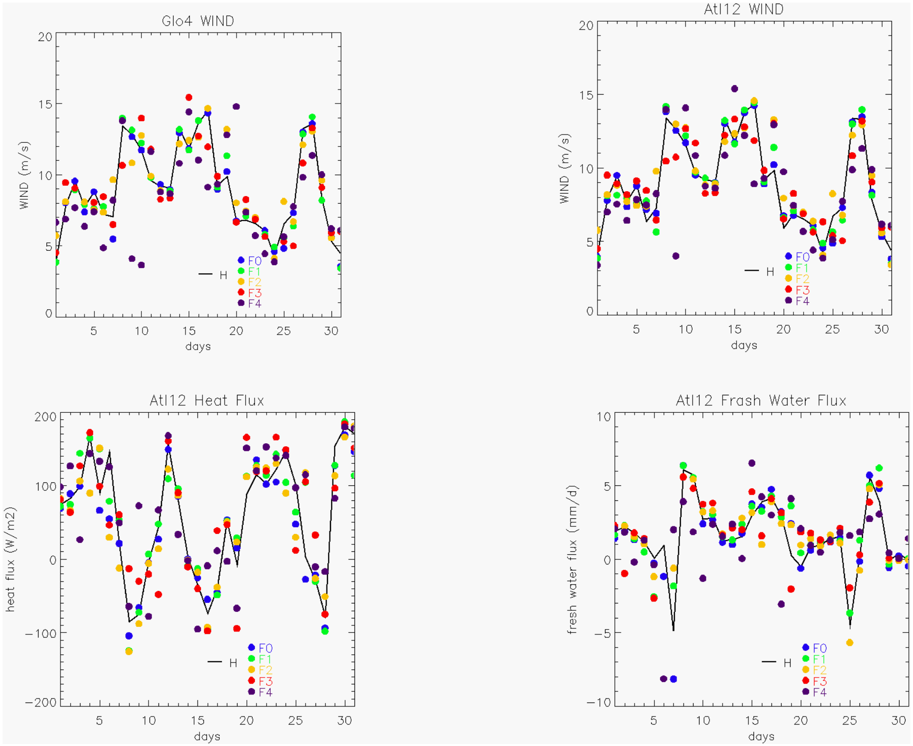

Figure 8. Temporal evolution of atmospheric forcing for hindcast (black line) and forecasts (coloured dots). Top panels: evolution of wind speed for Glo4 (left) and Atl12 (right) systems. Bottom panels: heat flux (left) and fresh water flux (right) for the Atl12 system.

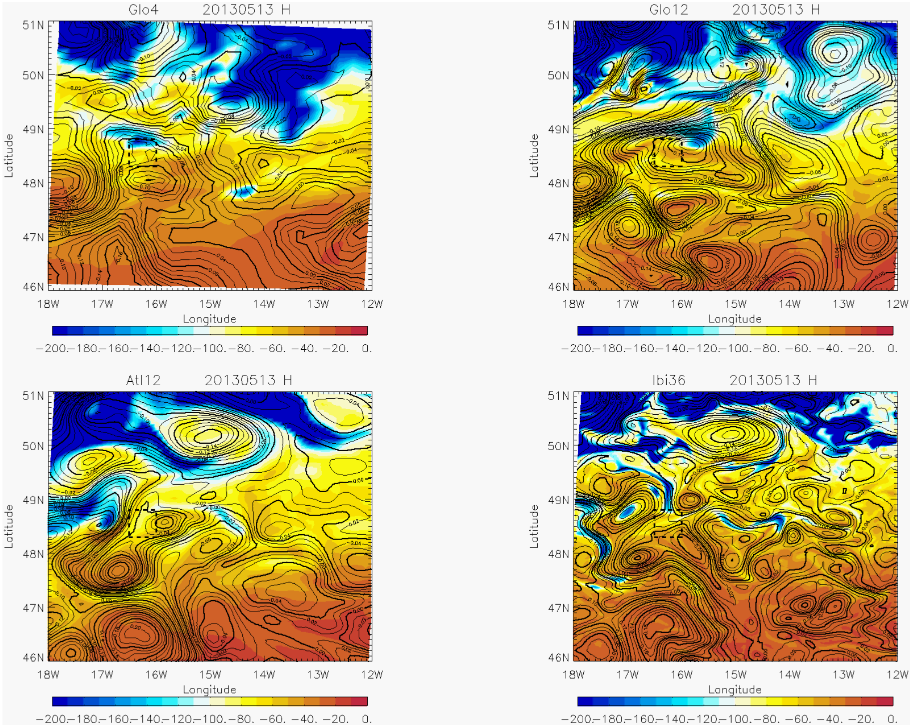

Figure 9. Mixed layer depth (colour field) and sea surface height (contours) simulated by the four systems for 13 May. The black dotted box shows the area in which the statistics are computed using the models and observations.

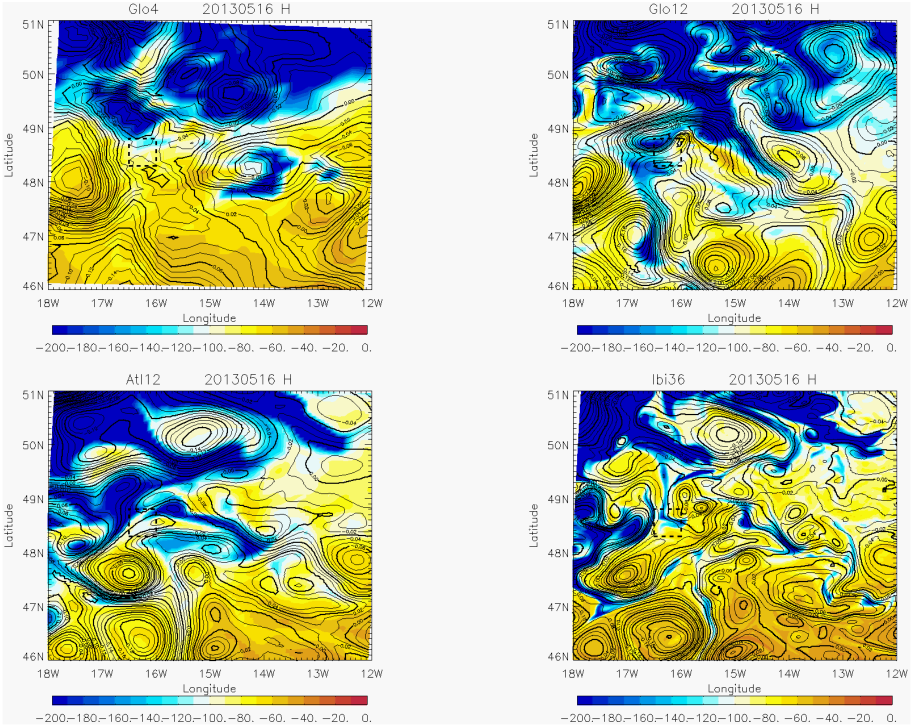

Figure 10. Same as Fig. 9 for 16 May.

OSD
11, 1435–1472, 2014

Forecasting the mixed layer depth in the north east Atlantic

Y. Drillet et al.

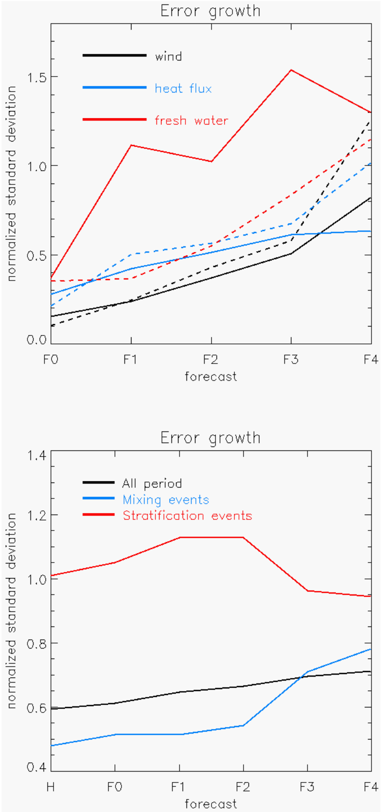

Figure 11. Standard deviation of the forecast normalised by the standard deviation of the observations. Top panel: atmospheric fields (wind in black, heat flux in blue and fresh water in red) where analyses are considered as observations. The solid line is for May 2013 and the dashed line considers only the mixing event. Bottom panel: ocean mixed layer depth forecast (for all the systems), in black for May 2013, in blue only during the mixing event and in red during the stratification event.

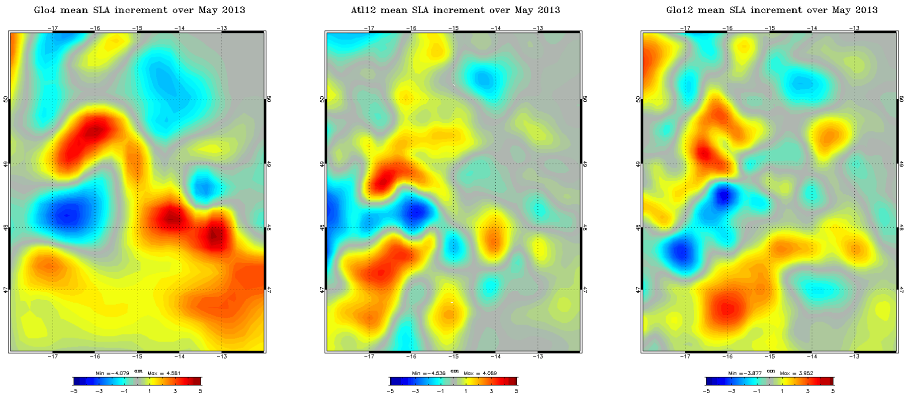

Figure 12. Glo4, Atl12, Glo12. Mean SLA increment computed over May 2013 for GLo4, Atl12 and Glo12 systems.

Title Page
Abstract
Introduction
Conclusions
References
Tables
Figures
◀|
|▶
◀
▶
Back
Close
Full Screen / Esc
Printer-friendly Version
Interactive Discussion

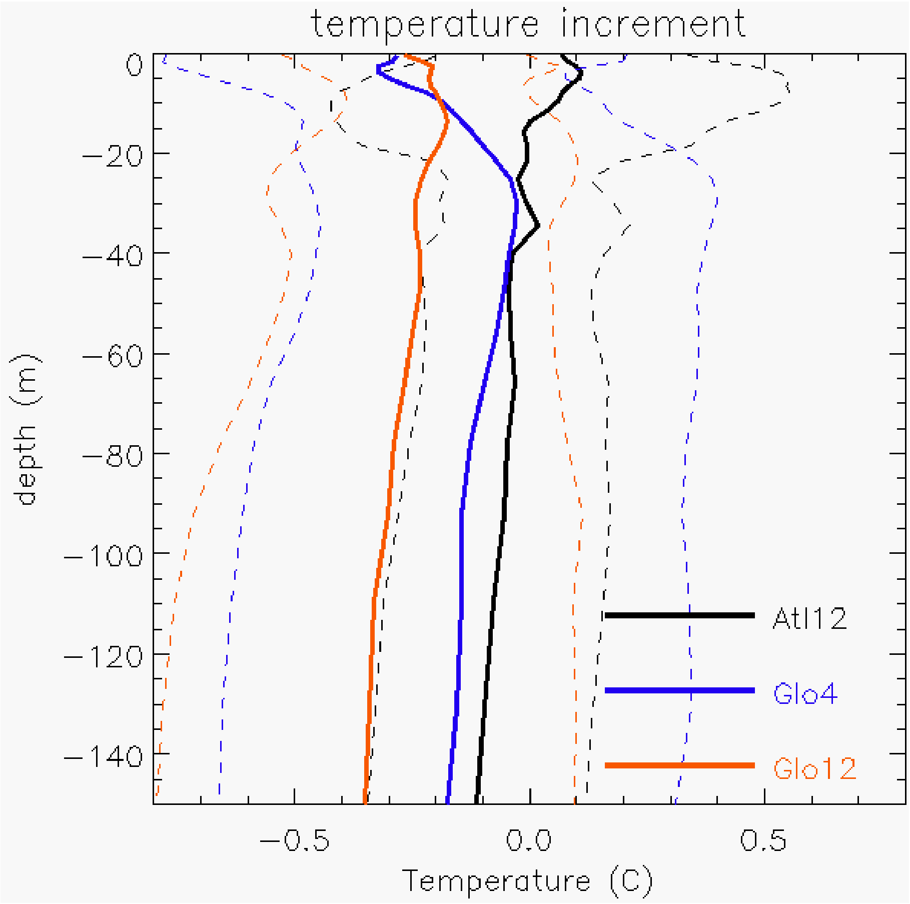

Figure 13. Mean temperature increment (solid line) and standard deviation (dashed line) for May for the three systems (Glo4 in blue, Atl12 in black and Glo12 in red).