Interactive comment on "Evaluation of real time and future global monitoring and forecasting systems at Mercator Océan" by J.-M. Lellouche et al.

Anonymous Referee #1

Received and published: 27 March 2012

The authors present a description of two analysis/forecasts systems developed at Mercator Ocean. They then present a series of comparisons between analyses and observations. I find the description of the model and data assimilation systems too brief to benefit the reader. Many important details are left out. I identify specific details that I would like to know below. The evaluation of the model is generally appropriate. However, the authors mostly present comparisons between analyses and observations, not forecasts and observations. The comparisons that do include forecasts (eg Figure 15) are generally not favorable. I would like to see more comparisons with the forecasts. The English is poor throughout the paper, often very loose, and occasionally misleading. I include some examples below, but there are many more. I recommend that a native English speaker review this document. This paper is worthy of publication, but several aspects of the paper and analysis, detailed below, should be considered first.

We thank anonymous Referee #1 for his careful reading of our manuscript and for his constructive remarks. Following his advices, we tried to make the manuscript clearer and more detailed, particularly for the description of the model and the data assimilation system. A native English speaker will review the document next week. All remarks detailed below by the referee were considered and/or discussed.

1. The most concerning aspect of this paper is the data assimilation. The authors say "the analysis is performed on a reduced horizontal grid (1 point every 4 in both directions)…" (page 1129, line 24). They also state that they "use a weighting function which sets the covariances to zero beyond a distance defined as twice the local spatial correlation scale" (page 1129, line 16). The correlation scales are shown in Figure 2. Both the zonal and meridional length-scales are less than 100 km for most of the globe. I presume a smooth localizing function is used (although this detail is not included in the paper), so even data that is 100 km from a grid point is down-weighted. So, for the 1/4° degree model, analyses are computed on a 1 degree grid; and data influence is set to zero beyond about 200 km, and down-weighted 100 km away (i.e., the next grid point in latitude). It sounds like this configuration is close to one-dimensional assimilation for each grid point. The horizontal spatial structures of the ensemble appear to be almost completely discarded by the localization. The authors might consider commenting on this in a revised version of the paper.

The reviewer's comment is entirely relevant. We fully agree with the fact that, in theory, both the zonal and meridional length-scales are not large enough for the ¼° model (IRG systems). However, they are correct for the 1/12° model (HRZ systems).

The problem raised by the reviewer may occur mainly at mid latitudes. Indeed, the radii are larger in the equatorial band and the distance between two points of analysis varying with the cosine of the latitude, the problem is reduced at high latitudes. Another detail, now mentioned in the paper, is that the analysis grid contains all the points near the coast and a point every two in the first 150 km from the coast. The problem mentioned by the reviewer does not penalize therefore the coastal analyses.

A one-week test simulation was also performed by imposing radii larger than 200 km (instead of 75 km in the IRG_DEV reference run). The increments obtained with both simulations have globally the same structures. Differences concern some extrema linked to the assimilation of temperature and salinity bad or isolated profiles 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. We initially tried to find the best compromise between getting the "better analysis" and having a sustainable system for operational use. The choice of the size of the radii is finally imposed by the cost of the system in an operational context. We are grateful to Referee #1 who definitely pointed out a flaw in the conception of the 1/4° system. This flaw does not alter the mean conclusions of the article but it will be corrected in further versions of the system.

2. Specific comments

2.1. Data assimilation: Page 1129, line 6: Can the authors be more precise about how they construct the anomalies that underpin their data assimilation? They describe it as "anomalies … with respect to a running mean so that they can give and estimate of the 7-day scale error…" To be of value to the ocean forecasting community, a precise formulation/definition would be beneficial.

We added a paragraph and a figure in order to explain more precisely how we construct the anomalies.

The sentence page 1129, lines 5-9 "In our case, the anomalies are computed from a long numerical experiment (typically around 10 yr) with respect to a running mean so that they can give an estimate of the 7-day scale error on 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)."

is completed by:

More precisely, each temporal anomaly M' corresponds to the difference between the model state M and a running mean $\langle \mathrm{M} \rangle_{-\tau}^{+\tau}$ over a fixed time window period which is ranged from $-\tau$ to τ (see Fig. 1a). Moreover, the signal at a few horizontal grid interval "Δx" in the model outputs on the native full grid is not physical but only numerical (Grasso, 2000). It is not relevant from an analysis point of view to consider and use this signal to build the analysis update. That's why several passes of a Shapiro filter are applied in order to remove the very short scales corresponding in practice 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 point every 2 in both horizontal directions and all the points along the coast) to limit the storage and the load cost at the analysis stage.

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

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

where ν are the temporal frequencies of the model state and $\nu_{\max}$ is the cut-off frequency (equal to 1/36 days^{-1} in our case). The main characteristic of the anomaly calculation is to filter out temporal scales at low frequencies in order to keep high frequencies for which the period is shorter than the assimilation cycle. For an assimilation cycle centred on the Nth day of a given year, ocean state anomalies falling in the window [N-Δn; N+Δn] of each year of the free run are gathered and define the covariance of the model forecast error (see Figure 1b). In our case, Δn is equal to 60 days, which means that anomalies are selected over 120-day length windows centred on the Nth day of each year of the free run. So in SAM, the forecast error covariances rely on a fixed basis seasonally variable ensemble of anomaly. This method implies that at each analysis step a sub-set of anomalies is used that improves the dynamical dependency. A significant number of anomalies are kept from one analysis to the other, ensuring error covariance continuity.

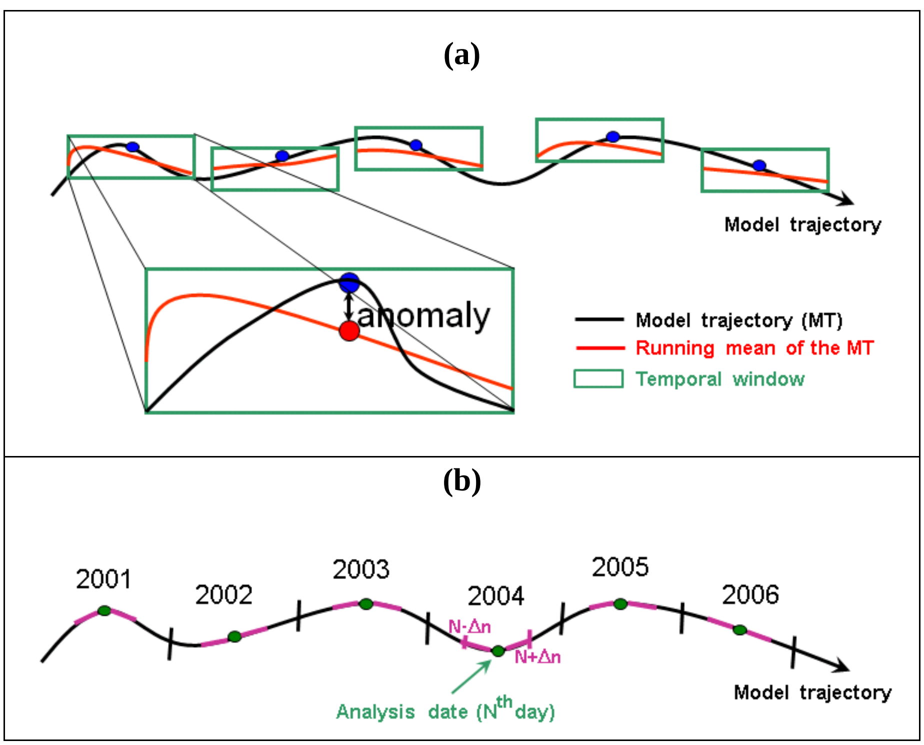

Fig.1. Schematic representation of the anomalies calculation along a model trajectory **(a)** and of the use of these anomalies to build the model forecast covariance **(b)**.

Reference: Grasso, L. D.: The differentiation between grid spacing and resolution and their application to numerical modelling, Bull. Amer. Meteor. Soc., 81, 579-580, 2000.

2.2. Page 1129, line 22: How are the correlation length- and time-scales computed? What are they based on?

For 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 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. Scales are computed from the temperature at 100 m and 300 m. At every point, space and time lag correlations are computed with neighbour points. The distances where 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, we increase the radii in order to remain the same order of magnitude on average as the IRG_V1V2 system radii.

2.3. Page 1129, line 24: What is the impact of producing analyses on a 1 degree grid instead of the native 1/4 degree grid of the model?

The anomalies are computed on a reduced horizontal grid (1 point every 2 in both horizontal directions and all the points along the coast). To save computing time, the native grid is sub-sampled during the analysis (1 point every 4 in both horizontal directions, all the points along the coast and 1 point every 2 in the first 150 km from the coast). Using more points (1 point every 2 in both horizontal directions and all the points along the coast) for the analysis grid does not bring any significant differences.

2.4. Page 1130, line 4: What are the details of the "balance operator". I presume a geostrophic balance is involved – what happens at the equator?

The sentence "A physical balance operator allows to deduce from these increments a physically consistent sea surface height increment." is not true for the systems discussed in this paper. This was the case with old systems. It is thus a mistake to have said that. For the systems discussed in this paper, the SSH increment is the sum of barotropic and dynamic height increments. Dynamic height increment comes from the temperature and salinity increments. In this calculus, there is no problem at the equator.

2.5. Page 1130, line 7: I understood by oral presentations from researchers from Mercator Ocean, that the initialization approach was a "double backwards analysis increment". Has that been abandoned?

This initialization approach has not been abandoned. We just upgraded the initialization of the model in order to better correct the model in the early days of the simulation.

The sentence page 1130, lines 6-9 "All these increments are applied progressively thanks to the Incremental Analysis Update (IAU) method (Bloom et al., 1996; Benkiran and Greiner, 2008) allowing avoiding model shock every week due to the unbalance between the analysis increments and the model physics."

is completed by:

This way, the IAU reduces spin up effects. It is fairly similar to nudging but it does not exhibit its 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 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 7 days (referred to as "BEST" second trajectory) with a tendency term added in the model prognostics 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 repartition function shown in Fig. 2. 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.

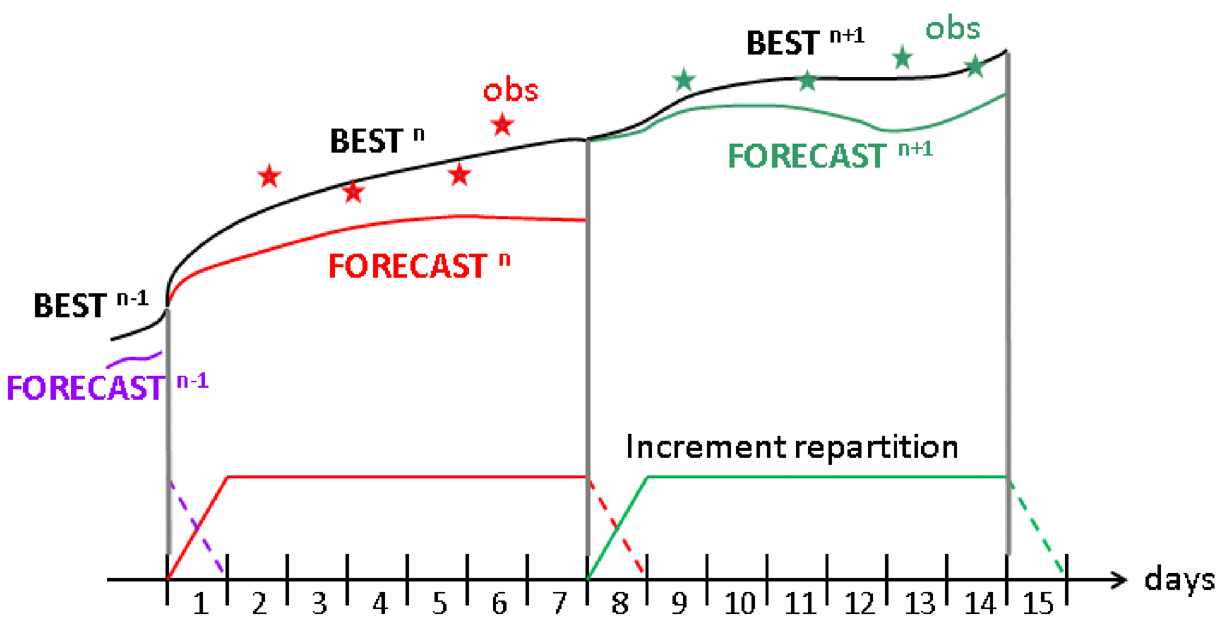

Fig.2. Schematic representation of the IAU procedure for three consecutive cycles n-1, n and n+1.

2.6. Page 1130, line 10: Is there a reference for the bias correction? The details in this paper are insufficient to stand alone.

We added a paragraph in order to explain more precisely how the bias correction works.

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° x 1°) grid. The two variables are treated separately because temperature and salinity biases are not necessarily correlated. Then, the 3D-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 error bias is large in the regions of sharp gradients) and is modelled by means of an anisotropic Gaussian 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.

2.7. Page 1130, line 23: The concept of "pseudo-observations" is clear. Where and when are these applied? Based on what criteria?

The sentence page 1130, lines 22-27 "The concept of "pseudo-observations" or "Observed-No Change" (innovation equal to zero) has also been introduced to overcome the deficiencies of the background errors, in particular for extrapolated and/or poorly observed variables. We apply this kind of parameterization on the barotropic height, the variables under the ice, on coastal salinity (runoffs), at the equator on the velocities and on open boundaries (for HRZ systems only)."

<u>is moved by</u>:

The concept of "pseudo-observations" or "Observed-No Change" (innovation equal to zero) has also been introduced to overcome the deficiencies of the background errors, in particular for extrapolated and/or poorly observed variables. We apply this kind of parameterization on the barotropic height and on the 3-D coastal salinity at the rivers mouth 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 200 m depth and near open boundaries of HRZ systems. These observations are geographically positioned on the analysis grid points and not on a coarser grid in order not to generate aliasing on the horizontal. Time of these observations is the same as the analysis, namely the fourth day of a 7-day assimilation cycle. Still in the concern about cost reduction in an operational context, the 3-D variables mentioned above were sampled on the vertical in order to keep only about ten model levels.

3. Loose and sometimes misleading language: Throughout the paper, the English is often poor.

3.1. Page 1123, Line 1: At first glance, the "evaluation of a future monitoring and forecast system", as the title suggests, would seem an impossibility.

We changed the title which becomes: "Evaluation of global monitoring and forecasting systems at Mercator Océan".

3.2. Page 1124, Line 15: The abstract says that the paper shows how the "validation impacts on the quality of the systems". The authors clearly mean that the paper shows how refinements or adjustments to the system based on validation impacts on the quality of the system". Validation itself is passive – with no direct impact on quality.

We agree with Referee #1 that this sentence was misleading. We used the sentence that is suggested here.

3.3. Page 1125, Line 26: "…Mercator Ocean, which is in charge of the global ocean…" is a poor choice of words. The mean the "…Mercator Ocean, which is responsible for the global ocean forecasting efforts under MyOcean…" I note that Mercator Ocean is neither "in charge" of the global ocean, nor are they the only forecast center responsible for global ocean forecasting. They merely take responsibility for this under a European Project. Other forecast centers within Europe, and outside of Europe also produce global forecasts.

We agree with Referee #1. The text has been modified.

3.4. Page 1126, line 2: "What is meant by "It is declined in different configurations."?

This means that we have several systems covering different geographical areas with various horizontal resolutions. The text has been modified.

3.5. Page 1126, line 11: Add reference for the GODAE OceanView "regions".

We added the reference "Hernandez et al., 2009".

3.6. Page 1126, line 6: The "global intermediate resolution" system is defined as the IRG – GIR would be logically more correct, and would be consistent with the definition of the "high resolution zoom" as HRZ.

We agree with Referee #1. However, we chose to follow a different logic, which is the symmetry between IRG and HRZ ("Intermediate" related to "High" and "Global" related to "Zoom"). Moreover, modify IRG into GIR would imply to remake numerous figures whose title reveals IRG. We would like for this reason not to consider this modification.

3.7. Page 1134, Section 3.2: It is not clear why this section is included in this paper. It adds very little.

We agree with Referee #1. Section 3 was too long. Some of the "history" information was moved to the introduction, and the rest was written more concisely.

3.8. Page 1136: The email quails@mercator-ocean.fr does not work. Does this email exist? I emailed it, but my email could not be delivered.

The right email is: qualif@mercator-ocean.fr. The text has been modified.

3.9. Page 1138, line 19, 20: "What does the "best analyses" refer to? Do the authors simply mean "analyses"?

We agree with Referee #1. The text has been modified.

3.10. Page 1140, line 5: "all the systems were closer to observations than climatology". An alternative formulation to the Skill Score in equation (2), has climatology as the denominator. The authors might consider this metric to confirm the above statement.

Actually, the blue curves on Fig. 5 and Fig. 6 already show that the departures from the observations are smaller for all the systems than for the climatology. We agree that using the climatology in equation (2) clearly confirms this result (not shown). We chose not to show the skill scores with respect to climatology as we were showing skill scores with respect to persistence. The latter illustrate the improvements we wish to make in the future in relation with the users needs. The results are not spectacular at the moment but the skill scores are positive and increasing from one version to the next, which is an encouraging result.

3.11. Page 1140, line 18: Regarding comparisons with "… OSTIA observations (not assimilated)…" Although OSTIA SST observations are not assimilated, the data used to construct OSTIA (noting that OSTIA is an analysis, not observations), including AVHRR and AMSRE data, are assimilated by Mercator. So these comparisons are not independent as implied by "(not assimilated)".

We agree with Referee #1. The text has been modified.

3.12. Page 1143, equation (1): could be simplified by simply using speed, instead of the absolute value of velocity. It is really the mean relative speed bias, not velocity bias.

We agree with Referee #1. The text has been modified.

3.13. Page 1143, line 20 and Figure 13: Why do the authors not show maps of the RMS residuals. The mean of the residuals is interesting, but it is only part of the story here.

Suggested maps have been added and commented.

3.14. Page 1146, line 17: Where is "the OVIDE repetitive section"?

The section appears on Fig. 19 but it is too small… We added the following panel in the Fig. 19.