

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

**A. Sellar (Referee)**

**alistair.sellar@metoffice.gov.uk**

**Received and published: 9 May 2012**

## **1 Overall assessment**

The authors describe an operational ocean forecast system, and present some results of validation of hindcasts of this system. The evaluation of the hindcasts is comprehensive and the conclusions are broadly justified by the evidence. The figures, while numerous, are well chosen and demonstrate the breadth of the validation performed.

I would like to see some clarification regarding the correspondence between these "forecast" results in section 4.2 and the accuracy of the operational forecasts (I expand on this point below). Finally, the quality of the writing is variable and needs to be improved before publication: many sentences are over-complicated or confusing.

We thank A. Sellar (Referee #2) for his careful reading of our manuscript and for his constructive remarks. Following his advices, we tried to make the manuscript clearer. A native English speaker will review the document next week. All remarks detailed below by the Referee were considered and/or discussed.

## **2 Major issues**

In the section on forecast accuracy (4.2), the statistics come from the data assimilation innovations. For readers who are not familiar with data assimilation, the authors should explain why the innovations give an indication of forecast quality. They should also include the caveats on this, i.e. briefly describe the differences between the "forecast" used as the DA background, and an operational forecast. For example, the surface fluxes are presumably from NWP analyses rather than NWP forecasts.

We agree with Referee #2. The text has been modified to better explain the differences between the forecast used as the data assimilation background, and an operational forecast. A schematic representation of the IAU procedure has also been introduced to assist in the understanding.

Also, it is conventional when presenting forecast scores to indicate the forecast lead time. Presumably in this work these are averaged over all 7 days of the run, in which case the results are indicative of the average performance over the first 7 days. This should be explained in section 4.2, and figures 13 to 17 should include some mention of the lead time they refer to, if it is possible to do this without making the captions too long.

We agree with Referee #2. The forecast lead time has been indicated in the text.

In section 4.4 I am a little confused about the drift in the IRG\_DEV system. The figures show that there is a difference between the model 2011 in 2007, and that the model is not biased relative to the observations in 2011. Therefore the model is either biased in 2007, or there is a similar change in the observations between these 2 years (presumably the former).

The global cumulative trend of temperature at 300 m is displayed in Figure 23. There is a noticeable cooling East of the Philippines, and two regions of warming West of Australia in IRG\_V1V2 (Figure 23a). These signals are also present in IRG\_DEV (Figure 23b) 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 (Figure 24). This cold bias was not present in the 2007 innovations (we added a figure to show that). This confirms the drift shown in Figure 18 for the North Pacific. There are other regions where the trend from IRG\_DEV is not reliable (South Pacific East of Australia, South Indian between Madagascar and Australia...).

Moreover, we have checked the time evolution of the innovations and detected cold biases at the end of the period in the regions where the cooling trend appears. The signals are consistent with those observed on average in the GODAE regions (as the Indian Ocean or the North Pacific). In the Pacific in IRG\_DEV a cold bias appears at the end of the period for instance in the “nino” region (this region is a square located off the Philippines). This bias is not present in IRG\_V1V2, and we have checked that the number of assimilated data is the same in this box in both experiments.

### 3 Specific comments

#### 3.1 Text

I highlighted some of the sentences which I found confusing, but there are more than I have listed here.

- p1126 line 2: "It is declined in different configurations" I don't know what this means.

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

- p1127 line 15: "IGR" -> IRG
- p1131 line 7: "specificities" -> specifics (also in tables 1 and 2)

The text has been modified.

- p1131 line 18: It would be useful to say which observations go into the RTG analysis, to help understand the difference from the "AVHRR+AMSRE" analysis.

We agree with Referee #2. The text has been modified.

- Section 3 is longer than is justified by its relevance. I would recommend moving some of the "history" information to the introduction, and to make the rest more concise: there is some information which could be removed with any detriment to the paper.

We agree with Referee #2. The text has been modified.

- p1134 line 20: remove "with"
- p1135 line 27: "control" as a verb has different meanings in English and French. "checked" or "monitored" would do here.

The text has been modified.

- p1136 line 4: "Most NWP centres publish quality reports on a regular basis". I'm not sure if this is true. -> "Some"?

We agree with Referee #2. The text has been modified.

- p1138 line 11: "It is therefore considered that the QC built from GLORYS2V1 may be applied to other systems". Is this referring to the QC method, or the QC flags for these observations, or something else?

We have added some details in the text. 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 systems suffer from the same kind of defects, more related to forcings, 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 systems, assuming that the forecast errors or system biases are of the same magnitude or even lower than those of GLORYS2V1.

- p1138 section 4.1: "Best analysis" is not a term I recognise. "Best estimate" is a term defined by the GODAE community, though just "analysis" would be fine here.

We agree with Referee #2. The text has been modified.

- p1140 line 5: "We initially checked that all the systems were closer to the observations than the climatology." Slightly ambiguous. Sounds like  $|\text{model} - \text{obs}| < |\text{model} - \text{clim}|$ . Presumably the intention was  $|\text{model} - \text{obs}| < |\text{clim} - \text{obs}|$ .

We agree with Referee #2. The text has been modified.

- p1140 line 10-16: These sentences are a little too confusing. I think they are making relevant points, but could be made clearer.

The text has been modified.

- p1140 section 4.1.2: Although there is some independent information in OSTIA because it assimilates observations not included in RTG or the AVHRR+AMSRE product, it should be pointed out that it is not completely independent because it shares many observations with these products. The "reduced bias" is to some extent a reflection of the fact that OSTIA shares more observations with the AVHRR+AMSRE product than with RTG. In the (cloudy) high latitude coastal regions where the bias is reduced by the move to AVHRR+AMSRE, both this product and OSTIA are likely to be dominated by AMSRE (the only microwave instrument outside of the tropics used in either analysis) and in-situ data and so will be very consistent. I am sure the AVHRR+AMSRE product is a better dataset to assimilate than RTG, but the comparisons to OSTIA don't necessary prove this. I suggest that the authors include some caveats on the conclusions which can be drawn from the OSTIA comparison.

We agree with Referee #2. The text has been modified.

- p1140 line 25: "This SST product has the same quality level as OSTIA and both display better performance than RTG especially in high latitudes." Include a reference.

A reference has been included.

- p1141 line 8: "increment ... rejected". This is an interesting metric. It should be defined. (is it = increment - analysis + background?)

The "reject increment" has been defined in the text.

- p1141 line 12: "The concurrent effects of bulk fluxes and of IAU correction are not efficient in this region." I don't know what this means. That the increments are not sufficient to control the effect of the fluxes?

The text has been modified.

The system is not efficient enough in correcting the SST because heat fluxes are not part of the estimated state. The IAU correction of surface temperature does not work in the same sense as the bulk forcing function. We checked that most of the SST correction is swept away by the bulk forcing function.

- p1144 line 19: "It is smaller than the different internal errors involved in the system." This is unclear. Smaller than the uncertainty in the observations? If so, please include a reference.

The text has been modified.

The RMS of the SLA innovation 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 prescribed in 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, where largest values are located on shelves, along the coast and mesoscale activity or sharp fronts areas).

- p1145 line 27: "is" -> though
- p1146 line 1: "default" -> defect?

The text has been modified.

- p1146 line 4: "We checked that the correction of the precipitations lead to a deficit in summer". I don't know what this means.

The text has been modified.

We checked that the correction of the precipitations (see Sect. 2.3) actually leads to a deficit in precipitation in summer. It 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 neglects the seasonal cycle of wave mixing. It is strongly marked in this region. This contributes to an excess mixing in summer.

- 1146 section 4.3: It might be helpful to the reader to point out that the results in this section are based on analyses.

We agree with Referee #2. The text has been modified.

- p1146 line 22: "water masses charecteristics" -> water mass characteristics
- p1147 line 15: "which" -> whose; "teleconnexions" -> teleconnections
- p1147 line 18: "As" -> While?

The text has been modified.

- p1149 line 9: Please define "cumulative trend".

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.

- p1149 line 23: "The latter confirms the seasonality of the cold bias that is observed," The authors could refer to the earlier section (or figure) where this was discussed.

We agree with Referee #2. The text has been modified.

- p1151 line 25: "The IAU prevent from keeping the correction of the initial condition in the model because of the bulk formulation." This sentence should be re-written.

Same comment as for p1141 line 12. The sentence has been re-written.

### 3.2 Figures

- Figure 3: the words "temporal and geographical" could be removed without changing the meaning.
- Figure 15: A very short caption. Please include something like "skill relative to persistence", and include the "(x100)" in the caption - not just in the plot title.
- Figure 19: The map on each plot showing the location of the section is too small to see on the printed page. Please either include a larger map, or describe the position of the section more precisely in the text.
- Figure 24: The caption says 2010 but the plot title says 2011.

All these comments were taken into account.