

Review of “How essential are Argo observations to constrain a global ocean data assimilation system?” by Turpin, Remy and Le Traon.

The manuscript describes observing system experiments (OSEs) carried out with the Mercator-Ocean assimilation and forecasting system to demonstrate the impact of the Argo array on analyses and forecasts. A number of one-year experiments are carried out with all data assimilated, all data except Argo, and all data except half the Argo floats, as well as a free run. The demonstration of the impact of Argo data is important for funders of the observing system, and can also be beneficial to improve understanding of the data assimilation and forecasting system. There are two major areas which should be addressed before the paper is published, as described below, followed by some more minor comments.

Major comments

1. The main assessment strategy in the paper is to compare with Argo data. Presumably that will over-estimate the impact of Argo data. Some assessment using independent data, or data which are assimilated in all the assimilative runs would give a clearer picture of the impact. There needs to be at least some discussion of whether this assessment strategy is objective/fair. Only impact on temperature and salinity is shown – what is the impact of Argo on other variables? Also, in parts of section 3, the *analysis* is compared to observations, and some of the experiments are assimilating Argo data - the observations could in principle be bad, we could fit them closely and we would see a positive impact on the analyses, but that wouldn't necessarily be a good thing.
2. The description of the system used in the experiments is not very clear in section 2. Some things are repeated a few times (observations assimilated), and other important details are not described clearly. Aspects of the data assimilation, which are obviously crucial to the impact of the Argo data, are not clear from the description. For instance, it would be useful to have an idea of the effective length-scales, and to know how the temperature/salinity data affect other variables. What is the time window of the data assimilation? Crucially for interpretation of the results, is there any relaxation to climatology, and if so does it vary with depth? For the observations which are assimilated, are they the “real-time” version of the Argo data or have they undergone delayed-mode quality control? What about the other data-sets (e.g. SLA) – are they real-time versions of the data?

Minor comments

Abstract

- “half of Ago data sets” should be “half of the Argo data”

- “reducing observation-model forecast error” should be “reducing model forecast error” or “reducing observation-model forecast differences”
- Lines 16-17: There also seems to be a large impact in the Gulf Stream region.
- Line 19: “even with a 1/4° model resolution”. Not sure what is implied by this statement.
- Line 20: This last sentence seems to be making a general statement for all global data assimilation systems, but the study is only about the Mercator-Ocean one.

Page 1146, Line 26: on first reading I thought you meant that the Argo was array is measuring T & S mainly for operational oceanography, but I think you mean it is reporting in *real time* mainly for operational oceanography. More generally the first paragraph of the introduction doesn’t read very well.

Page 1147, Line 11: “optimizing the design” seems a bit strong here. Maybe “improving the design” would be better.

Page 1147, para starting on line 22. Two papers by Oke et al, 2015, in the Journal of Operational Oceanography provide a good overview of recent efforts on OSEs and should be included in the introduction.

Page 1148, line 9. The impact of Argo data on T and S only are assessed. What is the impact on e.g. SSH?

Page 1148, Line 13. “conclusions are”.

It is worth highlighting in the introduction that the results obtained from this study on the impact of Argo are specific to the Mercator-Ocean system, and that they can be put into the context of other similar studies.

Page 1148, Line 18. SLA should be Sea Level Anomaly. Also, this list of in situ data is repeated later in section 2.2.

Page 1149, line 3. “Momentum and sea surface fluxes”. Do you mean “Momentum, heat and freshwater fluxes”?

Page 1149, line 6. It should be made clear that the error covariances are static.

Page 1149, line 10. Many readers will not know IAU, so it is worth giving a little bit more detail and referring to a paper on it.

Page 1149, paragraph 2. There is no way to understand from this how the MDT was modified.

Table 1 doesn’t add much as it is full of acronyms which aren’t all described in the text (e.g. SAM). Perhaps the table could be removed?

Paragraph starting on line 19, page 1149. This seems more like an acknowledgement so should go at the end of the paper.

Page 1150, line 3. “important” should be changed to “numerous”. An assessment of their importance is the focus of the paper.

Page 1150. Line 14. There are other regions where the marine mammals also provide useful information such as the Kerguelen region.

Page 1150, line 16. Perhaps it would be worth mentioning that there is some coordination of gliders through EGO?

Page 1150, line 22-23. This approach is similar to that used by Fujii et al 2015 – “Evaluating the impacts of the tropical Pacific observing system on the ocean analysis fields in the global ocean data assimilation system for operational seasonal forecasts in JMA”.

Page 1150, line 26. Be more specific about which regions are more densely sampled by the No Argo platforms.

Page 1151 line 1. The time distribution is described as fairly regular yet fig 2 seems to show variations in terms of global observation numbers through the time period. What are these due to?

The last 6 months is chosen throughout the results sections. There doesn't seem to be much justification other than saying that there is a “spin-up” of the impact. It would be interesting to know how long the spin-up really takes.

Page 1151. The term “Other No Argo in situ data” seems quite convoluted. How about “non-Argo in situ data”?

Page 1152, line 25. Why are the model's western boundary currents warmer than observations?

Page 1152, line 27. The model's tropical Pacific is also warmer in the east which could imply an error in the slope of the thermocline. Some discussion of why that might occur would be useful to have.

In fig 3 the tropical differences appear quite large given that the moored buoys are assimilated in those regions. Is there an explanation for that?

Page 1154, line 19. Fig 5 is shown for one date in December. Would we see larger impacts in the southern hemisphere rather than northern hemisphere if we looked in June?

Page 1155. Line 4-7. It seems strange not to show any results from the 300-700m depth range.

Page 1155, lines 20-23. I got confused in this discussion about which layer you are commenting on.

Page 1156. Line 26. Are the differences smaller in the 0-300m layer because SST data are assimilated?

Page 1157. Line 6. Heat Content anomaly is only reduced in the 700-2000m layer.

Page 1157, Line 26. There do appear to be significant differences in parts of the South Atlantic.

Page 1158, line 13. It could be worth mentioning here whether the large differences are due to issues with the modelling of the Mediterranean outflow.

It is quite difficult to compare Figs 4 and 11 when they are quite far apart in the paper. Could you show the ratio of the differences shown in fig 11 to fig 4 to show the improvement more clearly?

Page 1159, line 1. "operational" PSY3 – presumably here you mean the Run-ref experiment analyses?

Page 1161, top paragraph. It would be helpful to the reader if you put these percentage improvements into some sort of context. For example, how much improvement would you expect from a normal upgrade to the model or data assimilation?

Page 1162. Line 6. It sounds from this that the experiments were carried out in real time, so perhaps leave out the "real time" part.

Page 1163, final paragraph. I don't quite understand what you mean by "our statements can be generalized to include other systems".

Table 2. At the moment, the table uses "x" to denote that an observing system is assimilated and a blank to denote that it isn't assimilated. It might be clearer to use a tick and a cross.

It is often very difficult to read text on the figures, particularly the color scales.

Figure 1. It isn't clear from the caption what the terms "odd" and "even" mean.

Figure 3 and subsequent figures. The time period of these comparisons should be included in the caption.

Figures 8 and 12. The light and dark blue lines are difficult to distinguish – perhaps use a different color. Also, in the caption the sub-plot letters follow the description for that sub-plot whereas in the other plots they precede the description. "South Ocean" should be "Southern Ocean".

Figures 13 and 15. The figures are very small. Perhaps split them onto two rows.

Figures 14 and 16. The green and yellow lines are difficult to distinguish. Also the “normalized RMS of innovations” statement is not very precise. In the bottom axis of plot (b) it has “(% of performance)” but presumably this is not a percentage.