Anonymous Referee #2

The paper provides a summary of the sea ice thickness anomalies found in the most recent ocean-ice analyses. It considers the impact of the assimilation of sea ice observations within some of these systems to spatial and temporal scales of the anomalies.

In my opinion the paper has lots of potential and is almost there, but at the moment it is missing some extra synthesis/analysis which would make it a really useful reference for the observation and modelling community. The impact of sea ice assimilation was considered (but did not state if any of the models assimilate anything other the concentration, one may nudge the thickness too?); it would be helpful to understand how the impact of other choices in the system may also influence the sea ice thickness anomalies. The sea ice models may have very different methods for modelling thickness. Do they have ice thickness distributions or a single category, does this have an impact? Different atmospheric forcing sets (with different imposed sea ice cover) may influence the local energy balance and in turn affect the ice thickness. I think considering a few other key elements of these systems would make this paper very helpful in guiding the use of reanalyses and their future development (both in terms of the type of data that is used but also the set up of the assimilation systems)

We thank the referee for evaluating our manuscript with such richness of details. We appreciated very much her/his comments, which were all taken into account in the revised version of the manuscript. We have tried to identify the impact of several other reanalyses' choices (10 in total) on the time and length scales of the sea ice thickness anomalies. We did identified other parameters that potentially affect the referred scales, as for instance the air-ice drag coefficient. For the atmospheric forcing the results are not conclusive as detailed in our answer for your second specific comment. None of the reanalyses assimilate sea ice thickness.

It is worthwhile mentioning that in order to better incorporate the referees' suggestions, the manuscript's structure has changed. New figures were added, while others were replaced. Thus, all tables, figures, pages and lines referred to in this rebuttal letter are directed to the updated version of the manuscript, unless otherwise stated.

Specific comments:

The paper provides some details of the reanalyses you have studies but I think it is missing a table summarising the reanlayses - some form of synthesis would be beneficial to the reader, the papers that do tabulate some of this (e.g. Chevallier et al) are not complete for this reanalysis set. A table with clear information about the forcing data set and types of data that are assimilated and what methods are used to do this. Given that we also know that the strength parameter impacts the thickness it would be good to tabulate P* or equivalent as well. As the paper addresses the timescales of anomalies it would also help to determine the assimilation window length and see if this has an impact.

We have now added Table 1 (pg. 4) to the manuscript. The table reproduces the info already presented by Chevalier et al. (2017), but it also brings info for the other four reanalyses not considered by these authors (GECCO2, GloSea5, PIOMAS and TOPAZ4). The following parameters and specifications are included in Table 1: nominal horizontal resolution, ocean-sea ice model, source of atmospheric forcing data, number of ice-thickness categories, EVP or VP dynamics, ice strength parameter (P*) or frictional dissipation coefficient (Cf), air-ice drag coefficient (C_W), ocean-ice drag coefficient (C_A), source of sea ice data assimilated and method used for assimilating sea ice data.

Since the atmospheric forcing and many other parameters/specifications (see further in this rebuttal letter) do not seem to play a clear role in the time and length scales, and mainly due to the fact that we are working with monthly time series, we don't expect to see large changes in the fields of time and length scales associated to the "assimilation window length". At least, not for this kind of comparison where each product has several different parameter/specification compared to the other systems.

Other parameters could be tested, nevertheless it is not always easy and straightforward to retrieve all information from all reanalyses, as well as to compare products with several varying parameters. Nevertheless, we believe that the list of parameters in Table 1 is comprehensive and **reinforce that time and length scales are mainly driven by the fact of whether or not the reanalyses assimilate sea ice concentration data** (see below).

One reason for requesting the table is when looking at your first figure. You present results but there is not much discussion about the differences that are present in the time series. Are you able to stratify the impact of certain assimilation choices other than whether sea ice data is assimilated.

I would suggest that you may see an impact of the different forcing sets that are used, many are forced with ERA-Interim but MOVE(CORE and G2), GECCO, EDCA, MERRA are not. The ice fields that the atmospheric forcing fields have "seen" will have a impact on the forcing they provide. Differences due to SST relaxation or model parameter choices may also play a role – it would be good to at least see if there are other reasons for the differences other than they include sea ice assimilation or not.

Figs. A and B (attached to this rebuttal letter) show how time and length scales (hereafter, TS and LS) are related to the reanalyses' parameters and specifications displayed in Table 1.

Fig. A shows a comparison of TS and LS against a set of 'uncountable' specifications (sea ice data assimilation, atmospheric forcing, sea ice model, dynamics, EVP or VP dynamics, and sea ice forcing). The results reinforce that TS and LS are indeed linked to the fact of whether the reanalyses assimilate (DA) or not (NA) sea ice data (Fig. Aa,b,i,j), while the other specifications do not seem to have a large impact on the studied scales (Fig. Ac–j). For the atmospheric forcing case, notice that we can divide the reanalyses into three groups: ERA-Interim (7 out of 14), NCEP/NCAR (3 out of 14) and Others (4 out of 14). So, an effective comparison can be made only between ERA-Interim and NCEP/NCAR. Nevertheless, from NCEP/NCAR 1 system assimilates sea ice data, while 2 other systems do not assimilate sea ice data. This makes it difficult to evaluate the impact of the atmospheric forcing on the studied scales.

Fig. Bg-p compares TS and LS against the 'countable' reanalyses parameters. A part of the horizontal grid resolution and the air-ice drag coefficient (C_A), which show a certain correlation with LS (Fig. Bh) and TS (Fig. Bm), respectively, the number of thickness categories, the ice strength parameter (P^*), and the ocean-ice drag coefficient (C_W) do not show a strong correlation with the studied scales.

In addition, Fig. Ba–f also compares TS and LS against the mean state (mean Sea Ice Volume; SIV), the interannual of variability (std SIV anomaly), and the Sea Ice Drift (SID). All these parameters show a certain correspondence with the TS and/or LS. The most pronounced is the case Mean Sea Ice Drift x TS shown in panel Fig. Be.

The most meaningful results from Figs. A and B are incorporated in the new version of the manuscript (Figs. 12 and 13; Sec. 4)

Section 2.3:

It was not entirely clear to me how you were treating draft and SIT differently from the reanalyses when comparing to observations. Did you use the snow cover from reanalyses to compute this when comparing to observations or just disregard for both reanalyses and observations?

We decided not to use the snow cover from the reanalyses for converting sea ice thickness to sea ice draft. This decision was taken in order to avoid adding uncertainties to the SIT fields. So, when comparing SIT from the reanalyses against Draft from the observations we made use of **two metrics**:

(i) **the correlation coefficient (R)**, as a measure of the **linear correlation** between SIT (reanalyses) and Draft (observations);

(ii) **the Mean Residual Sum of Squares (MRSS)**, as an indicator of whether SIT values from the reanalyses are good **predictors** for Draft observations;

We have improved this info in the text in order to make it clearer to the reader (pg. 5, lines 25-33).

Section 3:

Figure 4: some of the scatter diagrams look like the model thickness stops at a particular value e.g. ECDA are you missing some data from thickness categories? Where the draft data do not show a similar relationship to the SIT is there something different about the way snow on ice is treated in the systems? Is this dependent on the assumptions you made about how you compare draft and SIT?

The data used in this work was previously compiled and published by Chevalier et al. (2017) and Uotila et al. (2018). These authors collected the data as they were made available by the providers. So, it is really unlike that some thickness categories are missing.

In ECDA this feature (Fig. 4) takes place due to the fact that this system is characterized by relatively thin ice. See Fig. C at the end of this rebuttal letter. Notice that the same is also valid for GloSea5 and UR025-4.

Do you have an understanding of why the largest differences are near the Greenland coast and Canadian Archipelago? is this down to model physics differences?

The regions near Greenland coast and Canadian Archipelago are marked by the thickest sea ice over the studied domain (please, see Fig. B from the rebuttal letter to the first reviewer). So, proportionally, the same differences in these regions are amplified when calculating the errors. This is a simple, but interesting, effect that deserves some explanation in the text. Thanks for your observation. (pg. 19, lines 8-10)

Section 3.3:

You note that the GloSea systems have shorter timescales than others - is this persistence also linked to mean state? If you have thinner ice on average you may lose it over the summer and it will reduce your persistence.

Overall, the systems which do assimilate sea ice are marked by a certain correspondence between TS (persistence) and the mean state as shown in Fig. 13a (or Fig. Ba in this rebuttal letter). The same for the LS and mean state (Fig. 13b or Fig. B*b*). Nevertheless, this is an overall observation

for the cloud of points. Looking carefully at the GloSea systems, GloSea5 is indeed characterized by thin ice (second thinner mean state of the ensemble). On the other hand, GloSea5-G05 has thicker ice compared to GloSea5. Compared to the ensemble of the reanalyses which do assimilate sea ice, the GloSea5-G05 presents an intermediate mean state. In Sec. 4 (pg. 22, lines 11-20), we raise some aspects that could potentially impact the scales of the GloSea systems.

pg 12, lines 5-6: these sentences were not enitrely clear - would suggest you consider rephrasing to make sure the meaning is clear. do you mean that G2V3, ORAP5, PIOMAS, TP4 and UR025-4 are just similar and if so why hightlight this group - is it for a particular region of the Arctic. It wasn't totally clear what you wanted to point out here.

We agree that this sentence was confusing, and we have reconsidered it. (pg. 14, lines 5-6)

pg 12, lines 17-18: do you mean years or months?

We meant months. Corrected in the text.

Does your wavelet analysis give an indication how robust your findings might be based on the limited time series? If so I think this would be worth commenting on in the text.

The "cone of influence" shown in Fig. 7c,g (cross-hatched areas) highlights the region of the wavelet spectrum where edge effects become important due to the length of the time series. In this region of the spectrum, the results should be interpreted carefully, since the time series is short for solving some periods (y-axis) in some specific time spans (x-axis). This information is found in the Fig. 7's caption.

Section 4:

pg 17, lines 3-6: You are comparing c_GLORS05-G2V3 - do both of these systems assimilate sea ice data in the same way? You also note that G2V3 and ORAP5 could be considered similar – they may be similiar in some respects but potentially have different forcing G2V3 its not clear to me if G2V3 uses operational analysis from ECMWF rather than ERA-Interim - which may lead to some differences. ORAP5 uses a non-standard value of P* compared to standard LIM2 setup.

Indeed, C-GLORS05 and G2V3 assimilate sea ice data in a different way (Table 1), and we agree with referee's comment that this may be a reason for the relatively high RMSE between these two systems. The three referred reanalyses (C-GLORS05, G2V3 and ORAP5) apply the same atmospheric forcing (ERA-Interim). C-GLORS05 and G2V3 assume $P^* = 2.0 \times 10^4$, while ORAP5 uses $P^* = 1.5 \times 10^4$.

Based on referee's observations, we brought this discussion to the text (pg. 18, lines 5-12).

Figure 11 shows the reduction in time scale in one system with the application of sea ice assimilation but the change seems somewhat smaller than the differences with the no ice assimilation across the ensemble - would you expect this given you results? where would G2V1 be in your fig 13?

Following referee's 1 suggestion, the new version of Fig. 11 shows the TS and LS differences between the systems (G2V3 – G2V1). Fig. 11 suggests that the fact that time and length scales are shorter for systems with data assimilation is valid in terms of pan-Arctic averages, but not necessarily at every grid cell. The pan-Arctic AWM-TS and AWM-LS from G2V3 are 5 months

and 728.2 km, respectively. Without sea ice data assimilation (G2V1), the AWM-TS and AWM-LS increase to 5.5 months and 745.3 km, respectively. For this case, TS increases by 10% from G2V3 to G2V1, though the LS scale isn't strongly impacted (pg. 20, lines 1-6). In Sec. 4, we are discussing that TS is more sensitive to the systems' parameters and specifications.

Fig. 14 (empty red circle; highlighted in the gray rectangle) displays where G2V1 takes place in the diagram TS *x* LS.

Figure 13 and discussion pg19, lines 15-17. I was not convinced that the ice volume anomaly correlated well with the time and length scales - without the GloSea systems it seems less clear.

We agree that in the way the figure was plotted, it wasn't easy to identify the correlation between mean state and the scales (TS and LS). So, we have now plotted two different diagrams: Mean SIV *x* TS (Fig. 13a) and Mean SIV *x* LS (Fig. 13b). Besides the spread among the points and the relatively weak correlations (which are not significant at the 5% level), an increase in the mean SIV generally leads to longer time and length scales, taking into account that each reanalysis has different sets of specifications and parameters (pg. 21-22).

Technical comments:

Abstract: line 4: intend rather than intent

Corrected.

section 2.2:

pg4, line3: not sure what you mean by ponctual - is this "point" measurements?

Corrected.

pg4, line 14: run by Environment.…

Corrected.

section 3.1: pg8, line 2: Don't you mean MERRA rather than MOVE?

Indeed, we meant MERRA. Corrected in the new version of the manuscript.

section 3.2: pg8, line 22 and pg11, line8: do you mean "error" or "difference" ?

We meant error (RMSE). Clarified in the text.

Section 3.4:

the last sentence of the section "..for what a quantification is presented in the conclusions of this study" - the phrasing is somewhat awkward and hard to understand exactly what you mean
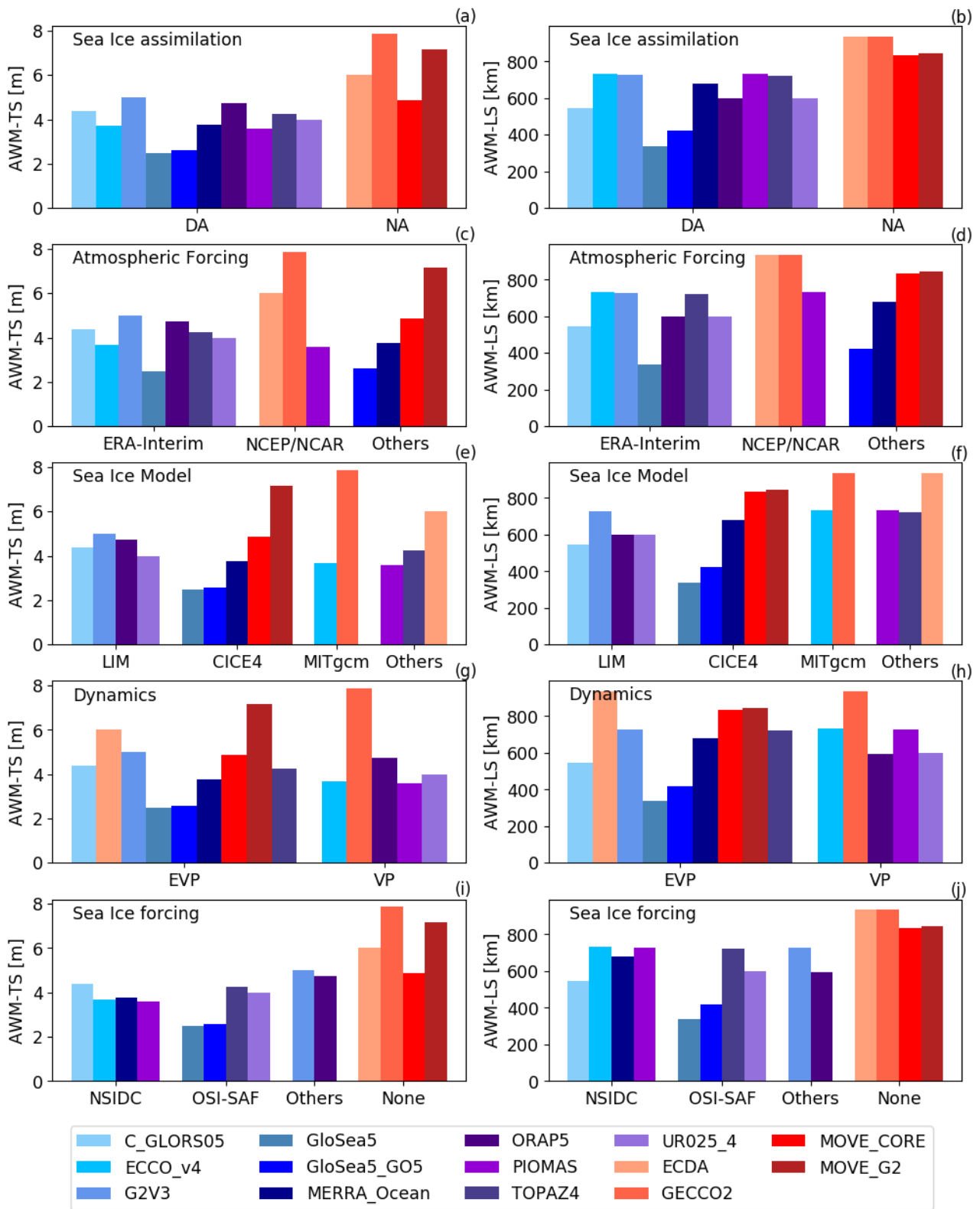
Corrected.

Fig. A: Histograms showing how the AWM-TS (left panels) and AWM-LS (right panels) are related to other "uncountable" reanalyses' parameters and/or specifications, as follows: (a-b) whether or not the system assimilates sea ice data; (c-d) the source of atmospheric forcing data; (e-f) the used sea ice model; (g-h) the used dynamics (Viscous-Plastic or Elastic-Viscous-Plasctic) to account for ice-ice interactions that control ice deformation'; and (i-j) the source of assimilated sea ice concentration data.
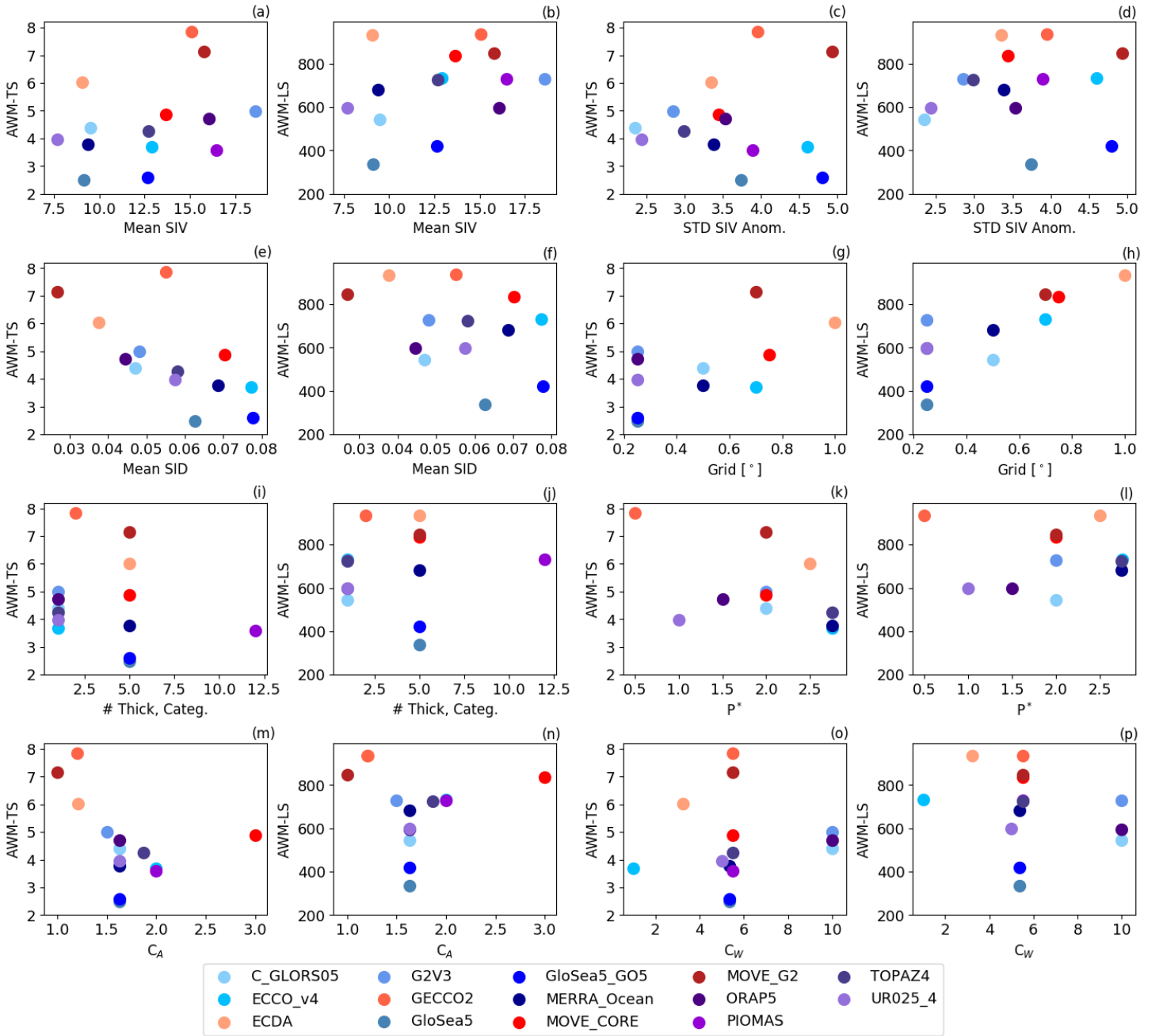
Fig. B: Scatter plots showing how the AWM-TS [months] (first and third columns) and AWM-LS [km] (second and fourth columns) are related to other reanalyses' parameters, as follows: (a-b) Mean Sea Ice Volume [m] (SIV); (c-d) standard-deviation of the SIV Anomaly [m]; (e-f) Mean Sea Ice Drift [m/s] (SID); (g-h) grid resolution [degrees]; (i-j) number of thickness categories for discretization of ice thickness; (k-l) ice strength parameter (P*) [N/m] for the ice strength formulation following Hibler (1979); (m-n) air-ice drag coefficient [10^{-3}] (C_A); and (o-p) ocean-ice drag coefficient [10^{-3}] (C_W).
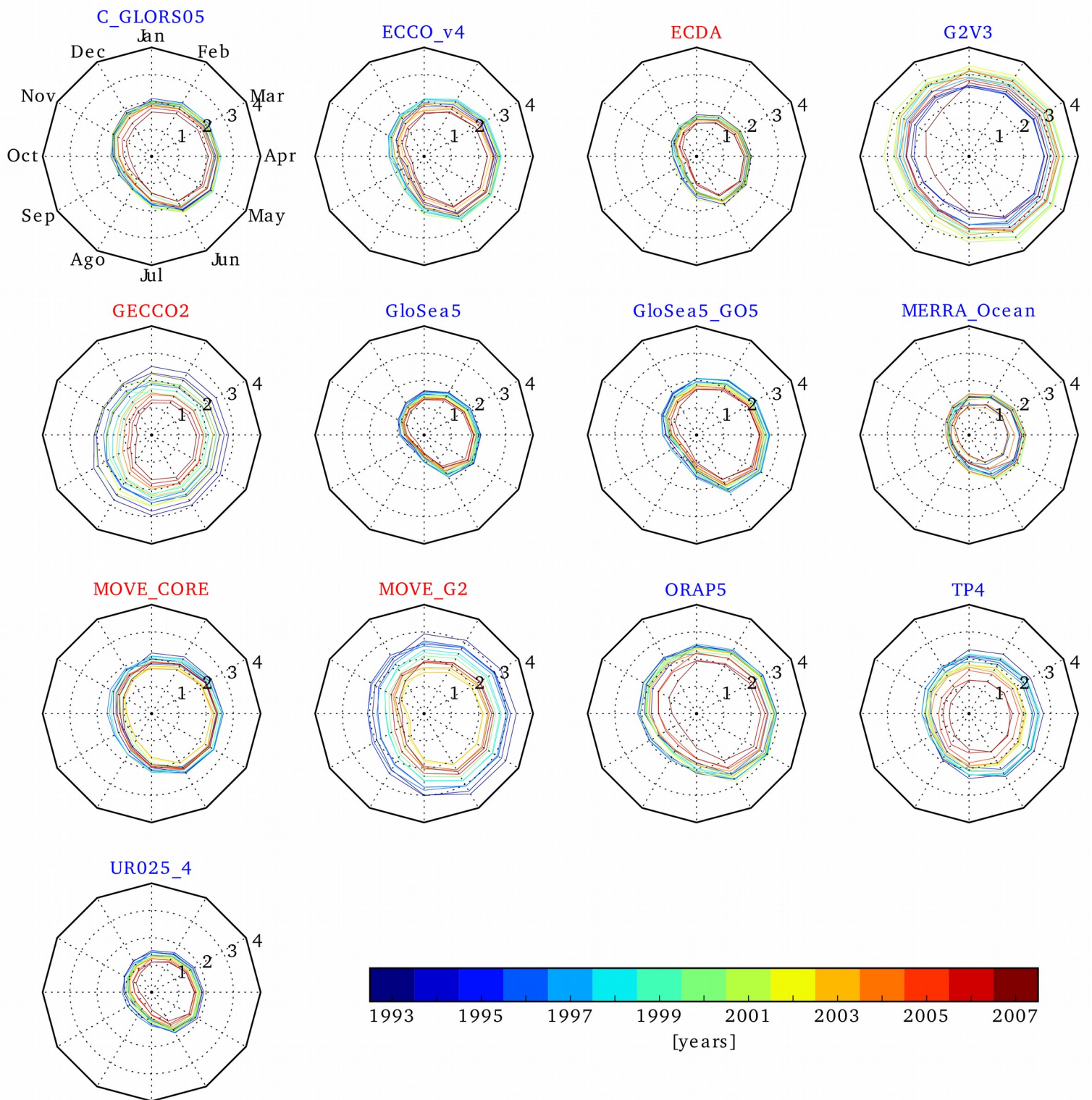
Fig. C: Radar plots for the Averaged-Weighted-Mean Sea Ice Thickness. Every color represents a different year from 1993 to 2007.