Response to the referee 2

Thank you very much for a helpful and constructive review. Please see below for a point-by-point response on your corrections and suggestions (with our replies marked with blue colour and indentation).

This manuscript assesses four variables in the CARRA data: ice surface temperature, ice albedo, ice thickness and snow depth on sea ice, against satellite data, in-situ observation, and the ERA5 data. It is found that both improvement and deterioration, mainly regional dependent, exist in the performance of the target variables in the CARRA data in comparison to the ERA5 data.

In general, the manuscript is significant and valuable for the users of the CARRA data, in this sense, I recommend major revision for this version of the manuscript. Substantial revision is needed before the manuscript can be accepted for publication: 1) the manuscript lacks possible mechanism analysis linking the observed improvements of the four variables to the differences in sea ice scheme between the CARRA and ERA5 systems, this part could be arranged in the last section. 2) The manuscript needs English-editing, some sentences have illogical structures, which make the readers hard to follow.

> Similar concerns regarding limited analysis of the relations between differences in sea ice parameterisation schemes of CARRA and ERA5 and observed performance of the two reanalysis systems were raised by the editor. We agree that the submitted manuscript is somewhat lacking in that aspect, mainly because for two out of the four variables of interest (namely for ice thickness and snow depth) ERA5 does not represent the governing processes at all, and for the other two either the reasons for observed improvement in CARRA are covered by earlier studies (for ice surface temperature) or not clear improvement is found (for ice surface albedo). Nevertheless, we will update the manuscript to provide more discussion on how the observed improvement/degradation found in CARRA compared to ERA5 is connected to the differences in the representation of the sea ice cover in the two reanalysis systems.

> Concerning the language and sentence structure of the manuscript, we will revise the text to improve its clarity and readability.

Minor comments:

L14: Over the whole MS, the authors use some statements like "added value", however, such statements are unclear. It is better to directly use: the performance of CARRA on XXX is better/worse than the ERA5.

> In our paper we use the term 'added value' to highlight and acknowledge that, in fact, CARRA, as a regional reanalysis, is forced by the ERA5 data. However, this term is used only once in a sense 'CARRA is better than ERA5' while in other cases we refer to (or imply) '*potential* added value' and use the term in the introductory sections when performance of the two reanalyses is yet to be assessed. When performing actual comparison we already use direct terms and state whether CARRA is better/worse than ERA5.

L42: Can you explain why operational NWP systems are tuned to perform best in midlatitudes? Why not in Tropics?

> In the original manuscript when we state that operational NWP systems are tuned to perform best in mid-latitude regions we meant the cases when a model domain includes both mid-latitude and polar regions. Obviously, that statement does not apply for an operational weather forecasting system used over a tropical model domain where it makes little sense to tune that

NWP system based on its mid-latitude or polar performance. We will update the text to convey our point in a more clear way and to avoid confusing generalisations.

L50: Can you specify what is "simplified one-dimensional parameterisation schemes"? This statement is so blurry.

We used 'simplified one-dimensional parameterisation schemes' as an umbrella term to distinguish simple schemes (which could be as basic as prescribed ice surface temperature, or thermal balance of a thin ice layer) from more advanced thermodynamic and dynamic-thermodynamic sea ice models. In the following paragraph of the main text in the original version of the manuscript we provided a number of examples of such schemes. In the updated version of the manuscript we will additionally refer to Hines et al. (2015), Køltzow et al. (2019) and Solomon et al. (2023), should the readers be interested in simple sea ice parameterisation schemes applied in operational non-coupled NWP systems.

L65: "Additionally" is not needed.

We believe that sentence is more readable as it as is now, therefore we will keep it in the revised manuscript.

L85, 124, 238, 620: "at the moment of writing this manuscript" is not needed.

We believe that these remarks are well-justified and necessary to not confuse readers if some of the discussed products would be extended or cease operational production after this study is published, so we will keep them in the revised version of the manuscript.

L85: with a time interval of 3 hours

We agree that 'every 3 hours' in the original manuscript looks somewhat confusing. We will reword that sentence.

L85-87: "the CARRA data set includes the output from model integration...lead times over 6 hours". Does this mean the CARRA also include forecasting data? If so, this statement could be changed to "the CARRA data set also includes model forecasting data......."

Yes, the CARRA data set provides both objective analysis and forecasting fields. To make it more clear we will update the manuscript as suggested.

L101-102: "Sea ice albedo ....surface albedo". As you mentioned direct albedo, diffuse albedo, it is better to declare their definition, otherwise you can obsolete this sentence.

To us the terms 'direct albedo' and 'diffuse albedo' are pretty self-explanatory and do not require additional introduction. However, we agree that it might be somewhat confusing to see these terms for a reader more used to 'black-sky' and 'white-sky' albedo. We will update the manuscript to make it more clear, and refer to Lucht et al. (2000) in case readers are interested in additional details.

L114: Can you explain what is "fallback data set"?

In the manuscript we refer to OSISAF OSI-450 as a 'fallback data set' meaning that it is used for dates when ESA CCI SICCI data are not available. We will update the text to make it more clear.

L140: Please illustrate how to calculate verification score. Or give an reference.

Our intention was to simply state that ERA5 is used as a baseline when assessing performance of CARRA (where applicable) without referring to any specific metric. We will reword that sentence to make it more clear. As for the actual verification scores discussed in the our paper, we use pretty standard statistics such as mean error or standard deviation of error which, we believe, do not require additional explanations.

L143: for example

Will be corrected.

L147-148: "in the present study..... a denser model grid"....hard to follow. Please reorganize.

Indeed, that sentence is rather cumbersome and fits poorly within the rest of the paragraph. We will remove this sentence in the updated manuscript and extend the paragraph to be more clear and easy to follow.

L150: but the characteristics

We think that in this context 'and' would be a better replacement for 'although' used in the original version of the paper. We will update the manuscript accordingly.

L161: Where is White Sea? Not marked in Fig. 1.

Thank you for spotting a missing label, we will update the figure to include it.

L168: "of the the sea ice"......delete one of "the"

We will update the manuscript accordingly.

L169: those derived

Will be corrected as suggested.

L190: "MODIS retrieves .... in situ observations". Any reference ???

To support our statement that MODIS ice surface temperature retrievals used in the present study tend to show a cold bias when compared to in situ observations, we will refer to studies by Hall et al. (2004), Herrmannsdörfer et al. (2023) and Li et al. (2020) in the updated version of the manuscript.

L218: Why the data between 2016 and 2019 is not used?

In the present study most of the analysis is focused on assessing the performance of the CARRA system in representing the modern-day Arctic. Therefore a representative 15 year period (from 2000 to 2015) was chosen, despite more data being available for both CARRA and CLARA-A2. Also, the 2016-2019 part of CLARA-A2 has been done as an extension. In this extension aerosol quantities are taken as climatology values, meaning the intercalibration of the radiances is not quite as robust as in the non-extension part up to 2015. The difference between the main data set and the extension is not very big, but considering that characteristic biases of CARRA were already apparent from the initial comparisons, we decided to exclude the 2015 onwards extension.

L362: January-February and middle August-middle October in zone A, and middle August-September in zone B

Thank you for providing these corrections, we agree with all of them except the first one. In February over zone A CARRA shows slightly lower (in absolute value) ice surface temperature error compared to ERA5 (as can be seen from Table S3). We will update the manuscript accordingly.

L364: December to March

To make that sentence less ambiguous, we will change the text from '*from December to April*' to '*from December to the end of March*'.

L369: December to March

Similarly to the previous comment, we will explicitly state in the updated manuscript that the period spans from December to the end of March.

L411: no supporting evidence for the statement "which is attributed ... central Arctic"

We think that our word choice might have confused the referee, as we used 'central Arctic' to refer to the northernmost part of the CARRA model domains. We agree that, as it is written now, that sentence could be interpreted in a way that snow accumulation and temperature drop in the central Arctic *outside* the CARRA model domains influences the surface albedo evolution *within* the CARRA area. Indeed, such a statement would require strong supporting evidence. However, in our manuscript we do not speculate about potential impacts of the outside regions on the performance of the reanalysis system but simply explain the observed feature in the distribution of sea ice albedo errors in August based on our knowledge of the technical implementation of the sea ice parameterisation scheme applied in CARRA with two different albedo formulations for snow-free and snow-covered sea ice grid cells. In the beginning of August sea ice in the model is actively melting and and dark snow-free ice surface is exposed (there are also grid cells which still retain some amount of snow, but they also have relatively low albedo of melting snow even though not as low as for melting ice). However, by mid-August new snow starts cov-
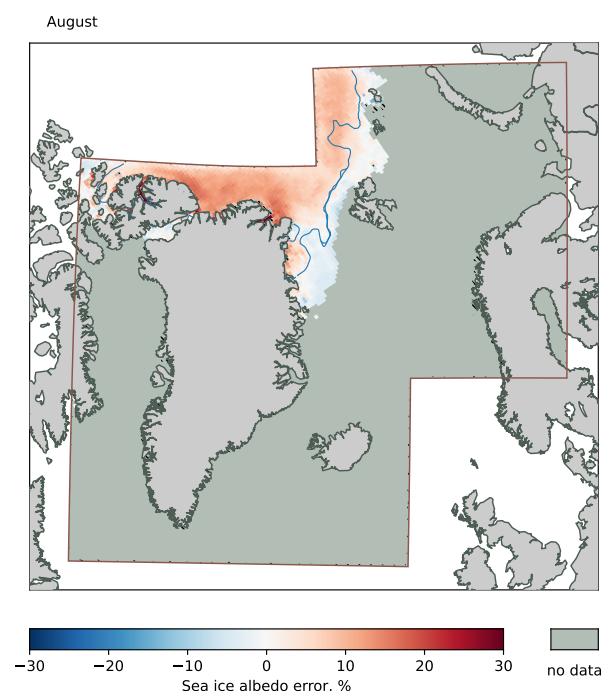

Figure 1: Mean error of the modelled surface albedo in August over sea-ice covered regions of CARRA, computed against the CLARA-A2 SAL product over the time period from 2000 to 2015. Median snow extent at the end of the month is outlined by a contour.

ering the ice resulting in increased surface albedo of snow-covered grid cells. Typically, by the end of August a considerable fraction of sea ice within CARRA model domains is covered by new snow, although the snow-free part is still not negligible. This results in a bimodal distribution of August sea ice albedo error in CARRA with one mode corresponding to bright fresh snow and another one — to grid cells with dark ice and old melting snow. This situation is illustrated in Fig. 1 showing the median snow extent over sea ice at the end of August. Therefore, to support our statement we will update the albedo error maps figure in the revised version of the manuscript to show the August snow extent. Additionally, we will clarify in the text that 'central Arctic' refers to the northernmost part of CARRA model domains to make the text more clear and avoid further confusion.

L440: delete ", for example"

We will update the manuscript accordingly.

L499: "first years"???

We meant the first half of the snow depth anomaly series there. We will reword that sentence to make it more clear.

References

Hall, D., Key, J., Casey, K., Riggs, G., and Cavalieri, D.: Sea ice surface temperature product from MODIS, IEEE Transactions on Geoscience and Remote Sensing, 42, 1076–1087, https://doi.org/10.1109/TGRS.2004.825587, 2004.

Herrmannsdörfer, L., Müller, M., Shupe, M. D., and Rostosky, P.: Surface temperature comparison of the Arctic winter MOSAiC observations, ERA5 reanalysis, and MODIS satellite retrieval, Elementa: Science of the Anthropocene, 11, 00 085, https://doi.org/10.1525/elementa.2022.00085, 2023.

Hines, K. M., Bromwich, D. H., Bai, L., Bitz, C. M., Powers, J. G., and Manning, K. W.: Sea Ice Enhancements to Polar WRF, Monthly Weather Review, 143, 2363–2385, https://doi.org/10.1175/MWR-D-14-00344.1, 2015.

Køltzow, M., Casati, B., Bazile, E., Haiden, T., and Valkonen, T.: An NWP Model Intercomparison of Surface Weather Parameters in the European Arctic during the Year of Polar Prediction Special Observing Period Northern Hemisphere 1, Weather and Forecasting, 34, 959–983, https://doi.org/10.1175/WAF-D-19-0003.1, 2019.

Li, N., Li, B., Lei, R., and Li, Q.: Comparison of summer Arctic sea ice surface temperatures from in situ and MODIS measurements, Acta Oceanologica Sinica, 39, 18–24, https://doi.org/10.1007/s13131-020-1644-7, 2020.

Lucht, W., Schaaf, C., and Strahler, A.: An algorithm for the retrieval of albedo from space using semiempirical BRDF models, IEEE Transactions on Geoscience and Remote Sensing, 38, 977–998, https://doi.org/10.1109/36.841980, 2000.

Solomon, A., Shupe, M. D., Svensson, G., Barton, N. P., Batrak, Y., Bazile, E., Day, J. J., Doyle, J. D., Frank, H. P., Keeley, S., Remes, T., and Tolstykh, M.: The winter central Arctic surface energy budget: A model evaluation using observations from the MOSAiC campaign, Elementa: Science of the Anthropocene, 11, 00 104, https://doi.org/10.1525/elementa.2022.00104, 2023.