Sea ice in the Baltic Sea during 1993/94–2020/21 ice seasons from satellite observations and model reanalysis

Anonymous Referee #2

https://editor.copernicus.org/index.php?_mdl=msover_md&_jrl=778&_lcm=oc108lcm109w&_acm=get_comm_sup_file&_ms=120859&c=270362&salt=1013438105556683178

We greatly appreciate your thorough review of our manuscript and the valuable feedback you provided. We have carefully addressed your comments and are happy to present our responses, along with additional analysis to support our study.

General comments:

Rev2.1)

I do have major concern in the period split: why you choose 2007 as the threshold for the date division, please provide some explanation.

Response: According to Pärn et al. 2022, A new regime of sea ice began in 2008 and in all basins of the Baltic Sea, a rapid warming during spring could be detected. Hence the study period is divided into two halves. It is mentioned in paper as well "Based on the regime shift detected by Pärn et al. (2022) we split our study period into two parts: 1993/94–2006/07 (preceding period) and 2007/08–2020/21 (recent period)" at line 206.

Rev2.2)

Another concern is the data usage in SST_BAL_SST_L4_REP_OBSERVATIONS_010_016 (satellite product) and SEAICE_BAL_SEAICE_L4_NRT_OBSERVATIONS_011_004 (SAR & ice charts-based product). How do you consider the uncertainty in the satellite and SAR & ice charts-based product considering you use the satellite product to determine/correct the sea ice fraction threshold in the reanalysis data, it is important to know the accuracy or uncertainty of the satellite product.

Response: The sea ice data in the satellite product is based on Copernicus SI-TAC products (which includes Ice charts and SAR based data) and high resolution SMHI data [RD.5] (CMEMS-SST-PUM-010-016-040, Product user manual: https://doi.org/10.48670/moi-00156). Although Ice charts based sea ice data can certainly have uncertainties, these datasets have consistently been used in studies to validate other products (Karvonen 2021; Leppäranta 2023; Pärn et al. 2021; Mäkynen et al. 2020) and are available for a long term period.
The product SEAICE_BAL_SEAICE_L4_NRT_OBSERVATIONS_011_004 is validated against icebreaker ice thickness data in the QUID file of the product (QUID for SEA ICE TAC Products 011_004, 011_011, 011_019, Ref. CMEMS-SEAICE-QUID-004_011_019, issue 1.1).
Since there is not yet a better automatic tool for interpretation of sea ice in high resolution imagery, and due to the lack of field observations, ice charts are considered being the data sets that best describes the true state of the sea ice, and thus other sea ice data sets use these as reference.

(CMEMS-SI-QUID-011-001to007-009to014, QUALITY INFORMATION DOCUMENT For SI TAC Sea Ice products 011-001, -002, -004, -006, -007, -009, -010, -011, -012, -013, -014, issue 2.10)
Required changes will be made in the manuscript to make it more clear.

Rev2.3)
And when I look at the Table 1, I am also wondering how is the RMSE and Bias look like during 0.15 and 0.25? What about other thresholds, such as 0.18 or 0.23? And since you've showed two criteria for threshold selection, how do you coordinate them together, such as in RMSE, 0.20 reanalysis threshold has the lowest value while in Bias, 0.25 seems to have the lowest Value.

Response: Bias is used as the primary criteria to select optimal threshold, RMSE and CC are calculated to confirm the results. Results from other thresholds are given below. The selection is made out of traditionally used ice fraction thresholds which are generally factors of 0.05 like 0.15, 0.20 etc., especially considering 0.20 gives reasonably low bias.

TH_SIF (Satellite)	TH_SIF (Model)	Bias (km^2)	RMSE (km^2)	CC
0.15	**0.15**	**1783**	**4887**	**0.994**
0.15	0.16	1490	4648	0.995
0.15	0.17	1208	4439	0.995
0.15	0.18	935	4260	0.995
0.15	0.19	671	4110	0.995
0.15	**0.20**	**416**	**3991**	**0.995**
0.15	0.21	168	3900	0.996
0.15	0.22	-73	3837	0.996
0.15	0.23	-309	3803	0.996
0.15	0.24	-539	3795	0.996
0.15	**0.25**	**-764**	**3810**	**0.996**

Due to formatting, the negative sign for bias values corresponding to 0.25 model TH_SIF was missing in the table which will be added. Required changes will be made to reduce ambiguity for the selection criteria.

Rev2.4)

And I am quite lost in Line 151, when you mentioned, "TH_SIF of 0.15 for the model dataset, provides more accurate estimates of maximum SIE", can you provide more clearly and statistically evidences in why 0.15 the accurate estimate of maximum SIE is.

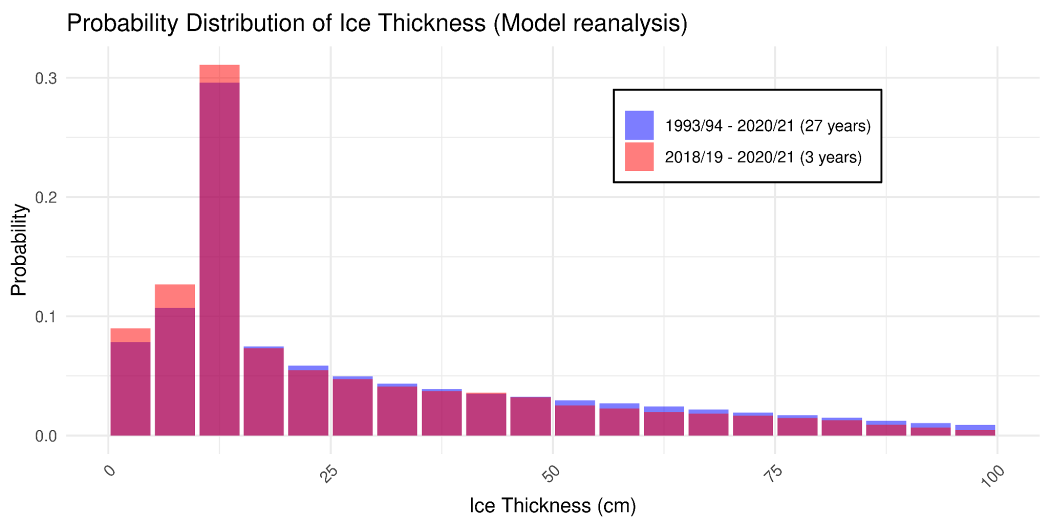
Response: It can be seen from figure 4 for max SIE. In terms of Statistics, Bias and RMSE are calculated between max SIE time series (similar to figure 3) for these two threshold combinations and shown below in the table. From the table, it is apparent that TH_SIF of 0.15 for the model dataset provides more accurate estimates of maximum SIE compared to TH_SIF of 0.20.

TH_SIF (Satellite)	TH_SIF (Model)	Bias (km^2)	RMSE (km^2)
0.15	0.15	-1165	14682
0.15	0.20	-5552	16143

Rev2.5)

In Section 4.2, when you are trying to correct the reanalysis sea ice thickness based on three years SAR images and ice chart product, I am not sure if it is statistically robust. Given that samplings for grid is large, but when you consider the annual changes, 3 years is quite short, and not long enough to support your ice thickness correction statements.

Response: Beside it being the common period available between the two datasets, the model reanalysis sea ice thickness data shows considerable variations within these 3 years and probability distribution plot shows similar thickness distribution for these 3 years compared to the complete 27 years reanalysis thickness data.

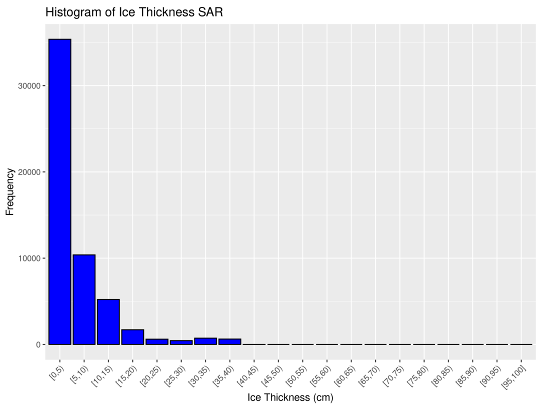


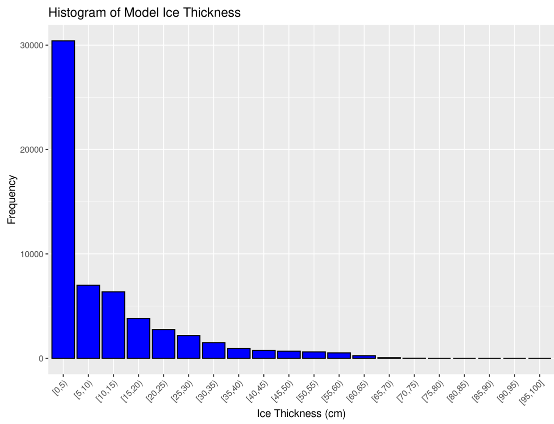
Required changes will be made to mention it, in the manuscript.

Rev2.6)

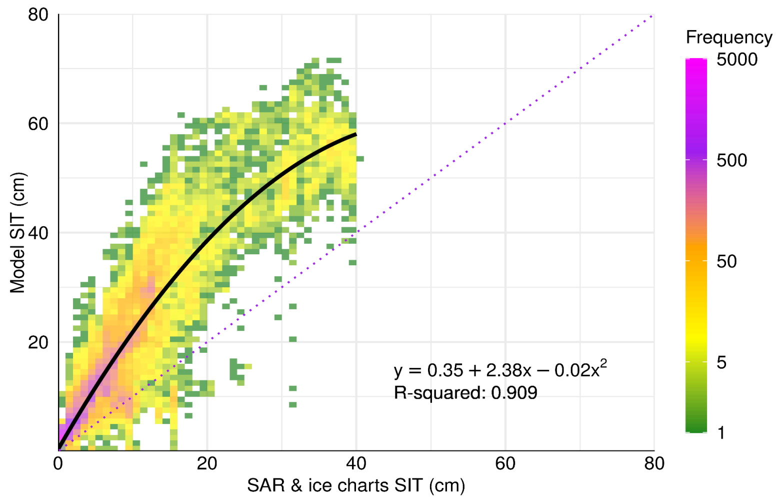
When I look at the Figure 5, it is quite obviously that Model SIT has the saturation stage in high value compared with the SAR images and ice chart. Then (1) how to explain this condition; (2) instead of the linear relationship, how about using the exponential lines to picture the fitting?

Response: (1) The SAR & ice chart based SIT does seem to get saturated at lower values compared to Model SIT, this could be related to how the data processing and filtering for the SAR & ice charts data is done. Required changes will be made to mention this drawback in the manuscript. The thickness distribution for the common data from SAR & ice charts based and model based data is given below.





(2) Using non linear fitting slightly improves the fit (Example for nonlinear fit shown below), but the difference of parameters such as R squared value is really small compared to linear fit. So to simplify the analysis, linear fitting is chosen (the below equation can be used to get a slightly better fit).



Rev2.7)

And I don't understand how to apply the correction coefficient in Line 168, did you overall divide the values by the 1.81?

Response: Yes, that is correct.

It will be mentioned in the manuscript.

Rev2.8)

My next concern is the motivation behind the assessment of the Baltic Sea ice product, which is missing in the Discussion section. For example, what are the limitations of using the current data? What insights can be provided to modelers for improving models? Which updates have improved the product performance compared to previous versions? The current discussion lacks depth and does not provide the audience and the community with sufficient information beyond the assessment Results.

Response: Suggested changes will be made in the discussion section of the manuscript.

Detailed comments:

Rev2.9)

Line 65, Dataset part: please provide detailed information on the temporal resolution and time span of the three products. Additionally, explain how you coordinate these products with different resolutions and specify the interpolation methods used.

Response: The products are interpolated to a common grid system, using bilinear interpolation. In case of satellite and model data, the common grid resolution used is 0.0277 x 0.02, while for Ice charts & SAR and model data, the common grid used is 0.0277 ° × 0.0166 ° (2 × 2 km).
Required description about temporal resolution and extent will be added in the manuscript.

Rev2.10)

Line 80, please fill in the reference.

Response: Reference "(Hersbach et al. 2020)" will be added in the manuscript.

Rev2.11)

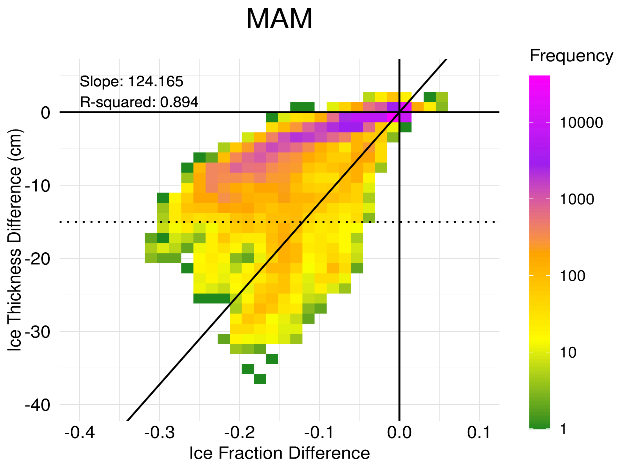
Figure 2, I suggest moving either panels (a) and (b) or (c) and (d) to the appendix, as they seem to replicate information.

Response: Suggested changes will be made in the manuscript.

Rev2.12)

Figure 10: specify the units in panels (a) and (d). I'm quite interested in the fitting process in panels (c) and (f). When focusing on the density plot, consider showing how the linear fitting looks when focusing on high-intensity values or averaging bin values, and then performing the linear fitting.

Response: Figure 10: panel (a) and (d) are ice fraction differences hence unitless, for panel (b) and (e) units are specified in figure caption. Figure for Linear fitting when focusing on high intensity values (ice thickness difference more than 15 cm) is shown below (For MAM density plot). The higher thickness difference values here correspond to mostly the Bothnian Bay basin where thickness changes have been the largest during spring (Fig. 10e).



Rev2.13)

Line 231: could you provide an interpretation of why the Gulf of Finland sub-basin exhibits the most significant reduction during the melting season compared to the freezing season?

Response: Although our study has not focused on the underlying reasons for these changes, it is evident that recent research has detected rapid warming during the spring season (Pärn et al. 2022). Given that the Gulf of Finland is a shallow basin, this warming could have contributed to the significant reduction in ice fraction observed during the melting phase (spring season). Required text will be added in the manuscript.

Rev2.14)

Figure 11 and 12, could you overlay the trend with the 95% significance level? For example, use stipples to indicate the 95% confidence level or plot only the trends that are above the 95% confidence level.

Response: Changes will be made to overlay the trend with the 95% significance level.

Rev2.15)

Figure 12(b): verify the values around 55°N, 21°E, and explain why this area shows the largest reduction in ice thickness.

Response: Around 55°N, 21°E (Curonian lagoon), the absolute trend values for ice thickness are really low for the area, but in terms of percentage of reduction the larger because the area has very thin ice and is separated by Curonian Spit from the Baltic Sea. But we are not concerned about this region in the study hence it is not mentioned in the manuscript.

Rev2.16)

Line 261, wrong reference format.

Response: Reference format for “Åström et al. 2024" will be corrected in the manuscript.