Author's point-to-point response on Referee Comment #1 to tc-2022-92. The reviewer comments appear in black. The responses are in blue and the proposed changes to manuscript are in *Arial*. L# refers to that in the track-changes file.

Introduction:

The current study presents a statistical method that uses SIC and SIT from a reanalysis dataset to construct a historical SIT dataset. The idea behind this is that the SIT of the source reanalysis dataset is not accurate in the melt season as no SIT measurements in the melt season are available to feed into that reanalysis dataset, and that incorporating statistical relationships between SIC and SIT leads to an improved SIT dataset. Detailed comparisons with in situ observations and other often-used SIT datasets show that the newly developed SIT performs well. In addition, assimilation runs are performed in which only SIC or both SIC and the newly constructed SIT dataset are assimilated, which are then used to initialize 7-day forecasts. The skill of forecasts initialized from the assimilation runs in which both SIC and the new SIT dataset are assimilated is shown to be higher than forecasts initialized from assimilation runs in which only SIC was assimilated. The analysis shown is detailed and interesting, but there are several major issues that the authors have to address before I can recommend publication.

Major comments:

1. Although I understand that the authors are not native English speakers, the English is poor, which makes it difficult to follow the text. I suggest the authors improve the language by consulting with an English native speaker.

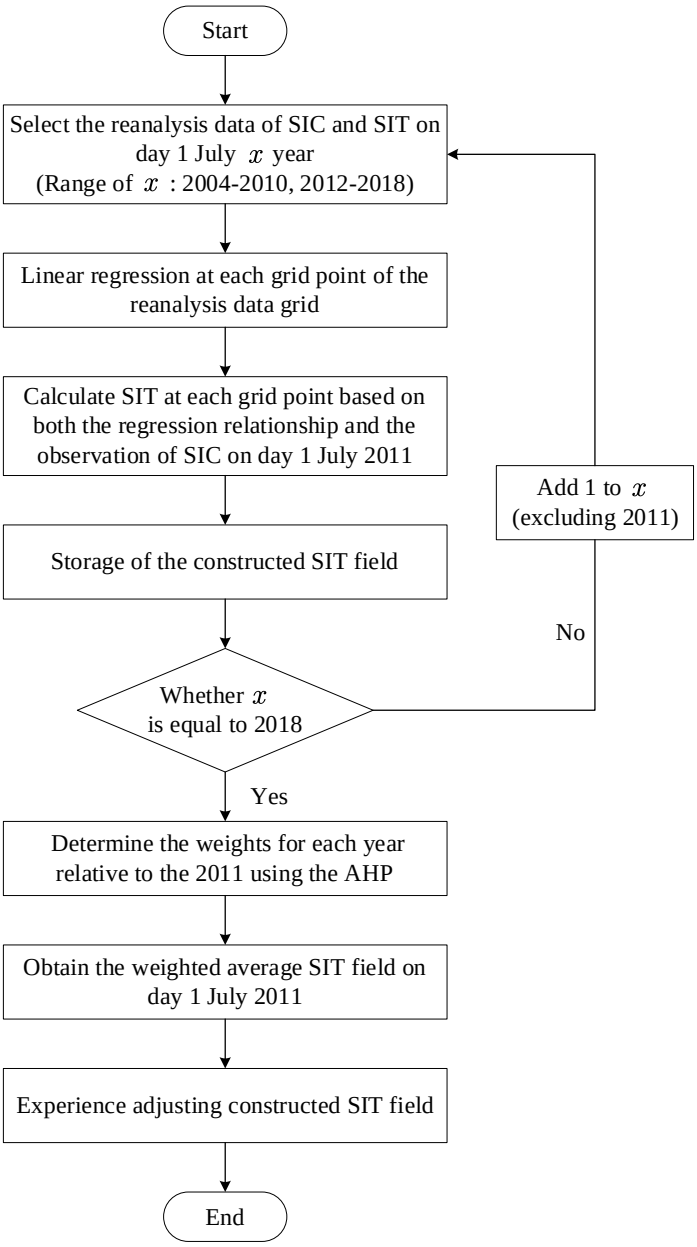
Thanks for your advice. We carefully modified the grammar of the sentence and the content of the manuscript in detail.

2. Perhaps it is due to my lack of expertise in the area, or the poor English (or a combination of these factors), but I do not fully understand the statistical model that is used to construct the SIT as described in section 3.1 and in figure 2. Other readers may

have similar problems and therefore this should be improved. In particular, I do not understand the 'linear regression for each grid point'. What is particularly confusing is that the authors write that (l. 145) 'the linear regression process is carried out at each grid point … for each year.' And (l. 147) 'the corresponding SIC-SIT regression ... can be obtained for each year'. This description suggest that the linear regression is done spatially for each year (i.e. regression of SIT at a location with SIC at all other locations in a fixed year), but other text later in the paper suggests the linear regression is done at a specific grid point over the time dimension (i.e. regression of SIT at a location with SIC at that same location over time). Figure 2 also suggests that the regression is done spatially for each year, but I don't think that is what the authors mean. Please clarify.

Thank you for pointing this out. The linear regression we proposed here is done at a specific grid point over the time dimension, and is not done spatially for each year. In fact, what L145 and L147 try to explain is that for the proposed sea ice thickness field on July 1, 2011, we need to carry out linear regression for each non null value point on the reanalysis grid on July 1 of each year (from 2004 to 2018, except 2011). In other words, the independent variable is the reanalysis sea ice concentration value of a grid point on July 1 of a certain year, and the dependent variable is the reanalysis sea ice thickness value of the corresponding grid point. In this way, the SIC-SIT regression relationship at each grid point can be obtained for each year. We modified it in the revised manuscript. In addition, what we want to express in Figure 2 is also the regression in the time dimension, not the spatial dimension. We modified the second step and the third step in Figure 2 in the revised manuscript.

"Then, starting from 2004, the linear regression process is carried out at each grid point (non null point) of the reanalysis data grid, with SIC of a grid point being as the independent variable and SIT of the grid point corresponding to SIC being as the dependent variable. The same linear regression process is performed in other years. The corresponding SIC-SIT regression relation at each grid point can be obtained for each year."



Revised Figure 2. A flow chart of bivariate regression model of SIT (using 1 July 2011 as an example).

3. The abstract should be improved, as I initially did not understand the method that the authors are introducing. I understood that the aim of the authors is to construct a historical SIT dataset, but they aim to do that based on gridded SIC and It should be explained more clearly how the SIT that is the input of the BRMT method differs from the SIT output. Also, the abstract contains several statements about improved performance, without specifying the baseline:

- 17: 'BRMT-constructed SIT is more accurate': more accurate than what?

- 19: ‘closer to observations’: closer than what?
- 21: ‘significantly improved’: compared to what?

The baseline should be specified. Finally, some more details on the forecasting experiments should be included in the abstract. In particular, it would be helpful to note that these pertain to 7-day forecasts (to contrast with seasonal forecasts that run for up to a year and in the context of which SIT initialization is often discussed).

Thanks for your suggestion.

First of all, SIC and SIT of TOPAZ4 reanalysis data are used to construct the statistical relationship between SIC and SIT in BRMT method. Then, satellite observed SIC data is used as input and BRMT-constructed SIT is the output. I’m not sure whether “the SIT that is the input of the BRMT method” you mentioned refers to the SIT of TOPAZ4. In fact, we compared the SIT of BRMT, the SIT of TOPAZ4 and the site SIT in Figure 5 and Figure 6 in Section 4.2. The results show that the SIT of BRMT is closer to the observation than that of TOPAZ4. In addition, according to your suggestion in the Minor comments 2, we added the forecast experiment results with both the SIT of TOPAZ4 and the SIC from satellite remote sensing as the initial fields in Section 6. We hope to show the difference between the SIT of TOPAZ4 and the SIT of BRMT in a more comprehensive way. The conclusion of this part is supplemented in the abstract.

Secondly, we modified the several statements about improved performance without specifying a baseline in the abstract.

Finally, we supplemented the conclusion of the forecast experiments in the abstract.

Minor comments (note: there are many more grammatical errors that I don't list below, see main comment #1):

1. Some key references are omitted and should be included: Dirkson et al 2015 (https://doi.org/10.1002/2015GL063930.1) develops 3 statistical methods to generate a SIT datasets, and Dirkson et al 2017 (https://doi.org/10.1175/JCLI-D-16-0437.1) shows that one of the 3 statistical methods leads to improved seasonal forecasts of SIC. The authors should include these references.

Thanks. We included quotations from these two papers in the introduction of Section 1.

"Dirkson et al. (2015) employed maximum covariance analysis to identify patterns of covariability between SIT and two predictors (SIC and lagged sea level pressure) by considering the thermodynamic and dynamic relationship among them, so as to realize the real-time estimation of SIT. On this basis, the prediction performance of sea ice area and regional SIC to some extent was improved by the SIT initialization field generated based on the further improved statistical model through extrapolation (Dirkson et al., 2017)."

2. The authors try to highlight the importance of the newly SIT dataset by comparing forecasts initialized from assimilation run in which it is used with forecasts initialized from assimilation runs in which it is not used. While this is interesting and worth reporting, it only highlights the importance of initializing SIT versus not initializing SIT. To investigate whether or not the newly developed SIT dataset provides additional value compared to other SIT datasets (e.g. that from the reanalysis dataset that it was derived from) in the context of forecasts, an additional set of forecasts would have to be presented in which an alternative SIT dataset is used for creating the initial conditions.

Thanks for your suggestion. The new forecast experiment in September 2011 with both the SIT of TOPAZ4 and the SIC from satellite remote sensing as initial fields is supplemented (named Exp_TOPAZ). Comparing the forecast results of the two

experiments, it can be seen that Exp_SIC&SIT is significantly closer to the observations than Exp_TOPAZ in terms of the RMSE between SIC forecast results and observations in the whole forecast period (Fig. 1), or in terms of 1d or 7d forecast results of sea ice extent (Fig. 2). In the comparison between the 1 d SIT forecast results of the two experiments and the four situ observation facilities (Table 1), the average absolute deviations between Exp_SIC&SIT and two observation facilities (BGEP_2011D and IMB_2011L) are 39% and 67% lower than that of Exp_TOPAZ, while the average absolute deviations between Exp_SIC&SIT and the other two observation facilities (BGEP_2011B and IMB_2011K) are 86% and 65% higher than that of Exp_TOPAZ, respectively. In general, the initial conditions provided by SIT from BRMT are expected to provide valuable reference results for Arctic sea ice forecast. The conclusion of this part is added to the discussion in Section 6.

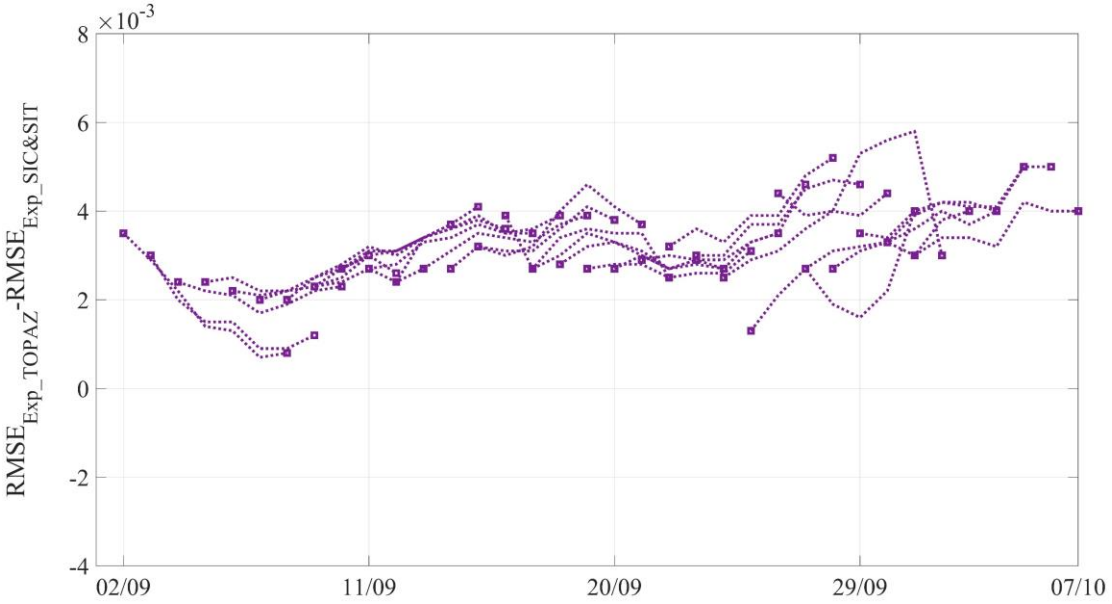


Figure 1. Differences between the RMSEs of sea ice concentration forecast results of Exp_TOPAZ and Exp_SIC&SIT relative to the SSMI observation during the period of 2 September to 7 October 2011 (each segment represents the 7 d forecast). Date format is dd/mm.

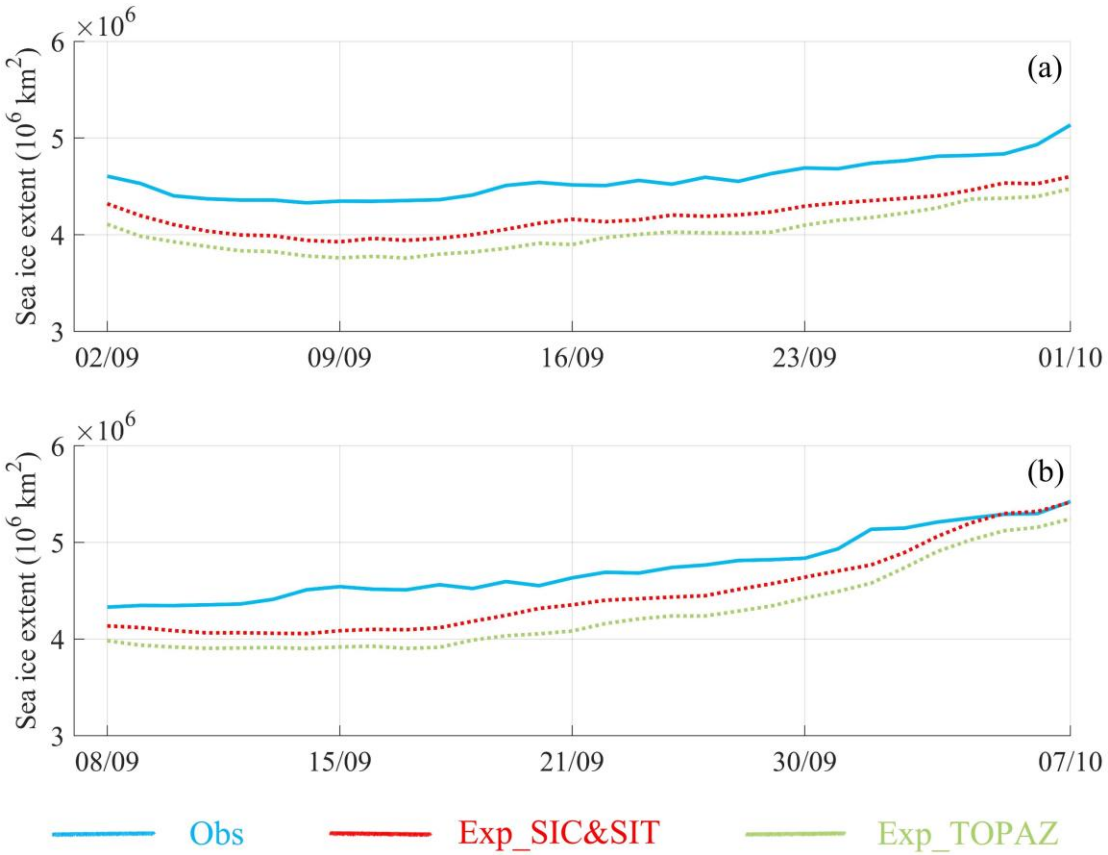


Figure 2. Comparison of sea ice extent (unit: 10^6 km^2) between 1 d (a) and 7 d (b) forecast results of Exp_SIC&SIT (red dotted line) and Exp_TOPAZ (light green dotted line) and the SSMI observations (solid blue line) in 2011, respectively. Date format is dd/mm.

Table 1. Average absolute deviations of sea ice thickness (m) between the 1d or 7d forecast results of Exp_SIC&SIT and Exp_TOPAZ and situ observations.

	1d		7d	
	Exp_SIC&SIT	Exp_TOPAZ	Exp_SIC&SIT	Exp_TOPAZ
BGEP_2011B	0.28	0.15	0.22	0.13
BGEP_2011D	0.32	0.53	0.27	0.43
IMB_2011K	0.43	0.26	0.39	0.24
IMB_2011L	0.14	0.43	0.12	0.35

3. 63,66: is à was

Agreed. We corrected this.

4. 255: will not à does not?

Agreed. We corrected this.

5. Figure 5 (bottom map): what do the colors represent?

The purpose of different mooring facilities marked by different colors is to make it easier to distinguish them in the figure. The colors have no specific meaning.

6. Figure 8: I suggest to use a non-linear scale for the Normalized standard deviation as a) the most interesting data is where normalized standard deviation is close to 1, and because most points are located there

Thanks for your suggestion. In order to better explain "In the part with both NSTD and NCRMSD less than 5.5" in Section 4.3, we added Figure S1 in the supplementary document. At the same time, this figure can more clearly show the dense data points in Figures 8a-8c, and also make the points with the normalized standard deviation close to 1 more obvious.

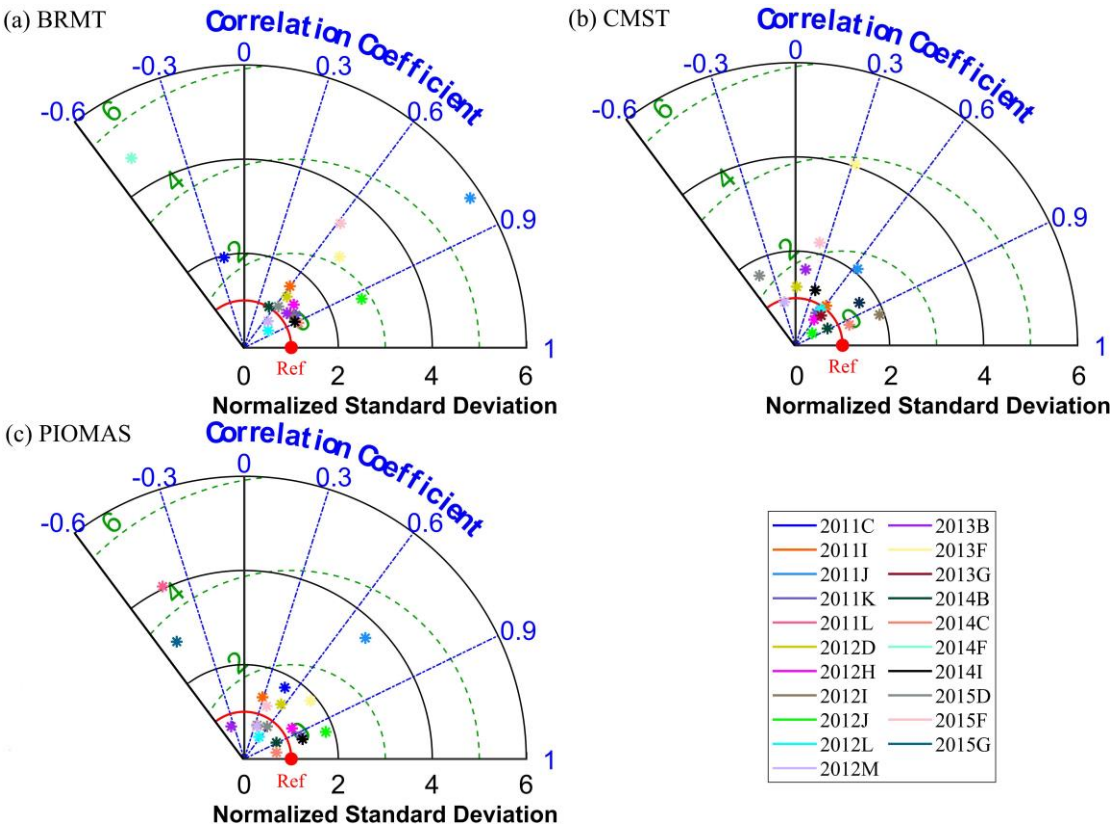


Figure S1. Taylor diagrams (the part of both NSTD and NCRMSD less than 5.5) of (a) BRMT, (b) CMST and (c) PIOMAS with respect to all available IMB buoy data during melting seasons from

2011 to 2015. The green dotted lines indicate the normalized CRMSD. The reference observations are indicated by Ref in red.

7. 365: 'In part with small deviation evaluation criteria value': not clear. In the following lines (including the quoted numbers for correlation coefficient), do you only use data points with 'small' standard deviation, and if yes, what is the cut-off value for 'small'?

Thank you for pointing this out. The small deviation evaluation criteria value means that both NSTD and NCRMSD are less than 5.5. In the following lines, we used the points with both NSTD and NCRMSD less than 5.5 to calculate the evaluation criteria. We changed *'In part with small deviation evaluation criteria value'* to *'In the part with both NSTD and NCRMSD less than 5.5 (Figure S1)'* in the revised manuscript.

8. 477 'significantly smaller': this is a bit hard to see from Fig. 13 as it is hard to compare panel a with panel b. Perhaps it would make sense to add a 3rd panel showing the difference between panel a and b? Also: it is not clear what the authors mean with 'a long-term stable effect' in l. 479

Thank you for pointing this out. We delated the Figure 13a and Figure 13b and added the figure (Figure 13 in the revised manuscript) showing the differences between the RMSEs of 1 d and 7 d sea ice concentration forecast results of Exp_SIC and Exp_SIC&SIT relative to the SSMI observation during the period of 2 September to 7 October in 2011-2013. It can be calculated from the Figure 13 that the average difference between the RMSEs of the 1 d forecast results of Exp_SIC and Exp_SIC&SIT is about 0.0043 in 2011, and the average difference in 2011 of 7 d is about 0.0038. These two values are about 0.0025 and 0.0029 in 2012, 0.0026 and 0.0028 in 2013 respectively. These indicate that the improvement of the SIT initial field not only plays a significant role in improving the forecast accuracy of SIC, but also this improvement will not weaken with the increase of forecast time. We changed the description in the revised manuscript.

9. Figure 14: it is hard to see what the authors refer to, as all the figures are so similar. Perhaps adding a contour line would help, but as it stands the current figure 14 does not add much to the paper. Figure 15 is much more informative

Thanks for your suggestion. Considering the low accuracy of SSMIS satellites in areas of low concentration or thin ice and the possibility that SSMIS may not be able to solve the problem of margin ice zones (MIZ), the MIZ analysis is deleted in the Section 5.2.1. In order to supplement the results of SIC, we added a new section to discuss the prediction results of sea ice extent (see Section 5.2.2 in the revised manuscript).

10. 530 ‘are largest’: except for Exp_Ctrl

Thank you for pointing this out. We changed this in the revised manuscript.

11. 533: ‘Variation law’: not sure what is meant with that

Thank you for pointing this out. We didn't describe it clearly. We think this should be “Variation trend”. It refers to that, according to Figure 10, the sea ice thickness of BRMT maintains the melting state at the end of September in the Beaufort Sea. We changed this in the revised manuscript.

Dirkson, A., Merryfield, W. J., Monahan, A: Real-time estimation of Arctic sea ice thickness through maximum covariance analysis, Geophys. Res. Lett., 42, 4869-4877, https://doi.org/10.1002/2015GL063930, 2015.

Dirkson, A., Merryfield, W. J., Monahan, A: Impacts of Sea Ice Thickness Initialization on Seasonal Arctic Sea Ice Predictions, J. Climate., 30, 1001-1017, https://doi.org/10.1175/JCLI-D-16-0437.1, 2017.