

## RESPONSE TO THE REVIEWERS

We express our sincere gratitude to the editor and both the reviewers for dedicating their time & thoughtful consideration, along with providing valuable feedback and suggestions for improvement of our manuscript.

Throughout this document, reviewer comments are indicated by plain blue text and our responses are presented in plain black text, addressing each comment individually.

### REVIEWER #1

This study evaluates Antarctic sea-ice thickness (SIT) and volume (SIV) from 39 CMIP6 models from 2002 to 2014, utilizing three sea-ice products, including satellite observations and two reanalysis datasets. It analyzes spatio-temporal variations in SIT and examines the seasonal co-variability between sea-ice area (SIA) and SIT, noting a negative relationship in summer and a positive winter one. Overall, the paper is well structured and relevant, though minor revisions are recommended for clarity and insight before publication in The Cryosphere.

#### **General Comments:**

1. The manuscript outlines the sea ice products in the data and methods section. However, the details regarding each product's parameters used in this study are inadequately described, and the process of calculating the SIV and SIA simultaneously. It is essential to provide a clearer explanation of the conversion procedure, and the parameters involved in obtaining the target variables. Ensure that all key variables (SIT, SIV, and SIA) are clearly defined in the calculation process in the manuscript.

Thank you for your comment.

We have updated our manuscript, and it now clearly outlines the details regarding each of the used sea-ice product. It provides clearer explanations of the conversion procedure and the parameters involved in obtaining the target variables such as SIA and SIV. The following lines and text have been added under each dataset section to address the above comment:

**Lines 113-118:**

“To eliminate mismatch in the spatial resolution, all the sea-ice products were regridded onto a common CMIP6 model grid. SIA is calculated by multiplying the monthly values of sea-ice concentration (SIC) by the corresponding grid cell area and summing over the Southern Hemisphere. SIV was computed as the product of the actual floe thickness, SIC and the grid cell area, summed over the circum-Antarctic region.”

**Lines 140-144:**

For the satellite product, SIA is calculated by multiplying the SIC with the corresponding grid cell area, both derived from the reanalysis dataset provided by the National Snow and Ice Data Center (NSIDC; Comiso, 2017) for the period 2002-2014. SIV is then computed by multiplying the resulting SIA with the satellite-derived floe thickness and summing over the entire Southern Hemisphere.

**Lines 176-179:**

For the reanalysis/synthesis datasets, SIA is calculated as the product of the SIC and the corresponding grid cell area obtained from the respective datasets. The SIA value so obtained is then multiplied with the “effective thickness” from each of the reanalysis/synthesis dataset to obtain their respective SIV.

2. The analysis shows that CMIP6 models better match satellite observations at thickness maxima (February) and reanalysis at minima. I recommend that the authors discuss potential causes, for example, whether the models’ parameterizations, snow-ice interactions, or ice growth processes could be driving the observed discrepancies.

Thank you for your comment. Our updated manuscript now includes a new section (Sec. 4), Discussions, which discusses the details and potential causes for why models simulate SIT differently across different seasons/months.

The following section has been added to address the above comment. **Lines 536-604:**

#### 4. Discussion

Antarctic sea-ice thickness (SIT) exhibits a seasonal cycle different from other surface sea-ice parameters with its maxima observed in February, when only the thickest multi-year ice survives the melt season. CMIP6 models generally capture this behavior, successfully simulating peak thickness in February. This agreement at summer maxima suggests that model thermodynamics allow a remnant of thick ice to persist through austral summer, similar to observations. During winter growth, however, Antarctic ice remains relatively thin on average. The thickness minimum occurs in late fall or early winter, depending on the type of SIT dataset.

Besides the thermodynamic factors, the observed discrepancies in the CMIP6 models with different sea-ice products can be due to a variety of factors such as model parameterizations and tuning, snow ice interactions or other sea-ice processes. CMIP6 models generally capture the spatial pattern of thicker coastal ice versus thinner ice elsewhere, but they tend to underestimate the absolute thickness in these deformed-ice regions (~58% less than satellite-based estimates; Hou et al., 2024). This bias points to limitations in model dynamics: if ridging schemes or ice strength parameters are not adequately representing pressure ridge formation, models will produce too little thick ice. Moreover, the ice thickness distribution in models is sensitive to mechanical parameters—e.g. the compressive strength and the number of thickness categories—which control how readily ice converges into thick ridges (Nie et al., 2023). Numerical experiments confirm that increasing the number of ice thickness categories or a higher horizontal model resolution improves the simulation of thick ice: more categories allow a better representation of high-thickness tails from ridging and finer grids reduce artificial numerical diffusion of ice thickness.

Model parameter tuning can also impact thickness. Sea-ice models are often tuned to get realistic ice extent or concentration, which can indirectly affect SIT. For instance, adjusting albedo or ice strength to match observed ice area might cause models to grow ice either too thick or too thin as a side effect. Similarly, the GIOMAS reanalysis (with  $\sim 0.8^\circ$  grid and concentration assimilation) is also known to underestimate thickness in deformed ice areas, partly due to smoothing and assimilation removing thick ice (Liao et al., 2022). This underestimation is echoed in many CMIP6 models. Studies of Arctic sea-ice in models have shown that tuning ice strength and improving ice drift physics can reduce biases in ice thickness and extent. In the Antarctic, this means capturing

features like the persistent Weddell Gyre (which retains most of the multi-year ice) and properly accounting for wave-induced ice break-up and floe size effects in the outer pack. Improving sea-ice dynamics in models (ridging parameterizations, drag coefficients, and interactions with ocean currents) is essential for simulating the observed thick ice in convergence zones. Model biases in atmosphere or ocean state can also translate into SIT biases, depending on how the sea-ice model is tuned. Therefore, diagnosing thickness biases requires careful attention to model configuration: ice strength parameters, snow thermal conductivity, melt pond and albedo schemes (both in the Arctic and Antarctic), and ocean mixing coefficients all may be tuned to some degree. Going forward, studies recommend reducing the need for ad-hoc tuning by improving physical realism—for example, ensuring accurate coupling of ice with atmosphere/ocean momentum and heat fluxes (Nie et al., 2023). In Arctic simulations, better coupling (e.g. improved ice–ocean drag and air–ice drag) was shown to reduce sea-ice bias, which may apply in the Antarctic as well.

High snowfall rates on thin sea-ice can lead to snow loading and flooding: the weight of snow depresses the ice floe, allowing seawater to inundate the snow layer, which subsequently freezes into snow-ice. Observations indicate that the snow-ice formation occurs across all Antarctic sectors during the growth season, contributing significantly to total ice thickness (Maksym and Markus, 2008). Climate models that include this process can build additional ice mass on top of the original ice, thickening the ice cover despite the insulating effect of snow. If a model omits or poorly parameterizes snow-ice formation, it will likely underestimate SIT, especially in heavy snowfall regions. At the same time, a deep snowpack insulates the ice from cold air, reducing basal ice growth. The net effect on SIT is a balance: more snow can mean slower growth, but also more snow-ice added. Model sensitivity studies confirm that uncertainties in snow processes strongly affect simulated SIT. For instance, altering the prescribed snow depth or thermal properties can change modeled SIT by about 10cm to nearly a meter (Nie et al., 2023). As mentioned in Sect. 2.1, snow cover is also a key factor in satellite thickness retrievals, which helps explain why models compare differently to satellite observations versus reanalyses in opposite seasons.

The presence of model deviations can hamper our understanding of climate-sea-ice interactions as well as biological feedback between the oceans and climate. For instance, due to the existing relationships between SIT and sea-ice motion, biases in the simulated thickness will also affect the

dynamics in the models which in turn will impact our understanding of the overall Antarctic sea-ice trends in the models (Lecomte et al., 2016; Sun and Eisenman, 2021). Additionally, lower SIT could create the misleading impression of lower albedo and increased light penetration, subsequently leading to increased Primary Production (Jeffery et al., 2020). Our study does not explore the reasons behind such continued biases in CMIP6. However, there are a variety of potential explanations (besides ones explained above) which may include cloud effects (Kay et al., 2016; Zelinka et al., 2020), spatial resolution that does not permit eddies, which are understood to be highly important for representation of Southern Ocean dynamics (Poulsen et al., 2018; Rackow et al., 2019), models lacking grounded icebergs as landfast ice (Fraser et al., 2023), biases in Southern Ocean stratification (Martinson and Iannuzzi, 1998), and temperature (Luo et al., 2023) and, the lack of coupled ice sheet interactions, which have relevance for the entire Antarctic climate system (Bronselaer et al., 2018; Golledge et al., 2019; Purich & England, 2023).

5. The description and analysis of Figure 2 are confusing. The manuscript mentioned that “we observe positive trends in SIT/SIV during the cooler seasons, which are absent in SIA.” However, the trend in the description of Figure 2 is not clear regarding which data’s trend is being referenced. Meanwhile, it seems that there are also positive trends in SIA during the cooler seasons, as indicated in the last figure in the fourth line.

Thank you for pointing out the confusion in the text. We have updated the lines in the section to avoid any further confusion with our statements. See the text below. Furthermore, we have also updated the figure, caption and serial numbers of Figure 2 to improve its clarity.

**Lines 278-285:**

Next, we look at the seasonal trends in the anomalies of the three sea-ice variables (SIT, SIV and SIA) across different CMIP6 models and the sea-ice products. In general, CMIP6 models simulate negative trends in Antarctic SIA and extent (Roach et al., 2020 and Shu et al., 2020), which contrast with the observed positive trend until mid-2015 (Li et al., 2023; Shu et al., 2015; Turner et al., 2013). Figure 2 demonstrates a similar pattern in the simulated SIT/SIV with a negative trend noticed across all the sea-ice variables. However, differences emerge in seasonal trends for

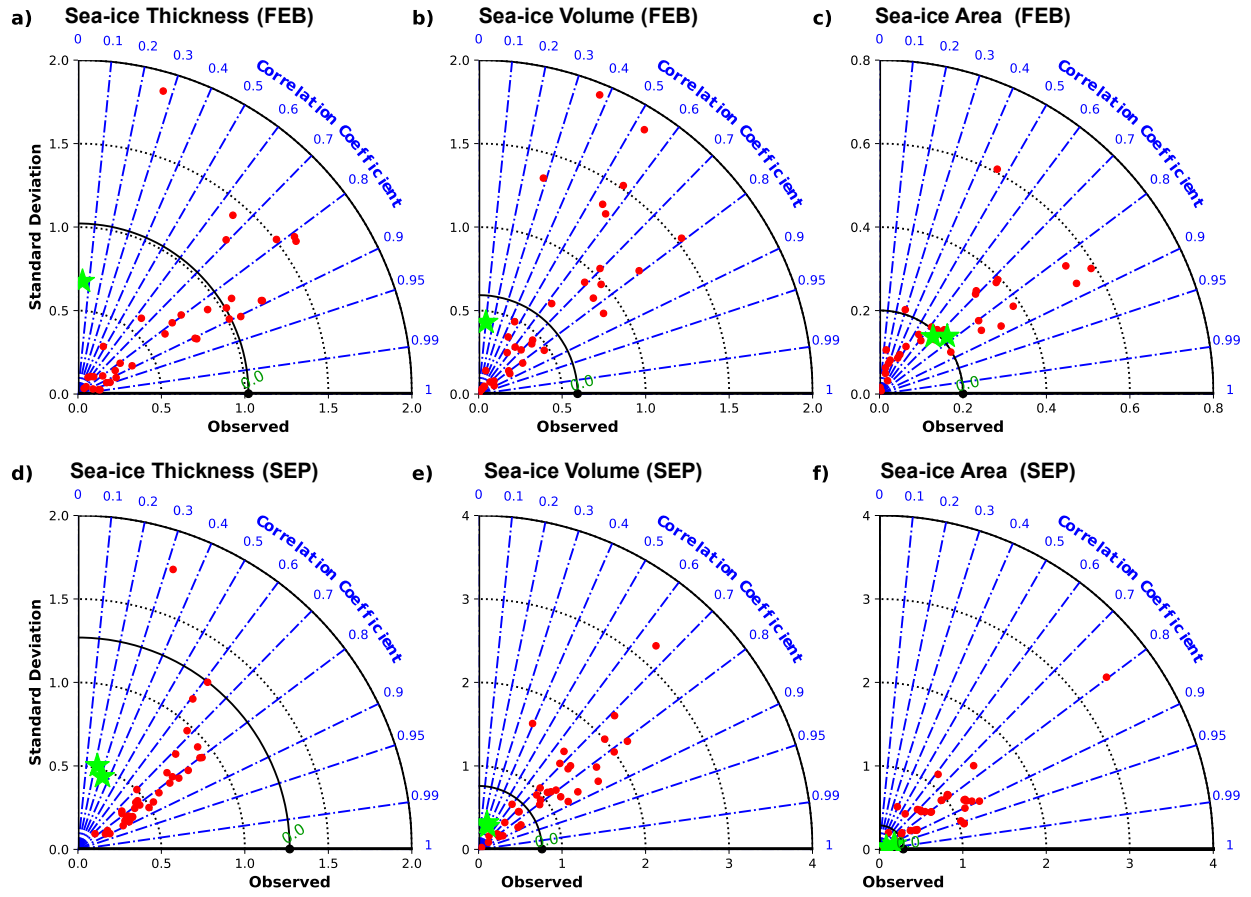
different sea-ice products (blue lines for GECCO3; green for GIOMAS and red for the satellite products in Fig.2) starting in the early 2000s.

In particular, significant positive trends in SIT/SIV are observed during the cooler seasons (winter and spring; Fig.2g-h,j-k), while no such trends (or very weak trends in SIV) are detected in the warmer seasons (Fig.2a-b,d-e). For SIA, small positive trends are observed during the warmer seasons (Summer and Fall), consistent with the previous studies (Martinson, 2012; Eayrs et al., 2019), while these trends are markedly reduced during the cooler seasons. This reduction is attributed to the thermodynamic constraints imposed by the Southern Ocean polar front, which limits ice edge expansion during the colder months. The observed positive trends in SIT/SIV during winter and spring imply that changes in Antarctic SIT may contribute to variability in total sea-ice mass/volume during these colder months. This could be linked to the presence of robust land-ocean temperature gradients during winters, which may result in high-intensity winds, one of the significant contributors to SIT/SIV fluctuations in the Southern Ocean (Zhang, 2014).

#### **Minor Comments:**

Some figures do not have specific sub-diagram serial numbers, which makes it impossible to properly match the description with the diagram when describing the diagram (e.g., Figure 2). It is recommended to add an ordinal number to each subgraph as well as additional annotations such as legend. Meanwhile, the quality of some figures needs to be optimized. For example, the black squares in Figure 3 are not clearly shown.

Thank you for pointing out this detail. Figures 2 and 3 and their captions are updated based on the comments provided.



**Figure 3: Taylor Diagrams representing spatial correlation and standard deviation using time-averaged means between CMIP6 models and different sea-ice products. For each model and synthesis product, two statistics are plotted: the Pearson correlation coefficient, related to the azimuthal angle (blue contours); and the standard deviation viz proportional to the radial distance from the origin (black dotted contours). Black solid contour corresponds to the standard deviation of the reference dataset. Red dots represent individual CMIP6 models and green stars represent synthesis datasets. The period used for comparison is 2002-2014 for February (a,b,c) and September (d,e,f). For February (a and b), GECCO3 is not included as it had very small negative correlation coefficients. Reference datasets used for SIT/SIV and SIA are Envisat-CryoSat-2 and NSIDC, respectively. Higher correlation coupled with a lower RMSD indicates greater accuracy of CMIP6 models in simulating the sea-ice variables.**

## **REVIEWER #2**

This study provides a systematic evaluation of Antarctic sea ice simulations in CMIP6 models by integrating satellite observations and reanalysis data. The paper is well-structured, employs rigorous methodology, and presents comprehensive data analysis. It contributes meaningfully to the field of Antarctic sea ice modeling.

### **Below are my detailed review comments:**

1. The manuscript mentions GIOMAS and GECCO3 reanalysis products but does not explicitly clarify whether these products assimilate observational data (e.g., sea ice concentration, satellite retrievals).

Thank you for pointing it out. We have added product details for both the reanalysis/synthesis products (GIOMAS and GECCO3). The updated lines now provide information about the data assimilations used in the above sea-ice products.

### **Lines 155-173:**

GECCO3 ocean synthesis is an improved version of GECCO2 based on MITgcm which employs the adjoint method to fit the model to a large variety of data over a multidecadal period (1948-2018). Unlike GECCO2, it is optimized over only one assimilation window. GECCO3 has 40 levels and uses the horizontal and vertical grid of the ocean component of MPI-ESM in the MR/HR configuration, providing a global eddy-permitting synthesis at a nominal resolution of  $0.4^\circ$  (Köhl, 2020). While GECCO3 does not directly assimilate sea-ice data, its ocean-ice coupling is influenced by the model's oceanic state. Hence, this synthesis primarily assimilates oceanic data, including temperature and salinity (EN4.2.1), along-track sea-level anomalies (AVISO), and sea surface temperature (HadISST), using a 4D-VAR (adjoint) method (Köhl, 2020).

GIOMAS uses the Parallel Ocean Model coupled with a 12-category thickness and enthalpy distribution ice model at a horizontal resolution of  $0.8^\circ$  (Zhang & Rothrock, 2003). In GIOMAS, the modeled SIC is nudged by assimilating satellite derived SIC from the Special Sensor Microwave Imager (SSM/I) launched by the Defense Meteorological Satellite Program (Weaver et al., 1987). Subsequently, other modeled variables including SIT are adjusted accordingly. This



process reduces the root-mean-square difference and improves the correlation between modeled SIT and observed SIT, while also causing the thinning of the mean SIT. This assimilation method has demonstrated good agreement of modeled SIT with satellite observations in the Arctic (Lindsay & Zhang 2006) and is useful for studying long-term variations in Antarctic sea-ice (Liao et al., 2022; Shi et al., 2021).

2. Lines 211–213 mention that the IPSL and EC-Earth3 models produce anomalously thick ice, but there is no explanation of why these models behave differently (e.g., ocean/ice dynamics, resolution, parameterizations). I suggest adding a paragraph comparing key models, such as IPSL vs. CESM2, to highlight how differences in physical schemes (e.g., ridging, thermodynamics) lead to divergent sea ice thickness simulations.

Thank you for the comment. As suggested, our updated text includes a paragraph which outlines potential reasons for the different SIT simulations in IPSL and EC-Earth3 and CNRM models.

**Lines 231-243:**

A few models, namely the IPSL and EC-Earth3-models (characterized by significant warm Southern Ocean biases; Döscher et al., 2022), exhibit anomalously thick sea-ice (>3m) in February. This can be attributed to the dynamically-related reasons causing the drift and deformation of the ice. This thick ice is captured in these models mainly due to the well-resolved wind patterns (Vancoppenolle et al., 2009) which result in the sea-ice drift and accumulation of the multi-year ice which survives the summer melt and is found mainly in the Weddell Sector during February. Additionally, in models with NEMO-LIM3 sea-ice modules, such as IPSL- and the EC-Earth3-models, the frazil ice formation dominates the total ice growth which may again lead to an accumulated SIT during February (Lie et al., 2021). By contrast, CNRM-models exhibit anomalously low thickness throughout the year. This is primarily due to high negative biases in their SIC simulations and their inability to realistically simulate thick sea-ice during the austral summer in the Weddell Sea (Voldoire et al., 2019).

Additionally, we have now added a new section (Sec. 4: Discussions) which outlines the detailed description of the potential reasons for seasonal differences in SIT simulations in different CMIP6 models.

3. Although the paper discusses biases in CMIP6 models when simulating Antarctic sea ice, the analysis of the underlying causes of these biases is somewhat brief. I recommend a more detailed discussion of the sources of model discrepancies, particularly with respect to how different models perform in different regions and the specific impacts these factors have on sea ice simulations.

Thank you very much for the inquiry. Our updated manuscript now includes a new section (Sec. 4), Discussions, which discusses the details and potential causes for why models simulate SIT differently across different seasons/months.

Following section has been added to address the above comment. **Lines 536-604:**

#### **4. Discussion**

The Antarctic sea-ice thickness (SIT) exhibits a seasonal cycle different from other surface sea-ice parameters with its maxima observed in February, when only the thickest multi-year ice survives the melt season. CMIP6 models generally capture this behavior, successfully simulating peak thickness in February. This agreement at summer maxima suggests that model thermodynamics allow a remnant of thick ice to persist through austral summer, similar to observations. During winter growth, however, Antarctic ice remains relatively thin on average. The thickness minimum occurs in late fall or early winter, depending on the type of SIT dataset.

Besides the thermodynamic factors, the observed discrepancies in the CMIP6 models with different sea-ice products can be due to a variety of factors such as model parameterizations and tuning, snow ice interactions or other sea-ice processes. CMIP6 models generally capture the spatial pattern of thicker coastal ice versus thinner ice elsewhere, but they tend to underestimate the absolute thickness in these deformed-ice regions (~58% less than satellite-based estimates; Hou et al., 2024). This bias points to limitations in model dynamics: if ridging schemes or ice

strength parameters are not adequately representing pressure ridge formation, models will produce too little thick ice. Moreover, the ice thickness distribution in models is sensitive to mechanical parameters— e.g. the compressive strength and the number of thickness categories —which control how readily ice converges into thick ridges (Nie et al., 2023). Numerical experiments confirm that increasing the number of ice thickness categories or a higher horizontal model resolution improves the simulation of thick ice: more categories allow a better representation of high-thickness tails from ridging and finer grids reduce artificial numerical diffusion of ice thickness.

Model parameter tuning can also impact SIT. Sea-ice models are often tuned to get realistic ice extent or concentration, which can indirectly affect SIT. For instance, adjusting albedo or ice strength to match observed ice area might cause models to grow ice either too thick or too thin as a side effect. Similarly, the GIOMAS reanalysis (with  $\sim 0.8^\circ$  grid and concentration assimilation) is also known to underestimate SIT in deformed ice areas, partly due to smoothing and assimilation removing thick ice (Liao et al., 2022). This underestimation is echoed in many CMIP6 models. Arctic studies have shown that tuning ice strength and improving ice drift physics can reduce biases in ice thickness and extent. In the Antarctic, this means capturing features like the persistent Weddell Gyre (which retains most of the multi-year ice) and properly accounting for wave-induced ice break-up and floe size effects in the outer pack. Improving sea-ice dynamics in models (ridging parameterizations, drag coefficients, and interactions with ocean currents) is essential for simulating the observed thick ice in convergence zones. Model biases in atmosphere or ocean state can also translate into SIT biases, depending on how the sea-ice model is tuned. Therefore, diagnosing SIT biases requires careful attention to model configuration: ice strength parameters, snow thermal conductivity, melt pond and albedo schemes (both in the Arctic and Antarctic), and ocean mixing coefficients all may be tuned to some degree. Going forward, studies recommend reducing the need for ad-hoc tuning by improving physical realism— for example, ensuring accurate coupling of ice with atmosphere/ocean momentum and heat fluxes (Nie et al., 2023). In Arctic simulations, better coupling (e.g. improved ice–ocean drag and air–ice drag) was shown to reduce sea-ice bias, which may apply in the Antarctic as well.

High snowfall rates on thin sea-ice can lead to snow loading and flooding: the weight of snow depresses the ice floe, allowing seawater to inundate the snow layer, which subsequently freezes

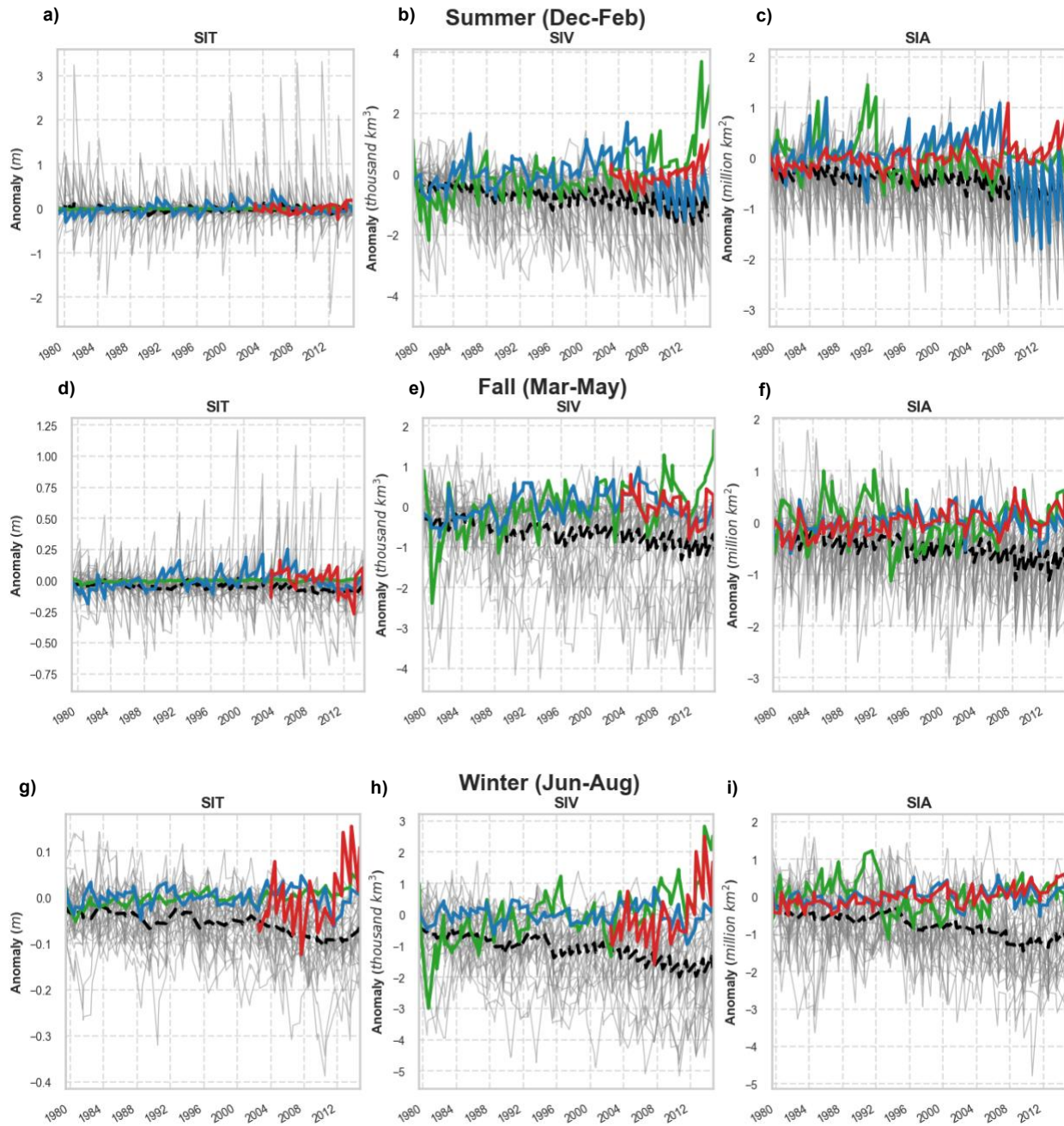
into snow-ice. Observations indicate that the snow-ice formation occurs across all Antarctic sectors during the growth season, contributing significantly to total ice thickness (Maksym and Markus, 2008). Climate models that include this process can build additional ice mass on top of the original ice, thickening the ice cover despite the insulating effect of snow. If a model omits or poorly parameterizes snow-ice formation, it will likely underestimate SIT, especially in heavy snowfall regions. At the same time, a deep snowpack insulates the ice from cold air, reducing basal ice growth. The net effect on thickness is a balance: more snow can mean slower growth, but also more snow-ice added. Model sensitivity studies confirm that uncertainties in snow processes strongly affect simulated thickness. For instance, altering the prescribed snow depth or thermal properties can change modeled SIT by about 10cm to nearly a meter (Nie et al., 2023). As mentioned, snow cover is also a key factor in satellite thickness retrievals, which helps explain why models compare differently to satellite observations versus reanalyses in opposite seasons.

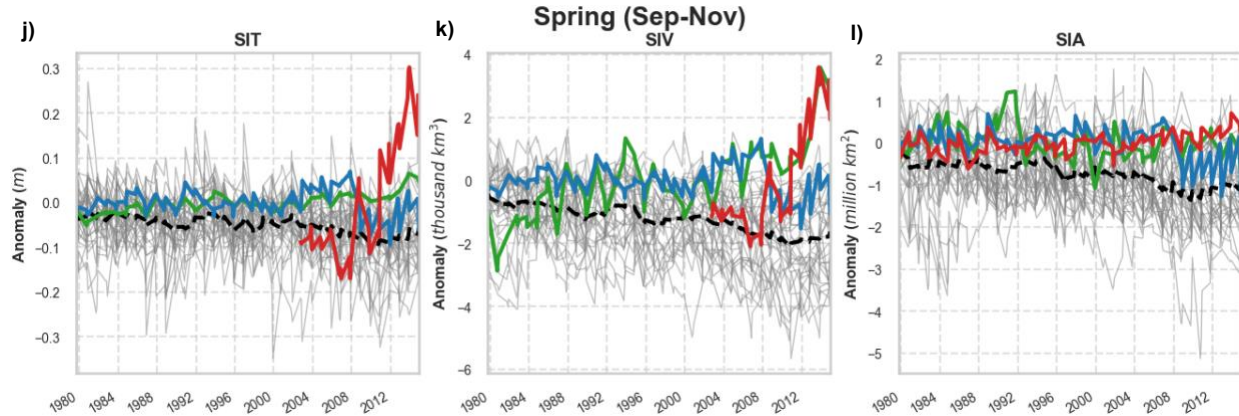
The presence of model deviations can hamper our understanding of climate-sea-ice interactions as well as biological feedback between the oceans and climate. For instance, due to the existing relationships between SIT and sea-ice motion, biases in the simulated thickness will also affect the dynamics in the models which in turn will impact our understanding of the overall Antarctic sea-ice trends in the models (Lecomte et al., 2016; Sun and Eisenman, 2021). Additionally, lower SIT could create the misleading impression of lower albedo and increased light penetration, subsequently leading to increased Primary Production (Jeffery et al., 2020). Our study does not explore the reasons behind such continued biases in CMIP6. However, there are a variety of potential explanations (besides ones explained above) which may include cloud effects (Kay et al., 2016; Zelinka et al., 2020), spatial resolution that does not permit eddies, which are understood to be highly important for representation of Southern Ocean dynamics (Poulsen et al., 2018; Rackow et al., 2019), models lacking grounded icebergs as landfast ice (Fraser et al., 2023), biases in Southern Ocean stratification (Martinson and Iannuzzi, 1998), and temperature (Luo et al., 2023) and, the lack of coupled ice sheet interactions, which have relevance for the entire Antarctic climate system (Bronse laer et al., 2018; Golledge et al., 2019; Purich & England, 2023).

5. The right column of Figure 2 (SIA anomalies) excludes GECCO3 and GIOMAS, unlike the left and middle columns (SIT/SIV). I suggest either adding the SIA data from

GECCO3 and GIOMAS or providing justification for their exclusion (e.g., NSIDC is the primary SIA reference).

Thank you for this keen observation and your comment. We have updated our figure which now includes SIA data from GECCO3 and GIOMAS.

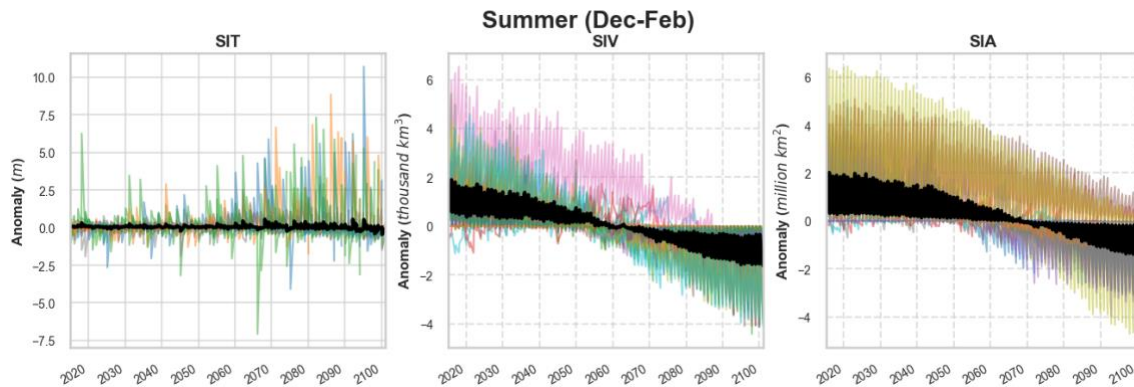




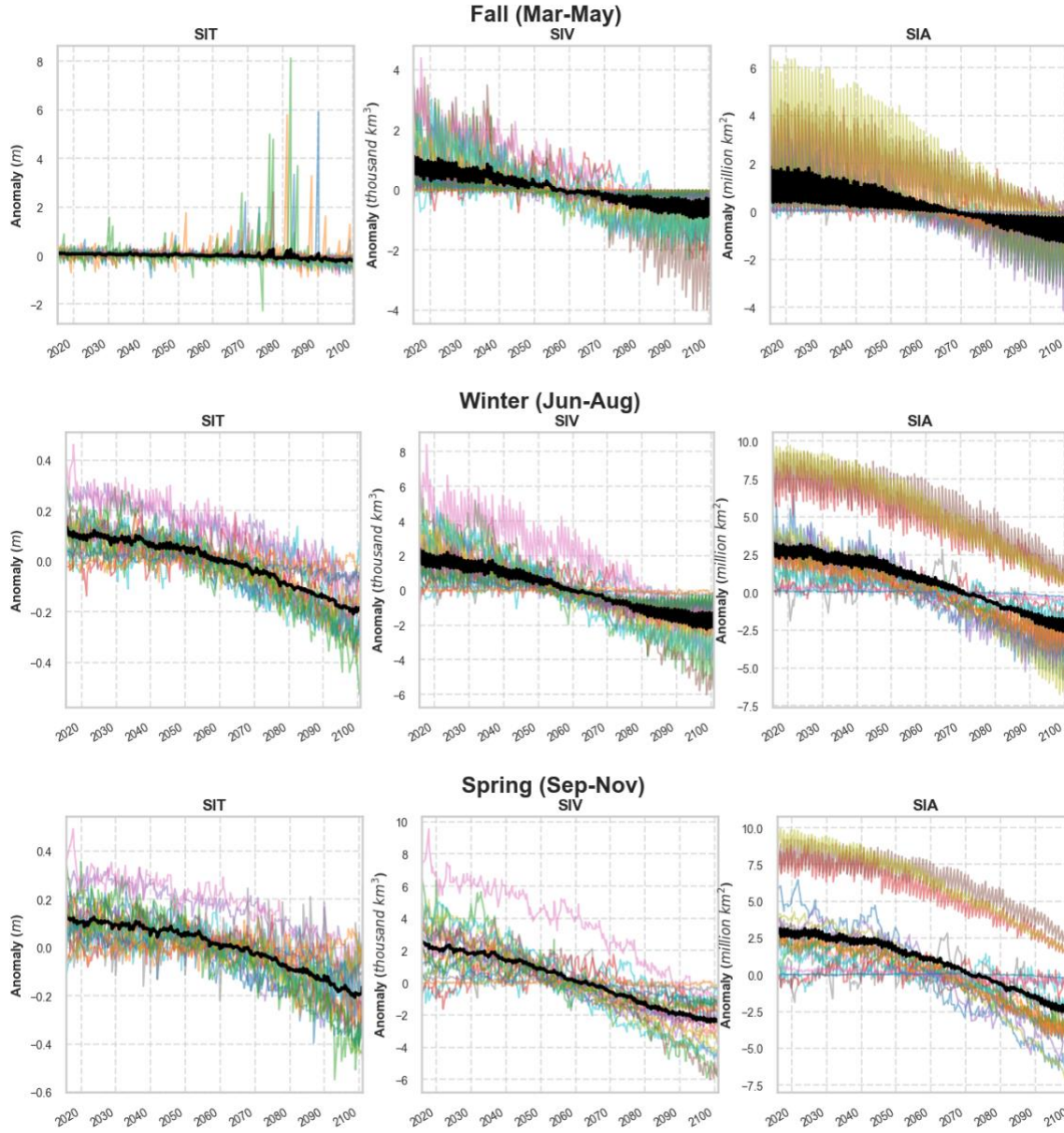
**Figure 2: Anomalies for four seasons: Summer and Fall (Warm Seasons; a-f), Winter and Spring (Cold Seasons; g-l) of SIT (left), SIV (middle) and SIA (right) of the circumpolar Antarctic. All the CMIP6 models are shown as grey lines, Multi-model mean in dashed line, GECCO3 in blue, GIOMASS in green, and NSIDC/ Envisat-CS-2 in red. All the sea-ice products extend for the time-period 1979-2014 except for satellite products which are only available for 2002-2014.**

6. The analysis stops at 2014, while Antarctic sea ice has shown a significant decline since 2015 (Raphael & Handcock, 2022). It would be helpful to briefly discuss whether CMIP6 models (e.g., SSP scenarios) capture this reversal, even if reanalysis data does not include post-2014 information.

Thank you for your keen interest in the analysis during the future scenarios. We have included a supplementary figure for the same (Fig. S1) and have discussed it in our manuscript as well.







**Figure S1: Anomalies for four seasons: Summer and Fall (Warm Seasons), Winter and Spring (Cold Seasons) of SIT (left), SIV (middle) and SIA (right) of the circumpolar Antarctic. All the CMIP6 models are shown as colored lines, and the Multi-model Mean in black line. The time-period is 2015-2100 for the SSP585 scenario. Since sea-ice products end in 2014, they could not be included here. Note that 3 models (namely ACCESS-CM-2, ACCESS-EM2-1 and CESM2-WACCM show slightly higher anomalies (especially for SIA), particularly evident during the cooler seasons (lower two panels).**

**Lines 297-303:**

We further examined seasonal anomalies in sea-ice variables under the high-emission SSP5-8.5 scenario for the period 2015–2100 (Fig. S1). The results are consistent with those in Fig. 2, showing pronounced seasonal trends in SIT, particularly during the cooler seasons (winter and spring). Under the warmer scenario, all CMIP6 models project a significant decline in SIT during these cooler months, while trends in SIT are largely absent during the warmer seasons (summer and fall). In contrast, SIV and SIA exhibit a persistent year-round decline, with negative anomalies becoming more pronounced from approximately 2060 onward (Fig. S1).

7. The Taylor diagrams in Figure 3 use “satellite products” as the reference, but it is unclear whether Envisat-CryoSat-2 is averaged or used separately. I recommend explicitly stating which dataset is used as the reference for each variable and justifying any averaging.

Thank you for this clarification comment. We now explicitly state in our text how the reference datasets are used for Taylor Diagram calculations. Additionally, we have also updated Fig.3 and its caption to reflect the same.

**Lines 312-315:**

We use the satellite dataset as the observation reference for SIT and SIV, calculating RMSDs and correlation values across the grids. For SIA, the reference dataset used is NSIDC. All the calculations were performed based on area-integrated spatial averages of reference datasets of SIT, SIV and SIA over the circum-Antarctic for February and September.

8. Section 3.4 relies on GIOMAS as the primary reference due to satellite biases in September. However, GIOMAS shows large errors in February (Figure 1c, early maxima). I suggest acknowledging the limitations of GIOMAS and cross-validating key results with GECCO3 and satellite data where possible.

Thank you for the comment. We have updated the first paragraph of Sec. 3.4 to accommodate the above comment. The following lines have been added for the same:



**Lines 446-466:**

In this section, we perform a spatial comparison for circum-Antarctic SIT across 39 individual CMIP6 model ensembles using different sea-ice products as observational references. However, we only discuss in detail GIOMAS as the observational reference. Here, GIOMAS is selected for a detailed discussion instead of other sea-ice products, primarily due to its relative overall closer agreement with the models (See Sec.3.1) and the very high SIT biases observed in the satellite products during September (Fig.5 and S2). In general, during both the months the mean spatial distribution of SIT in models shows lower biases with respect to the reanalyses/syntheses when compared to the satellite product (See Sect.3.3). GECCO3 produced results comparable to the GIOMAS reanalysis product (not shown). Although GIOMAS displays an earlier seasonal maximum, the MMMs exhibit the lowest biases relative to GIOMAS for this month, compared to GECCO3 and satellite-derived products (Fig.5).

Spatially, the satellite as well as the reanalysis/synthesis products show that thickest sea-ice resides in the western Weddell Sea along the Antarctic Peninsula and along the coastal edges of the ABS in the form of multi-year ice (Fig.8 and S3). There is relatively thinner sea-ice observed in the eastern Antarctic (Kurtz & Markus, 2012). Our analysis reveals that most of the CMIP6 models capture a similar spatial pattern in SIT around the Antarctic however, they do exhibit biases and underestimate SIT (Fig.5 and Fig.S2-S4). In February, over half of the models simulate thinner sea ice in the Weddell Sea compared to the GIOMAS product (Fig.S3) while thin biases relative to the satellite products are simulated almost across all the models (Fig.S4). On the circum-Antarctica scale, about 38% of the models simulate thicker ice compared to GIOMAS, with only 5 out of 39 models showing SIT greater than that estimated by satellite products.

Some of the legends are not clear enough. It is recommended to directly mark them in the figures or explain them in detail in the captions.

Thank you for pointing it out. We have updated Figures 2 and 3 along with their captions to better explain the details and to avoid confusion in differentiating various metrics.

9. Although the paper proposes directions for model improvements (e.g., improving ice-edge simulation accuracy, enhancing dynamic process representations), these suggestions are rather general and lack specific implementation details.

Thank you for the comment. We acknowledge the importance of this line of inquiry.

Here, our study provides a broad evaluation and comparison of CMIP6 climate models with three distinct sea-ice products, with the primary aim of assessing overall performance rather than analyzing individual model biases or improvements in detail. As such, we only briefly address model biases and suggest general directions for improvement, without proposing specific correction strategies. However, we believe it is best that such correction strategies are pursued in future studies, dedicated to detailed model diagnostics and development.