Response to the reviews of egusphere-2026-362 "Conservation of Heat in the Coupled Arctic Prediction System (CAPS v1.1): Comprehensive model evaluation based on the MOSAiC observations" by Chao-Yuan Yang, Fengguan Gu, Jiping Liu, Annette Rinke, Hu Yang, and Xiaoxu Shi

<u>Responses to comments by Reviewer #1</u>

We would like to thank the reviewer for the helpful comments on the manuscript.

General comments:

The authors present an updated version of the Coupled Arctic Prediction System (CAPS) with a corrected coupling scheme. The model experiments show good agreement between satellite observations, MOSAiC observations, and the spectrally nudged model run. Additional experiments without nudging with the corrected and uncorrected coupling schemes demonstrate the importance of the update. The paper demonstrates the utility of CAPS for Arctic coupled simulations and is an important reminder of the need for consistency across model components. The paper is well written, and I have only a few general comments and line-by-line suggestions.

In the introduction, the authors present broad trends about a selection of sea-ice and upper-ocean related topics. What I don't see clearly from the introduction is the motivation for the development of CAPS. I would like to see more of an emphasis on what the state-of-the-art is for coupled Arctic climate modeling and forecasting. For example: How is CAPS different from other systems? What need is being filled that wasn't filled before your updates to CAPS? What aspects of climate physics for the Arctic are uncertain, and how will improving CAPS help improve our knowledge? Not all of these specific questions need to be addressed, however the introduction should do a better job providing a context for the model improvements.

Response: We thank the reviewer for this helpful comment. We agree that the original Introduction did not sufficiently explain why improving heat-flux conservation is important beyond being a technical coupling modification. We have revised the Introduction to clarify that the inconsistent and duplicated surface-flux calculations in the original CAPS not only affect long-term heat conservation but also hinder the consistent implementation of additional surface processes into CAPS. We now explain that this issue is particularly relevant for future CAPS developments involving wave–ice interactions, where wave-modified surface roughness and sea spray-mediated heat fluxes directly affect atmosphere–sea ice–ocean exchanges. We kept this discussion concise because the present manuscript focuses on the atmosphere–ocean–sea-ice

coupling framework of CAPS rather than the wave component. The relevant paragraphs have been revised as follows, corresponding to Lines L81–L127 in the revised manuscript:

"… Atmosphere–sea ice–ocean interactions through variations in surface fluxes are also crucial for climate variability in Arctic sea-ice cover (e.g., Deng and Dai, 2019), and remote impacts of Arctic sea-ice on El Niño–Southern Oscillation (e.g., Deng and Dai, 2024; Liu et al., 2022). Conservation in coupled models can refer to several related but distinct issues. In many coupled modeling systems, coupler-level conservation is primarily associated with conservative remapping, temporal averaging, and area-weighted merging of fields exchanged between model components on different grids (Craig et al., 2017). In some systems, selected interface fluxes are also calculated centrally in the coupler or mediator (e.g., Danabasoglu et al., 2020; Karsten et al., 2024). These approaches are essential for reducing artificial sources or sinks associated with interpolation, coupling frequency, and component-grid mismatch. However, they do not by themselves guarantee that different model components use physically consistent surface fluxes if those fluxes are re-calculated independently within multiple model components. The issue addressed in CAPS is closer to latter problem. Because all CAPS components use the same horizontal grid, the revised coupling approach presented in this study targets the flux inconsistency caused by parallel calculations of related surface heat fluxes in CAPS (see Section 2 for details). Similar issues have recently been highlighted in regional polar coupled modeling. For example, Malyarenko et al. (2023) developed P-SKRIPS for the Ross Sea region by importing heat and mass fluxes calculated by the atmospheric component directly into the ocean-sea ice component under the same horizontal grid, thereby avoiding duplicated flux calculations and double scaling over mixed sea ice grid cells. Their study demonstrated that such inconsistencies can produce substantial flux mismatches in polar coupled simulations.

In addition to the importance of long-term energy conservation, a unified treatment of surface heat fluxes is also essential for the further development of CAPS as a process-oriented coupled Arctic framework. The duplicated calculation of surface heat flux makes the implementation of additional surface processes difficult, because new parameterizations must be introduced independently in the atmosphere and ocean–sea ice components while still maintaining consistency across the coupled interface. This issue becomes particularly relevant for future and ongoing CAPS developments involving wave-ice interactions, in which wave-induced surface heat flux changes can directly affect atmosphere–sea ice–ocean exchanges. Therefore, improving the conservation and consistency of surface heat fluxes is a necessary step toward both reducing artificial coupled model-drift and enabling more physically consistent extensions of CAPS.

The Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC; Nicolaus et al., 2022; Rabe et al., 2022; Shupe et al., 2022) was conducted from September 2019 to October 2020 aboard the Research Vessel Polarstern (Knust, 2017). MOSAiC was the largest-scale Arctic research expedition to date and provided comprehensive observations of atmosphere, sea ice, snow, and oceanic physical, chemical, and biological properties in the Arctic climate system. The MOSAiC campaign took place 22 years after the previous year-long Surface Heat Budget of the Arctic Ocean campaign (SHEBA; Uttal et al., 2002). Unlike SHEBA, which was conducted over multiyear sea ice, MOSAiC was conducted over thinner and more dynamic first-year sea ice, with the aim of improving our understanding and model representation of the coupled Arctic climate system under global climate change.

In this study, we aim to resolve a legacy surface heat-flux inconsistency in CAPS to minimize the inconsistency of heat-flux exchange between the atmosphere and ocean–sea ice components, and to evaluate the revised CAPS comprehensively. Compared with previous regional polar flux-conservation studies, the present work focuses on a pan-Arctic atmosphere–ocean–sea-ice coupling framework, applies the revised coupling over a year-long MOSAiC drift, and evaluated the atmosphere, ocean, and sea-ice components against satellite-derived sea ice products and MOSAiC observations. This paper is structured as follows ..."

In the Method section, we have also added a relevant paragraph for limitations in the representation of physical processes in the Arctic as follows, corresponding to Lines L259–L265 in the revised manuscript:

"The revised coupling approach presented in this study does not guarantee that other physical parameterizations in CAPS are free of biases in polar regions. Previous studies have shown that Arctic weather and climate prediction remains challenging because of limitations in the representation of stable boundary layers, clouds and radiation, atmosphere–ocean–sea-ice exchange, sea-ice/snow processes, and data assimilation under sparse polar observations (e.g., Ortega et al., 2022; Inoue, 2021; Guemas et al., 2016; Vihma, 2014; Hines et al., 2015)."

Craig, A., Valcke, S., and Coquart, L.: Development and performance of a new version of the OASIS coupler, OASIS3-MCT_3.0, Geosci. Model Dev., 10, 3297–3308, https://doi.org/10.5194/gmd-10-3297-2017, 2017.

Danabasoglu, G., Lamarque, J.-F., Bacmeister, J., Bailey, D. A., DuVivier, A. K., Edwards, J., Emmons, L. K., Fasullo, J., Garcia, R., Gettelman, A., Hannay, C., Holland, M. M., Large, W. G., Lauritzen, P. H., Lawrence, D. M., Lenaerts, J. T. M., Lindsay, K., Lipscomb, W. H., Mills, M. J.,

Neale, R., Oleson, K. W., Otto-Bliesner, B., Phillips, A. S., Sacks, W., Tilmes, S., van Kampenhout, L., Vertenstein, M., Bertini, A., Dennis, J., Deser, C., Fischer, C., Fox-Kemper, B., Kay, J. E., Kinnison, D., Kushner, P. J., Larson, V. E., Long, M. C., Mickelson, S., Moore, J. K., Nienhouse, E., Polvani, L., Rasch, P. J., and Strand, W. G.: The Community Earth System Model Version 2 (CESM2). J. Adv. Model. Earth Sy., 12, e2019MS001916. https://doi.org/10.1029/2019MS001916, 2020.

Malyarenko, A., Gossart, A., Sun, R., and Krapp, M.: Conservation of heat and mass in P-SKRIPS version 1: the coupled atmosphere–ice–ocean model of the Ross Sea, Geosci. Model Dev., 16, 3355–3373, https://doi.org/10.5194/gmd-16-3355-2023, 2023.

Ortega, P., Blockley, E. W., Køltzow, M., Massonnet, F., Sandu, I., Svensson, G., Acosta Navarro, J. C., Arduini, G., Batté, L., Bazile, E., Chevallier, M., Cruz-García, R., Day, J. J., Fichefet, T., Flocco, D., Gupta, M., Hartung, K., Hawkins, E., Hinrichs, C., Magnusson, L., Moreno-Chamarro, E., Pérez-Montero, S., Ponsoni, L., Semmler, T., Smith, D., Sterlin, J., Tjernström, M., Välisuo, I., and Jung, T.: Improving Arctic Weather and Seasonal Climate Prediction: Recommendations for Future Forecast Systems Evolution from the European Project APPLICATE. Bull. Amer. Meteor. Soc., 103, E2203–E2213, https://doi.org/10.1175/BAMS-D-22-0083.1, 2022.

Inoue, J.: Review of forecast skills for weather and sea ice in supporting Arctic navigation, Polar Sci., 27, 100523, https://doi.org/10.1016/j.polar.2020.100523, 2021.

Guemas, V., Blanchard-Wrigglesworth, E., Chevallier, M., Day, J. J., Déqué, M., Doblas-Reyes, F. J., Fučkar, N. S., Germe, A., Hawkins, E., Keeley, S., Koenigk, T., Salas y Mélia, D. and Tietsche, S.: A review on Arctic sea-ice predictability and prediction on seasonal to decadal time-scales. Q. J. R. Meteorol. Soc., 142: 546-561. https://doi.org/10.1002/qj.2401, 2016.

Vihma, T.: Effects of Arctic Sea Ice Decline on Weather and Climate: A Review. Surv. Geophys., 35, 1175–1214, https://doi.org/10.1007/s10712-014-9284-0, 2014.

Glisan, J. M., Gutowski, W. J., Cassano, J. J., and Higgins, M. E.: Effects of Spectral Nudging in WRF on Arctic Temperature and Precipitation Simulations. J. Climate, 26, 3985–3999, https://doi.org/10.1175/JCLI-D-12-00318.1, 2013.

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

The conclusions and discussion raises discussion of the temperature inversion structure. It seems to me that this topic should be raised first in the results section, along with the other discussion of the atmospheric structure.

Response: Thanks for the reviewer's comment. We have described the temperature inversion structure in the results section first after the comparison between the simulations and the observations of temperature and humidity profiles. The relevant paragraph in the results section has been revised as follows, corresponding to Lines L627–L649 in the revised manuscript:

"Figure 14 shows the vertical profiles of atmosphere temperature and relative humidity (RH) measured by radiosondes launched during the MOSAiC expedition, and the difference between the observations and the experiments along the track of MOSAiC drift. For atmospheric temperature profiles, Exp_FREE simulates profiles with biases oscillating between warm and cold, and the magnitude of biases is larger from October 2019 to April 2020 (Fig. 14b) compared to the rest of periods. From December 2019 to April 2020, Exp_FREE shows excessively warm biases, exceeding 10 °C above ~8 km altitude. In contrast to Exp_FREE, Exp_NUD shows smaller temperature differences relative to the radiosonde observations, with most differences confined within approximately ±10 °C (Fig. 14c). However, the remaining non-zero differences indicate that constraining the simulated atmosphere toward ERA5 does not eliminate all biases against the MOSAiC radiosondes. These residual differences suggest that the radiosonde comparison is also influenced by uncertainties in the large-scale atmospheric state used for nudging and by sub-grid-scale processes not fully constrained by spectral nudging. Low-level temperature inversions are common features over Arctic sea ice (e.g., Serreze et al., 1992; Devasthale et al., 2010), The inversion structures in the radiosonde observations and all experiments are evaluated in Fig. S6. Exp_FREE reproduces the frequent occurrence of low-level temperature inversions during the cold season, but differs from the radiosondes in their timing and vertical distribution (Figs. S6a, c). With spectral nudging applied, Exp_NUD shows a closer overall representation than Exp_FREE, but still does not fully capture the observed inversion evolution (Figs. S6d). In addition, Exp_FREE_OLD shows stronger inversions during the cold season compared with Exp_FREE, indicting the inconsistency in heat-flux exchange induced by the original coupling approach (Fig. 6) alters the low-level temperature structure (Fig. S6b)."

Serreze, M. C., Kahl, J. D., and Schnell, R. C.: Low-Level Temperature Inversions of the Eurasian Arctic and Comparisons with Soviet Drifting Station Data. J. Climate, 5, 615–629/630. https://doi.org/10.1175/1520-0442(1992)005<0615:LLTIOT>2.0.CO;2, 1992.

Devasthale, A., Willén, U., Karlsson, K.-G., and Jones, C. G.: Quantifying the clear-sky temperature inversion frequency and strength over the Arctic Ocean during summer and winter seasons from AIRS profiles, Atmos. Chem. Phys., 10, 5565–5572, https://doi.org/10.5194/acp-10-5565-2010, 2010.

In Figure 12, the choice of the color bar limits for panels b and c demonstrates substantial improvement between the free and nudged experiments, however it is difficult to see how strong the low bias is in the profiles in the nudged experiments. To my eye it looks like the profiles are 5-10 C too cold through much of the year. Perhaps seasonal average biases plotted as profiles (e.g., like the mean profiles shown in Fig. S6) would help clarify the issue.

Response: We thank for the reviewer's helpful comment. Your minor comment to the original Figure 4, 5, and S3 points out the use of colorblind friendly color bars and Reviewer 2 suggests the use of consistent style for all figures. Therefore, the revised Figure 12, corresponding to Figure 14 in the revised manuscript and Figure R1 below, changes the color bars of panel (a) and (d) to a perceptually uniform and colorblind friendly one. Also, the color bars of panel (b), (c), (e), and (f) are consistent with the blue-yellow-red color bar used in the original Figure 4, corresponding to Figure R4 below. We have also adjusted the data range of panel (c) from ±30 degrees to ±10 degrees to achieve similar color spanning between panel (b) and panel (c).

The representation of the boundary layer is an important point, as I would expect that changing the performance of surface fluxes in the coupler would have a strong effect on this part of the atmosphere.

Response: We thank the reviewer for raising this point. In this revision, we first compare the simulated profiles in Exp_FREE_OLD with the radiosonde observations, and calculate the temperature lapse rate below 2 km altitude for Exp_FREE_OLD as shown in Figure R2 below. The simulated temperature and relative humidity biases in Exp_FREE_OLD (Figs. R2a–b) show similar magnitude as those in Exp_FREE (Figs. R1b–c). The most prominent difference between the original and revised coupling approaches is evident in the low-level temperature lapse rates as shown in Figure R3. The lower-level temperature inversion in Exp_FREE_OLD is generally stronger than that in Exp_FREE_OLD during November 2019 to April 2020 (Figs. R3b–c). We have revised the descriptions and discussions of the temperature inversion reflect this point, corresponding Lines L646–L649 in the revised manuscript and referring to the response for the major comment above. The original Figure S10 is replaced by Figure R3, corresponding to Figure S6 in the revised supplement.

Line-by-line comments:

25 “showing the drifting behaviors” — It isn’t clear if “drifting” refers to sea ice drift or about the model accumulating errors, please rephrase for clarity

Response: We thank the reviewer for this helpful comment. We agree that the original term was ambiguous. Also, the “drift” behavior is usually used for longer-term climate-scale simulation, and this term was not an appropriate term for the present 14-month simulation. We have therefore revised the sentence as follows: “The revised coupling substantially reduces interfacial heat-flux residuals and improves the simulation of Arctic sea-ice conditions relative to the original coupling, which produces enhanced winter ice growth and reduced summer ice retreat during the 14-month simulation.”

37 “Arctic, a climate system” — the structure of this sentence is awkward

Response: We thank the reviewer for pointing this out. We have revised the sentence to avoid the awkward appositive structure. Now it reads “The Arctic climate system has been undergoing drastic changes and has transitioned toward a new state.”

50 “Arctic amplification related to sea ice changes” — perhaps missing a comma?

Response: Thank you for the comment. We agree that the original phrasing was unclear. Rather than only adding a comma, we have rephrased the sentence to clarify the relationship between Arctic amplification and sea ice changes. Now it reads “In addition to dramatic changes in sea-ice properties, near-surface air temperature in the Arctic regions is warming three to four times faster than the global average, a phenomenon known as Arctic amplification, which is closely linked to sea-ice changes (Holland and Bitz, 2003, Rantanen et al., 2022; Screen and Simmonds, 2010).”

54 “Besides of the” — the “of” is unnecessary

Response: We thank the reviewer’s helpful comment, and have corrected the sentence. Now it reads: “Besides the increases of freshwater content in the Beaufort Gyre, sea ice decline also plays a role in recent warming of Atlantic Water layer in the Arctic Eurasian basin, a phenomenon known as Arctic atlantification (Polyakov et al., 2017; Wang et al., 2024).”

78-83 This sentence should be broken into at least two sentences — as is, it reads as if you are saying that MOSAiC is the first expedition from 2019-2020, rather than first overall.

Response: Thank you for this helpful comment. We agree that the original sentence was too long and could lead to ambiguity regarding the timing and the "first year-round expedition" description. We have divided the sentence into sentences for clarity.

83 It is not clear to me that we can call MOSAiC the first when SHEBA exists. In what sense is MOSAiC the first? Largest, certainly.

Response: We thank the reviewer for this helpful comment. We agree that the original wording was too broad, especially because SHEBA was also a year-long Arctic field campaign. Our intension was to refer to the unprecedented scale and modern research-icebreaker-based year-round drift of MOSAiC, but we agree that the sentence could be misleading. We have therefore revised the text to describe MOSAiC as the largest-scale Arctic research expedition to date, while retaining the comparison with SHEBA in the follow sentence. Now the relevant sentences read as "The Multidisciplinary drifting Observatory for the Study of Arctic Climate (MOSAiC; Nicolaus et al., 2022; Rabe et al., 2022; Shupe et al., 2022) was conducted from September 2019 to October 2020 aboard the Research Vessel *Polarstern* (Knust, 2017). MOSAiC was the largest-scale Arctic research expedition to date and provided comprehensive observations of atmosphere, sea ice, snow, and oceanic physical, chemical, and biological properties in the Arctic climate system. The MOSAiC campaign took place 22 years after the previous year-long Surface Heat Budget of the Arctic Ocean campaign (SHEBA; Uttal et al., 2002). Unlike SHEBA, which was conducted over multiyear sea ice, MOSAiC was conducted over thinner and more dynamic first-year sea ice, with the aim of improving our understanding and model representation of the coupled Arctic climate system under global climate change."

242 The L-site network and CO fit the 15 km description, but the M sites are further out.

Response: We thank the reviewer for pointing this out. We agree that the original description was imprecise, as the 15 km description mainly applies to the CO and L-site network, whereas the M sites were distributed more broadly. We have revised the sentence to clarify this point. We also note that the SIMBA buoy from the M4 site is used in our evaluation because M4 was located close to the CO site and had relatively long temporal coverage. As shown in Figure R4 below, the distance between the M4 and CO sites was approximately 3 km, which support its use together with the CO and L-site SIMBA buoys for representing the local snow and sea ice conditions along the MOSAiC drift. Now the revised sentence read "These observations were distributed within the MOSAiC observational network, including the central observatory (CO), three large sites

(L1, L2, and L3) located approximately 15 km from CO, and several medium sites (M1–M8) distributed over a broader area." The sentences describing the selection of the SIMBA buoy at the M4 site has also been revised as follow: "SIMBA buoys with relatively long temporal coverage were selected for averaging over their shared period from December 2019 to June 2020. These buoys were located at the CO site (T56, T62, and T66), the L1, L2, and L3 site (T63, T67 and T70), and the M4 site (T58), which was close to CO."

244 "is referred to" — grammar is off here, please rephrase the sentence.

Response: Thank you for pointing this out. We have rephrased the sentence for clarity and grammatical correctness. Now it reads "This distribution of the MOSAiC sites is shown in Fig. 2 of Bliss et al. (2023)."

325 ULSIE/OLSIE are only defined in a figure caption — please move the definitions into the main text. As these are less commonly used acronyms, I found myself needing to look them up many times — consider whether there is a more direct way to describe it.

Response: We thank the reviewer for this helpful suggestion. We agree that defining ULSIE and OLSIE only in the figure caption made the main-text discussion difficult to follow, especially because these acronyms are not commonly used. We have therefore moved the definitions into the main text and revised the subsequent discussion to use the more direct phrases "underestimated local sea ice extent" and "overestimated local sea ice extent" instead of relying heavily on the acronyms. The relevant paragraph has been revised as follows, corresponding to Lines L370–L379, L383–L389 in the revised manuscript:

"The monthly averaged March and September SIC fields from the observations in 2020, together with the difference between the simulations and the observations, are shown in Fig. 4. Following the IIEE definition of Goessling et al. (2016), the lined areas in Figs. 4b–d and 4f–h indicate grid cells where the simulated and observed local SIE disagree based on the 15 % SIC threshold. Specifically, underestimated local SIE refers to regions where the observed SIC exceeds 15 % but the simulated SIC does not, whereas overestimated local SIE regions where the simulated SIC exceeds 15 % but the observed SIC does not. Corresponding to the SIE shown in Fig. 3a, the underestimated SIE of Exp_FREE and Exp_NUD in March 2020 mainly arise from underestimated local SIE in the Sea of Okhotsk, the Barents Sea, the Labrador Sea, and Davis Strait (Figs. 4b–c)."

"In September 2020, the comparable areal coverage of overestimated local SIE along the coasts

of Alaska, central Siberia, and eastern Siberia and underestimated local SIE near Fram Strait and the Canadian Archipelago in Exp_FREE contributes to the smaller difference in total SIE relative to the observations (Figs. 3a, 4f). In contrast, underestimated local SIE in the central Arctic Ocean and Canadian Archipelago, partly compensated by overestimated local SIE in the Chukchi Sea, is the main contributor to the underestimated SIE in Exp_NUD (Figs. 3a, 4g).”

433 “show a throughout snow depth” — rephrase this for clarity

Response: We thank for the reviewer’s comment. We have rephrased the sentence to improve clarity and grammar. Now it reads “However, in contrast to the nearly constant snow depth observed from December 2019 to Man 2020, both experiments show a continuous increase in snow depth during this period.”

Figure 4: The color scale for panel a and e is not print or colorblind friendly, however the blue-yellow-red color scale is good. Please update the color map for panels a and e. It is difficult to see the vertical and horizontal lines. While I recognize that the simulation domain is large, the study focus is on the central Arctic. Hence, I think that zooming in to the area north of, say, 60 or 70 latitude (as in Fig S9) would be justified, and would enable readers to see the details discussed in the text.

Response: Thanks for the reviewer’s comments. We have revised the original Figure 4, corresponding to Figure 4 in the revised manuscript and Figure R5 below. First, the color bar of panel (a) and (e) has changed to a perceptually uniform and colorblind friendly one. Second, the density of vertical and horizontal lines has been increased. Third, the boundaries of summer panels (e–h) have been adjusted, zooming in to the area north of 65N.

Figure 5: As with fig 4, look for a color scale which is perceptually uniform. See the instructions at the Copernicus guide for authors: https://publications.copernicus.org/for_authors/manuscript_preparation.html#figurestables

Response: Thanks for the reviewer’s suggestion. We have changed the color bar of the original Figure 5, corresponding to Figure 5 in the revised manuscript and Figure R6 below, to a perceptually uniform and colorblind friendly one.

As with Figure 4, Figure S3 is difficult to see. You may consider changing the figure from a 2x6 plot to 4x3, as there is a lot of extra unused white space on the page.

Response: Thanks for the reviewer’s suggestion. The revised Figure S3, corresponding to Figure

R7 below, changes the panel layout from 2x5 to 3x4, and the color bar to a perceptually uniform and colorblind friendly color bar same as Figure R5–R6. We have also adjusted the layout of original Figure S6, corresponding Figure S7 in the revised supplement and Figure R8 below, to 2x2 to better utilize unused white space on the page.

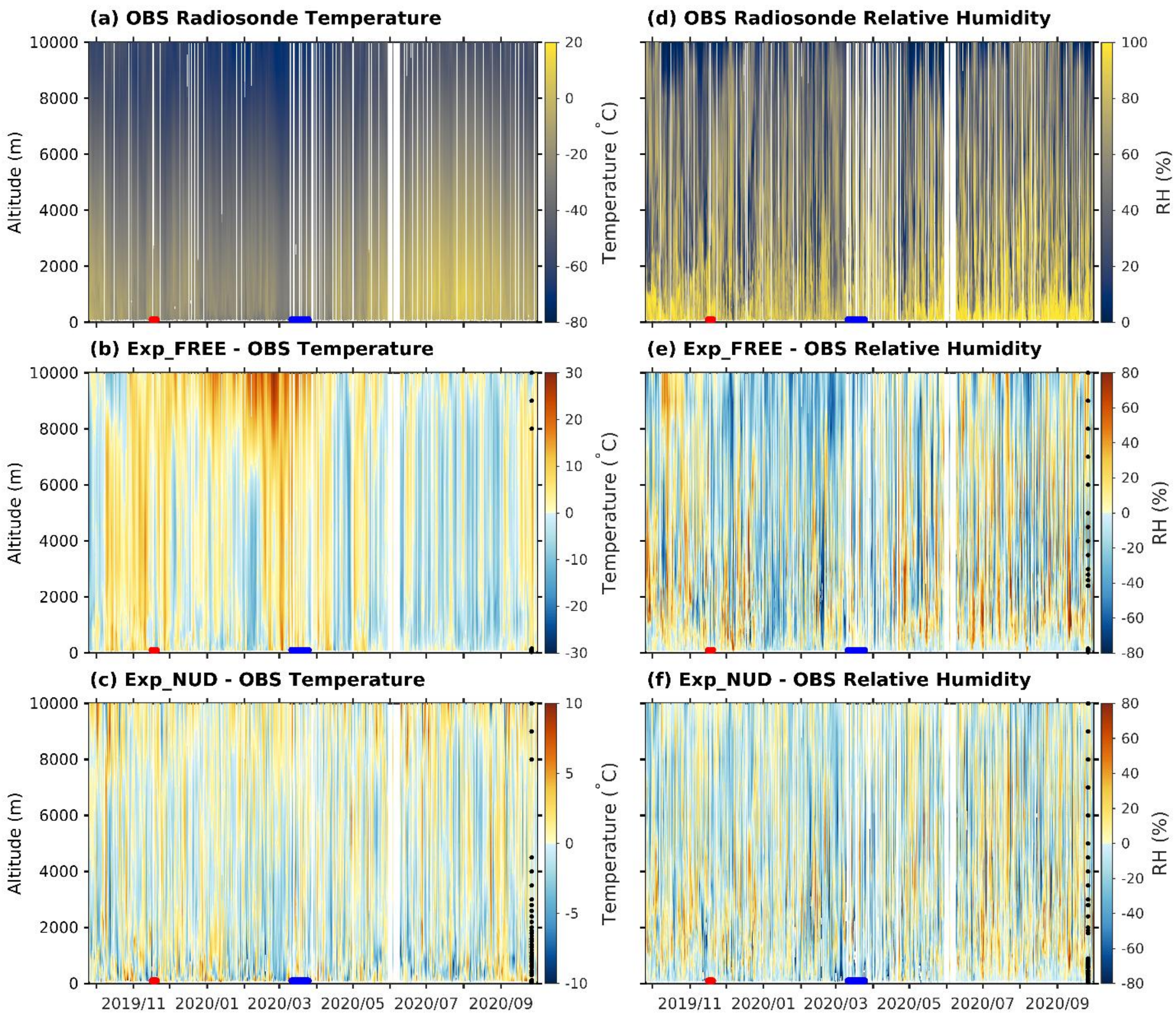

Figure R1 Time-height distributions of radiosonde observations for (a) temperature and (d) relative humidity (RH), the corresponding simulated temperature biases for (b) Exp_FREE, and (c) Exp_NUD, and the RH biases for (e) Exp_FREE, and (f) Exp_NUD along the track of MOSAiC drift. Note: red and blue lines on time-axis represent the STORM-1 and STORM-2 periods; black dots near the right edge of panels (b), (c), (e), and (f) indicate altitude levels where the full-period mean bias is significantly different from zero based on a pair moving-block bootstrap with 3-day blocks and Benjamini-Hochberg false discovery rate correction at $q = 0.05$. The horizontal position of the black dots is for visualization only and does not correspond to a specific time.

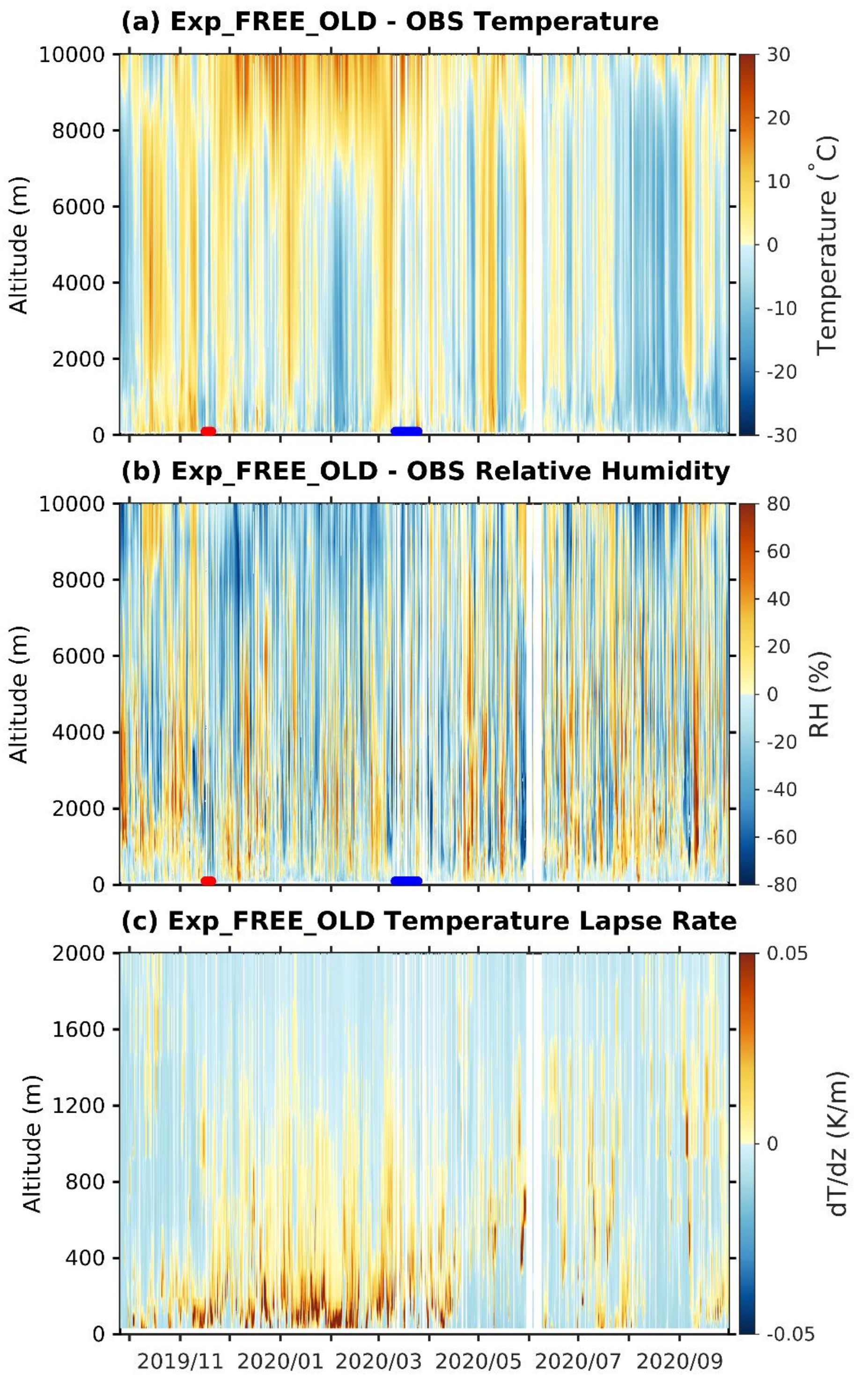

Figure R2 Time-height distributions of simulated (a) temperature and (b) relative humidity biases in Exp_FREE_OLD with respect to the radiosonde observations, and (c) distributions of temperature lapse rate below 2 km altitude along the track of MOSAiC drift. Note: red and blue lines on time-axis represent the STORM-1 and STORM-2 periods.

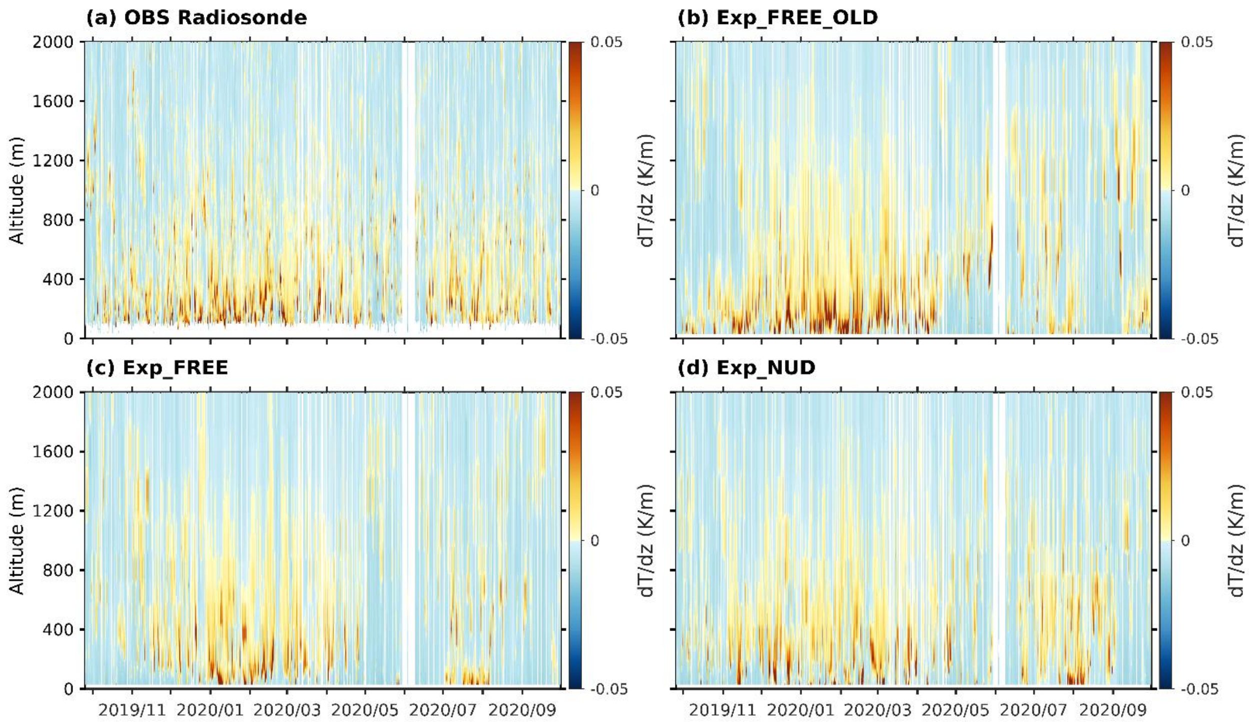

Figure R3 Time-height distributions of temperature lapse rate below 2 km altitude for (a) radiosonde, (b) Exp_FREE_OLD, (c) Exp_FREE, and (d) Exp_NUD along the track of MOSAiC drift.

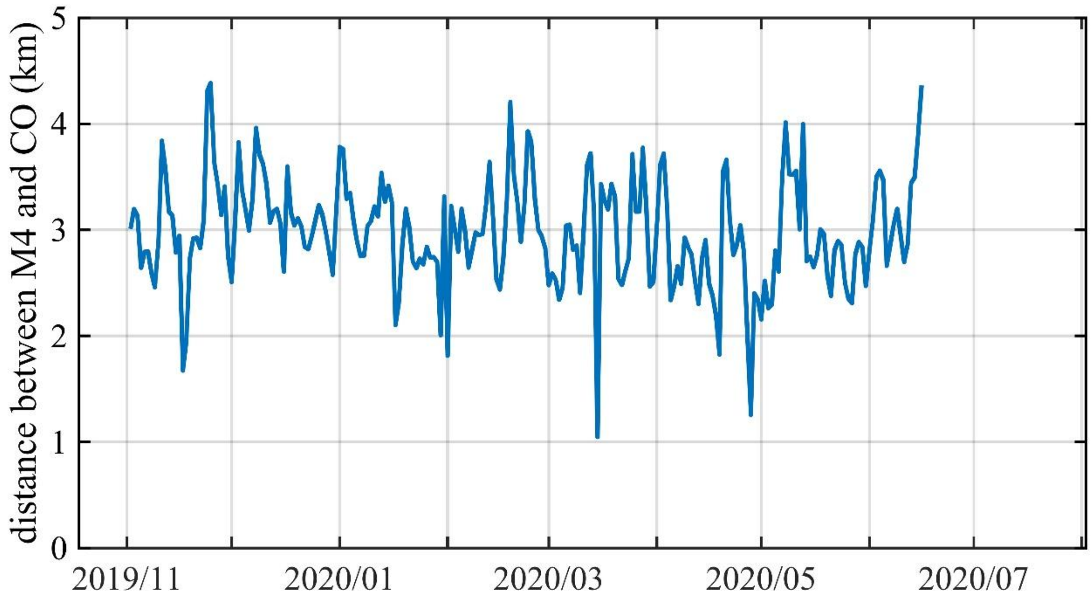

Figure R4 Time-series of the distance between the M4 and CO sites.

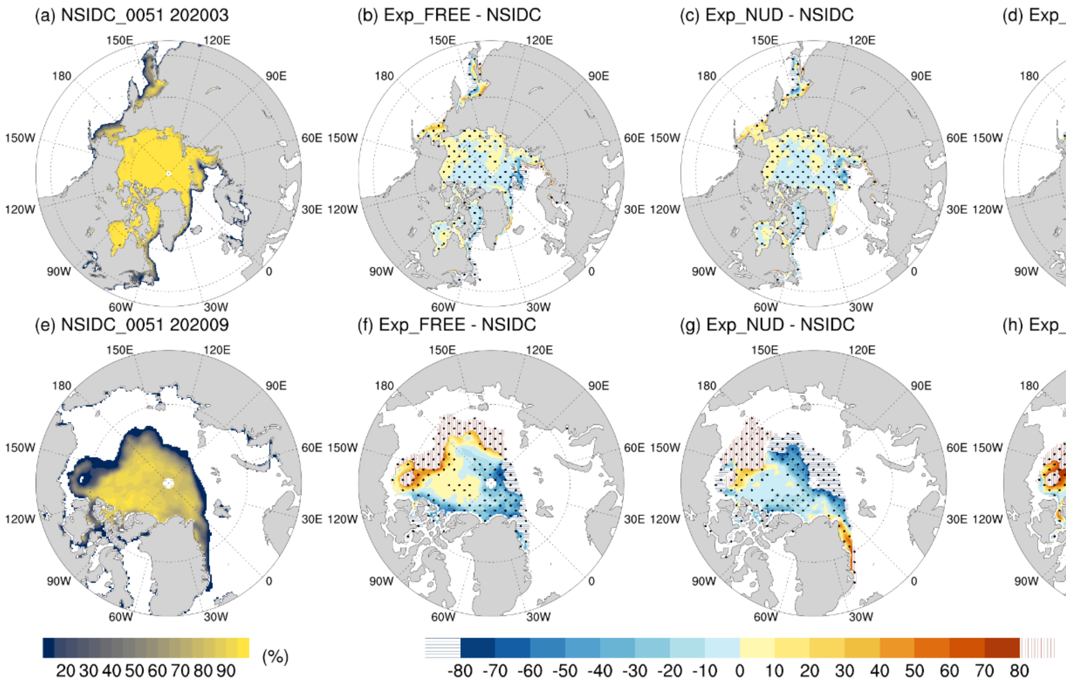

Figure R5 Monthly mean sea ice concentration (SIC) for the observations in (a) March 2020, and (e) September 2020, and the difference between the experiments and the observations for Exp_FREE, Exp_NUD, and Exp_FREE_OLD in (b–d) March 2020, and in (f–h) September 2020. Note: vertical/horizontal-lining represents the overestimated/underestimated local sea ice extent with respect to the observed local sea ice extent, following the definition of IIEE (Goessling et al., 2016); black dotted area denotes grid cells where the mean paired SIC difference is significantly different from zero based on a two-sided paired *t*-test, after Benjamini–Hochberg false discovery rate correction at $q = 0.05$.

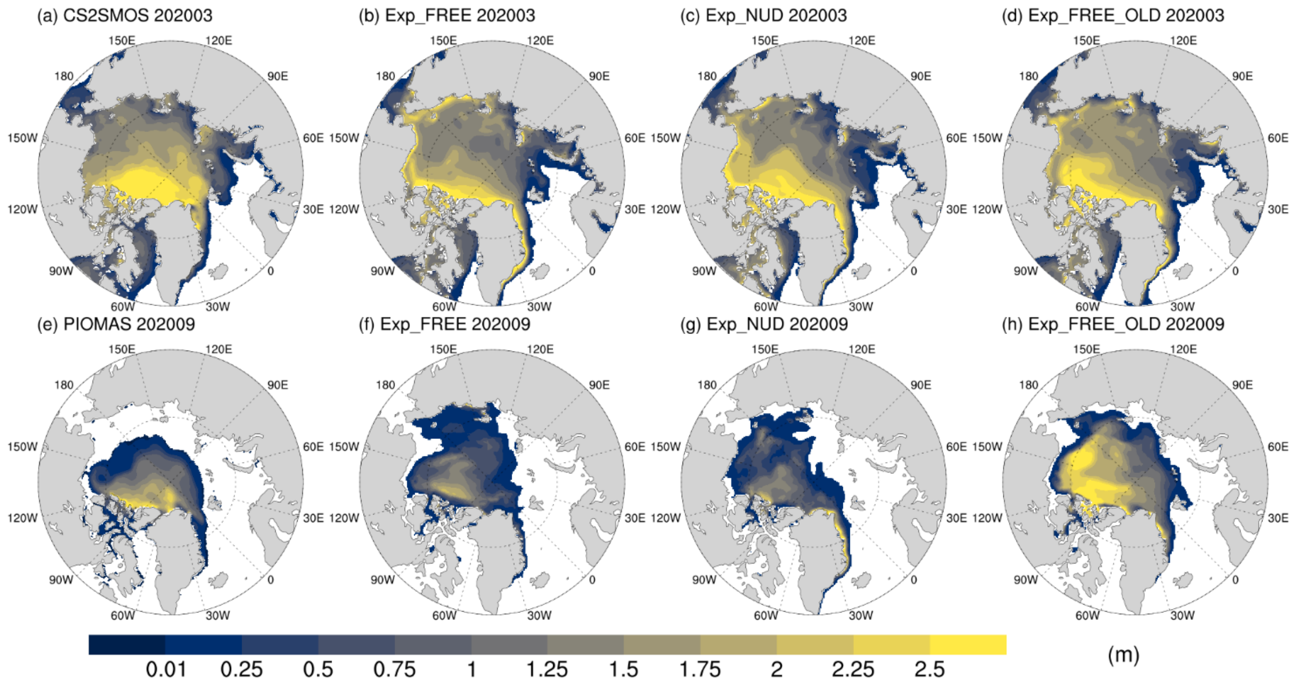

Figure R6 Monthly mean sea ice thickness for the reference data (CS2SMOS and PIOMAS), Exp_FREE, Exp_NUD, and Exp_FREE_OLD in (a–d) March 2020, and (e-h) September 2020.

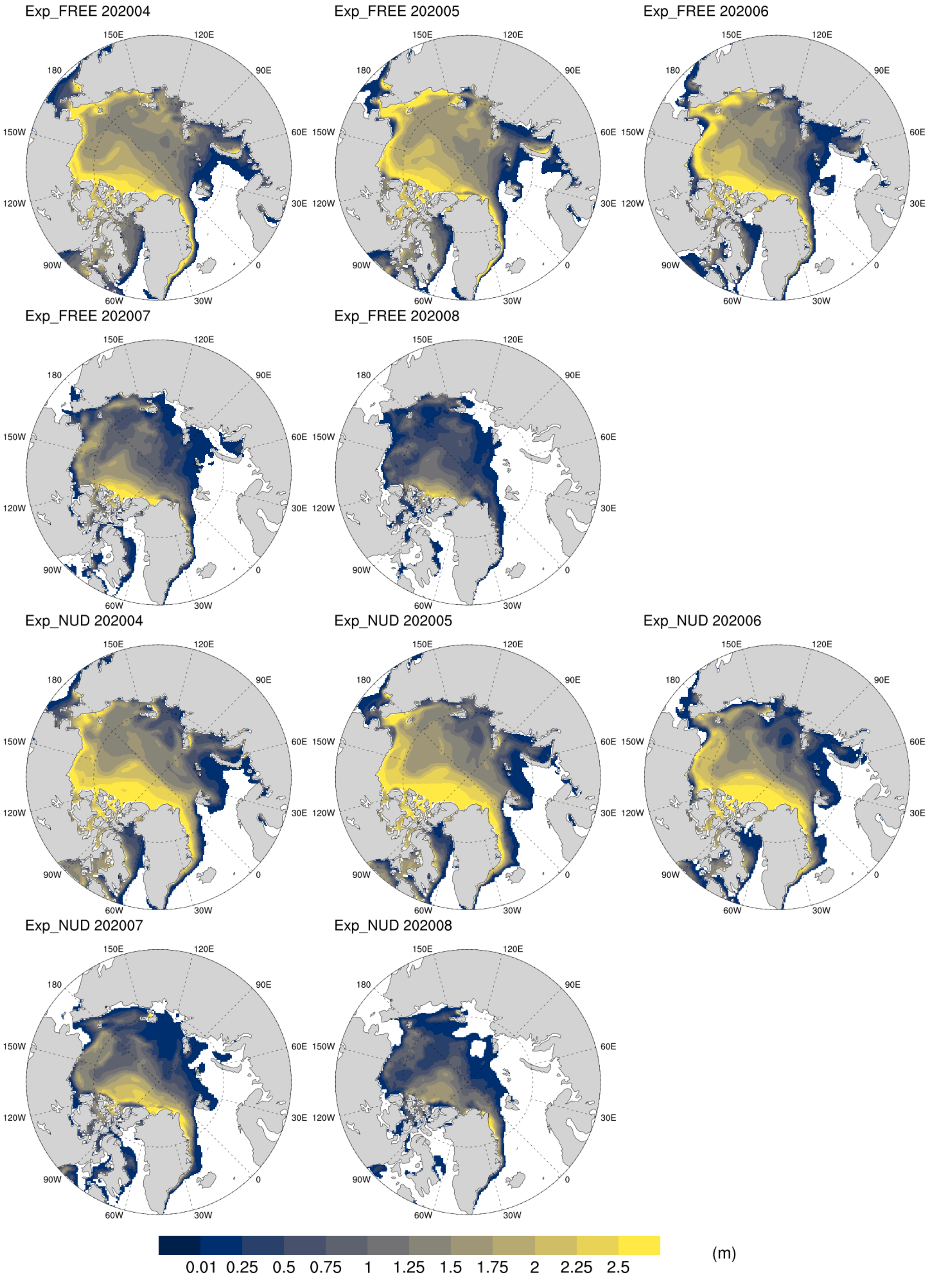

Figure R7 Monthly mean sea ice thickness of Exp_FREE and Exp_NUD from April 2020 to August 2020.

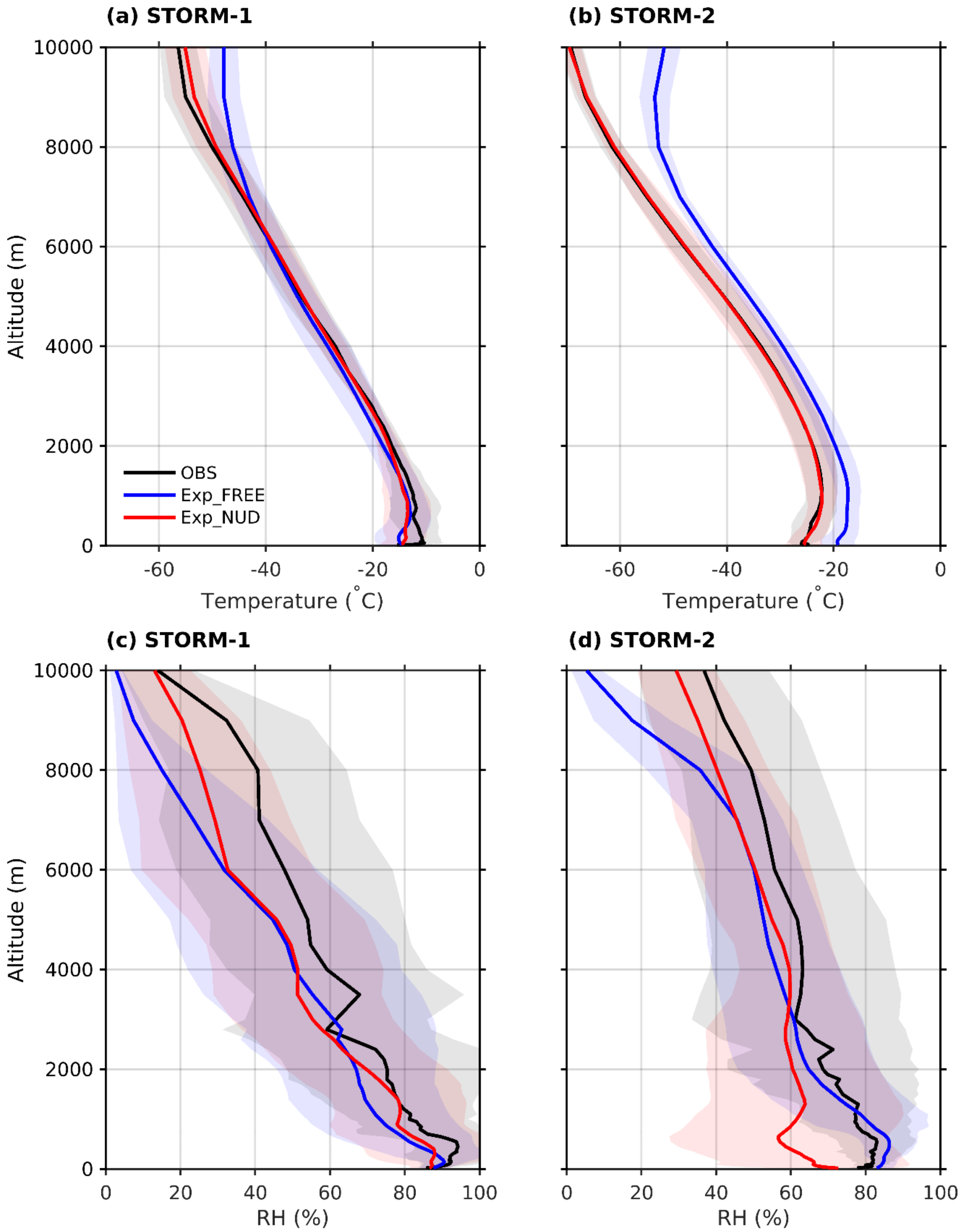

Figure R8 Mean (a, b) temperature and (c, d) relative humidity profiles for the observations (black line), Exp_FREE (blue line), and Exp_NUD (red line) during the STORM-1 period, and the STORM-2 period. Note: the shading represents one standard deviation range of corresponding variables.