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

Response to comments by Reviewer #2

We would like to thank the reviewer for the helpful comments on the manuscript. In addition to the revisions described in our point-by-point responses below, we have comprehensively revised the Abstract to reflect the updated scope and interpretation of the study. The revised Abstract now describes the model-development contribution in terms of improved consistency of interfacial heat-flux exchange, replaces subjective references to model "drift" with time-bounded description of cumulative sea-ice biases in the original coupling, incorporates the newly added quantitative evaluation against MOSAiC observations, and clarifies that the improved agreement in the spectrally nudged experiment is sensitive to the nudging configuration and should not be attributed to the revised coupling or model physics alone.

This manuscript presents an updated version of the Coupled Arctic Prediction System (CAPS v1.1), in which the atmosphere–ice–ocean coupling strategy has been revised to ensure conservation of heat fluxes exchanged among model components. Specifically, the study introduces an improved heat-conserving coupling framework between WRF and ROMS-CICE and evaluates its performance using observations from the MOSAiC campaign. The authors show that the revised coupling substantially reduces heat-flux inconsistencies and leads to improved simulations of sea-ice conditions during the MOSAiC period. In addition, the manuscript evaluates atmospheric, oceanic, and sea-ice performance and examines the impact of spectral nudging on simulated synoptic variability.

Overall, the topic is highly relevant to the GMD community, as coupled-model consistency and energy conservation are fundamental requirements for reliable Arctic climate simulations. The manuscript is generally well organized, and the evaluation dataset is comprehensive.

However, several aspects require clarification, and there is significant room for improvement in the current form. In particular, the novelty of the study should be more clearly articulated and better motivated. For example, the introduction should more explicitly position the proposed coupling improvement relative to existing coupling frameworks. In addition, the quantitative diagnosis of energy conservation and the attribution of remaining model biases need to be substantially strengthened.

I therefore recommend major revision before the manuscript can be considered for publication.

Major Comments

1. Novelty of the heat-conservation framework

The manuscript states that CAPS v1.1 achieves conservation of heat exchange between WRF and ROMS-CICE. However, it remains unclear how this implementation differs from, and improves upon, existing approaches used in other coupled systems. The authors should explicitly describe how the new implementation compares with standard coupler-based conservation methods. Without such a comparison, it is difficult to assess the methodological novelty and broader relevance of CAPS v1.1.

Response: We thank the reviewer for requesting a clearer explanation of how the revised CAPS coupling differs from coupler-based conservation methods. We agree that the original manuscript did not sufficiently distinguish between different meanings of "conservation" in coupled models. We now clarify that coupler-based conservation often refers to conservative remapping, temporal averaging, and area-weighted merging of fields exchanged between model components, particularly when the atmosphere, ocean, sea-ice, and land models use different grids. In some coupled systems, selected interface fluxes are also calculated centrally in the coupler. However, this is not the primary issue addressed in CAPS, because WRF, ROMS, and CICE in CAPS use the same horizontal grid, and therefore horizontal conservative interpolation is not the dominant source of inconsistency. We have also added a discussion of the closely related P-SKRIPS development of Malyarenko et al. (2023). Their study addressed non-conservative heat and mass exchange in a regional polar coupled model by importing fluxes calculated by PWRF directly into MITgcm and avoiding duplicated flux calculations and double scaling in mixed sea-ice grid cells. The revised CAPS coupling follows the same general concept of avoiding duplicated, component-dependent interface-flux calculations, which also preserves the overall WRF–ROMS–CICE coupling structure at the same time. The novelty of the present manuscript is therefore not the introduction of a new coupler-conservation algorithm, but the implementation of this flux-consistent coupling approach in the pan-Arctic WRF–ROMS–CICE framework, the correction of a legacy heat-flux inconsistency, and the comprehensive evaluation of the resulting coupled Arctic modeling system during the MOSAiC period. In the revised manuscript, we modify the Introduction and Methods to clearly reflect the above responses.

The relevant paragraph in the Introduction has 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. Section 2 provides …"

The relevant paragraph in the Method has been revised as follows, corresponding to Lines L193–L200 in the revised manuscript:

"… Table 1 summarizes the variables that WRF, ROMS, and CICE exchange through the couplers with the improved coupling approach. The variables exchange with the original coupling approach are also summarized in Table S1 in the supplement.

This revised coupling approach is similar in the concept of the flux-conservation approach used in P-SKRIPS (Malyarenko et al., 2023), where fluxes computed by the atmospheric component are imported into the ocean-sea ice component to avoid repeated and inconsistent flux calculations. The implementation in CAPS differs in the flux ownership assigned by interface: WRF is used as the source of atmosphere–ocean turbulent heat fluxes, whereas CICE is used as the source of atmosphere–ice turbulent heat fluxes and transfers these fluxes back to WRF. This design preserves the surface fluxes exchanged between the atmosphere and ocean–sea ice components while retaining the existing WRF–ROMS–CICE coupling structure."

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.

2. Quantitative heat-budget analysis

A central claim of the manuscript is improved heat conservation. However, the evaluation focuses primarily on model performance metrics and sea-ice state variables, rather than directly demonstrating closure of the heat budget.

In addition, the manuscript would benefit from a more explicit discussion of the current limitations and their broader implications. I strongly recommend that the authors incorporate a more comprehensive coupled-system heat-budget analysis into the Discussion section—specifically addressing atmospheric heat gains and losses, oceanic heat uptake, sea-ice heat storage variability, and the residual heat imbalance. Such an analysis would provide a valuable foundation for future work and help contextualize the present findings within the broader energy-cycle framework.

At present, improved heat conservation is inferred only indirectly through sea-ice performance, which is insufficient to fully support the main claim.

Response: We thank the reviewer for requesting a more comprehensive heat-budget analysis. We agree that the original manuscript did not make the heat-conservation evidence sufficiently explicit beyond sea-ice response. In the original Figure 6, it shows the "difference" of interfacial fluxes between the WRF-side and the ROMS-CICE side, and the "difference" was not clearly defined in the original manuscript. In this revision, we define the inconsistency in heat-flux exchange between the atmosphere and ocean-sea ice components as the interfacial residual at

the atmosphere–ocean–sea-ice coupling cells, $R_{coupling}$, formulated as,

$$R_{coupling}(x, y, t) = F_{surface}^{WRF}(x, y, t) - F_{surface}^{ROMS/CICE}(x, y, t)$$

where $F_{surface}^{WRF}$ is the flux component used in the WRF physical processes, and $F_{surface}^{ROMS/CICE}$ is that in the ROMS/CICE processes. The flux component includes net surface heat flux, net shortwave radiation, net longwave radiation, sensible heat flux, and latent heat flux. We calculated the mean residual, absolute mean residual, and RMSE of interfacial heat-flux residuals for all flux components in Exp_FREE_OLD, Exp_FREE, and Exp_NUD as shown in Figures R1 and R2, corresponding to Figures 6 and S4 in the revised manuscript. Although the main heat-conservation conclusion is drawn from the comparison between Exp_FREE_OLD and Exp_FREE, we also include Exp_NUD in Figure 6 in the revised manuscript because it uses the same revised coupling under a constrained-atmosphere configuration. The negligible residuals in both Exp_FREE and Exp_NUD indicate that the revised coupling maintains small interfacial heat-flux residuals in both free-running and nudged simulations. We have revised the relevant paragraph in Section 3.1.2 as follows, corresponding to Lines L419–L454 in the revised manuscript:

"To diagnose the consistency of heat-flux exchange between the atmosphere and the ocean-sea-ice components with the original and revised coupling approaches, we define the residual of interfacial heat-flux exchange at the atmosphere–ocean–sea-ice coupling cells, which can be expressed as

$$R_{coupling}(x, y, t) = F_{surface}^{WRF}(x, y, t) - F_{surface}^{ROMS/CICE}(x, y, t)\,, \qquad (3)$$

where $F_{surface}^{WRF}$ is the flux component used in the WRF physical processes, and $F_{surface}^{ROMS/CICE}$ is that in the ROMS/CICE processes. Both $F_{surface}^{WRF}$ and $F_{surface}^{ROMS/CICE}$ are the weighted averaged fluxes based on Eq. (1). The flux component is defined as positive downward and includes net surface heat flux (NetHF), with the components of net shortwave radiative flux (NetSW), net longwave radiative flux (NetLW), sensible heat flux (SH), and latent heat flux (LH). Figure 6 shows the domain-mean residual, domain-mean absolute residual, and root-mean-square error (RMSE) of interfacial heat-flux residuals for all flux components in Exp_FREE_OLD, Exp_FREE, and Exp_NUD. Exp_NUD is included in this residual analysis as an additional check that the revised coupling maintains similar interfacial heat-flux residuals under the constrained-atmosphere configuration. Mean residuals of NetHF are 0~15 W/m2 for most of the simulation

period with the original coupling approach (Exp_FREE_OLD), whereas those with the revised coupling approach (Exp_FREE and Exp_NUD) are negligible (less than 0.1 W/m2, Figs. 6a, S4a). Since all fluxes are defined positive downward, positive (negative) residuals of NetHF represents that the ocean-sea ice component receives less (more) energy than it supposes to receive, which can be also viewed as an artificial cooling (warming) effect in the ocean-sea ice coupled system due to the inconsistency in heat-flux exchange. Mean residuals of NetHF in Exp_FREE_OLD are mainly contributed by turbulent components rather than radiative components (Figs. 6a–e). Mean residuals of SH are mostly positive throughout the simulation, with the magnitude ranging from ~0 to ~12.5 W/m2 (Fig. 6c). Mean residuals of LH are mostly positive from September 2019 to April 2020 (up to ~11.5 W/m2) and then negative afterward (up to ~8 W/m2) (Fig. 6d). For Exp_FREE and Exp_NUD, the mean (Fig. 6a–e), mean absolute (Figs. 6f–j), and RMSE residuals (Figs. 6k–o) remain much smaller than those in Exp_FREE_OLD, confirming that the revised coupling substantially reduces both the systematic and local/intermittent interfacial heat-flux residuals. The remaining residuals are generally small and mainly reflect numerical and diagnostic differences rather than the large systematic inconsistency seen in Exp_FREE_OLD. The contrast between Exp_FREE_OLD and Exp_FREE demonstrates the effect of the revised coupling itself, because both experiments are free-running and differ primarily in the coupling approach. The similarly small residuals in Exp_NUD indicate that the revised coupling remains consistent when spectral nudging is applied.”

Furthermore, we have also added a daily integrated surface/interfacial energy-exchange analysis for Exp_FREE_OLD and Exp_FREE shown in Figure R3 and corresponding to Figure 7 in the revised manuscript. This analysis places the interfacial residuals in Figure R1 into a coupled-system energy-cycle context. For sea ice, the net exchange is decomposed into atmosphere–ice and ice–ocean contributions. For the atmosphere, lower-boundary exchange is separated into ice-covered cells, open-ocean cells, and land cells, For the ocean, upper-boundary exchange is separated into ice-covered cells and open-ocean cells. Ice-covered cells are defined as cells with SIC ≥ 1%, and open-ocean cells are defined as the remaining ocean cells. The masks are diagnosed daily from each experiment, so the decomposed exchange reflect both flux differences and the evolving areal coverage of each surface type. We note that this diagnostic represents the surface/interfacial energy exchange among model components, rather than a fully closed three-dimensional prognostic heat budget for each component, because atmospheric and oceanic heat storage, lateral boundary transports, and internal source-sink terms are not explicitly included. Figure R3 focuses on Exp_FREE_OLD and Exp_FREE, because these two experiments differ

only in the coupling approach and therefore provide the cleanest comparison for diagnosing the effect of the coupling revision itself in the energy-cycle context. We have added a new paragraph in Section 3.1.2 describing these energy cycle as follows, corresponding to Lines L455–L488 in the revised manuscript:

"To place the interfacial residuals shown in Fig. 6 into a coupled-system energy-cycle context, we further diagnose the daily integrated energy exchange through the lower boundary of the atmosphere, the atmosphere–ice and ice–ocean interfaces of sea ice, and the upper boundary of the ocean for Exp_FREE_OLD and Exp_FREE (Fig. 7). Positive values represent energy gain by the component or interface indicated in each panel. Because the purpose of Figure 7 is to isolate the effect of the coupling revision on the component-level energy pathways, we compare Exp_FREE_OLD and Exp_FREE only; Exp_NUD is not included in this diagnostic because it additionally differs by the use of spectral nudging. Both experiments reproduce a pronounced seasonal surface energy cycle. The atmosphere gains energy from the surface during the cold season and loses energy during the melt season, while the ocean shows the opposite seasonal behavior, with heat loss in autumn–winter and heat uptake in late spring–summer (Figs. 7b–c). These domain-integrated atmospheric and oceanic exchanges are mainly controlled by open-ocean grid cells, as indicated by the close correspondence between the total atmospheric/oceanic energy terms and their open-ocean contributions (Figs. 7h–i), while the ice-covered and land-cell contributions are smaller in magnitude (Figs. 7e, f, j).

The sea-ice net exchange is roughly one order of magnitude smaller than the atmospheric and oceanic exchanges but is more directly linked to the simulated sea-ice thermodynamic response. Sea ice loses energy from autumn to early spring and gains energy during the melt season, with a peak in early summer (Fig. 7a). This seasonal transition is mainly controlled by the atmosphere–ice interface: sea ice loses energy to the atmosphere during the cold season and gains energy from the atmosphere during late spring and summer (Fig. 7d). Compared with Exp_FREE, Exp_FREE_OLD generally shows stronger negative atmosphere–ice exchange during winter and weaker positive atmosphere–ice exchange during summer, indicating enhanced cold-season energy loss from sea ice and reduced melt-season atmospheric energy input to sea ice in the original coupling approach. Although differences in the domain-integrated atmospheric and oceanic exchanges are relatively small because they are dominated by open-ocean cells, the differences over ice-covered cells are more evident and directly affect the sea-ice energy pathway. This helps explain why the original coupling approach leads to a cumulative overestimation of sea-ice thickness and coverage during the 14-month simulation, characterized

by enhanced winter ice growth and reduced summer ice retreat (Figs. 3–5). This diagnostic represents the surface/interface energy exchange among model components rather than a fully closed three-dimensional prognostic heat budget for each component, because atmospheric and ocean heat storage, lateral boundary transports, and internal source-sink terms are not explicitly included."

3. MOSAiC evaluation

The evaluation against MOSAiC observations is extensive; however, much of the discussion remains descriptive. The manuscript would benefit from the inclusion of quantitative skill metrics, such as RMSE, bias, correlation coefficients, and/or Taylor diagrams.

These metrics should be reported consistently across atmospheric, oceanic, and sea-ice variables to enable a more objective assessment of model performance.

Response: We thank the reviewer for this helpful suggestion. In the revised manuscript, we have added quantitative skill metrics to provide a more objective evaluation of model performance against MOSAiC observations. Specifically, model outputs were paired with observations at representative vertical levels, including altitudes of 1, 3, 5, 7, and 9 km for radiosonde profiles and depths of 10, 20, 50, 75, and 100 m for CTD profiles. For single-level variables, including snow depth, ice thickness, and ice surface fluxes, the corresponding model and observational time-series were directly paired. Based on these paired samples, we constructed Taylor diagrams to summarize the relative model skill across atmospheric, oceanic, and sea-ice variables for Exp_FREE and Exp_NUD, as shown in Figure R4 below, corresponding to Figure 10 in the revised manuscript. We also added bias and RMSE statistics for Exp_FREE and Exp_NUD in Tables R1–R2 below, corresponding to Tables 4–5 in the revised manuscript. These additions allow a more consistent and quantitative comparison of the free-running and nudged simulation against MOSAiC observations across the atmosphere, ocean, and sea-ice components. We have also revised Section 3.2 by adding concise quantitative summaries to the corresponding subsections, while retaining the original time-series and profile-based interpretation. These revisions can be found on Lines L550–L738 in the revised manuscript.

4. Impact of spectral nudging

The manuscript shows that spectral nudging improves storm representation and reduces several biases, which is an interesting aspect, particularly for Arctic synoptic variability.

However, this raises an important question: to what extent are the reported improvements due to improved model physics versus relaxation toward large-scale reanalysis?

The authors should discuss (i) whether nudging may mask deficiencies in the coupled system, (ii) the sensitivity of results to nudging strength, and (iii) whether the main conclusions regarding heat conservation remain valid without nudging.

This distinction is important, as GMD papers should clearly separate model development from externally imposed constraints.

Response: We thank the reviewer for raising this important point. We agree that the improved agreement in Exp_NUD must be interpreted in the context of spectral nudging. The purpose of the revised coupling is to reduce the diagnosed inconsistency in heat-flux exchange between the atmosphere and ocean-sea ice components. This correction does not guarantee that other physical parameterizations in CAPS are free of biases in polar regions. In this revision, we now explicitly distinguish the coupling-revision experiment from the constrained-atmosphere experiment: Exp_FREE_OLD and Exp_FREE are used to assess the effect of revised coupling, whereas Exp_NUD is used to examine how the MOSAiC evaluation changes when the large-scale atmospheric state is constrained toward ERA5 by spectral nudging. The relevant paragraph has been revised as follows, corresponding to Lines L250–L271 in the revised manuscript:

"In this study, three experiments covering the entire period of the MOSAiC expedition, starting from 00 UTC September 1st, 2019, to 00 UTC November 1st, 2020 were conducted, and the three primary experiments serve different purposes. Exp_FREE_OLD provides the baseline simulation using the original coupling approach. Exp_FREE uses the revised coupling approach without spectral nudging and is used to evaluate the free-running behavior of the revised CAPS after correcting the heat-flux inconsistency. Exp_NUD uses the revised coupling approach with spectral nudging and is designed as a constrained hindcast experiment in which the large-scale atmospheric state is constrained toward ERA5. The spectral nudging in Exp_NUD is applied to wind, temperature, and humidity, with horizontal wavenumbers 1–3 nudged toward ERA5 data, and only at grid points above the boundary-layer height. 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). Therefore, differences between Exp_FREE_OLD and Exp_FREE primarily reflect the effect of the revised coupling, whereas differences between

Exp_FREE and Exp_NUD reflect the sensitivity to the representation of the large-scale atmospheric state. The improved agreement with observations in Exp_NUD should therefore be interpreted as the combined effect of the revised coupled model and the imposed large-scale atmospheric constraint, rather than as an improvement caused by the revised coupling or model physics alone."

During the revision, we also identified that the intended upper-level-only nudging design in Exp_NUD was affected by a WRF namelist option that deactivated nudging below the planetary boundary layer. As a result, the nudging in the original Exp_NUD configuration was applied above the boundary-layer height, rather than only above approximated 6.7 km as originally described. We have corrected the description of Exp_NUD throughout the revised manuscript. To assess the sensitivity of the results to nudging design and strength, we conduct two additional experiments. Exp_NUD_UPD applies spectral nudging only to the upper 25 of the 50 WRF model levels, with a 10-level transition zone from zero to the full nudging weight of 0.0003 s^{-1}, following a design similar to Cassano et al. (2011). Exp_NUD_UPW is identical to Exp_NUD_UPD except that the full nudging weight is reduced by one order of magnitude to 0.00003 s^{-1}, motivated by Glisan et al. (2013), who showed that weaker nudging can still reduce biases in Arctic WRF simulations.

These additional experiments show that the MOSAiC evaluations are sensitive to the vertical extent and strength of spectral nudging (Figs. R5–R7). We have therefore clarified the interpretation of Exp_NUD throughout the manuscript. Specifically, Exp_NUD is now described as a constrained-atmosphere sensitivity experiment used to assess the sensitivity of CAPS to the representation of the large-scale atmospheric state. Its improved agreement with MOSAiC observations is not attributed to the revised coupling or model physics alone, because spectral nudging relaxes the atmospheric circulation toward ERA5. At the same time, the main conclusion regarding improved heat-flux consistency remains valid without nudging, because it is primarily supported by the comparison of Exp_FREE with Exp_FREE_OLD (Figs. R1–R3). We have added new discussion paragraphs on the role and limitations of spectral nudging and have included nudging-sensitivity diagnostics in the revised manuscript and/or Supplement. The relevant paragraph in the discussion section now reads as, corresponding to Lines L786–L815 in the revised manuscript:

"… Regardless of the challenges posed to CAPS, the consistent energy exchange between the atmosphere and ocean–sea ice components in CAPS provides a solid foundation for future improvements in the representation of physical processes, especially for processes related to

surface heat flux calculations.

The improved coupling reduces the interfacial heat-flux residuals and improves the simulated sea-ice evolution relative the original coupling approach in the free-running configuration. In addition to this coupling-focused comparison, the MOSAiC evaluation shows that the constrained-atmosphere experiment, Exp_NUD, agrees more closely with observations in several quantitative metrics, especially for radiosonde temperature/RH profiles and ocean temperature. However, this improved agreement reflects the combined effect of the revised coupled model and the imposed large-scale atmospheric constraint. Moreover, non-negligible biases remain in Exp_NUD, indicating that spectral nudging can reduced errors associated with large-scale atmospheric evolution but cannot fully compensate for biases in physical parameterizations, coupled surface exchange, or the reanalysis fields used for nudging. Therefore, Exp_NUD is used as a sensitivity experiment to examine how the coupled system responds when large-scale atmospheric state is constrained, rather than as evidence that the revised coupling or model physics alone removes the remaining biases.

Because Exp_NUD applies spectral nudging above the boundary-layer height, it represents a relatively strongly constrained experiment. To assess whether the MOSAiC evaluation depends on the vertical extent and strength of this constraint, we conducted two additional sensitivity experiments. Exp_NUD_UPD applies spectral nudging only to the upper 25 of the 50 WRF model levels (approximately above 6.7 km altitude), with a 10-level transition zone from zero to the full nudging weight of 0.0003 s-1, following a design similar to Cassano et al. (2011), whereas Exp_NUD_UPW uses the same vertical design but reduces the nudging weight by one order of magnitude (Glisan et al., 2013). The additional nudging experiments further indicate that the MOSAiC evaluation remains sensitive to the nudging configuration. In the radiosonde comparison, Exp_NUD_UPD and Exp_NUD_UPW both reduce the large temperature departures seen in the free-running atmosphere, but residual temperature and RH differences remain throughout the troposphere, especially in the lower atmosphere and during late summer (Fig. S11). The snow/sea-ice comparison also shows that changing the vertical extent and strength of nudging modifies snow accumulation and melt timing, while the simulated SIT remains within the broad observational uncertainty (Fig. S12). Surface fluxes are likewise sensitive to the nudging setup, in particular turbulent heat fluxes during spring and summer (Fig. S13).In this study, the implementation of our revised coupling approach is based on the shared horizontal grid between the atmosphere component and the ocean-sea ice component. …”

The Abstract has also revised to reflect this clarified interpretation of Exp_NUD and the

additional nudging-sensitivity experiments. It now distinguishes the improvement in interfacial heat-flux consistency from the observation-based improvements obtained under the constrained-atmosphere configuration and nots that non-negligible biases remain.

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.

5. Computational cost

As CAPS v1.1 is presented as a new model version, the manuscript would benefit from including information on computational performance, including (i) additional computational expense, (ii) coupling overhead, and (iii) scalability relative to CAPS v1.0. Such information is highly relevant for a model development study.

Response: We thank the reviewer for suggesting that the computational cost of the revised coupling should be documented. In this revision, we have added a computational-cost

benchmark comparing the original and revised CAPS coupling configurations under the production-resolution setup used in this study. To make the comparison as fair as possible, we conducted 31-day benchmark simulations from 00 UTC October 1st, 2019 to 00 UTC November 1st, 2019 using identical input/output settings, the same node list, and the same processor layout for WRF and ROMS/CICE as shown in Table R3.

The benchmark indicates that the revised coupling introduces negligible additional wall-clock cost. The wall time changes from 3 h 04 min 29 s with the original coupling to 3 h 04 min 33 s with the revised coupling, corresponding to a wall-time overhead of only 0.036%. The throughput remains essentially unchanged, decreasing only from 0.6629 to 0.6627 simulated years per wall-clock day. The estimated computational cost changes only from 11584.8 to 11588.9 core-hours per simulated year. These results indicate that the additional heat-flux exchange operations introduced by the revised coupling do not impose a meaningful computational burden under the configuration used in this study. We have also added SLURM-reported memory indicators for the model execution step. The maximum resident set size (MaxRSS) increases from 2.56 to 3.41 G, while the average resident set size (AveRSS) remains 0.19 G. The maximum virtual memory size (MaxVMSize) increases from 4.77 to 6.47 G.

Regarding scalability, the current pan-Arctic CAPS configuration has a moderate horizontal grid size of 320x440. Under this problem size and the production processor layout used here, the model is already close to the practical scaling range where further speed-up is increasing limited by communication, load balance, and input/output overhead rather than by the additional surface-flux exchange operations introduced in this revision. Therefore, we did not conduct a full strong- or weak-scaling assessment in this revision. Instead, we report a controlled production-layout benchmark to access whether the revised coupling changes the computational cost relative to the original coupling. Because the revised coupling changes the ownership and exchange of already diagnosed surface-flux variables without adding new horizontal interpolation, remapping, or additional model components, we do not expect it to fundamentally alter the parallel scalability of CAPS. We have clarified this limitation in the revised manuscript and note that a comprehensive scalability assessment for larger or higher-resolution CAPS configurations will be an important part future model-development work.

We have added a new section (Section 3.1.3) for reflecting the computational-cost estimation, corresponding to Lines L521–L549 in the revised manuscript, and the relevant section reads as:

"To document the computational cost of the revised coupling, we conducted 31-day benchmark simulations using the original and revised coupling approach. Both simulations were integrated

from 00 UTC October 1st, 2019 to 00 UTC November 1st, 2019 using identical input/output settings, the same node list, and the same processor layout. Wall time and memory usage were obtained from the SLURM accounting record for the model execution step.

The throughput shown in Table 3 is reported as simulated years per wall-clock day (SYPD), calculated as

$$SYPD = \frac{D_{sim}/365}{T_{wall}/24}$$

where D_{sim} is the number of simulated days and T_{wall} is the wall time in hours. The computational cost is also reported as core-hours per simulated year (CHPY),

$$CHPY = T_{wall} \times N_{core} \times \frac{365}{D_{sim}}$$

where N_{core} is the number of allocated cores.

The revised coupling introduces negligible additional wall-clock cost relative to the original coupling. The wall time increases from 3 h 04 min 29 s to 3h 04 min 33 s, corresponding to an overhead of 0.036 %. The throughput remains nearly unchanged, from 0.6629 to 0.6627 SYPD. The estimated computational cost also changes slightly, from 11584.8 to 11588.9 CHPY. These results indicate that the revised heat-flux exchange does not impose a meaningful computational burden under the benchmarked configuration. The present benchmark should be interpreted as a production-configuration overhead estimation rather than a comprehensive scaling test. The current CAPS domain has a moderate horizontal grid size of 320x440, and the production layout already uses 320 cores. For this problem size, further speed-up is expected to be increasingly limited by communication, load balance, and input/output overhead. Because the revised coupling only changes the ownership and exchange of surface-flux variables and does not introduce additional horizontal interpolation, remapping, or new model components, the negligible overhead shown in Table 3 suggests that the revised heat-flux exchange does not materially affect computational performance under the configuration used in this study. A more complete scalability assessment for larger of higher-resolution CAPS configurations is left for future work."

Minor Comments

The formatting of references is inconsistent. For example, "J. Geophys. Res.-Oceans" is abbreviated (Line 691), whereas "Monthly Weather Review" is written in full (Line 701). Please ensure consistency throughout the reference list. In addition, consider limiting the author list (e.g., first five authors followed by "et al.") where appropriate (e.g., Line 911).

Response: We thank the reviewer for pointing this out. We have reviewed the reference list and made the formatting of journal titles consistent by using abbreviated journal names where standard abbreviations are available. Regarding the author lists, we have retained the complete list of authors because this follows the reference style required by Copernicus/EGU journals, which specifies listing all authors in the bibliography rather than truncating the list with "et al.". The corresponding changes are reflected in the revised manuscript.

The font size of legends and axis tick labels in Figures 6–8 (Lines 1068, 1075, 1085) is relatively small. Please increase the font size to improve readability and maintain consistency with Figures 9 and 11.
Response: Thanks for the reviewer's suggestion. We have increased the font size in the original Figures 6–8, matching the font size used in the original Figures 9 and 11. The revised figures correspond to Figures 6, 8, and 9 in the revised manuscript and Figures R1, R8 and R9 below.

In Figure 13 (Line 1123), the colorbar labels are overly denseand the label-text are too busy. Consider increasing the length of the colorbar and adjusting the labeling for improved clarity.
The presentation of the figure at Line 1135 requires significant improvement. (i) The colormap and value range should be adjusted, as the current visualization is unclear. (ii) Font sizes should be increased. (iii) There are unexplained white gaps in the figure, which interrupt continuity and may confuse readers.
Response: We thank the reviewer for this suggestion. We combine the suggestion of Reviewer #1 and have adjusted color bars and font sizes, and revised the figure caption of the original Figure 13, corresponding to Figure 15 in the revised manuscript and Figure R10 below. The revised figure caption now reads as: "Figure 15 The averaged mean sea level pressure of (a, d) ERA5, (b, e) Exp_FREE, and (c, f) Exp_NUD, and the averaged 2-m temperature and 10-m wind fields of (g, j) ERA5, (h, k) Exp_FREE, and (i, l) Exp_NUD during the period of STORM-1 and STORM-2. Note: white star represents the averaged location of radiosondes during the corresponding period, and white gap in panels showing simulated fields represents out of model domain (Fig. S1)."

The colorbar styles in Figures 14 and 15 are inconsistent with those used in Figure 13 and other figures. Please standardize the formatting across all figures.
Response: We thank for the reviewer's helpful comment. Reviewer #1 comment points out that a perceptually uniform and colorblind friendly color bar is needed. Therefore, the raw maps in

all revised figures use perceptually uniform and colorblind friendly color bars, which are adjusted for different variables, and the difference maps use the same blue-yellow-red color bar. The revised Figures 14 and 15, corresponding to Figures 16 and 17 in the revised manuscript are shown in Figures R11 and R12.

In Supplementary Figure S9, the buoyancy location is marked in gray, which is difficult to distinguish. A darker color (e.g., black) would improve visibility.

Response: We thank the reviewer for this helpful suggestion. The revised Figure S9, corresponding to Figure S10 in the revised supplement and Figure R13 below, changes the color bar to a colorblind friendly one, in which a white star marker has better visibility.

Terms such as "drifting behavior" (Line 25) are subjective. Please replace them with more precise, quantitative descriptions where possible.

Response: We agree with the reviewer that "drifting" is subjective and not an appropriate term for the present 14-month simulation, because it is usually used for longer term climate-scale integrations. The revised sentence in the abstract now reads as: "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."

Several figures (e.g., Figures 4, 12, and 14) show differences between experiments without indicating statistical significance. Please include significance testing where appropriate.

Response: We thank the reviewer for this suggestion. We have added statistical-significance assessment to the monthly SIC difference maps, and the bias panels of atmospheric and oceanic profiles as shown in Figures R14, R15, and R11 below, corresponding Figure 4, 14, and 16 in the revised manuscript. For the monthly SIC difference maps (Fig. R14), daily NSIDC SIC was first regridded to the model grid, after which paired daily simulation-minus-observation differences were calculated at each grid cell. A two-sided paired *t*-test was used to assess whether the mean daily difference with each selected month differed from zero. The resulting spatial *p*-values were corrected using the Benjamini–Hochberg false discovery rate (BH-FDR) procedure at $q = 0.05$, separately for each month and experiment. Black dotted area indicates grid cells that remain significant after BH-FDR correction. For the atmospheric and ocean profiles (Figs. R15, R11), we tested the full-period mean bias at each vertical level rather than applying a pointwise test to

every time-height or time-depth grid cell, because the model and the observations are paired in time but only one realization is available for each experiment. We used 3-day moving blocks to account for temporal autocorrelation. Each block contains 12 paired radiosonde samples for the four-times-daily radiosonde data and 3 paired samples for the daily CTD data. Bootstrap resampling with 5000 resamples was then performed over these blocks to estimate the sampling distribution of the full-period mean bias at each vertical level. For each height or depth level, the null hypothesis is that the full-period mean bias is zero. The resulting *p*-values were further adjusted using BH-FDR procedure to control for multiple vertical-level tests. Significant levels are marked by black dots near the right edge of panels.

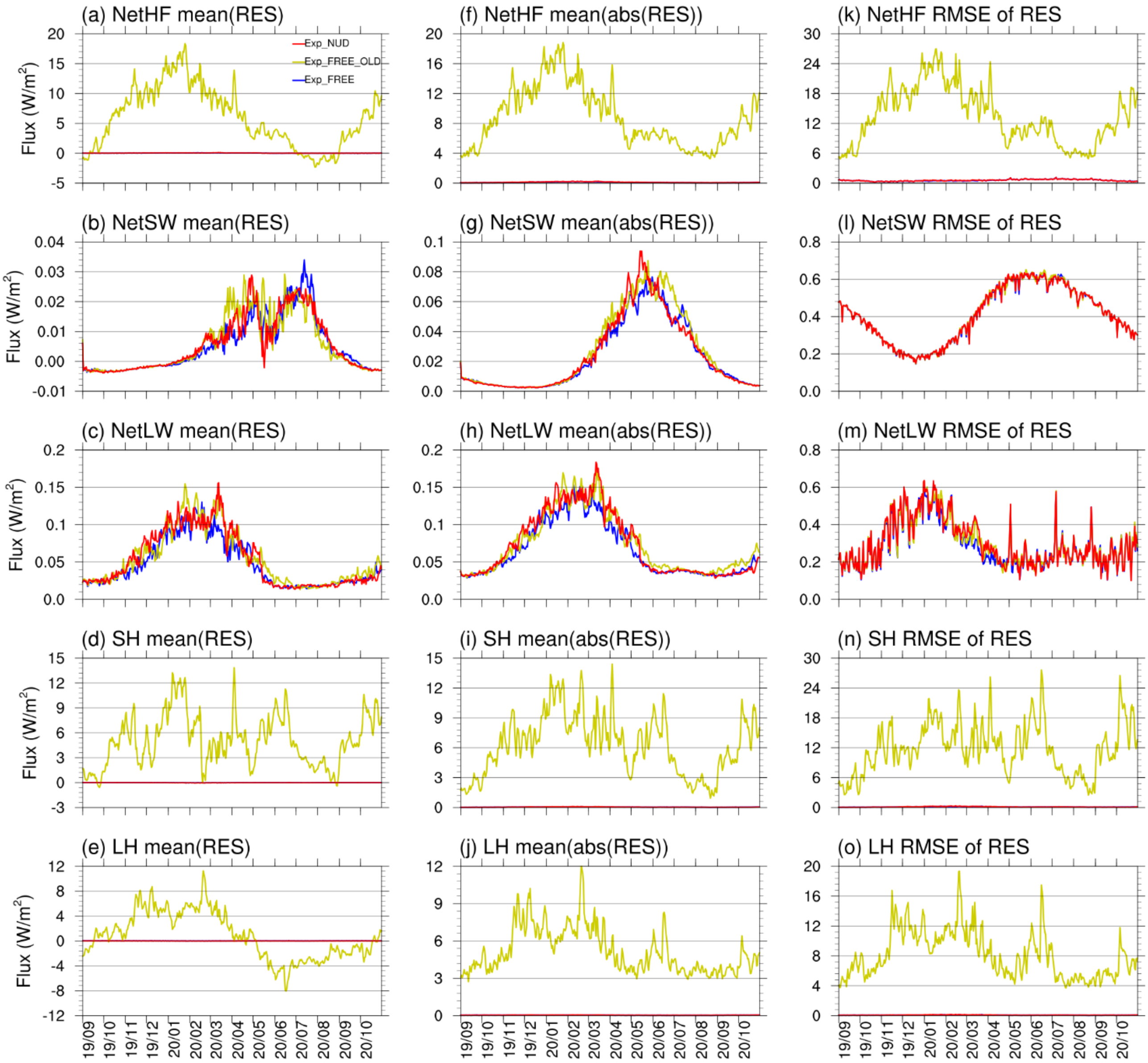

Figure R1 Domain-averaged statistics of interfacial heat-flux residuals at atmosphere–ocean–sea-ice coupling cells for all flux components in Exp_FREE_OLD (yellow line), Exp_FREE (blue line), and Exp_NUD (red line). The statistics include the (a–e) mean residual, (f–j) mean absolute residual, and (k–o) root-mean-square error (RMSE), with residuals defined by Eq. (3).

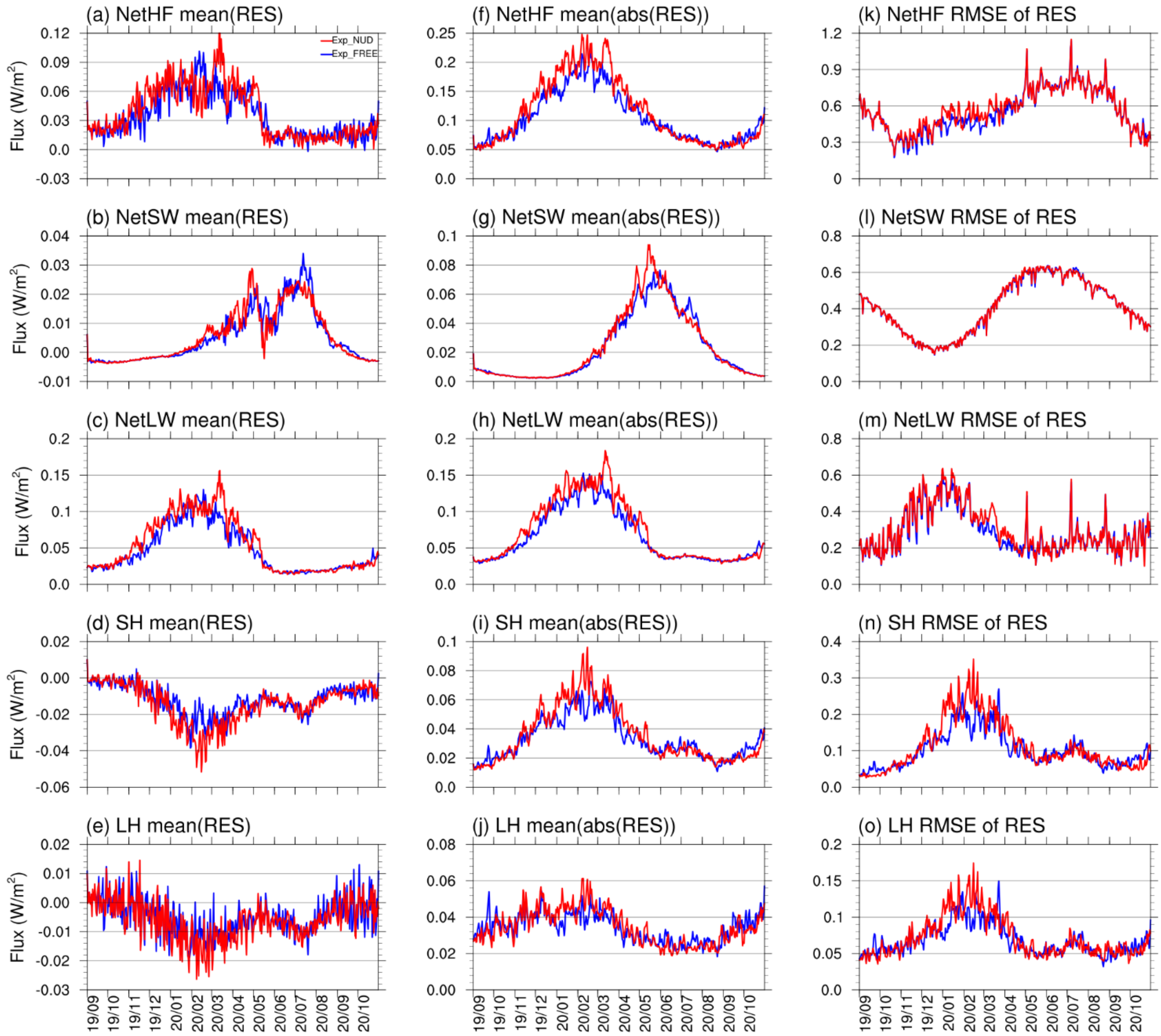

Figure R2 Same as Figure R1, but only includes Exp_FREE and Exp_NUD.

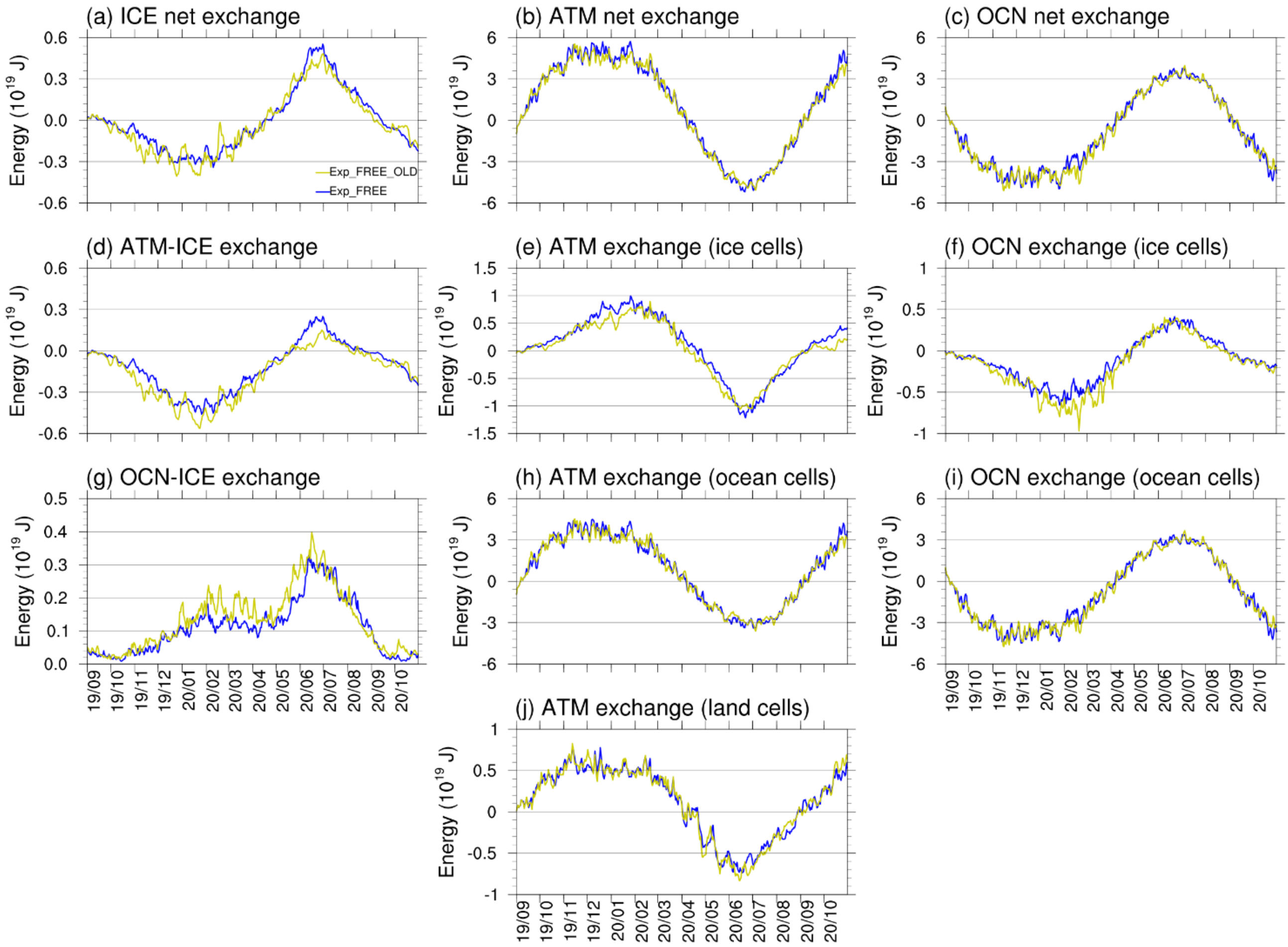

Figure R3 Daily integrated surface/interfacial energy exchange for the (a) sea-ice, (b) atmosphere, and (c) ocean in Exp_FREE_OLD (yellow line) and Exp_FREE (blue line). Sea-ice energy exchange is decomposed into (d) atmosphere–ice and (g) ice–ocean contributions; atmospheric lower-boundary exchange is separated into (e) ice-covered, (h) open-ocean, and (j) land cells; and ocean upper-boundary exchange is separated into (f) ice-covered and (i) open-ocean cells. Note: positive values indicate energy gain by the component or interface named in each panel; ice-covered cells are defined by SIC ≥ 1%.

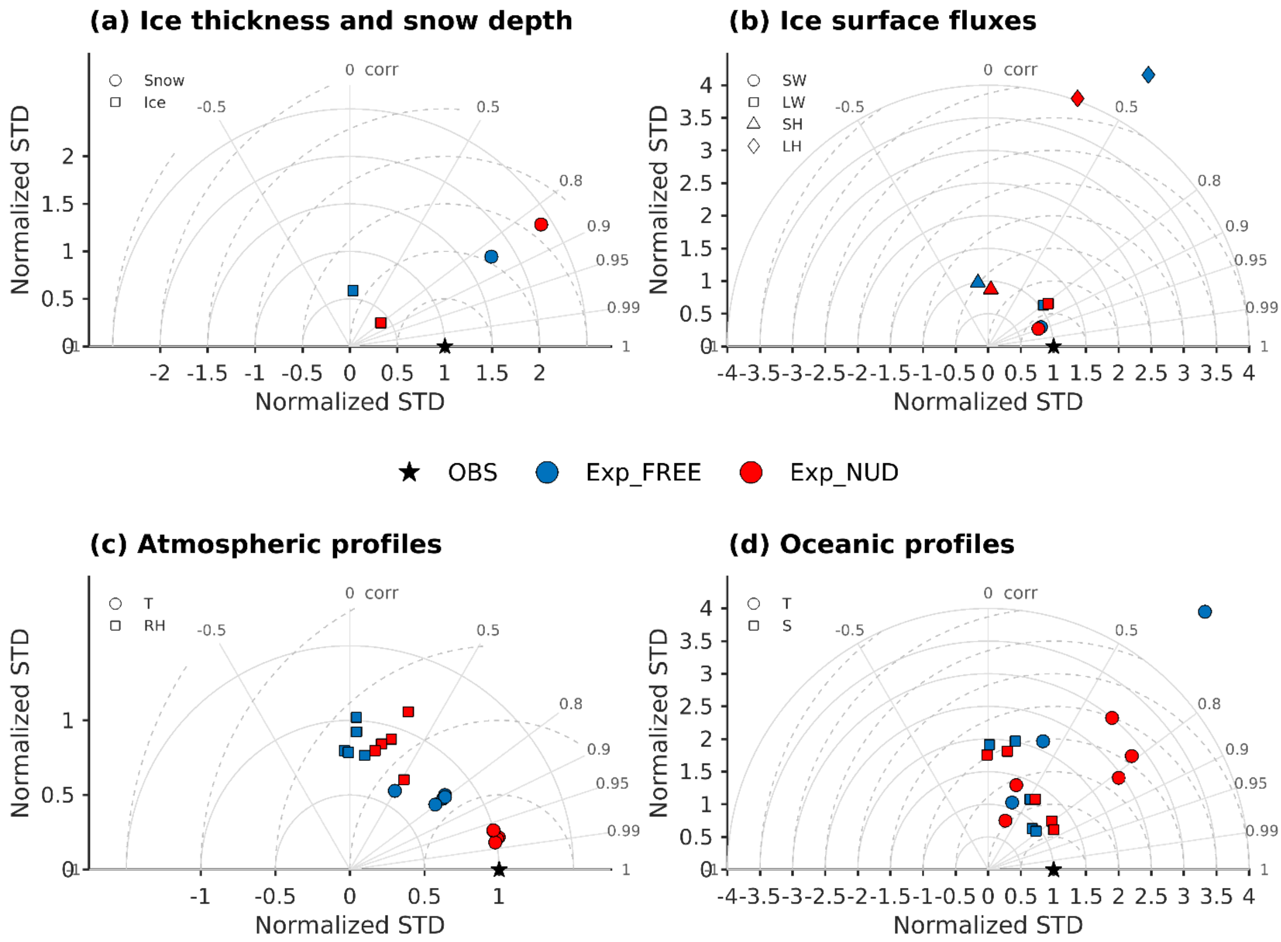

Figure R4 Taylor diagrams comparing simulated and observed (a) ice thickness and snow depth, (b) ice surface heat fluxes, (c) atmospheric profiles, and (d) ocean profiles. Atmospheric variables are evaluated at 1, 3, 5, 7, 9 km using radiosonde observations, oceanic variables at 10, 20, 50, 75, 100 m using CTD observations, and single-level sea-ice and surface variables using their corresponding observation levels. Statistics are calculated from paired model and observation samples for Exp_FREE (blue) and Exp_NUD (red). The Taylor diagrams summarize the correlation coefficient, normalized standard deviation, and normalized centered RMSE, while the corresponding bias and full RMSE are listed in Tables R1 and R2.

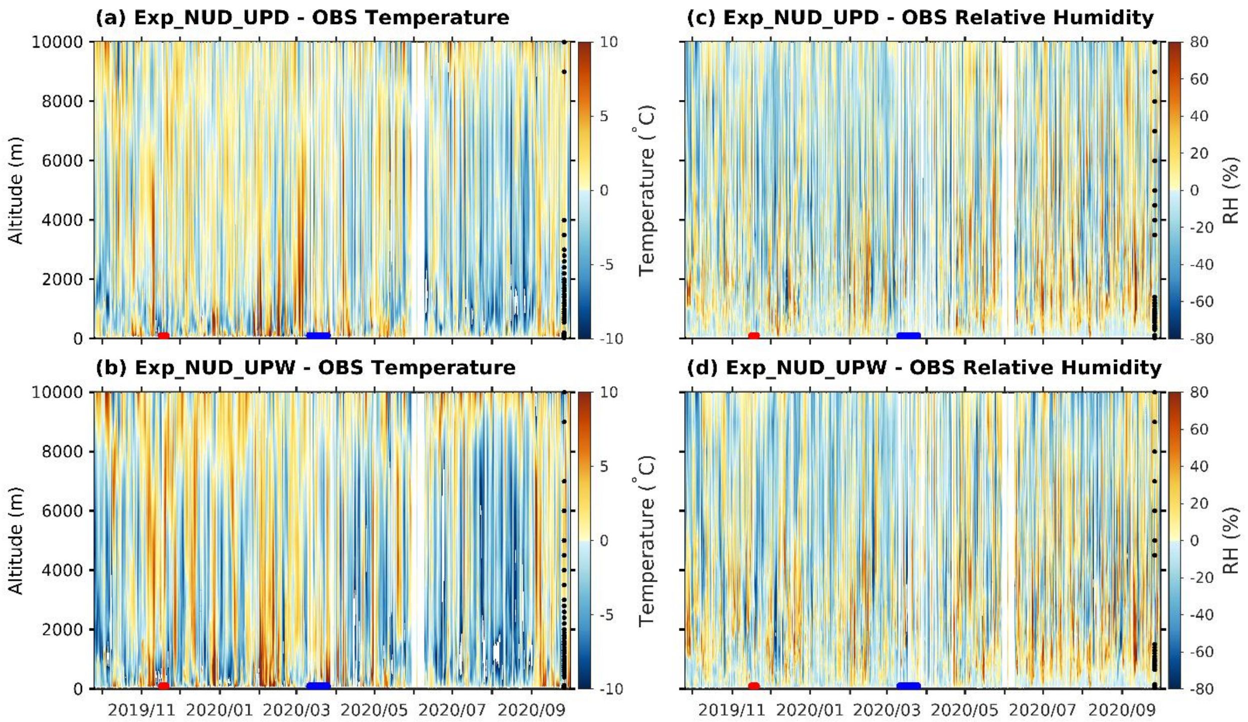

Figure R5 Time-height distributions of simulated temperature biases for (a) Exp_NUD_UPD, and (b) Exp_NUD_UPW, and the RH biases for (c) Exp_NUD_UPD, and (d) Exp_NUD_UPW with respect to the radiosonde observations along the track of MOSAiC drift. Note: red and blue lines on time-axis represent the STORM-1 and STORM-2 periods; black dots near the right edge 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.

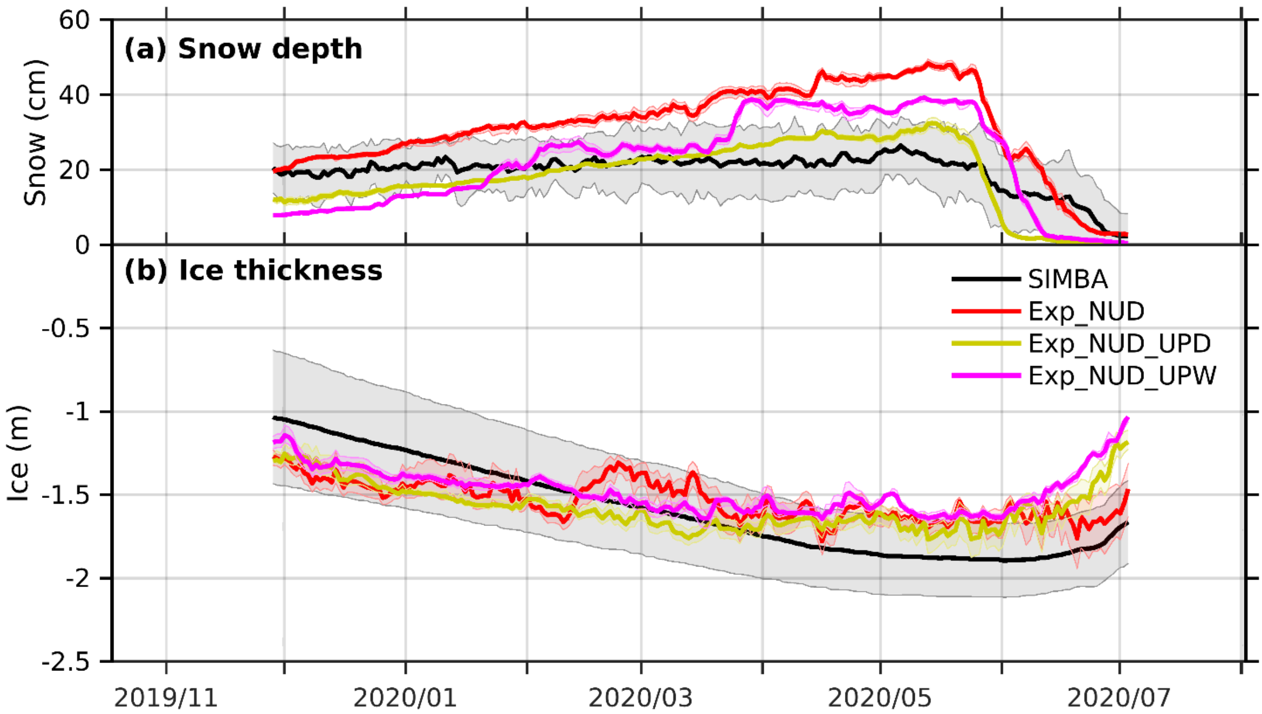

Figure R6 Time-series of daily-mean (a) snow depth and (b) sea ice thickness for SIMBA buoys (black line), Exp_NUD (red line), Exp_NUD_UPD (olive-yellow line), and Exp_NUD_UPW (purple line) along the track of MOSAiC drift. Note: the shadings represent the uncertainty range of the simulated and observed variables. For SIMBA buoys, the uncertainty range is ±1 standard deviation. For the simulations, the uncertainty range is the range of the nearest grid cell to the observed latitude/longitude and four nearby grid cells.

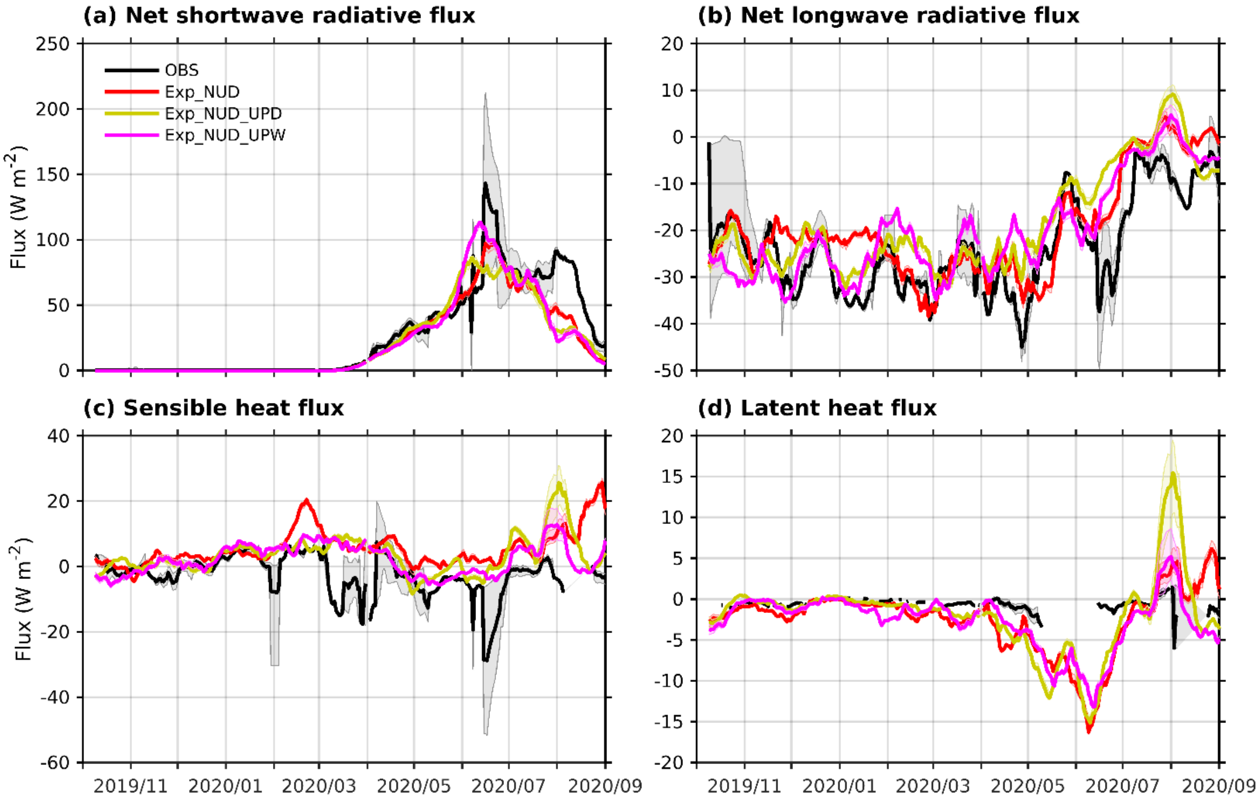

Figure R7 Time-series (15-day running-averaged) of (a) net shortwave radiative flux, (b) net longwave radiative flux, (c) sensible heat flux, and (d) latent heat flux for the observations (black line), Exp_NUD (red line), Exp_NUD_UPD (olive-yellow line), and Exp_NUD_UPW (purple line) along the track of MOSAiC drift. Note: all fluxes are defined as positive downwards. The shadings represent the uncertainty range of the simulated and observed variables. For the observations, the uncertainty range is the range of available sites with non-missing values. For the simulations, the uncertainty range is the range of the nearest grid cell to the observed latitude/longitude and four nearby grid cells.

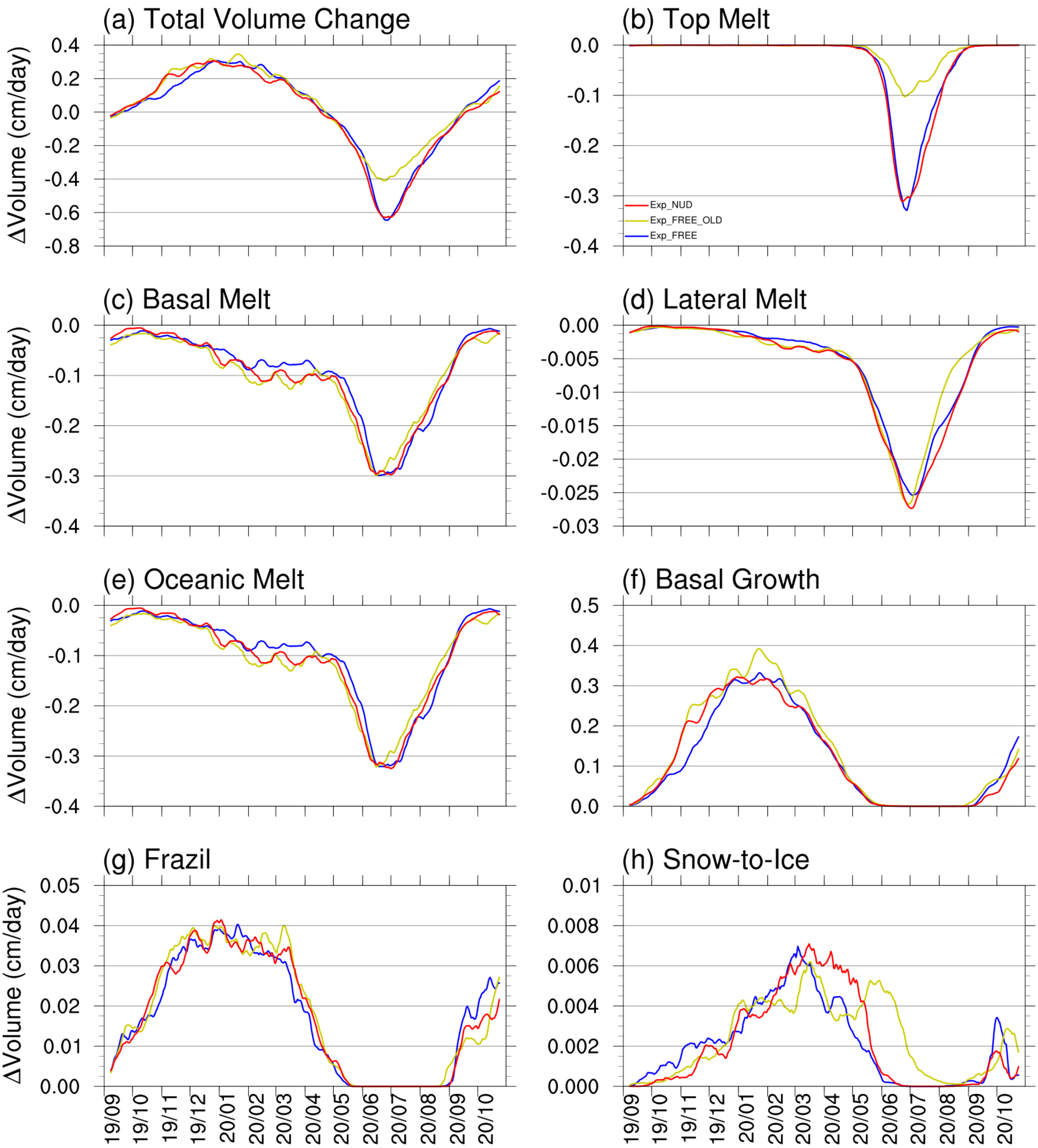

Figure R8 Time-series (15-day running-averaged) of sea ice volume budget for Exp_FREE (blue line), Exp_FREE_OLD (yellow line), and Exp_NUD (red line). Sea ice volume budgets include: (a) total volume change, (b) sea ice melt at the air-ice interface (top melt), (c) sea ice melt at the bottom of the ice (basal melt), (d) sea ice melt at the sides of the ice (lateral melt), (e) sea ice melt from oceanic energy (basal plus lateral melt), (f) sea ice growth at the bottom of the ice (basal growth), (g) sea ice growth by supercooled open water (frazil), and (h) sea ice growth due to transformation of snow to sea ice (snowice). Note: (a)-(h) are averaged over the model domain (Fig. S1).

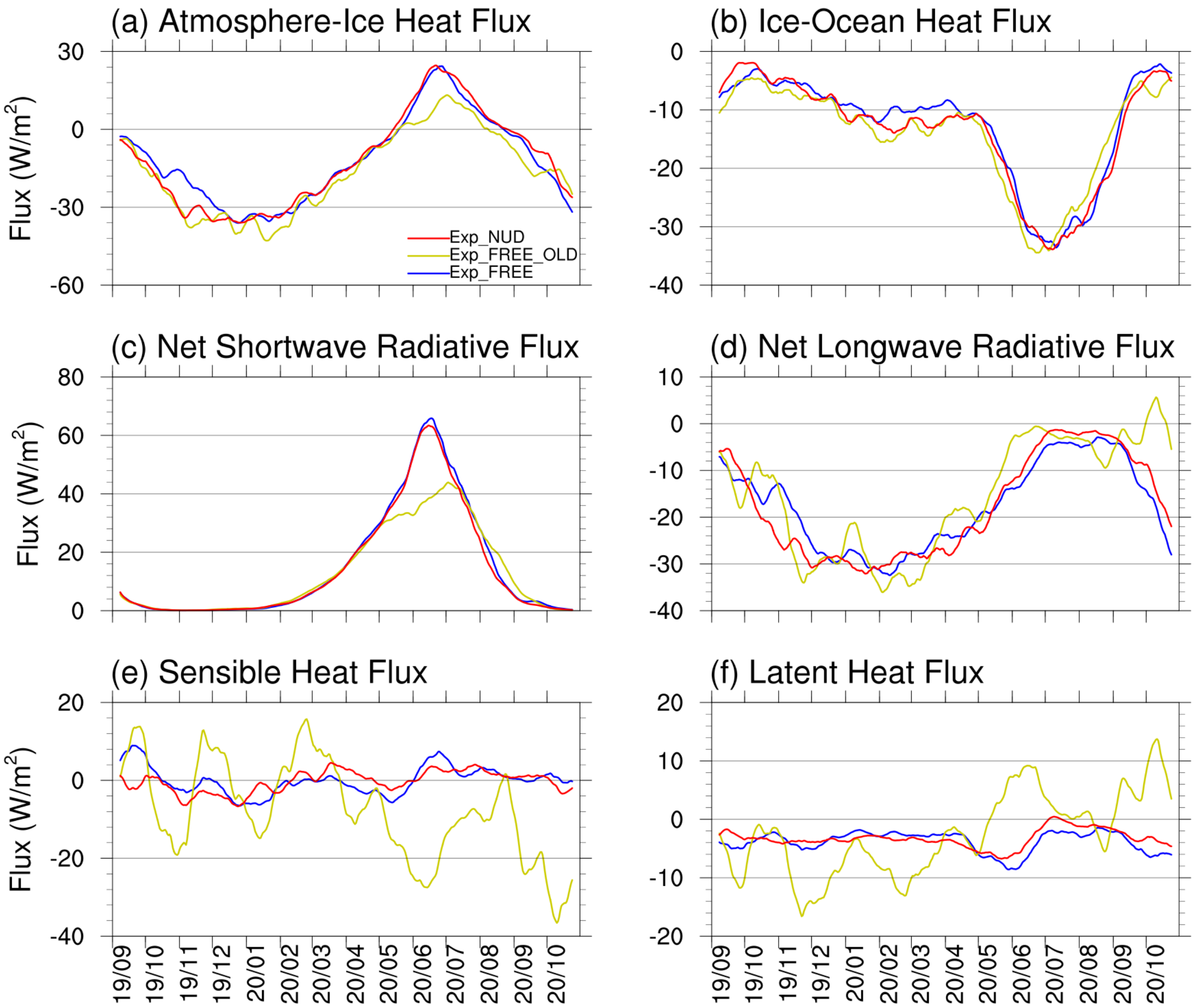

Figure R9 Time-series (15-day running-averaged) of (a) atmosphere–ice heat flux, (b) ice–ocean heat flux, (c) net shortwave radiative flux, (d) net longwave radiative flux, (e) sensible heat flux, and (f) latent heat flux for Exp_FREE (blue line), Exp_FREE_OLD (yellow line), and Exp_NUD (red line). Note: (a)-(f) are averaged with SIC-weighting over all grid cells with at least 1% SIC to represent the mean value of the grid cell, are defined as positive downward.

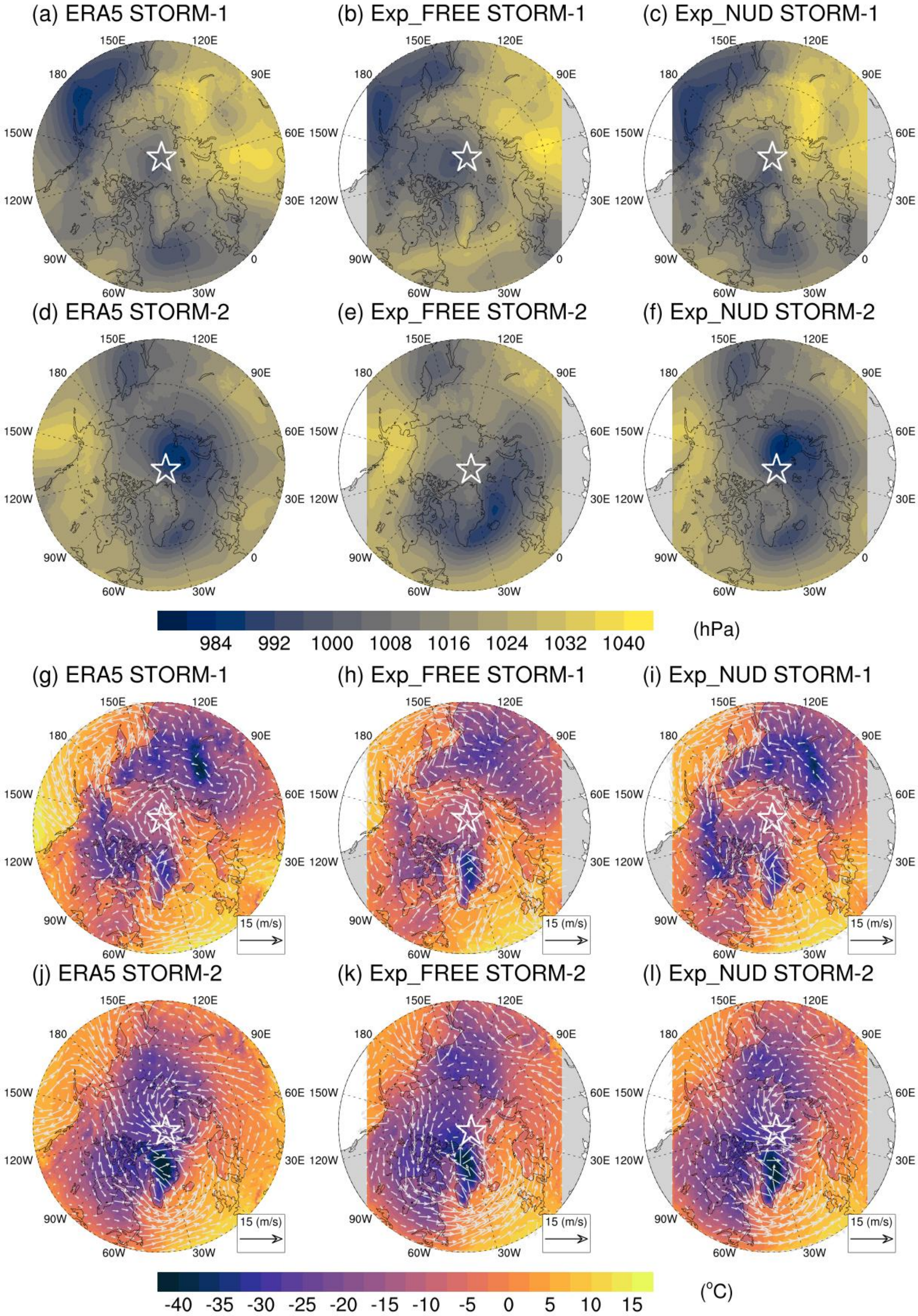

Figure R10 The averaged mean sea level pressure of (a, d) ERA5, (b, e) Exp_FREE, and (c, f) Exp_NUD, and the averaged 2-m temperature and 10-m wind fields of (g, j) ERA5, (h, k) Exp_FREE, and (i, l) Exp_NUD during the period of STORM-1 and STORM-2. Note: white star represents the averaged location of radiosondes during the corresponding period, and white gap in panels of simulated fields represents out of model domain (Fig. S1).

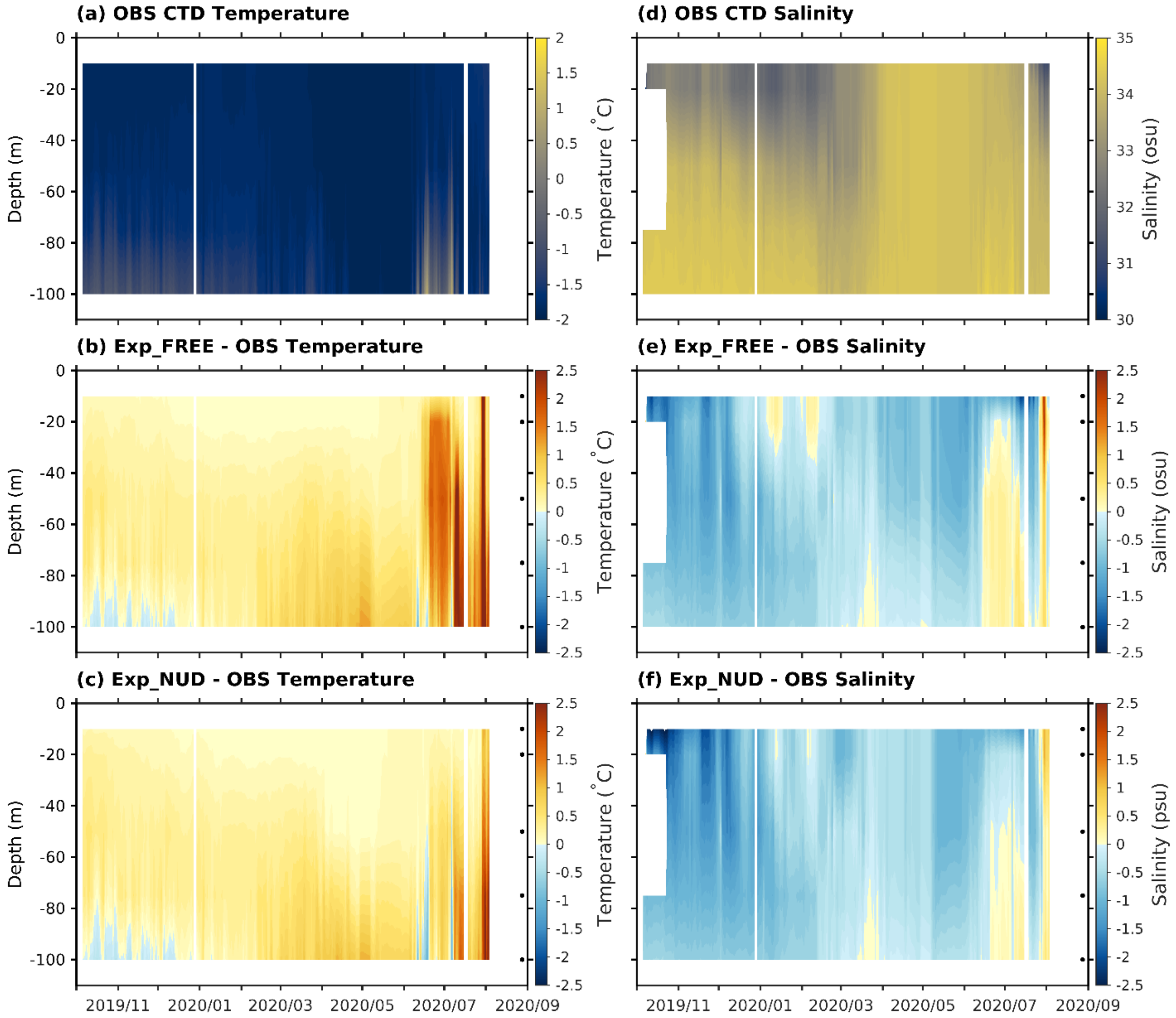

Figure R11 Time-depth distributions of CTD observations for (a) temperature and (d) salinity, the corresponding simulated temperature biases for (b) Exp_FREE, and (c) Exp_NUD, and the salinity biases for (e) Exp_FREE, and (f) Exp_NUD along the track of MOSAiC drift. Note: black dots near the right edge of panels (b), (c), (e), and (f) indicate depth 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.

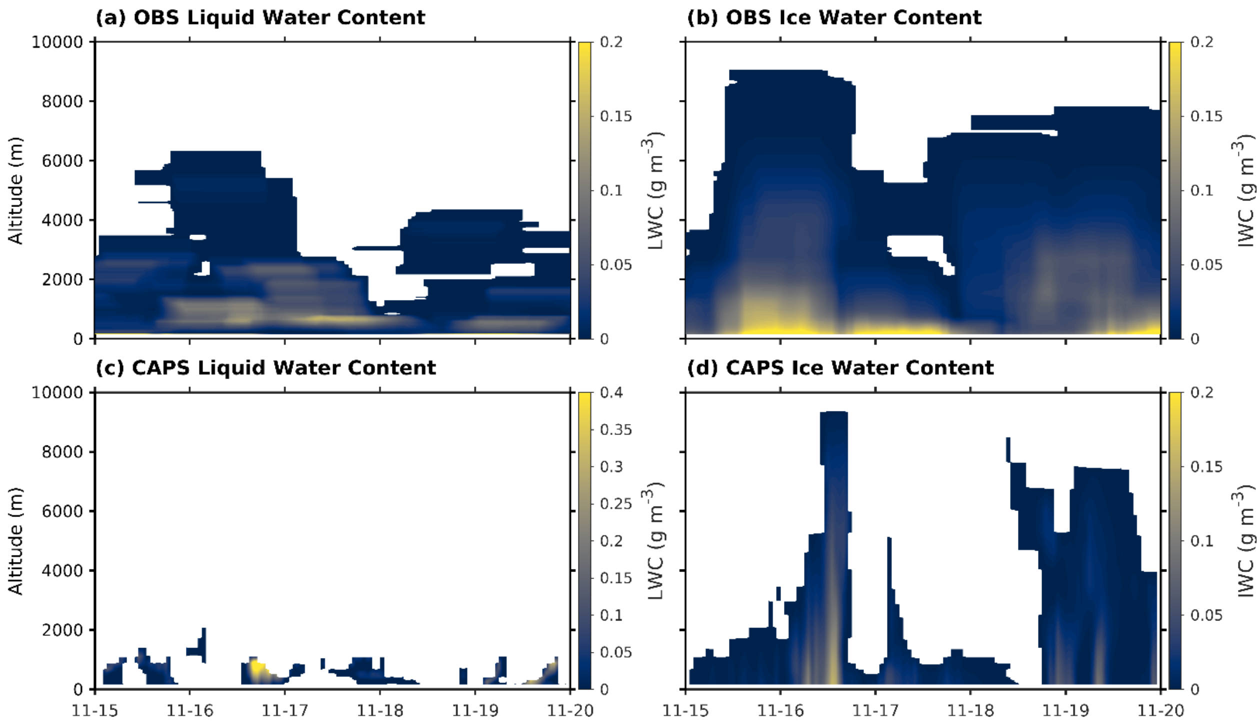

Figure R12 The evolutions of (a, c) liquid water content profile, and (b, d) ice water content profile for the observations and the model simulations during the STORM-1 period.

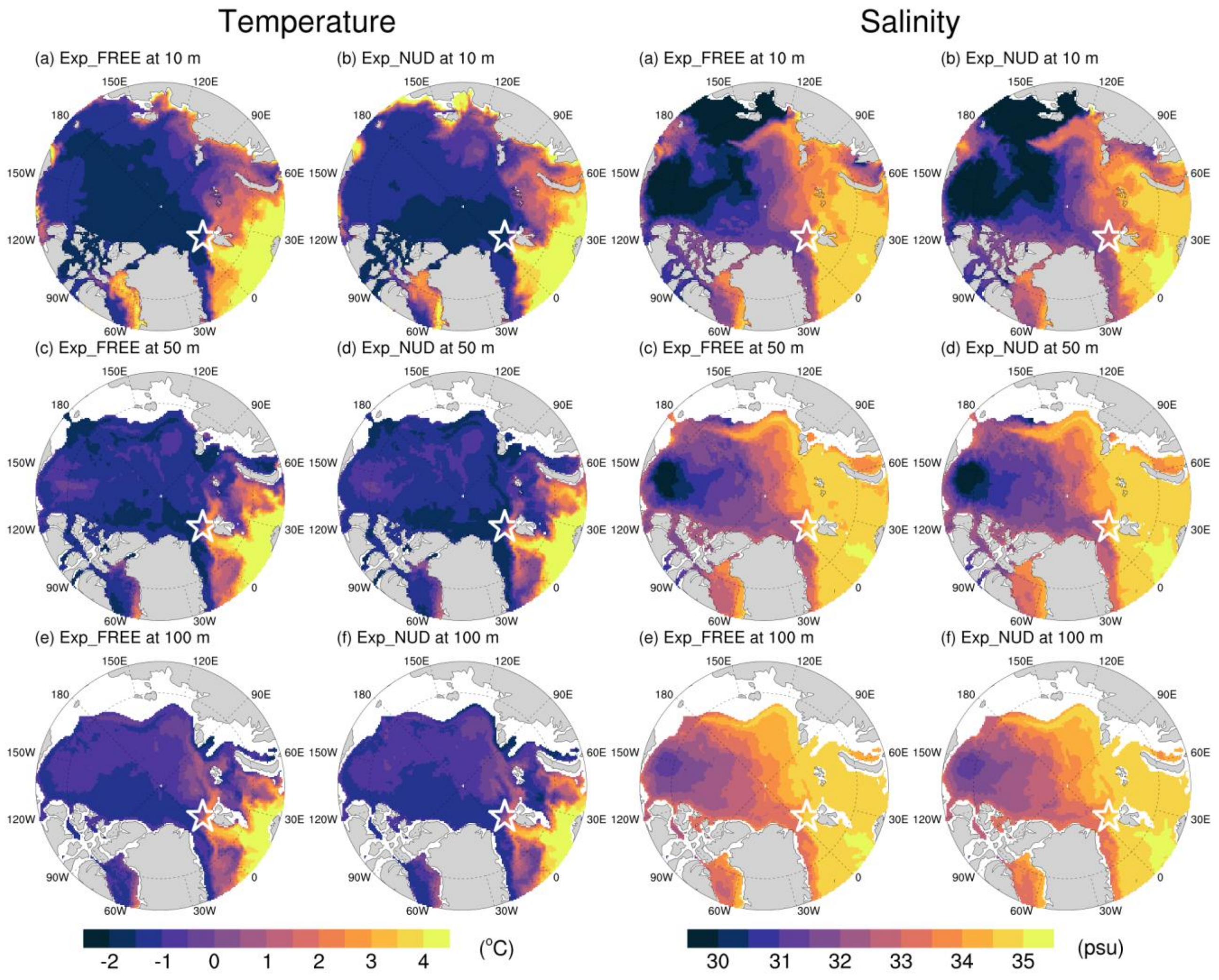

Figure R13 Daily mean temperature (left panel) and salinity (right panel) of Exp_FREE and Exp_NUD at (a, b) 10, (c, d) 50, and (e, f) 100 m depth on July 1st, 2020. Note: white star represents the averaged location of CTD buoys

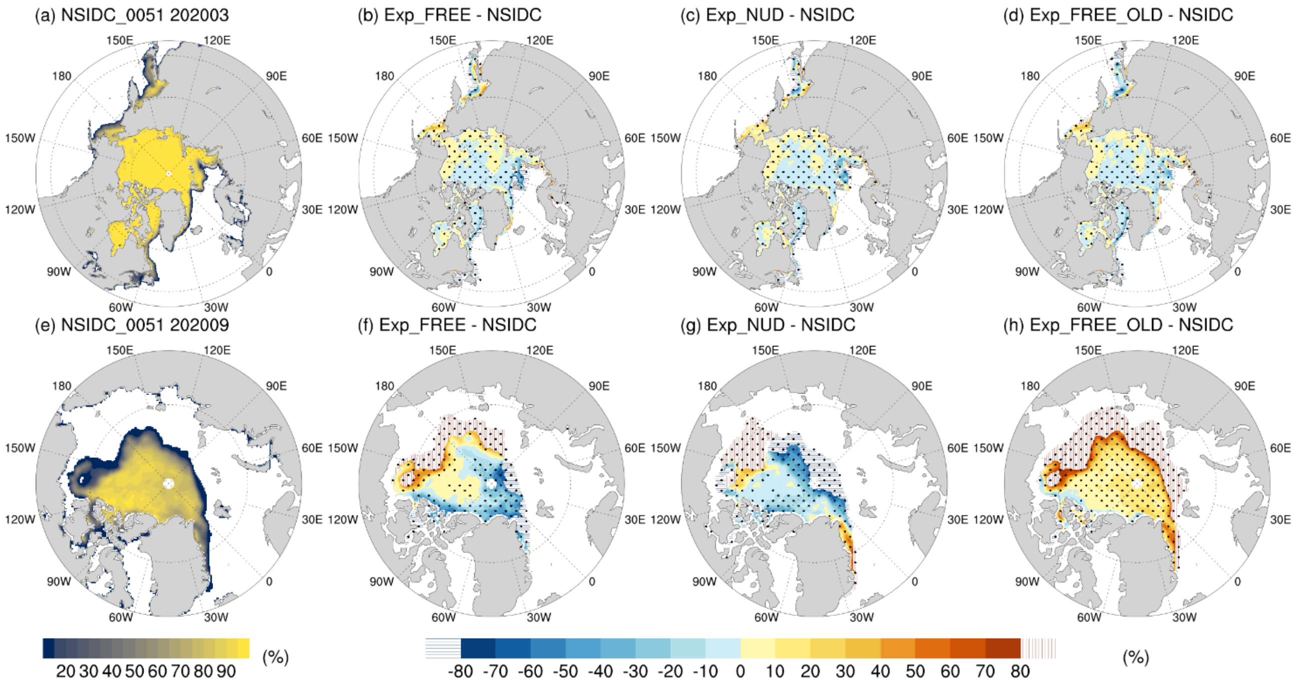

Figure R14 Monthly mean sea ice concentration (SIC) for the observations in (a) March 2020, and (e) September 2020, and the difference between the experiments and the observations for Exp_FREE, Exp_NUD, and Exp_FREE_OLD in (b–d) March 2020, and in (f–h) September 2020. Note: vertical/horizontal-lining 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 $q = 0.05$.

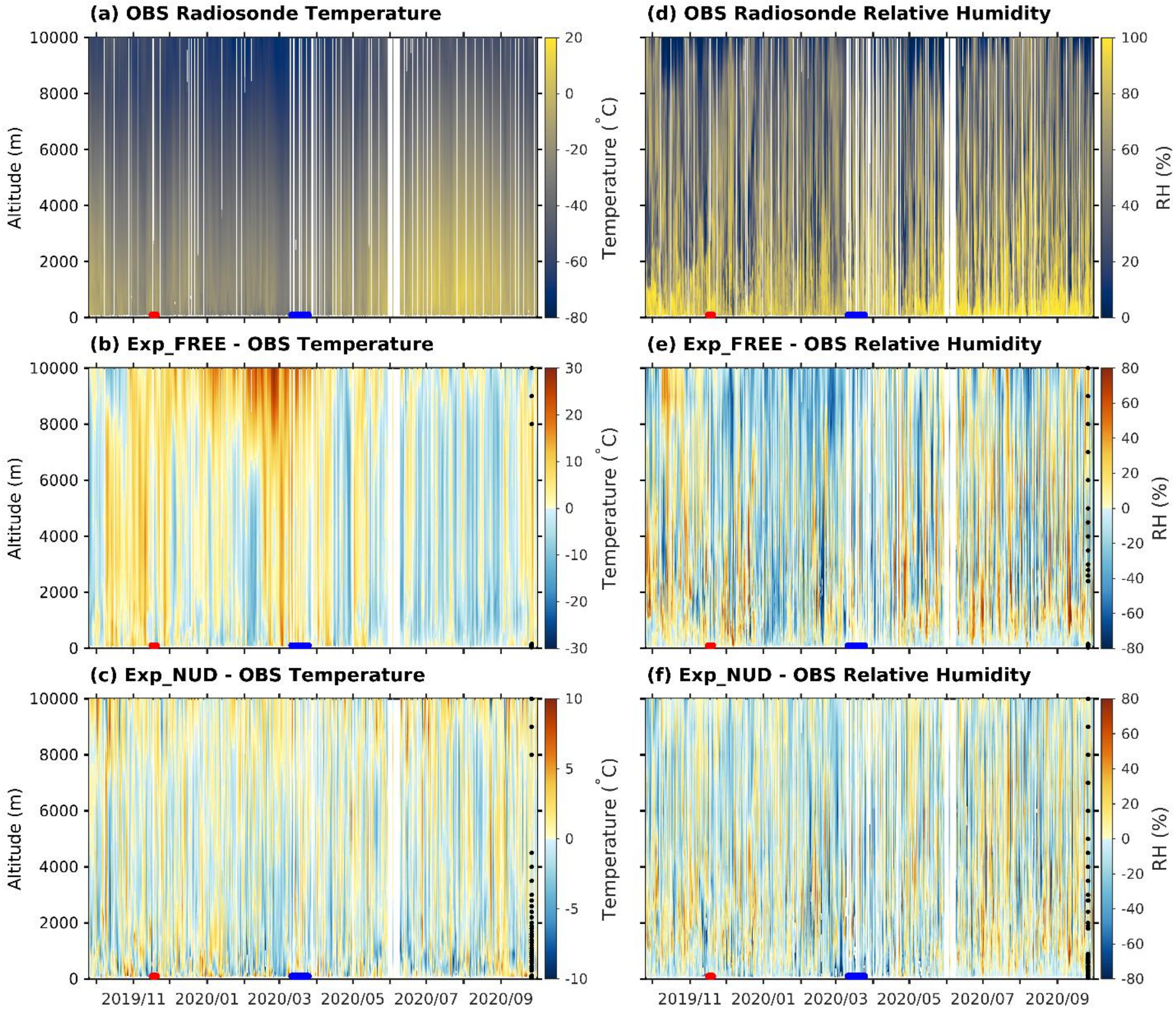

Figure R15 Time-height distributions of radiosonde observations for (a) temperature and (d) relative humidity (RH), the corresponding simulated temperature biases for (b) Exp_FREE, and (c) Exp_NUD, and the RH biases for (e) Exp_FREE, and (f) Exp_NUD along the track of MOSAiC drift. Note: red and blue 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.

Table R1 Bias and RMSE of Exp_FREE against MOSAiC observations for atmospheric, oceanic, and sea-ice variables. Atmospheric profile variables are evaluated at altitudes of 1, 3, 5, 7, and 9 km, oceanic profile variables at depths of 10, 20, 50, 75, and 100 m, and single-level variables at their corresponding observations levels.

Variable	Level	Bias	RMSE
SIT (m) / snow depth (cm)	mean	-0.38 m / 1.51 cm	0.49 m / 5.30 cm
NetSW / NetLW (W/m^2)	mean	-4.67 / 5.19 W/m^2	12.7 / 8.16 W/m^2
SH / LH (W/m^2)	mean	6.29 / -1.49 W/m^2	11.67 / 4.18 W/m^2
Air temperature (˚C) / RH (%)	1 km	-0.89 ˚C / 1.82 %	6.27 ˚C / 29.00 %
	3 km	0.21 ˚C / -5.20 %	6.00 ˚C / 35.46 %
	5 km	0.45 ˚C / -4.58 %	6.13 ˚C / 33.03 %
	7 km	1.06 ˚C / -6.42 %	6.12 ˚C / 28.38 %
	9 km	3.98 ˚C / -13.15 %	8.24 ˚C / 29.28 %
Ocean temperature (˚C) / salinity (%)	10 m	0.15 ˚C / -0.76 psu	0.39 ˚C / 0.98 psu
	20 m	0.29 ˚C / -0.58 psu	0.50 ˚C / 0.81 psu
	50 m	0.52 ˚C / -0.58 psu	0.72 ˚C / 0.74 psu
	75 m	0.66 ˚C / -0.46 psu	0.83 ˚C / 0.60 psu
	100 m	0.58 ˚C / -0.25 psu	0.87 ˚C / 0.40 psu

Table R2 Same as Table R1, but for Exp_NUD.

Variable	**Level**	**Bias**	**RMSE**
SIT (m) / snow depth (cm)	**mean**	**-0.05 m / 11.72 cm**	**0.20 m / 14.10 cm**
NetSW / NetLW (W/m^2)	**mean**	**-5.54 / 4.70 W/m^2**	**13.16 / 7.94 W/m^2**
SH / LH W/m^2	**mean**	**8.99 / -1.90 W/m^2**	**12.31 / 3.89 W/m^2**
Air temperature (˚C) / RH (%)	**1 km**	**-0.61 ˚C / -2.07 %**	**2.77 ˚C / 25.34 %**
	3 km	**-0.42 ˚C / -2.44 %**	**2.14 ˚C / 30.08 %**
	5 km	**-0.16 ˚C / -4.87 %**	**2.09 ˚C / 29.24 %**
	7 km	**0.01 ˚C / -6.56 %**	**1.82 ˚C / 25.70 %**
	9 km	**0.61 ˚C / -5.74 %**	**2.28 ˚C / 20.22 %**
Ocean temperature (˚C) / salinity (psu)	**10 m**	**0.11 ˚C / -0.87 psu**	**0.17 ˚C / 1.09 psu**
	20 m	**0.15 ˚C / -0.70 psu**	**0.21 ˚C / 0.88 psu**
	50 m	**0.33 ˚C / -0.62 psu**	**0.41 ˚C / 0.77 psu**
	75 m	**0.49 ˚C / -0.53 psu**	**0.61 ˚C / 0.65 psu**
	100 m	**0.46 ˚C / -0.30 psu**	**0.73 ˚C / 0.43 psu**

Table R3 Computational-cost benchmark for the original and revised CAPS coupling configurations on an HPC system with two 32-core Intel Xeon Platinum 8358P processors at 2.60 GHz and 512 GB RAM per node. SYPD: simulated years per wall-clock day. CHPY: Core-hours per simulated year. MaxRSS: the maximum resident set size. AveRSS: the average resident set size. MaxVMSize: the maximum virtual memory size.

	The original coupling	The revised coupling
Configuration	Exp_FREE_OLD	Exp_FREE
Simulated period	31 days	31 days
Core allocated	320 in total 240 for WRF 80 for ROMS/CICE	320 in total 240 for WRF 80 for ROMS/CICE
Wall time	3:04:29	3:04:33
SYPD	0.6629	0.6627
CHPY	11584.8	11588.9
Wall-time overhead	N/A	0.036 %
MaxRSS	2.56 G	3.41 G
AveRSS	0.19 G	0.19 G
MaxVMSize	4.77 G	6.47 G