1 Supplementary Material
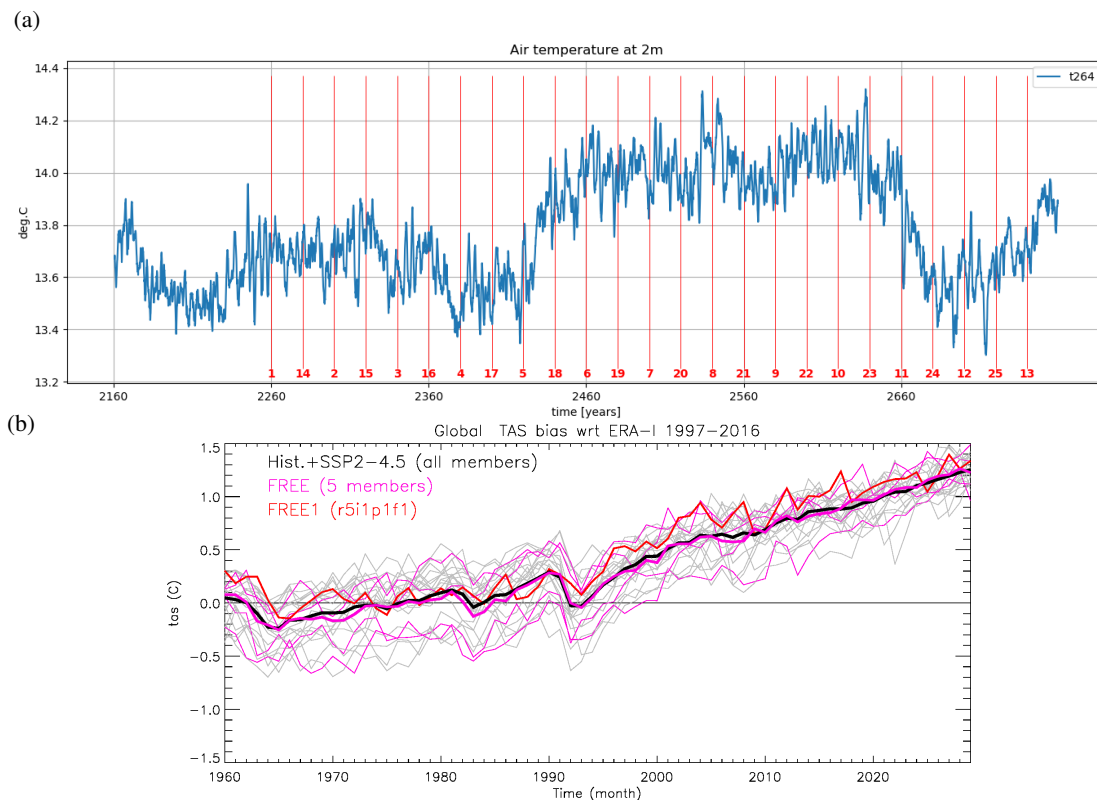
Figure S1. (a) Modelled global mean surface air temperature at 2 m (TAS) with EC-Earth3 in the 500-year pre-industrial control run (as 12-year moving averages). Numbers 1-25 indicate the initial states selected every 20 years for 25 realizations of EC-Earth3 CMIP6 historical simulations (i.e. r1-r25). In the present study a subset of 5 members (e.g. r1, r4, r5, r8 and r18) of the CMIP6 historical (1850-2014) and the consecutive SSP2-4.5 scenario (2015-2100) experiments, referred to as FREE in the text, are used for assessment of the prediction skills of the initialized decadal experiments. The 5 members are selected to represent historical simulations stated from different state in the time series of the global mean TAS, i.e., r1 - on an "average" state, r4 and r5 - on a state of relatively cold TAS, and r8 and r18 on a relatively warm state. Note a single simulation labeled r5 provides the model climatology for anomaly initialization in this study, so call FREE1. Credit: the EC-Earth3 development portal by Philippe Le Sager, Royal Netherlands Meteorological Institute (KNMI). (b) TAS bias in the EC-Earth3 CMIP6 historical and the corresponding SSP2-4.5 simulations with respect to the ERA-Interim 20-year average over 1997-2016. Thin lines present individual members (e.g. FREE1 in red, FREE in pink and all members in gray) and thick lines present ensemble means for FREE (in pink) and all members (in black), respectively. Figure (b) shows that, during the period of interest, there is no significant differences in the TAS ensemble mean and its variability between the FREE (5-member) and the full ensemble of 25 members, indicating that the sub-ensemble FREE can well represent the overall feature of the full ensemble.

Table S1. Summary of EC-Earth DCPP experiments with AI

Experiment name	Resolution	Atm ICs [Ensemble size]	Anomalous Ocn states [Ensemble size]	Initialized hindcasts	Years available [Ensemble size]
V2.3-CMIP5[†]	Atm: T159, 62 levels Ocn: ORCA 1°,42 levels	ERA40[1](-1978), ERAII[1](1979-)	ORAS4[5] for T&S, U&V SIC[1] from another model	Yearly start, Nov 1, 10-year long	1960-2005[5]
V3.3-CMIP6	Atm: T255, 91 levels Ocn: ORCA 1°,75 levels	ERAI [1+2 perturb.]	ORAS5[5] for T&S, SIC, SIT, SNT	Yearly start, Nov 1, 10-year long	1979-2018[15][‡]

T&S = temperature and salinity; U&V = horizontal velocities; Atm & Ocn = atmosphere and Ocean, respectively; [†] refers to Hazeleger et al. (2013);
[‡] Besides the present 5-member AI2 hindcasts, there are 10 additional members generated by combining two sets of perturbed atmospheric ICs and 5 sets of ocean and sea-ice ICs as a total of 15-member ensemble contributed to the CMIP6 DCPP with EC-Earth3-CPSAI; an extension for the period 1960-1978 is based on one single member of ORAS5 for ocean and sea-ice ICs together with 15 perturbed atmospheric ICs from ERA-40.
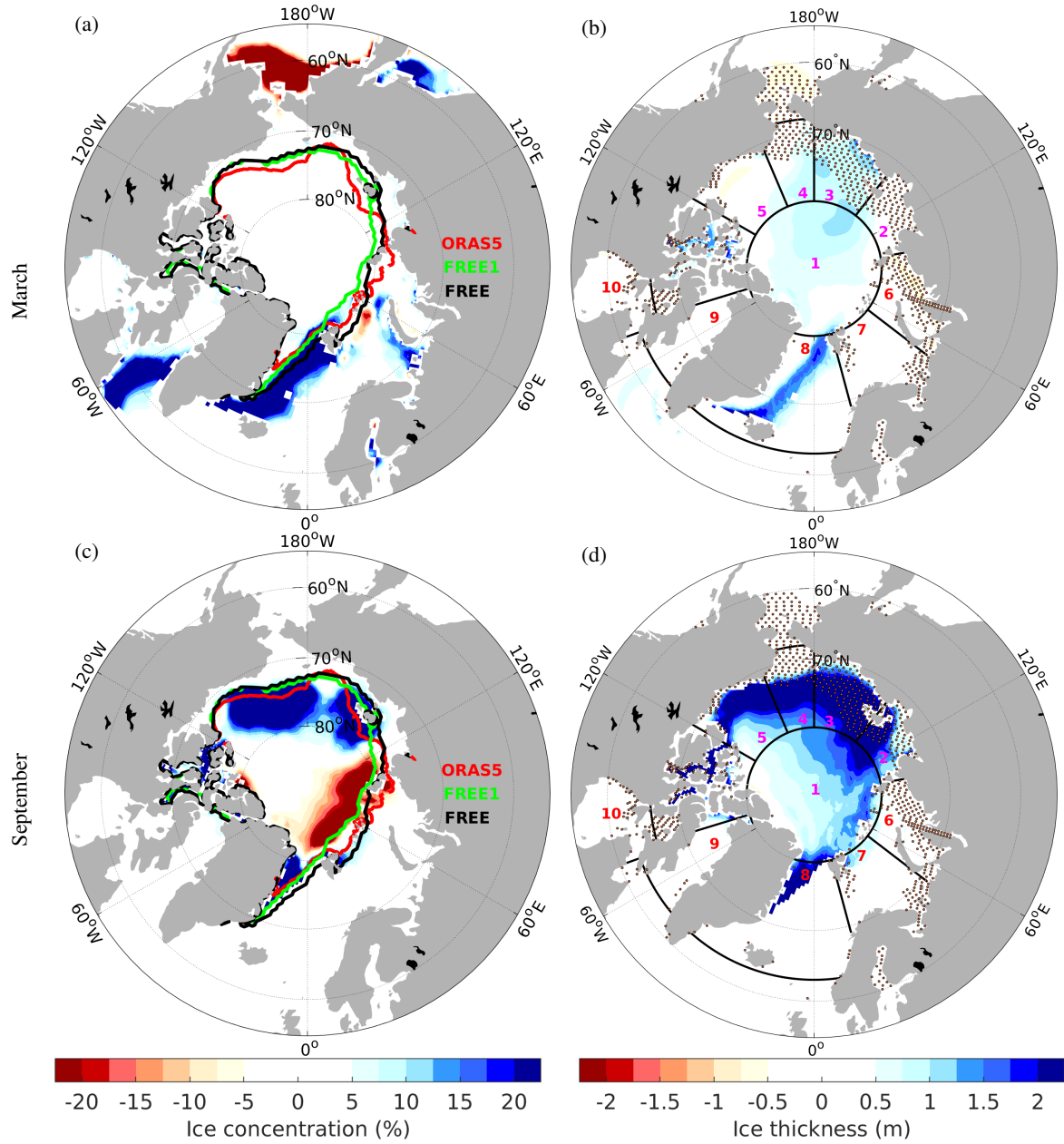
Figure S2. Difference of 20-year climatology of annual mean maximum (March, top panels) and minimum (September, bottom panels) sea ice states for the forecast assessment period 1997-2016, calculated as 5-member ensemble mean FREE minus REF. (a and c) SIC. Color lines indicate the September sea-ice extent (15 % SIC) climatology for ORAS5 (red), FREE1 (green) and FREE (black). (b and d) SIT with red dots indicating a mask (water depth <100 m) to neglect SIT initialization. All maps have the bounding latitude at 56°N. Note that, blue areas in SIC and SIT represent regions with more sea-ice in models than the reference, corresponding to cold bias in surface temperature. Ten sub-regions represent the sector for the Central Arctic Ocean (CAO) and its adjacent waters, namely (1) central Arctic (80°N north), (2) Laptev Sea, (3) East Siberian Sea, (4) Chukchi Sea and (5) Beaufort Sea, and the sector for the Arctic Marginal Ice Zone (MIZ) and the transition waters, namely (6) Kara Sea, (7) Barents Sea, (8) GIN (Greenland/Iceland/Norwegian) Seas, (9) Baffin Bay and (10) Hudson Bay.
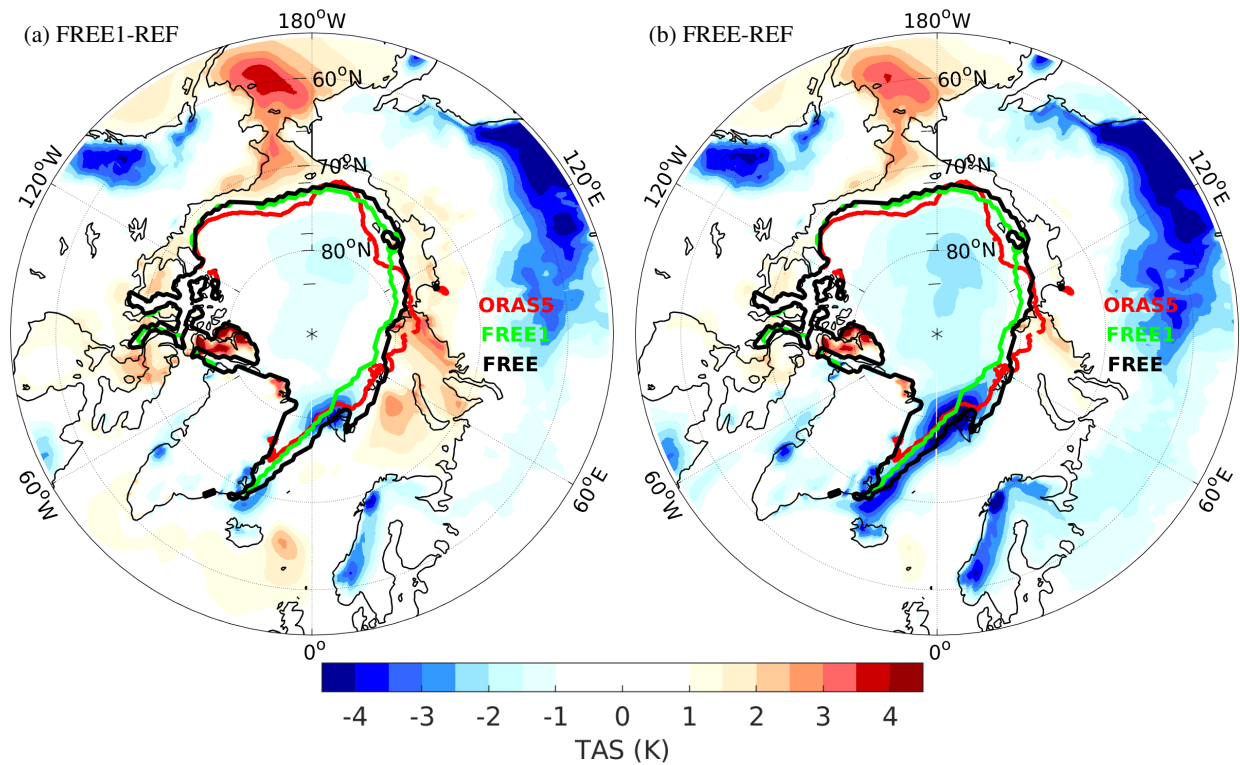
Figure S3. Difference of 20-year climatology of annual mean TAS for the forecast assessment period 1997-2016: calculated as single meanber FREE1 minus REF (a) and 5-member ensemble mean FREE minus REF (b), respectively. Color lines denote the September sea-ice extent ((15 % SIC) climatology for ORAS5 (red), FREE1 (green) and FREE (black) over the same period, indicating the domain with multiple year ice. Maps have the bounding latitude is 56°N and the 70°N circle indicates the polar cap domain.
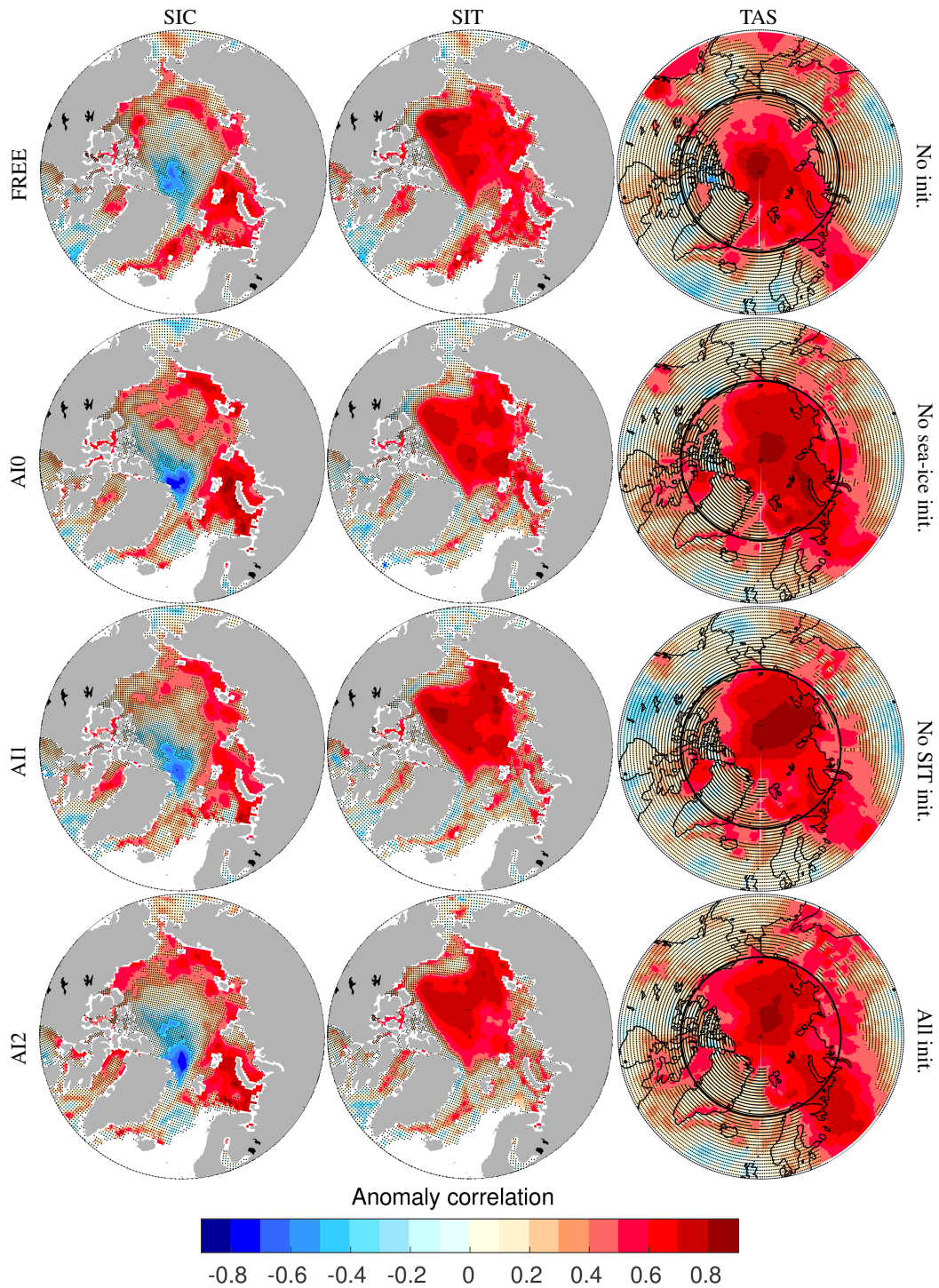
Figure S4. Anomaly correlation of SIC (left), SIT (middle) and TAS (right), respectively for FREE, AI0, AI1 and AI2 experiments from top to bottom. The average of forecast years 2-5 are evaluated. The reference data are taken from ORAS5 or ERAI over the period 1997-2016. Regions are stippled if not significant (p = 0.05). The black line illustrates the polar cap domain (north of 70° N).
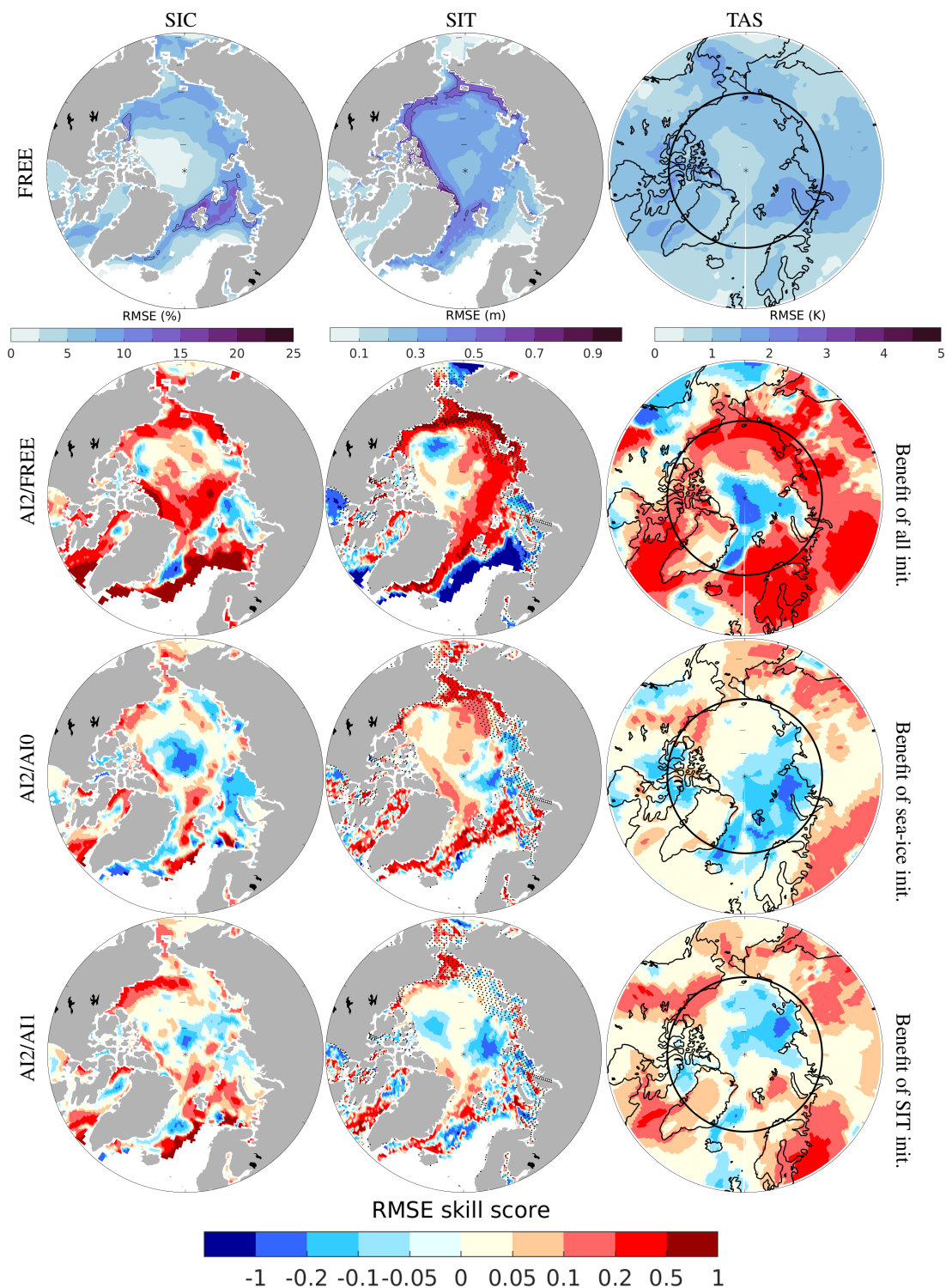
Figure S5. RMSE of FREE with respect to REF (top row) and RMSE skill score (lower rows) for the year 2-5 average: SIC (left), SIT (middle) and TAS (right). The contour lines mark RMSE$\geq$10% or 0.5 m for SIC and SIT, respectively. The RMSE skill score is calculated as 1-(RMSE$_{AI2}$/RMSE$_{INIT}$), where *INIT* denotes FREE (no init.), AI0 (no sea-ice init.) and AI1 (no SIT init.). Scores above 0 denote more accurate in AI2 than *INIT*, and vice versa. The regions discard SIT initialization in AI2 are stippled.
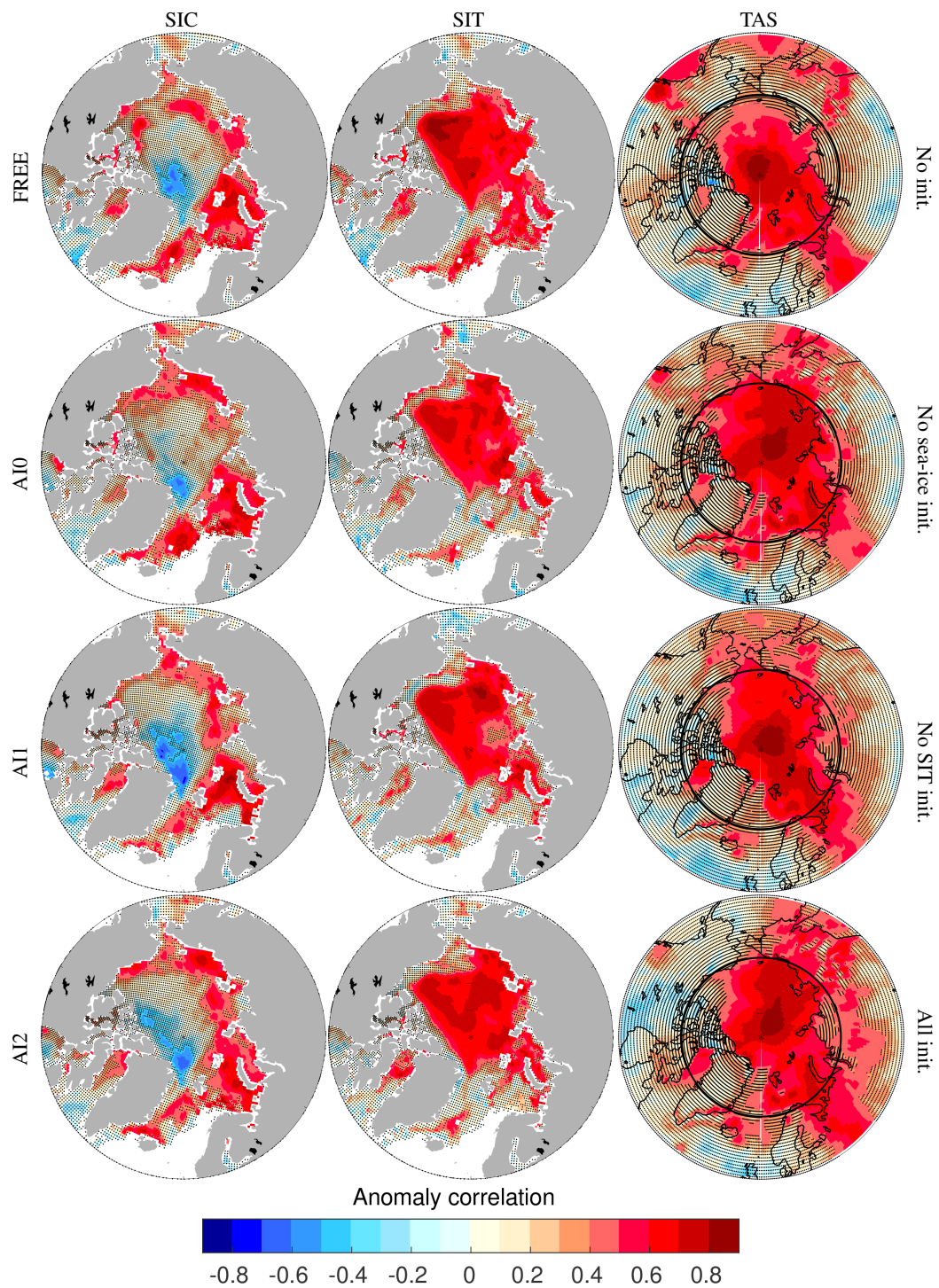
Figure S6. Same as Fig. S4, but for the year 6-9 average.
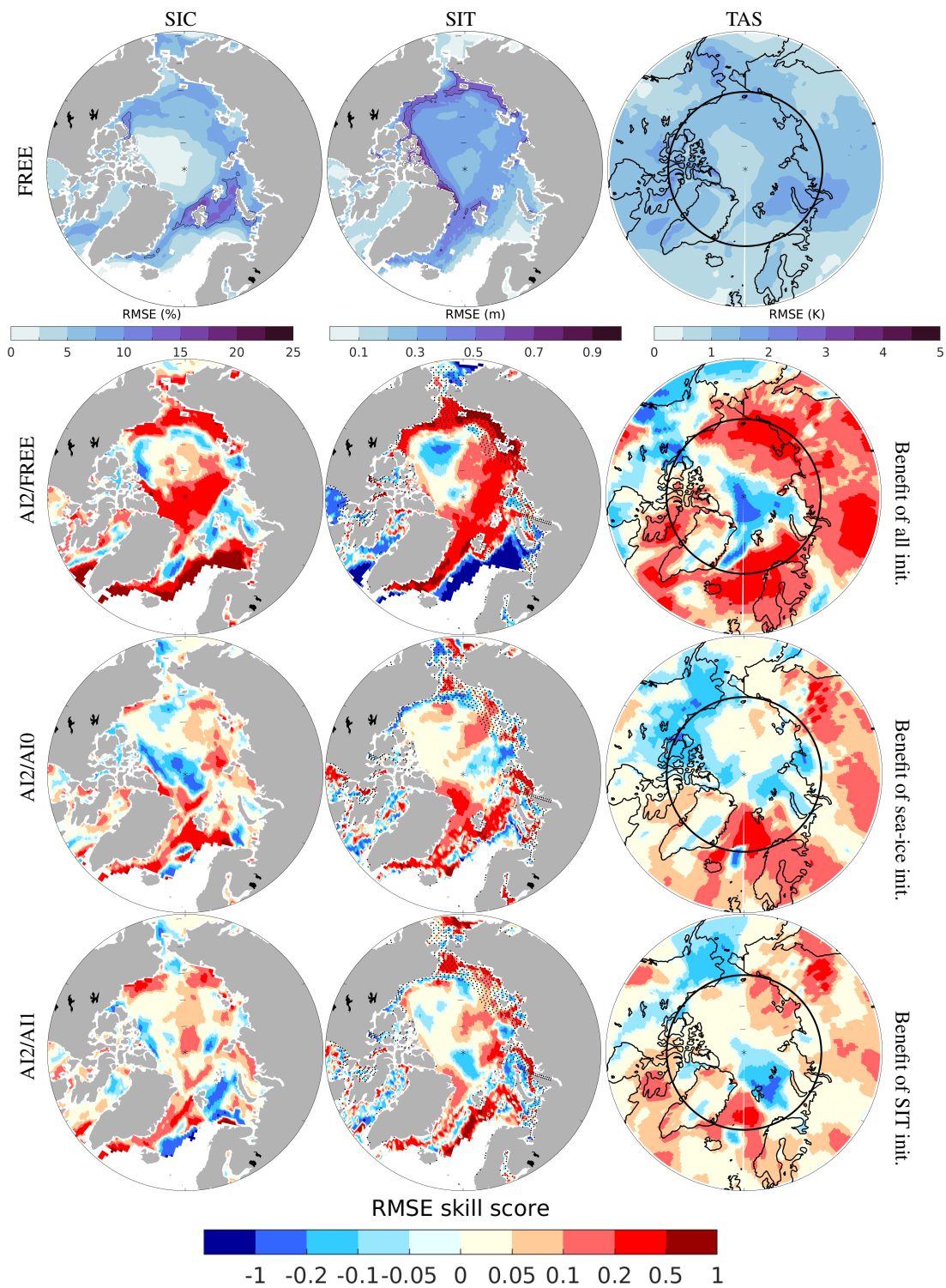
Figure S7. Same as Fig. S5, but for the year 6-9 average.
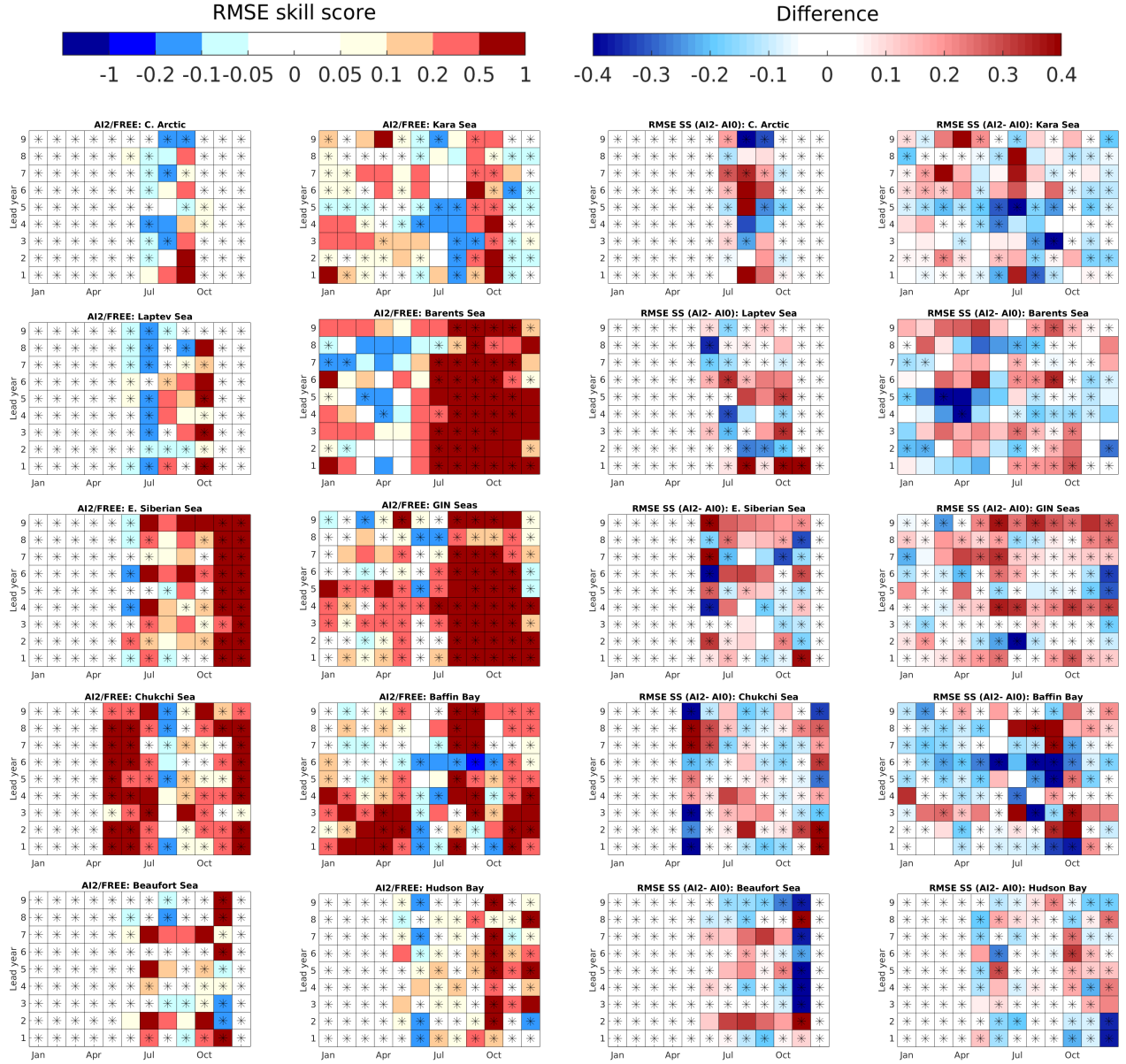
Figure S8. Prediction of regional SIE on decadal time-scales: RMSE skill score of AI2 relative to FREE (AI2/FREE, two columns on the left) and the difference of skill score between AI2 and AI0 (two columns on the right). RMSE is calculated based on area-integrated SIE with respect to REF for AI2 and FREE, respectively, and then the RMSESS is calculated as $1-(\mathrm{RMSE}_{AI2}/\mathrm{RMSE}_{FREE})$. The difference is calculated as AI2/FREE - AI0/FREE. Red (blue) colors denote higher (lower) skill score than the reference. White colors in AI2/FREE denote 0 score, meaning $\mathrm{RMSE}_{AI2}/\mathrm{RMSE}_{FREE}$=1. Boxes are stippled if not significant (p = 0.05).
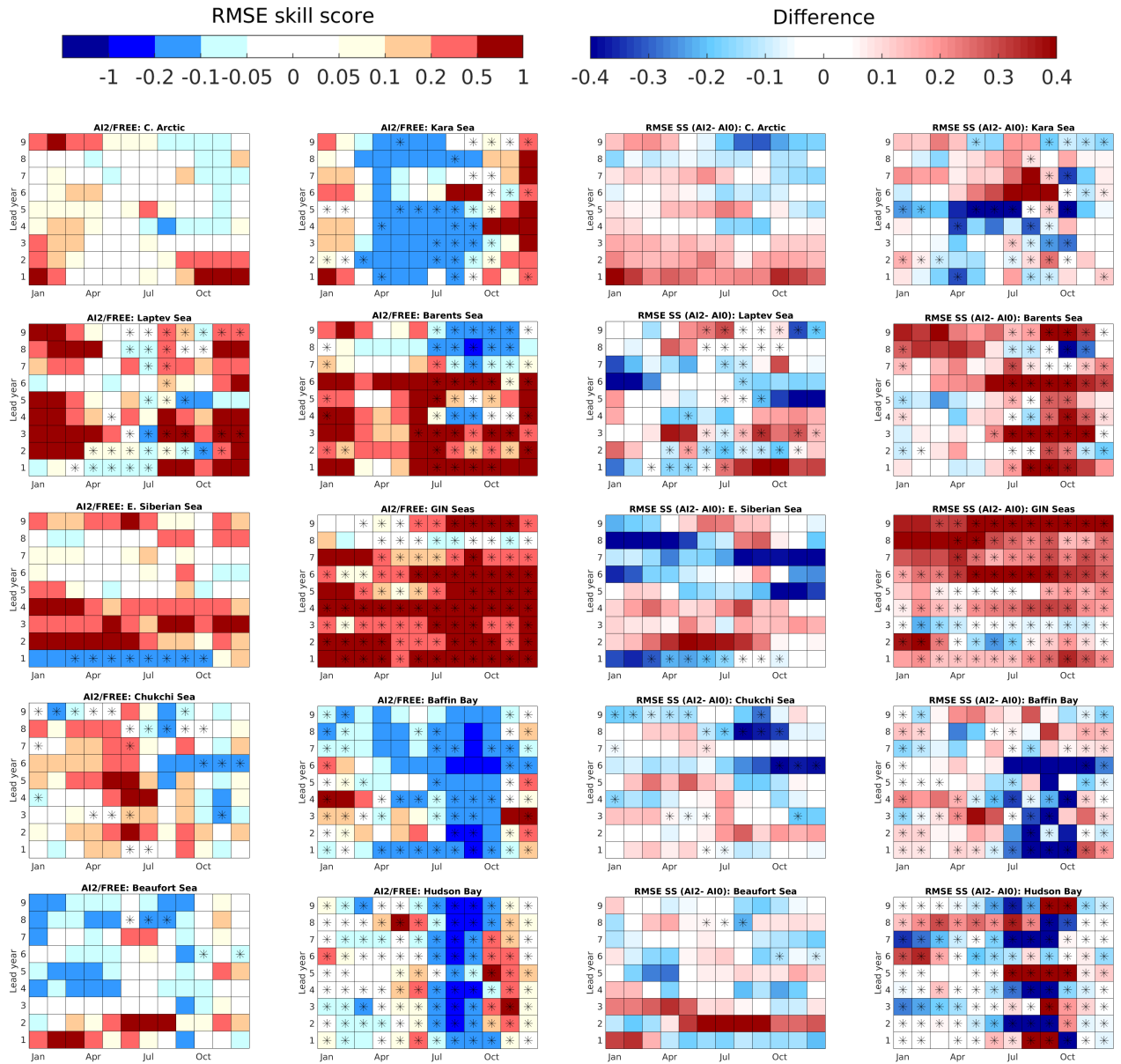
Figure S9. As Fig. S8, but for regional SIV.