

Table S1. Summary of the near surface mean bias, RMSE, and skill score for each product for the Extratropical Northern Hemisphere. Metrics are separated into an annual mean metric, and a metric for the cold and warm seasons. The best performing product is listed in bold for each metric, and season. The 95% confidence interval is also included. Multiple bold values may appear if the confidence interval of multiple products overlap for a particular metric and season.

Metric	Product	Annual	Cold Season	Warm Season
Bias	CFSR	-2.63 ± 0.13° C	-2.37 ± 0.19° C	-2.8 ± 0.19° C
	ERA-Interim	-2.83 ± 0.21° C	-4.92 ± 0.34° C	-0.69 ± 0.34° C
	ERA5	-0.73 ± 0.14° C	-0.91 ± 0.24° C	-0.58 ± 0.24° C
	ERA5-Land	0.29 ± 0.14° C	0.89 ± 0.22° C	-0.35 ± 0.22° C
	FLDAS	-2.62 ± 0.14° C	-3.42 ± 0.26° C	-2.18 ± 0.26° C
	JRA55	-2.15 ± 0.17° C	4.16 ± 0.31° C	-7.04 ± 0.31° C
	MERRA2	-1.55 ± 0.12° C	-1.88 ± 0.21° C	-1.33 ± 0.21° C
	Ensemble Mean	-1.68 ± 0.12° C	-2.1 ± 0.20° C	-1.32 ± 0.20° C
RMSE	CFSR	3.89 ± 0.14° C	3.74 ± 0.15° C	3.65 ± 0.15° C
	ERA-Interim	4.55 ± 0.26° C	5.88 ± 0.33° C	2.17 ± 0.33° C
	ERA5	2.75 ± 0.15° C	3.14 ± 0.18° C	2.15 ± 0.18° C
	ERA5-Land	2.51 ± 0.17° C	2.67 ± 0.21° C	2.13 ± 0.21° C
	FLDAS	4.01 ± 0.13° C	4.60 ± 0.17° C	3.12 ± 0.17° C
	JRA55	7.68 ± 0.17° C	5.87 ± 0.21° C	8.95 ± 0.21° C
	MERRA2	2.90 ± 0.13° C	3.30 ± 0.15° C	2.33 ± 0.15° C
	Ensemble Mean	2.85 ± 0.13° C	3.17 ± 0.14° C	2.31 ± 0.14° C
Pearson Correlation	CFSR	0.959 ± 0.003	0.655 ± 0.018	0.931 ± 0.018
	ERA-Interim	0.971 ± 0.002	0.731 ± 0.014	0.905 ± 0.014
	ERA5	0.970 ± 0.003	0.564 ± 0.021	0.950 ± 0.021
	ERA5-Land	0.970 ± 0.003	0.644 ± 0.015	0.954 ± 0.015
	FLDAS	0.970 ± 0.002	0.734 ± 0.014	0.939 ± 0.014
	JRA55	0.552 ± 0.008	0.570 ± 0.015	0.402 ± 0.015
	MERRA2	0.978 ± 0.002	0.674 ± 0.015	0.954 ± 0.015
	Ensemble Mean	0.982 ± 0.002	0.793 ± 0.011	0.957 ± 0.011
Skill Score	CFSR	0.946 ± 0.007	0.689 ± 0.019	0.902 ± 0.019
	ERA-Interim	0.896 ± 0.012	0.62 ± 0.022	0.955 ± 0.022
	ERA5	0.951 ± 0.008	0.685 ± 0.017	0.955 ± 0.017
	ERA5-Land	0.949 ± 0.009	0.601 ± 0.022	0.95 ± 0.022
	FLDAS	0.935 ± 0.008	0.691 ± 0.017	0.938 ± 0.017
	JRA55	0.536 ± 0.012	0.564 ± 0.017	0.647 ± 0.017
	MERRA2	0.954 ± 0.008	0.73 ± 0.016	0.960 ± 0.016
	Ensemble Mean	0.958 ± 0.007	0.805 ± 0.014	0.961 ± 0.014

Table S2. Summary of the near surface mean bias, RMSE, and skill score for each product for the Extratropical Northern Hemisphere. Metrics are separated into an annual mean metric, and a metric for the cold and warm seasons. The best performing product is listed in bold for each metric, and season. The 95% confidence interval is also included. Multiple bold values may appear if the confidence interval of multiple products overlap for a particular metric and season.

Metric	Product	Annual	Cold Season	Warm Season
Bias	CFSR	-3.12 ± 0.16° C	-1.63 ± 0.19° C	-4.48 ± 0.19° C
	ERA-Interim	-2.71 ± 0.22° C	-2.26 ± 0.29° C	-2.99 ± 0.29° C
	ERA5	-0.54 ± 0.16° C	0.27 ± 0.20° C	-1.34 ± 0.20° C
	ERA5-Land	0.10 ± 0.17° C	1.33 ± 0.21° C	-1.10 ± 0.21° C
	FLDAS	-3.09 ± 0.19° C	-2.72 ± 0.26° C	-3.71 ± 0.26° C
	JRA55	-1.82 ± 0.19° C	0.49 ± 0.27° C	-3.67 ± 0.27° C
	MERRA2	-2.08 ± 0.15° C	-0.70 ± 0.19° C	-3.30 ± 0.19° C
	Ensemble Mean	-1.91 ± 0.15° C	-0.95 ± 0.18° C	-2.82 ± 0.18° C
RMSE	CFSR	4.54 ± 0.16° C	2.98 ± 0.16° C	5.28 ± 0.16° C
	ERA-Interim	4.05 ± 0.22° C	3.42 ± 0.27° C	4.13 ± 0.27° C
	ERA5	2.73 ± 0.15° C	2.10 ± 0.17° C	2.98 ± 0.17° C
	ERA5-Land	2.76 ± 0.17° C	2.28 ± 0.20° C	2.82 ± 0.20° C
	FLDAS	4.68 ± 0.15° C	3.99 ± 0.18° C	4.83 ± 0.18° C
	JRA55	4.24 ± 0.16° C	3.26 ± 0.14° C	4.86 ± 0.14° C
	MERRA2	3.44 ± 0.16° C	2.17 ± 0.15° C	4.00 ± 0.15° C
	Ensemble Mean	3.19 ± 0.14° C	2.20 ± 0.13° C	3.69 ± 0.13° C
Pearson Correlation	CFSR	0.817 ± 0.012	0.77 ± 0.018	0.749 ± 0.018
	ERA-Interim	0.909 ± 0.007	0.857 ± 0.012	0.847 ± 0.012
	ERA5	0.906 ± 0.01	0.849 ± 0.014	0.871 ± 0.014
	ERA5-Land	0.901 ± 0.01	0.86 ± 0.011	0.881 ± 0.011
	FLDAS	0.857 ± 0.013	0.84 ± 0.014	0.833 ± 0.014
	JRA55	0.781 ± 0.009	0.865 ± 0.011	0.753 ± 0.011
	MERRA2	0.91 ± 0.008	0.856 ± 0.013	0.869 ± 0.013
	Ensemble Mean	0.922 ± 0.007	0.892 ± 0.011	0.871 ± 0.011
Skill Score	CFSR	0.792 ± 0.019	0.768 ± 0.019	0.768 ± 0.019
	ERA-Interim	0.852 ± 0.019	0.751 ± 0.019	0.794 ± 0.019
	ERA5	0.872 ± 0.017	0.825 ± 0.018	0.843 ± 0.018
	ERA5-Land	0.839 ± 0.02	0.781 ± 0.022	0.864 ± 0.022
	FLDAS	0.759 ± 0.022	0.697 ± 0.02	0.782 ± 0.02
	JRA55	0.794 ± 0.016	0.685 ± 0.018	0.762 ± 0.018
	MERRA2	0.81 ± 0.021	0.814 ± 0.02	0.803 ± 0.02
	Ensemble Mean	0.866 ± 0.017	0.853 ± 0.015	0.841 ± 0.015

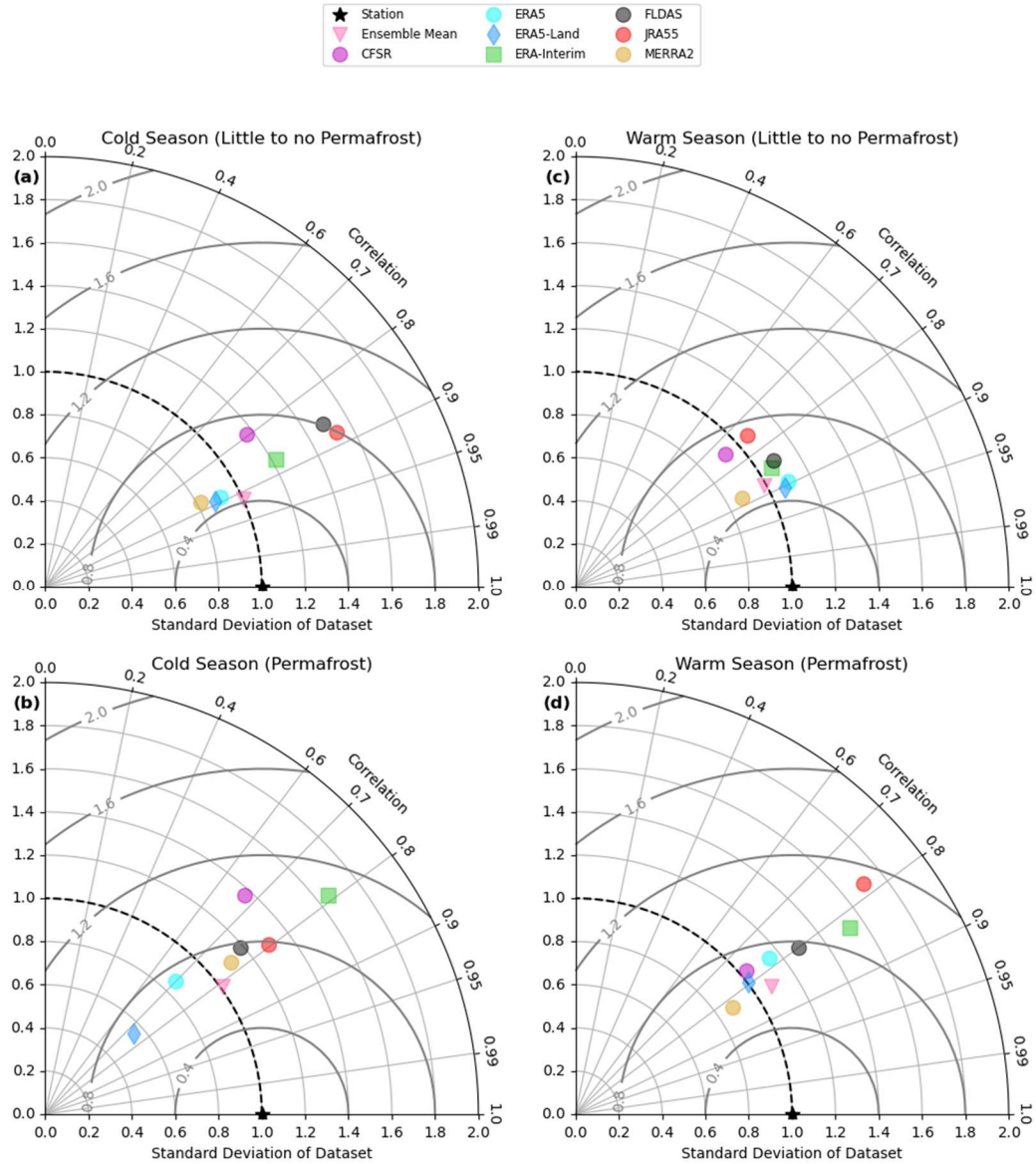


Figure S 1. Taylor Diagram of the cold season and the warm season performance of reanalysis products at depth. (a) and (c) refer to the zone with little to no permafrost, while (b) and (d) refer to the permafrost zone. The left panels, (a) and (b) are for the cold season while the right panels (c) and (d) refer to the warm season. The concentric rings (solid grey lines) refer to the centralized root mean square error (CRMSE), and a product would have a CRMSE of zero if the timeseries of the reanalysis matched the station data perfectly; with a normalized standard deviation of one, and a correlation of one.

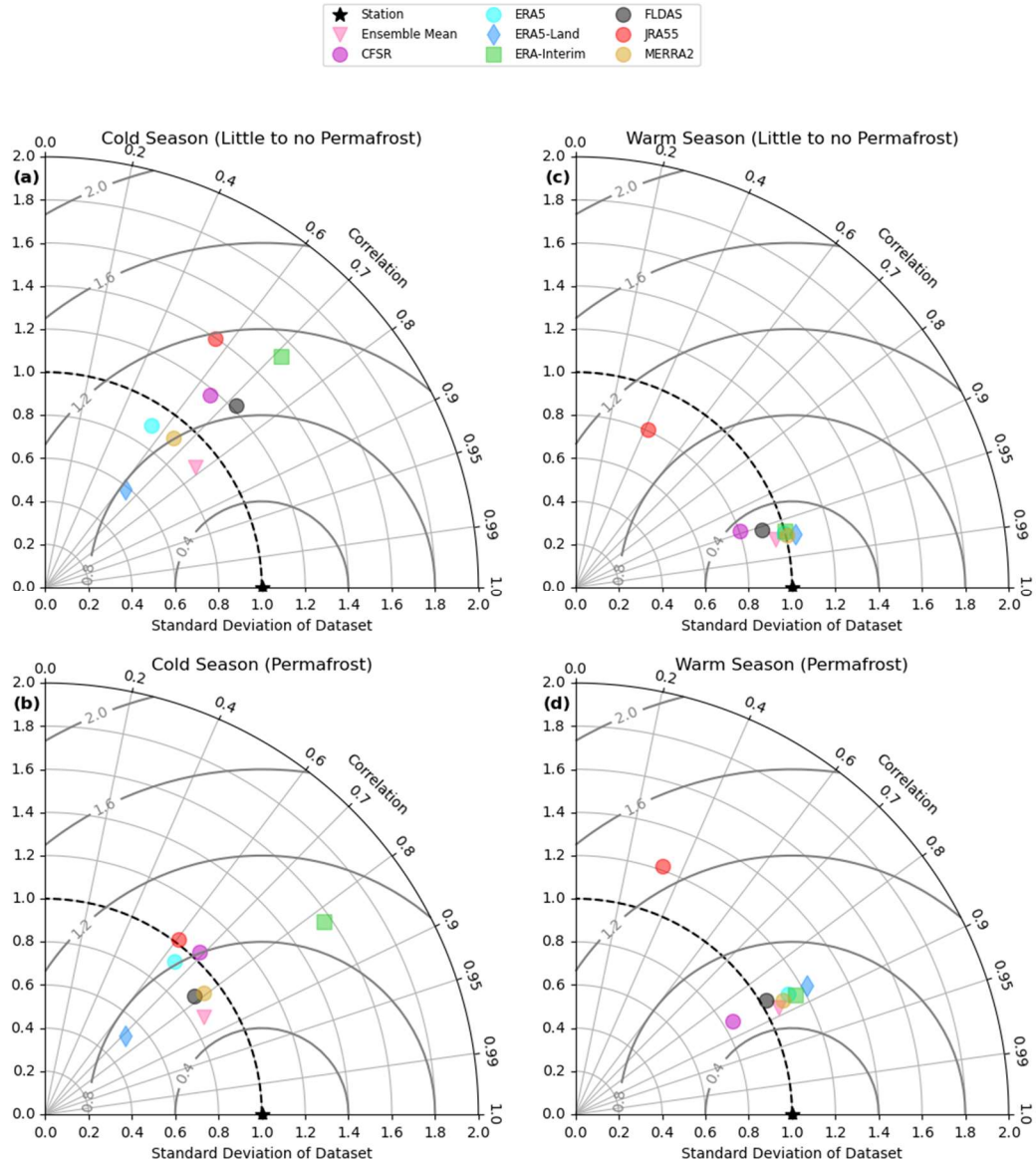


Figure S 2. Taylor Diagram of the near surface cold season and the warm season performance of reanalysis products. (a) and (c) refer to the zone with little to no permafrost, while (b) and (d) refer to the permafrost zone. The left panels, (a) and (b) are for the cold season while the right panels, (c) and (d) refer to the warm season. The concentric rings (solid grey lines) refer to the centralized root mean square error (CRMSE), and a product would have a CRMSE of zero if the timeseries of the reanalysis matched the station data perfectly; with a normalized standard deviation of one, and a correlation of one.

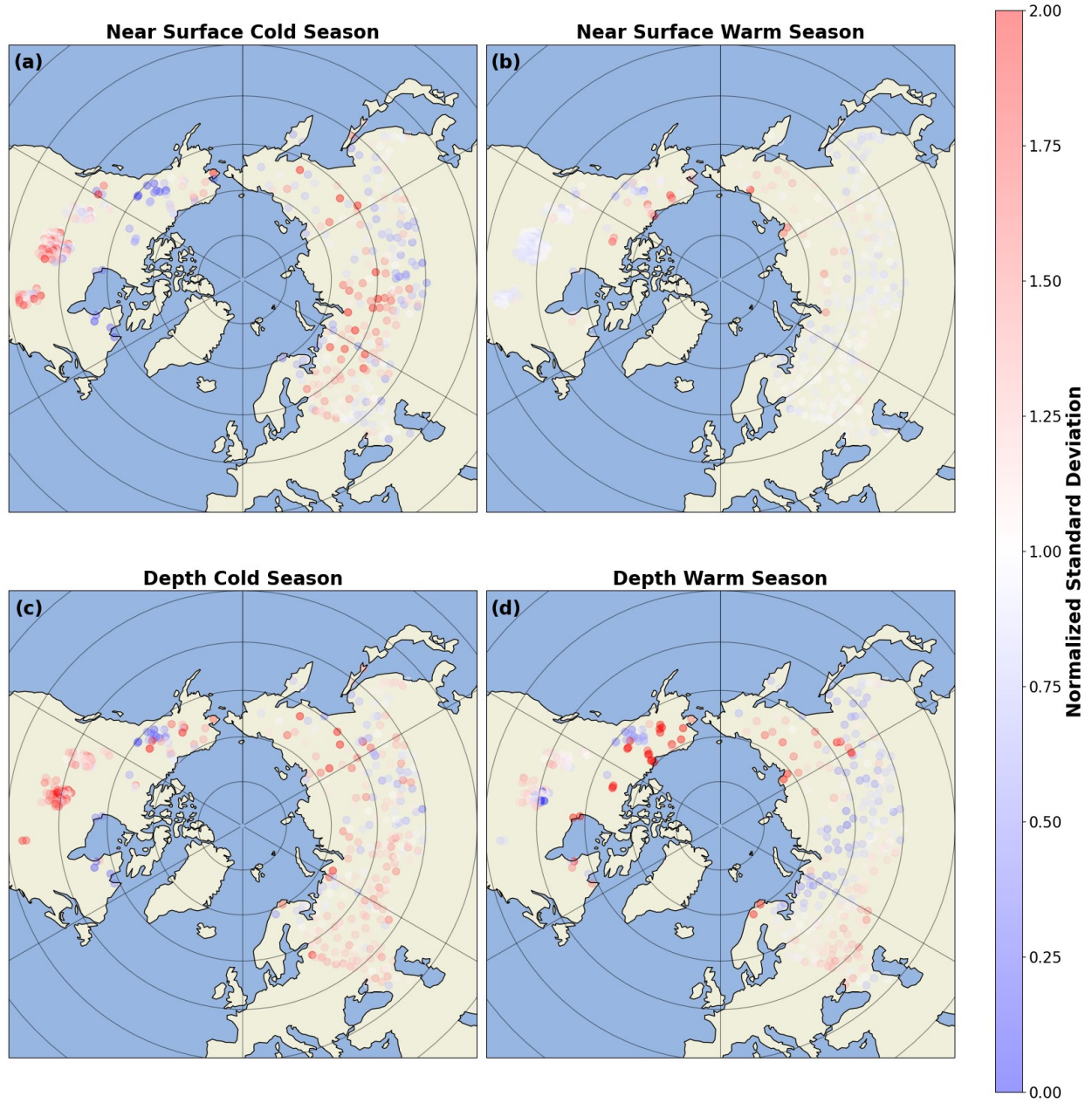


Figure S 3. The normalized soil temperature standard deviation (as a measure of variability), relative to in situ values, in the ensemble mean. The top panels, (a) and (b) refer to the near surface, while the bottom panels, (c) and (d) refer to biases at depth. Panels on the left (A and C) are for the cold season, while the right panels (B and D) are for the warm season.

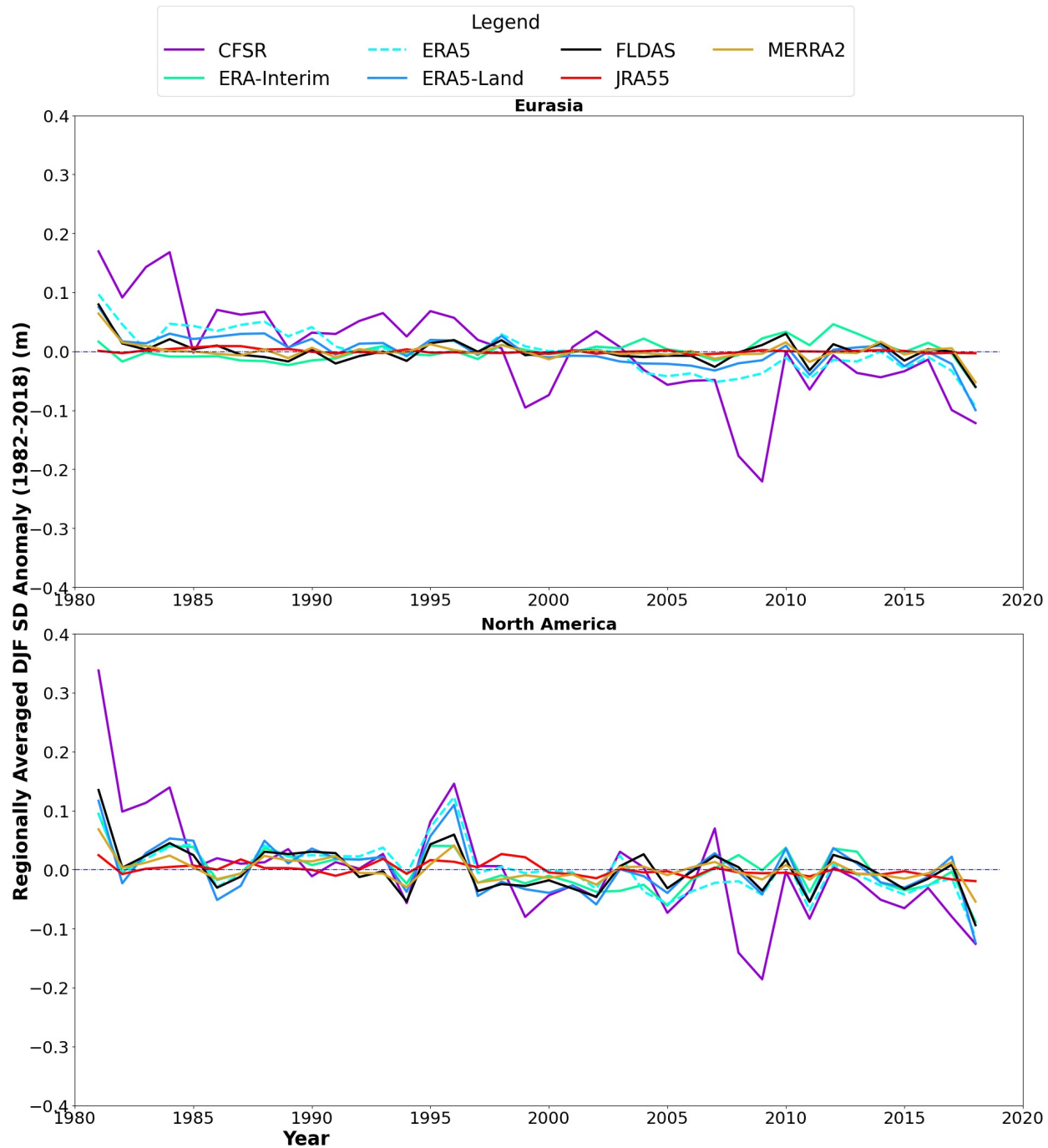


Figure S 4. DJF Snow depth anomalies for each of the reanalysis products. (a) displays the regionally averaged 1982-2018 DJF snow depth anomalies for each reanalysis product north of 40°N over Eurasia, while (b) displays the same, but over North America.

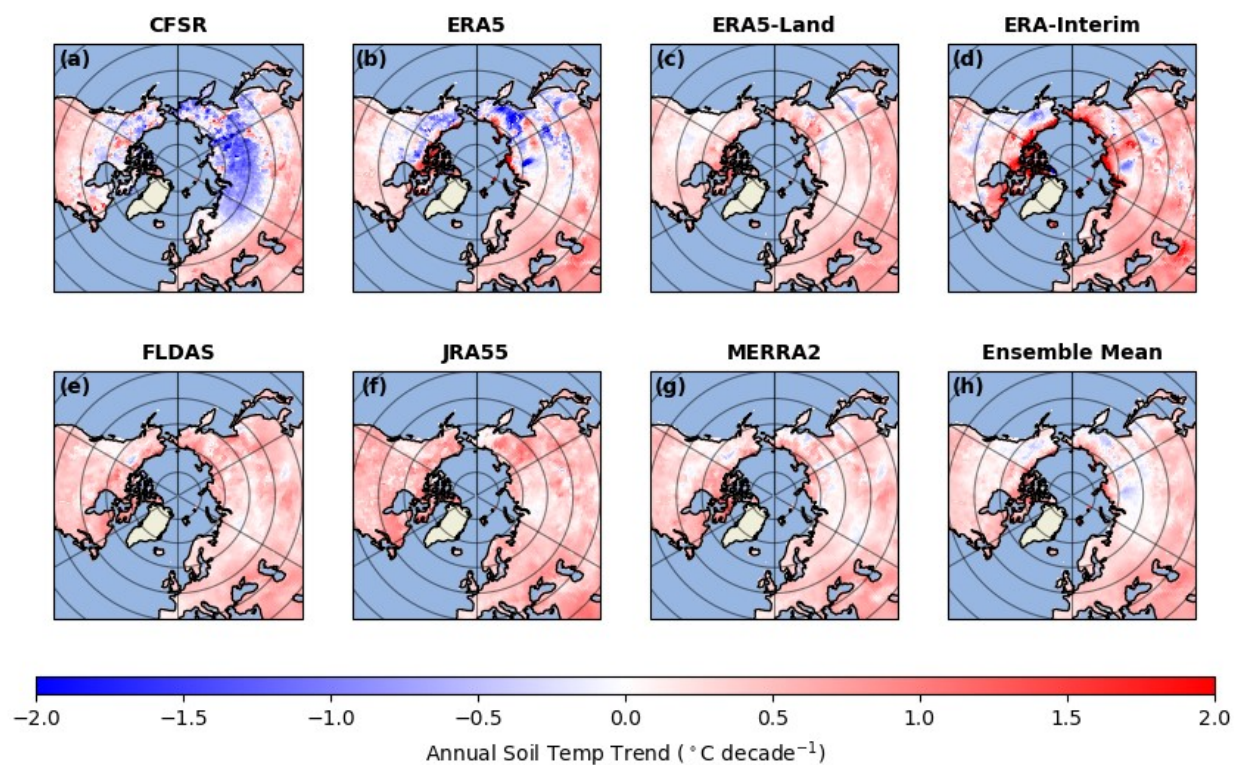


Figure S 5. Map of the near surface 1985-2010 annual mean soil temperature trends for each of the products and the ensemble mean.

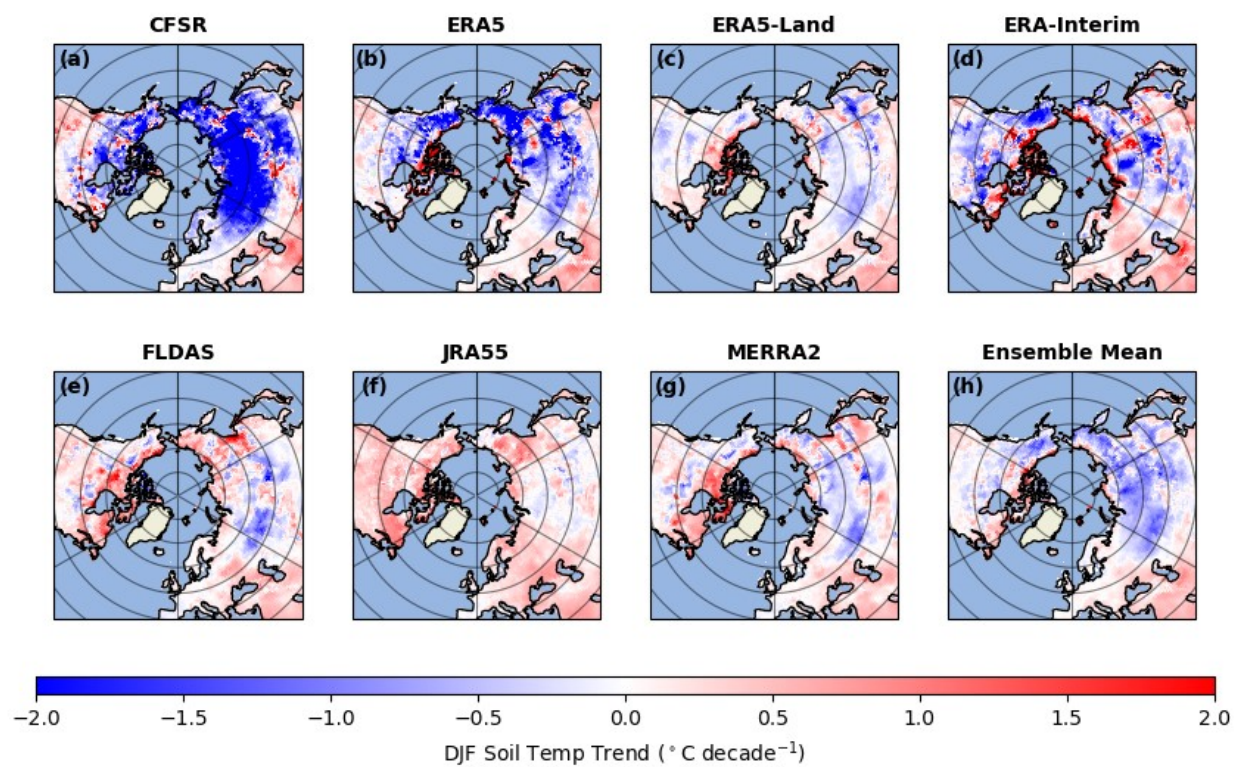


Figure S 6. Map of the near surface 1985-2010 DJF soil temperature trends for each of the products and the ensemble mean.

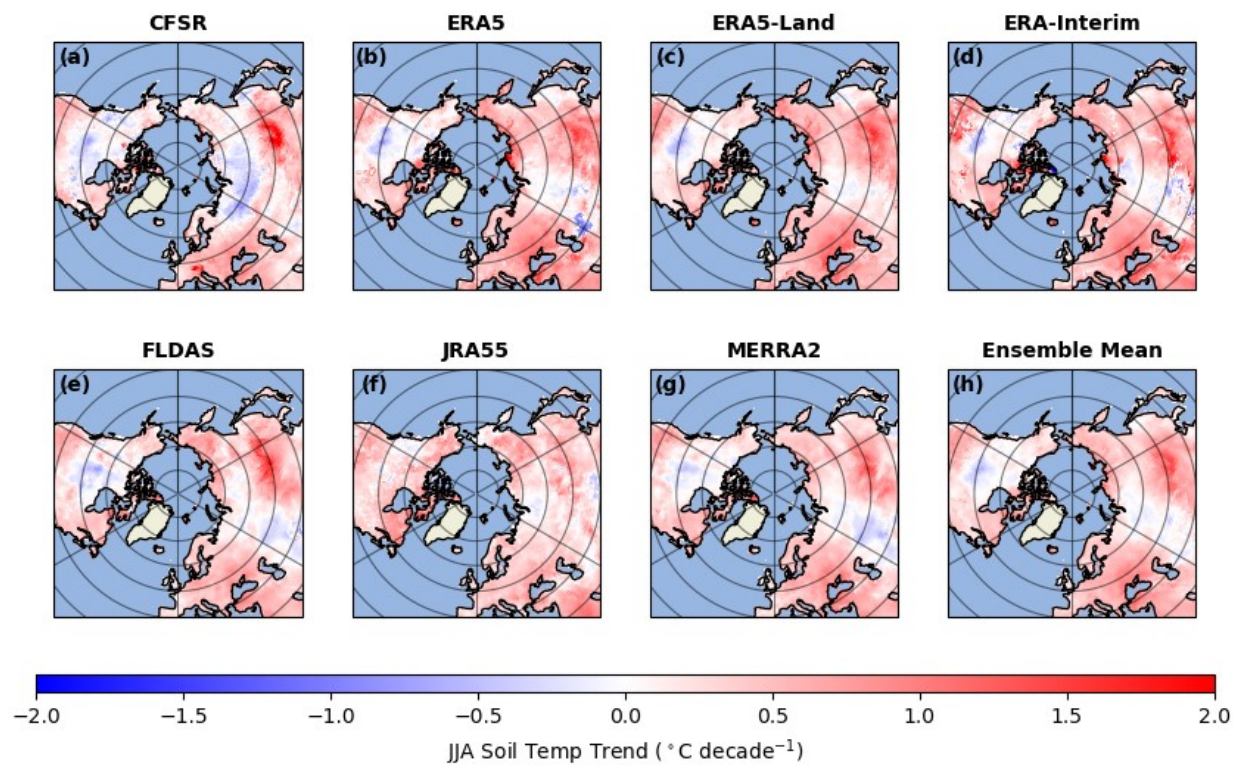


Figure S 7. Map of the near surface 1985-2010 JJA soil temperature trends for each of the products and the ensemble mean.

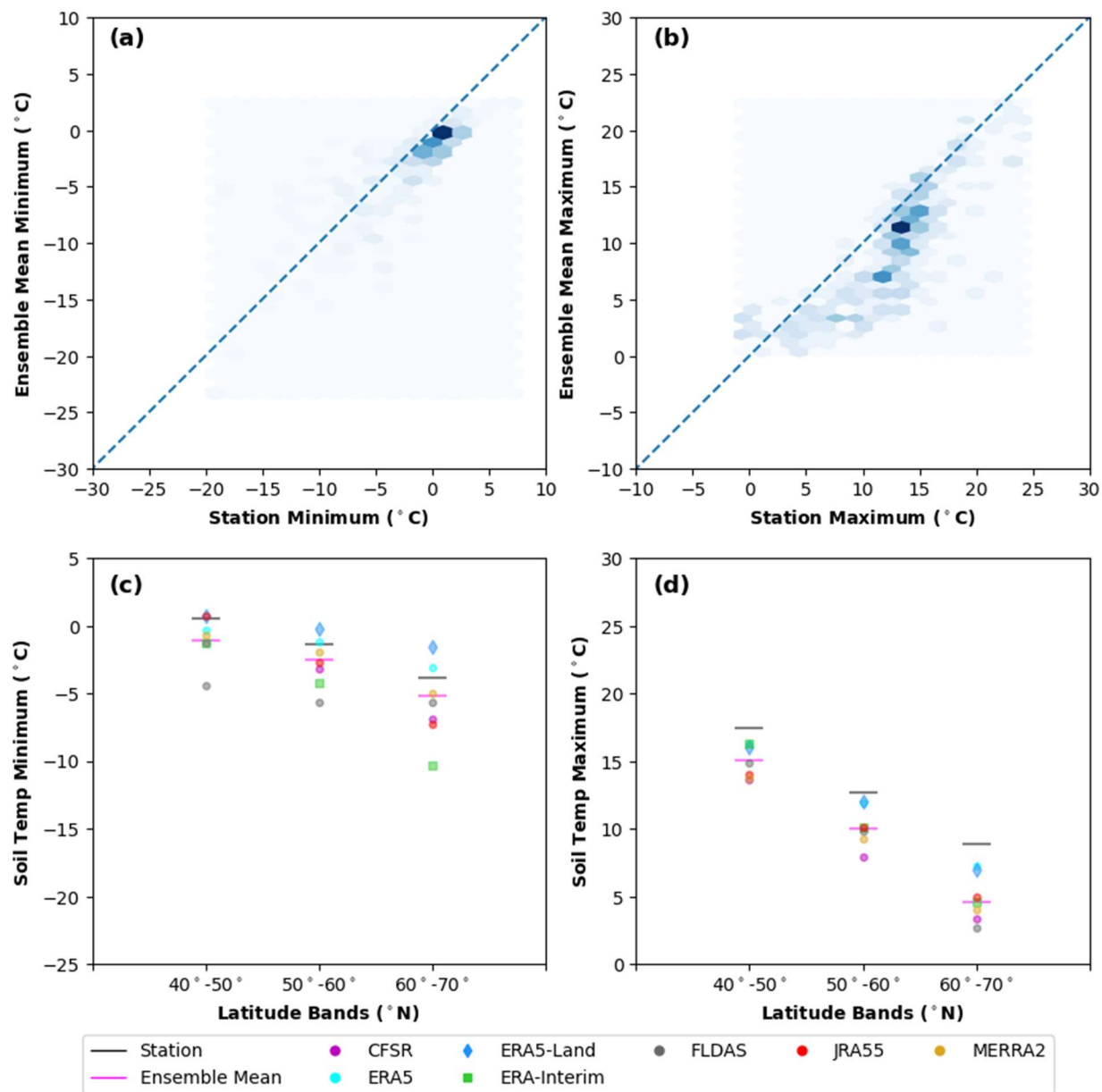


Figure S 8. Performance of the soil temperature variability, at depth, in the Ensemble Mean. (a): Scatterplot of the station and ensemble mean winter minimum soil temperature. (b): Scatterplot of the station and ensemble mean summer maximum soil temperature. (c): latitudinal averages (from Eurasian grid cells) of near surface soil temperature winter minimum for the ensemble mean and contributing products. (d): latitudinal averages (from Eurasian grid cells) of near surface soil temperature summer maximum for the ensemble mean and contributing products.