Supplementary information for "The Arctic has warmed nearly four times faster than the globe since 1979"

Mika Rantanen[1,*], Alexey Yu. Karpechko[1], Antti Lipponen[2], Kalle Nordling[1,3], Otto Hyvärinen[1], Kimmo Ruosteenoja[1], Timo Vihma[1], and Ari Laaksonen[1,4]

[1]Finnish Meteorological Institute, Helsinki, Finland

[2]Finnish Meteorological Institute, Kuopio, Finland

[3]CICERO Center for International Climate Research, Oslo, Norway

[4]Department of Applied Physics, University of Eastern Finland, Kuopio, Finland

[*]Correspondence to: mika.rantanen@fmi.fi

1 Supplementary Notes

Table S1: Observational datasets used in this study.

Dataset	Abbreviation	Resolution	Period	Reference
ERA5 reanalysis	ERA5	0.25°	1950-2021	[1]
NASA GISTEMP v4	GISTEMP	2°	1950-2021	[2]
Berkeley Earth	BEST	1°	1950-2021	[3]
HadCRUT.5.0.1.0.	HadCRUT5	5°	1950-2021	[4]

The three in-situ datasets (GISTEMP, BEST and HadCRUT5) are merged land-ocean indices, i.e. they use station data on land areas and sea surface temperatures (SSTs) above the oceans. GISTEMP relies largely on Global Historical Climatology Network Monthly version 4[5] on land areas. BEST has its own land surface record[6,7], and HadCRUT5 is based on the CRUTEM5 database[8]. For SSTs, GISTEMP uses Extended Reconstruction Sea Surface Temperature version 5 of NOAA[9]. BEST and HadCRUT5 use the HadSST4 dataset[10].

Table S2: Climate model ensembles used in this study. Note that CanESM5 is excluded from CMIP6 ensemble as CanESM5 was used as a separate large ensemble dataset.

Dataset	Forcing	Number of models	Ensemble size	Period	Reference
CMIP5	RCP4.5	36	36	1950-2040	Taylor et al.[11]
CMIP6	SSP2-4.5.	41	194	1950-2040	Eyring et al.[12]
MPI-GE	RCP4.5	1	100	1950-2040	Maher et al.[13]
CanESM5	SSP2-4.5.	1	50	1950-2040	Swart et al.[14]

Table S3: CMIP5 models used in this study. Only one realization per model was used.

1. ACCESS1-0	19. GFDL-ESM2M
2. ACCESS1-3	20. GISS-E2-H
3. BCC-CSM1-1	21. GISS-E2-R
4. BNU-ESM	22. HadGEM2-AO
5. CanESM2	23. HadGEM2-CC
6. CCSM4	24. HadGEM2-ES
7. CESM1-BGC	25. INMCN4
8. CESM1-CAM5	26. IPSL-CM5A-LR
9. CMCC-CM	27. IPSL-CM5A-MR
10.CMCC-CMS	28. IPSL-CM5B-LR
11. CNRM-CM5	29. MIROC-ESM
12. CSIRO-Mk3-6-0	30. MIROC-ESM-CHEM
13. EC-Earth	31. MIROC5
14. FGOALS-g2	32. MPI-ESM-LR
15. FGOALS-s2	33. MPI-ESM-MR
16. FIO-ESM	34. MRI-CGCM3
17. GFDL-CM3	35. NorESM1-M
18. GFDL-ESM2G	36. NorESM1-ME

Table S4: CMIP6 models and the number of realizations per model used in this study.

Model name	Members	Model name	Members
1. ACCESS-CM2	3	22. GFDL-CM4	1
2. ACCESS-ESM1-5	10	23. GFDL-ESM4	3
3. AWI-CM-1-1-MR	1	24. GISS-E2-1-G	19
4. BCC-CSM2-MR	1	25. HadGEM3-GC31-LL	1
5. CAMS-CSM1-0	2	26. IITM-ESM	1
6. CAS-ESM2-0	2	27. INM-CM4-8	1
7. CESM2	3	28. INM-CM5-0	1
8. CESM2-WACCM	3	29. IPSL-CM6A-LR	11
9. CIESM	1	30. KACE-1-0-G	3
10. CMCC-CM2-SR5	1	31. KIOST-ESM	1
11. CMCC-ESM2	1	32. MIROC6	3
12. CNRM-CM6-1	6	33. MIROC-ES2L	30
13. CNRM-CM6-1-HR	1	34. MPI-ESM1-2-HR	2
14. CNRM-ESM2-1	9	35. MPI-ESM1-2-LR	10
15. EC-Earth3	21	36. MRI-ESM2-0	1
16. EC-Earth3-CC	1	37. NESM3	2
17. EC-Earth3-Veg	7	38. NorESM2-LM	3
18. EC-Earth3-Veg-LR	3	39. NorESM2-MM	2
19. FGOALS-f3-L	1	40. TaiESM1	1
20. FGOALS-g3	4	41. UKESM1-0-LL	14
21. FIO-ESM-2-0	3		

2 Supplementary Methods

For observations, we define the uncertainty ranges in dT/dt_A and dT/dt_G as the 90 % statistical uncertainty range. For the CMIP5 and CMIP6 simulations, the uncertainty ranges reflect the 90 % confidence interval (5th-95th percentile range) derived from multi-model ensemble. For the MPI-GE and CanESM5 simulations, we use the 90 % confidence interval (5th-95th percentile range) derived from the ensemble members. The statistical significance of the observed trends was estimated at the 95 % level (P=0.05) using the nonparametric Mann-Kendall test[15,16] with pyMannKendall python package[17].

The uncertainty range in observed AA_{43} ratios reflects 90 % confidence intervals which

were estimated using the bootstrap method[18]. For the CMIP5 and CMIP6 simulations, the corresponding AA_{43} uncertainty range is the 90 % confidence interval derived from the multi-model ensemble, and for MPI-GE and CanESM5, derived from the ensemble members. The AA_{43} ratios in the climate models were first calculated separately for each realization and then averaged across them (i.e. as the mean of ratios, not a ratio of means), thus following the recommendation by Smith et al[19].

Following the methodology in earlier studies[20,21], we test whether the observed and simulated AA_{43} ratios are equal. By assuming exchangeability between models, the representation of the AA_{43} ratio is

$$M_{ij} = u_m + Eint_{ij} + Emod_i, i = 1, ..., N^m, j = 1, ..., N^i \tag{1}$$

and

$$O = u_o + Eint_o \tag{2}$$

where M_{ij} and O are AA_{43} ratios calculated from single model runs or the observations, respectively. Similarly, u_m and u_o are the true, unknown, deterministic AA_{43} ratios due to external forcing in the model runs and observations. u_m is common for all models and represents essentially the multi-model mean AA_{43}. $Eint_{ij}$ and $Eint_o$ are perturbations to M_{ij} and O, respectively, caused by the internal climate variability, and for the models, $Eint_{ij}$ is different for each realization. $Emod_i$ is the perturbation to Mij that is introduced by the model error in model i. Similarly as in[21], we assume that these perturbations are exchangeable. N^m stands for the number of models, and N_i is the number of realizations per model i.

The null hypothesis H_0, formulated as the observed and simulated AA ratios being equal,

$$H_0 : u_m = u_o \tag{3}$$

is tested by constructing distributions that includes the both sources of uncertainty in Eq. (1)

and (2). Again, we follow[21] and construct the empirical distribution as follows:

(a) Calculate the observed AA_{43} O over the 43-year period of 1979-2021

(b) Select a sample of N^m models with replacement, and for each model in that sample, select one realization at random from that model's available ensemble of realizations, and then average over those N^m realizations to obtain a version of $M_{..}$

(c) Select, at random, a single model i from those models with multiple realizations, and then select, at random, a single run j from that model's ensemble. Calculate the difference $M_{ij} - M_{i.}$ between the AA_{43} in that single run and the mean of the AA_{43} ratios from that model's ensemble. This difference represents an estimate of the deviation in the j-th trend for model i that is induced by the internal variability. Due to the small size of the model i ensemble, the deviations are smaller than would be representative of an infinitely large ensemble of runs for model i. Thus, to compensate for that loss of variance, multiply the difference $M_{ij} - M_{i.}$ by $[N_i/(N_i - 1)]^{0.5}$.

(d) Calculate $a - b + c$, as described above, and repeat many times to build a distribution for $a - b + c$.

As described in[20], the step (b) represents the variations in the multi-model mean estimate of u_m that arises from the choice of exchangeable models used to compute $M_{..}$ (term $Emod_i$ in Eq. 1) as well as the internal variability (term $Eint_{ij}$ in Eq. 1).

Step (c) demonstrates the uncertainty in O arising from internal variability. Because the observations (O) represents only a single realization of internal variability, step (c) is constructed by estimating single realizations of internal variability as they were realized in models. Because the interval variability is superimposed on the externally forced trends, step (c) is naturally added to step (b).

In our study, 10^6 repetitions were used to construct the empirical distributions for $a - b + c$, from which we calculated the p-values for the test of the null hypothesis H_0. Smaller proportion

of negative $a - b + c$ values (simulated trend is at least as large as observed) in the distribution implies smaller p-values and consequently stronger evidence against the null hypothesis.

The procedure above was performed for the CMIP6 ensemble. For CMIP5, MPI-GE, and CanESM5 we performed the procedure otherwise similarly, except that in point (b) there is only one set of model runs from which to resample. Thus (b) is a random sample with replacement of N_i CMIP5 models or model realizations (in MPI-GE and CanESM5) and c) is as above but always selecting, at random, a single run j from the ensembles of CMIP5, MPI-GE, and CanESM5. See more about the test Fyfe et al.[20] and Swart et al.[21].

3 Supplementary figures

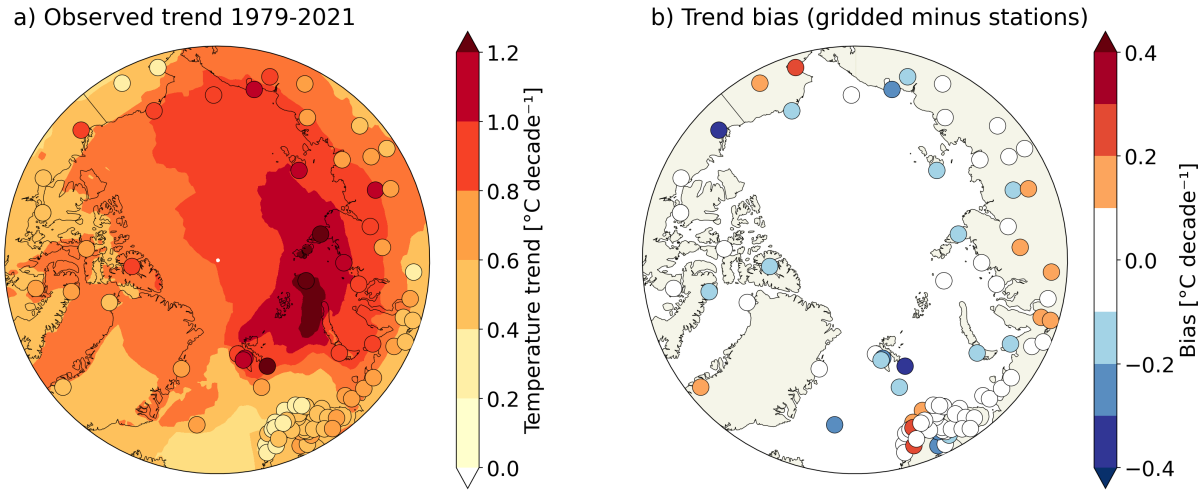

Figure S1: The annual warming trend in the Arctic over the 1979-2021 period (a), and the bias of the trend (b). The shading in a) depicts the observational average across the four datasets (GISTEMP, HadCRUT5, BEST, ERA5), and the circles indicate the station observations. The bias in b) is defined as the gridded data minus station observations.

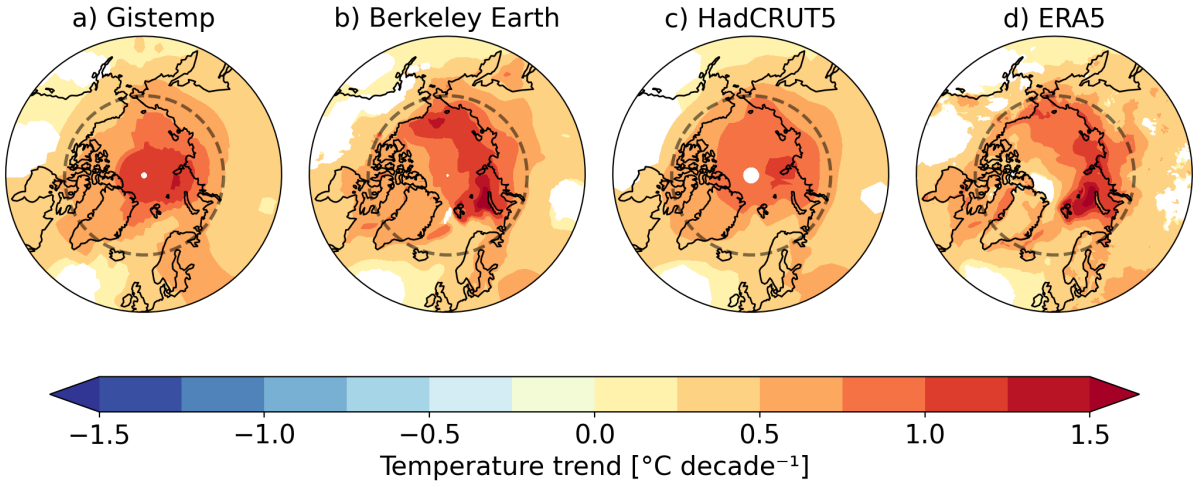

Figure S2: Annual mean temperature trends for the period 1979-2021 derived from a) NASA GISTEMP v4, b) Berkeley Earth, c) HadCRUT5, and d) ERA5 reanalysis. Dashed line indicates the Arctic Circle (66.5°N latitude). Areas without a statistically significant change (determined using a two-sided Wald t-test, with $P < 0.05$ indicating a significant difference) are masked out.

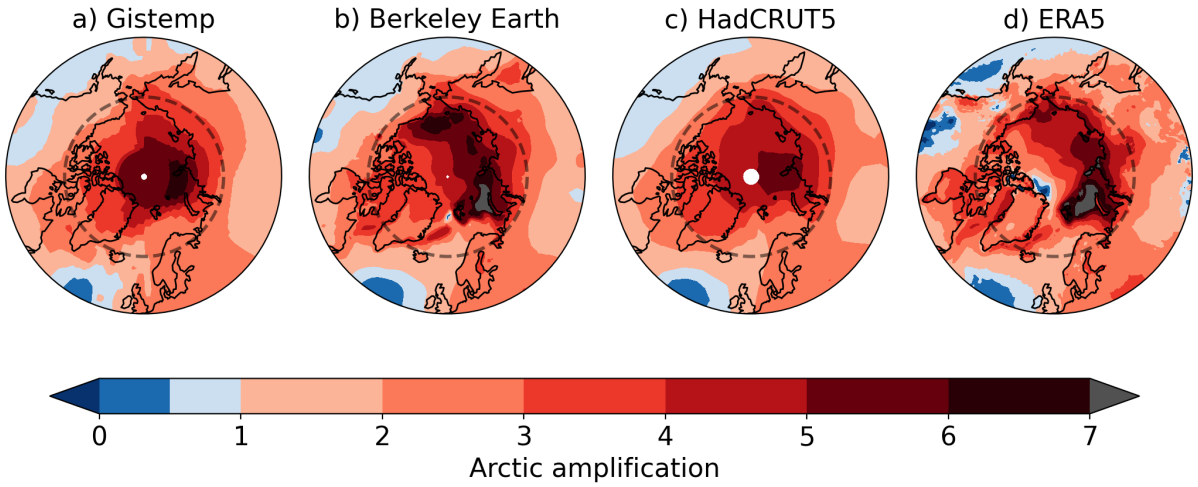

Figure S3: Arctic amplification for the period 1979-2021 derived from a) NASA GISTEMP v4, b) Berkeley Earth, c) HadCRUT5, and d) ERA5 reanalysis. Dashed line indicates the Arctic Circle (66.5°N latitude).

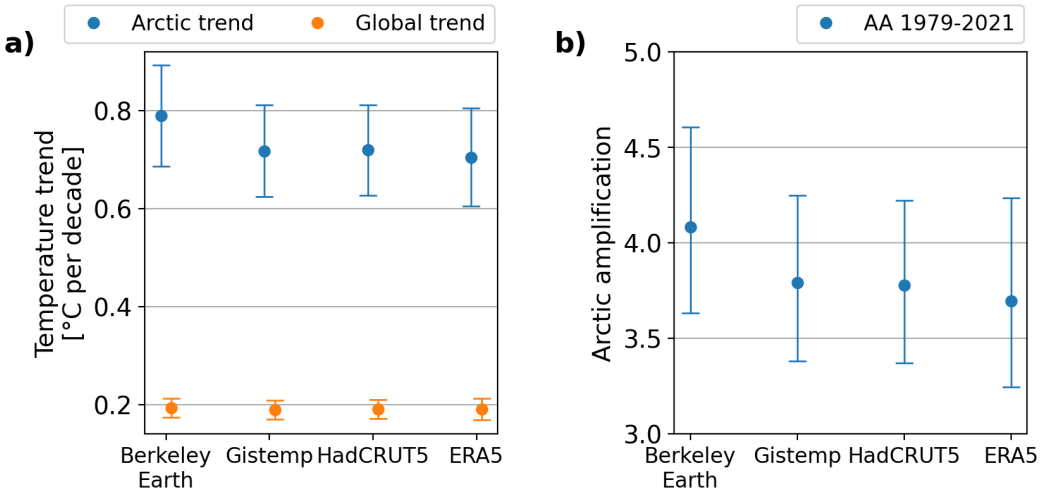

Figure S4: a) Observed annual-mean temperature trends in 1979-2021 in the Arctic (blue) and globally (orange). b) Observed 43-year Arctic amplification (AA_{43}) ratio calculated for the 1979-2021 period (blue). The dots represent the central estimate (mean), and the error bars show the 90 % confidence intervals. See Supplementary Methods for more explanation for the calculation of the uncertainty estimates.

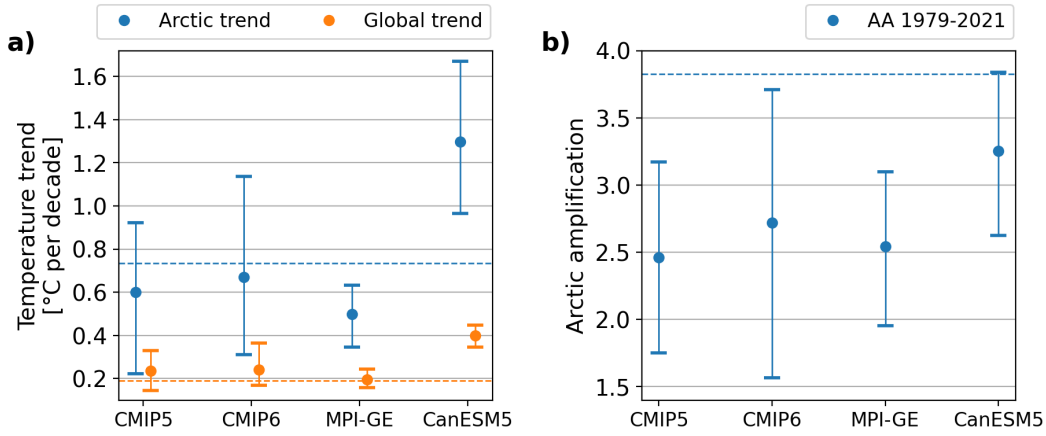

Figure S5: a) Simulated annual temperature trends in 1979-2021 in the Arctic (blue) and globally (orange). b) Simulated 43-year Arctic amplification (AA_{43}) ratio calculated for the 1979-2021 period. The horizontal dashed lines in a) show the observed Arctic (blue) and global (orange) trends, and in b) the observed AA_{43}. The dots represent the central estimate (mean), and the error bars show the 90 % confidence intervals. See Supplementary Methods for more explanation for the calculation of the uncertainty estimates. Note that CanESM5 model is excluded from CMIP6 ensemble.

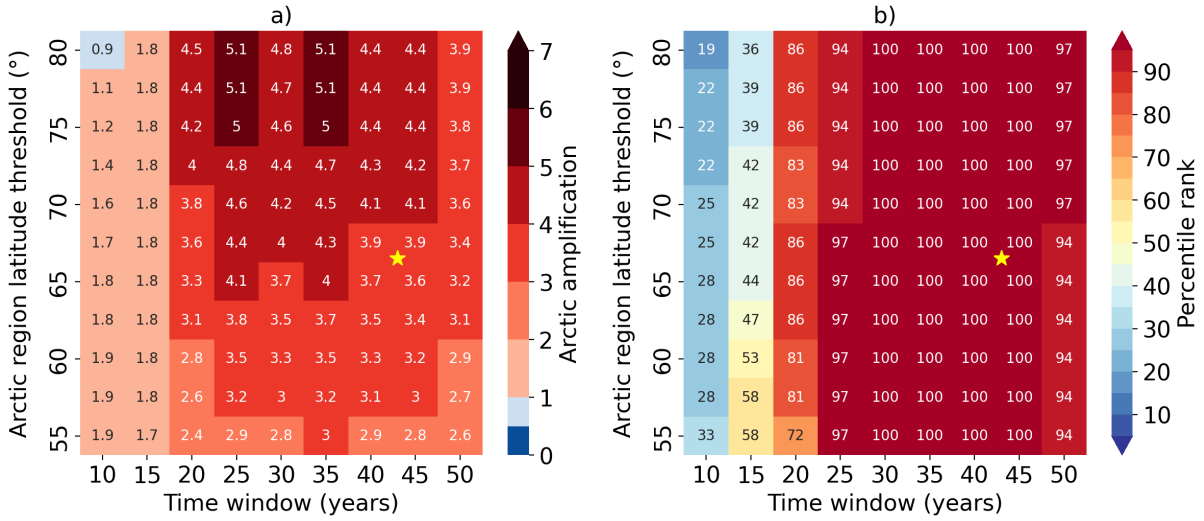

Figure S6: The sensitivity of AA (a) to the time window used in calculating the linear trends (x-axis) and the southern boundary of the Arctic (y-axis), and (b) the percentile rank of observed AA in the CMIP5 ensemble distribution. The end year of all linear trends is fixed to 2021. Thus, for example, 50 years on the x-axis corresponds to the trend calculated for 1972–2021. The star marks the baseline value used in the study, corresponding to the 43-year linear trend and the southern boundary of 66.5°N. The observed AA is derived from the average of the four observational datasets. Note that panel a) is identical with Fig. 2a in the main text.

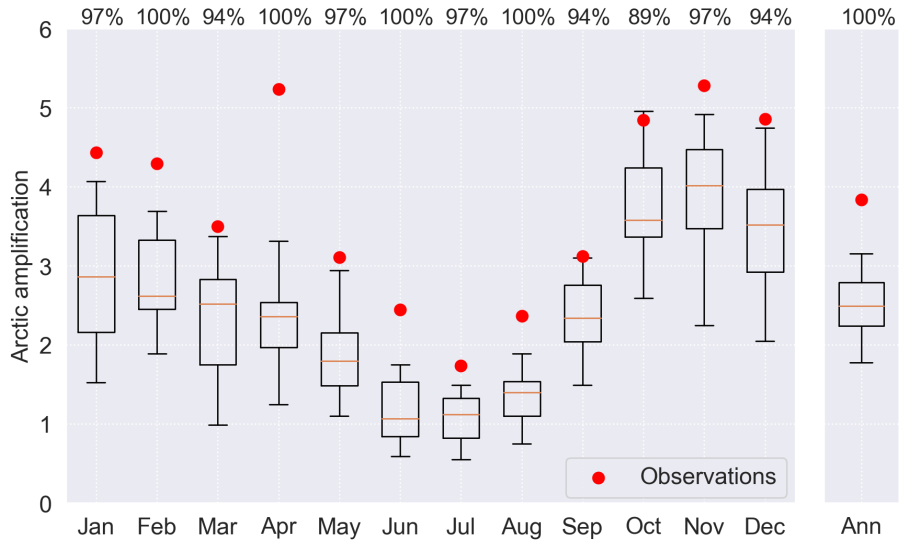

Figure S7: Seasonality of the 43-year (1979-2021) Arctic amplification ratio. The orange lines indicate the medians of CMIP5 realizations, boxes show the first and third quartiles, and whiskers extend to the 5-95th percentiles of the realizations. The red circles indicate the observed AA, as derived from the average of the four observational datasets. The numbers in the upper row give the percent rank of the observed AA in the CMIP5 ensemble distribution.

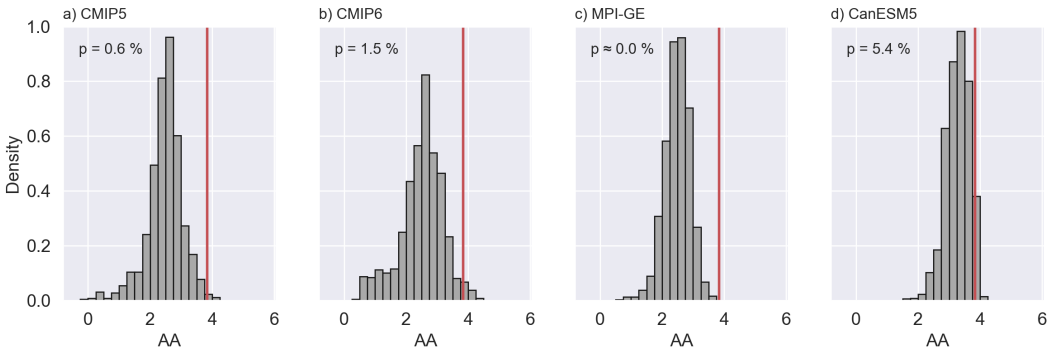

Figure S8: Frequency distributions of all possible 43-year AA ratios between 1970 and 2040 in a) CMIP5 ensemble, b) CMIP6 ensemble, c) MPI-GE ensemble and d) CanESM5 ensemble. The red line marks the observed AA ratio between 1979 and 2021. In this plot, only one realization per model in CMIP6 is used: we used the realization indexed as "r1i1p1" except for CESM2 for which "r4i1p1" was used. Note that panels a), c), and d) are identical with Fig. 3 in the main text.

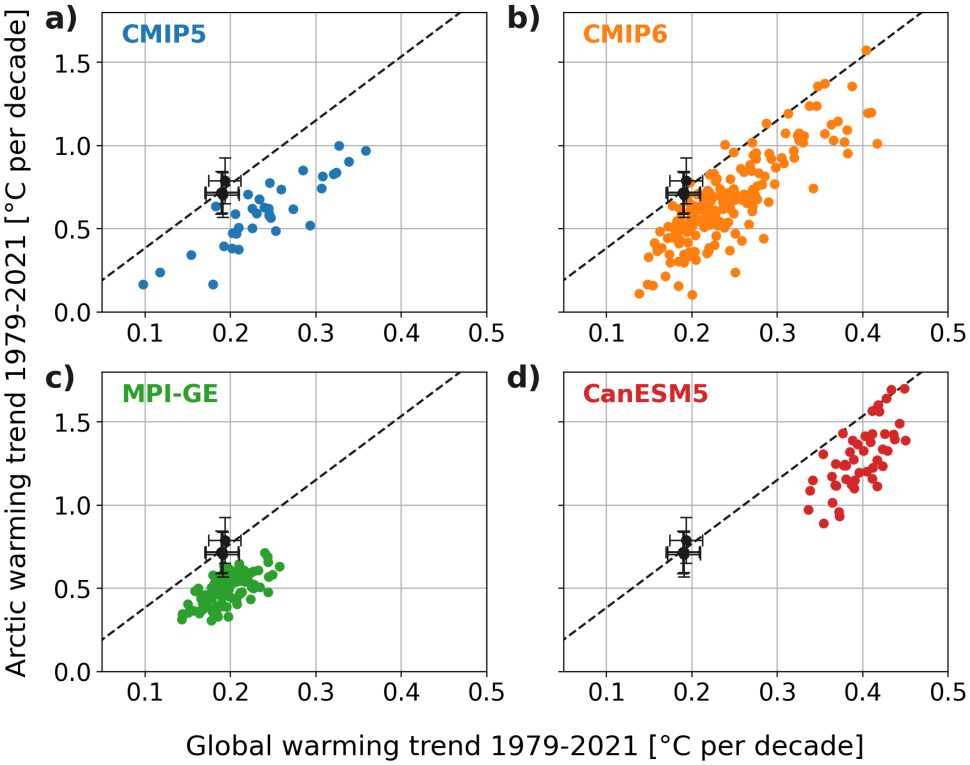

Figure S9: Arctic warming rate in 1979-2021 as a function of global warming rate in a) CMIP5, b) CMIP6, c) MPI-GE and d) CanESM5 ensembles. Black markers depict observations and their 90 % uncertainty ranges. The dashed line correspond to the observed AA as a function of global warming rate.

4 Supplementary References

[1] Hersbach, H. *et al.* The ERA5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society* (2020).

[2] Lenssen, N. J. *et al.* Improvements in the GISTEMP uncertainty model. *Journal of Geophysical Research: Atmospheres* **124**, 6307–6326 (2019).

[3] Rohde, R. A. & Hausfather, Z. The Berkeley Earth land/ocean temperature record. *Earth System Science Data* **12**, 3469–3479 (2020).

[4] Morice, C. P. *et al.* An updated assessment of near-surface temperature change from 1850: The HadCRUT5 data set. *Journal of Geophysical Research: Atmospheres* **126**, e2019JD032361 (2021).

[5] Menne, M. J., Williams, C. N., Gleason, B. E., Rennie, J. J. & Lawrimore, J. H. The global historical climatology network monthly temperature dataset, version 4. *Journal of Climate* **31**, 9835–9854 (2018).

[6] Rohde, R. *et al.* A New Estimate of the Average Earth Surface Land Temperature Spanning 1753 to 2011. *Geoinformatics & Geostatistics: An Overview* **7**, 2 (2013).

[7] Rohde, R. *et al.* Berkeley Earth Temperature Averaging process. *Geoinformatics & Geostatistics: An Overview* **1**, 20–100 (2013).

[8] Osborn, T. J. *et al.* Land surface air temperature variations across the globe updated to 2019: The CRUTEM5 data set. *Journal of Geophysical Research: Atmospheres* **126**, e2019JD032352 (2021).

[9] Huang, B. *et al.* Extended reconstructed sea surface temperature, version 5 (ERSSTv5): upgrades, validations, and intercomparisons. *Journal of Climate* **30**, 8179–8205 (2017).

[10] Kennedy, J. J., Rayner, N., Atkinson, C. & Killick, R. An ensemble data set of sea surface temperature change from 1850: The Met Office Hadley Centre HadSST.4.0.0.0 data set. *Journal of Geophysical Research: Atmospheres* **124**, 7719–7763 (2019).

[11] Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of CMIP5 and the experiment design. *Bulletin of the American Meteorological Society* **93**, 485–498 (2012).

[12] Eyring, V. *et al.* Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design and organization. *Geoscientific Model Development* **9**, 1937–1958 (2016).

[13] Maher, N. *et al.* The Max Planck Institute Grand Ensemble: enabling the exploration of climate system variability. *Journal of Advances in Modeling Earth Systems* **11**, 2050–2069 (2019).

[14] Swart, N. C. *et al.* The Canadian earth system model version 5 (CanESM5.0.3). *Geoscientific Model Development* **12**, 4823–4873 (2019).

[15] Kendall, M. G. Rank correlation methods. (1948).

[16] Mann, H. B. Nonparametric tests against trend. *Econometrica: Journal of the econometric society* 245–259 (1945).

[17] Hussain, M. & Mahmud, I. pymannkendall: a python package for non parametric Mann Kendall family of trend tests. *Journal of Open Source Software* **4**, 1556 (2019).

[18] Efron, B. Bootstrap methods: another look at the jackknife. In *Breakthroughs in statistics*, 569–593 (Springer, 1992).

[19] Smith, D. M. *et al.* The Polar Amplification Model Intercomparison Project (PAMIP) contribution to CMIP6: investigating the causes and consequences of polar amplification. *Geoscientific Model Development* **12**, 1139–1164 (2019).

[20] Fyfe, J. C., Gillett, N. P. & Zwiers, F. W. Overestimated global warming over the past 20 years. *Nature Climate Change* **3**, 767–769 (2013).

[21] Swart, N. C., Fyfe, J. C., Hawkins, E., Kay, J. E. & Jahn, A. Influence of internal variability on Arctic sea-ice trends. *Nature Climate Change* **5**, 86–89 (2015).