I would not have submitted this comment had our paper (Terhaar et al., 2025) not been explicitly discussed in the open review discussion around this preprint. In the first part of this comment, I therefore address the question of uncertainties in air-sea heat flux estimates from reanalysis products and their implications for the reconstruction of Terhaar et al. (2025), including an extended multi-product reconstruction. Since our paper also examined SST-based AMOC reconstructions, I use this opportunity in the second part to discuss uncertainties in this proxy that, in my view, have not been adequately acknowledged in the paper under discussion or in Caesar et al. (2018).

On the uncertainty of reanalysis-based air-sea heat fluxes and their use for AMOC reconstruction

I thank Mayer et al. for their detailed and constructive comment. I agree that direct air-sea heat flux estimates (F_s,direct) from atmospheric reanalyses carry uncertainties, both in absolute bias and temporal stability, as well documented in the literature they cite. At the same time, these fluxes are produced by atmospheric models that are strongly constrained by observations of the atmospheric state, and they are physically consistent with the observed large-scale circulation. They therefore carry real information and should not be dismissed on the grounds of weak observational constraint alone. In our paper (Terhaar et al., 2025), we acknowledge this uncertainty explicitly and welcome the opportunity to discuss it further here.

I do, however, take issue with the inference that a globally uniform adjustment of air-sea heat fluxes is the appropriate way to address these biases, and that such an adjustment necessarily implies an AMOC decline. As Mayer et al. acknowledge themselves, and as shown clearly in the preprint by Sohail and Zika (2025), the biases in reanalysis surface fluxes are not spatially uniform. Sohail and Zika demonstrate that adjustments are most strongly needed in the Southern Ocean and equatorial regions, while North Atlantic air-sea heat fluxes appear to be comparatively well represented. Applying a globally uniform correction as done in the studies that subsequently find an AMOC decline therefore introduces large and unphysical errors in the North Atlantic, thereby adding uncertainty rather than reducing it. To have any confidence in the trend after an adjustment, one must have confidence in the adjustment, and a globally uniform one is simply not physically justified. While the Sohail and Zika study is still a preprint and should be treated with appropriate caution, its results are physically compelling, have received positive referee comments, and point clearly toward the need for regionally resolved adjustments. Crucially, in the North Atlantic specifically, such an adjustment may be rather small, given the relatively good agreement between ocean heat content changes and integrated air-sea heat flux anomalies from reanalysis data shown by Sohail and Zika (2025), and would not result in an inferred AMOC decline (Sohail and Zika, 2025).

A key piece of supporting evidence that North Atlantic air-sea heat fluxes are relatively well constrained and can provide valuable information about past AMOC variability is the remarkable consistency among three independent reanalysis products (ERA5, JRA-55, and COREv2) in the North Atlantic. While these products differ in their global mean flux biases (Sohail and Zika, 2025), their agreement in the North Atlantic flux anomaly, both in terms of multi-decadal variability and trend, gives confidence that regional air-sea heat fluxes in this basin are indeed relatively well constrained. To further strengthen this point, I have now extended the reconstruction both in time and in the number of reanalysis products, adding COREv2 and extending the analysis to 3-year averages (Fig. 1 here, previously Fig. 9 in Terhaar et al. 2025, now updated). The extraordinary agreement among all three products across the full record, the good agreement with the reconstruction from hydrography by Fraser and Cunningham (2021), and the good correspondence with the RAPID array observations for the overlapping period, provides substantial confidence in the air-sea heat flux-based reconstructions and in the main conclusion of Terhaar et al. (2025).

I also greatly appreciate the alternative approach proposed by Mayer et al., which infers surface fluxes from the atmospheric energy budget constrained by top-of-atmosphere radiation (F_s,inferred). This method avoids some of the limitations of direct flux output and is arguably less susceptible to the biases described above. I consider it a valuable and potentially more accurate complementary approach. Importantly, I do not see it as contradicting our results (Terhaar et al., 2025). From what I have seen of F_s,inferred based reconstructions without a globally uniform bias correction, the reconstructed AMOC appears stable or shows at most a very weak decline of the order of ~3% over the period 1989 to 2016, which is not statistically significant and whose uncertainty range likely includes zero and positive values. I agree that extending F_s,inferred reconstructions to periods before 1989, prior to the availability of top-of-atmosphere radiation observations from CERES, would be an exciting and important advance, and I encourage such efforts

In summary, while I fully agree that air-sea heat flux estimates from reanalyses carry uncertainties that are not trivial, the available evidence suggests that these fluxes are better constrained in the North Atlantic than globally, and that regionally appropriate rather than globally uniform adjustments are what is needed. As it stands, direct and inferred fluxes air-sea flux estimates in the North Atlantic do not indicate a decline since the 1960s but can also not exclude it within the uncertainties. There is clearly still much work to be done, and I look forward to the continued development of better-constrained flux products.

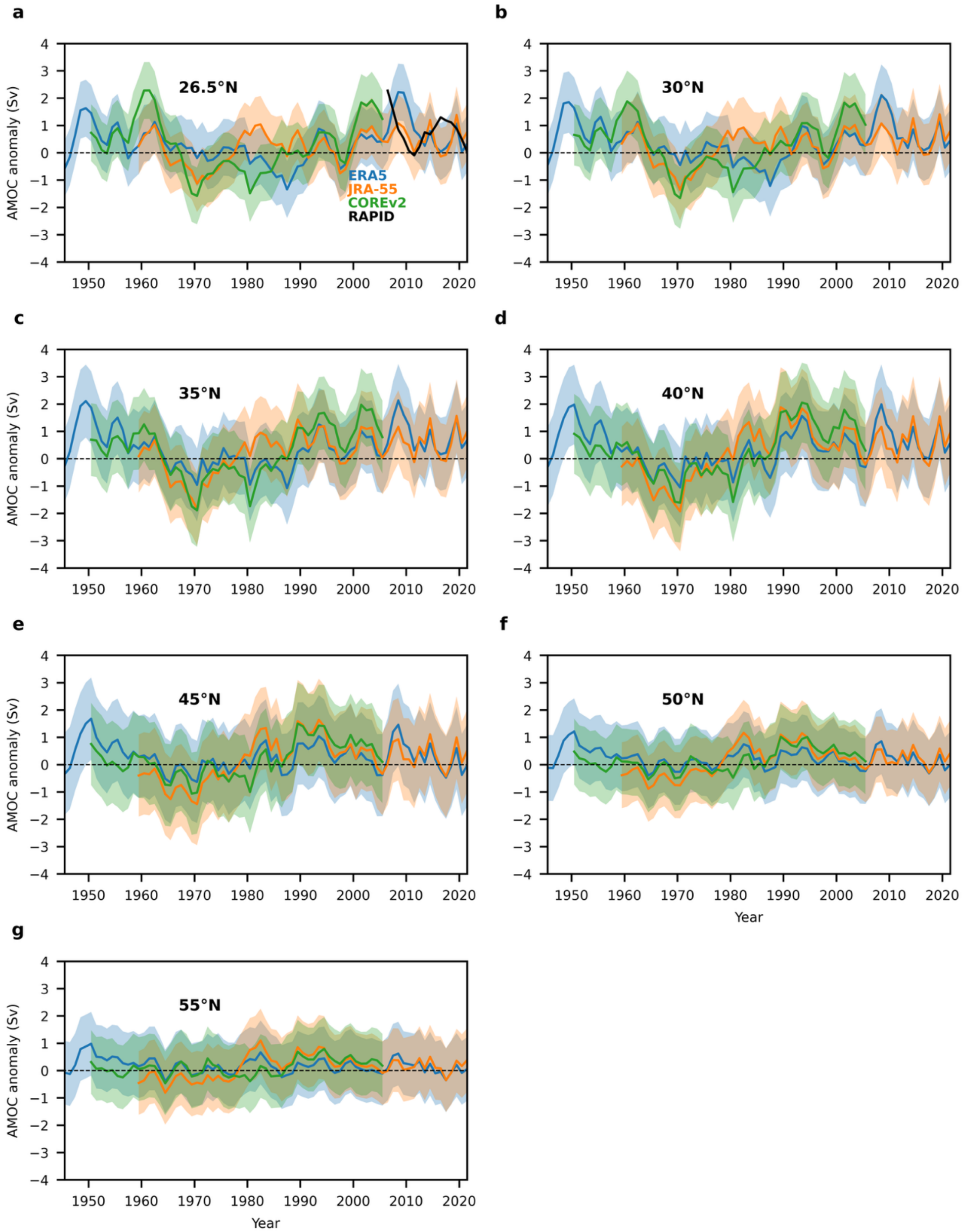

Fig. 1: AMOC anomaly (in Sv) reconstructed from air-sea heat fluxes from three reanalysis products (ERA5, blue; JRA-55, orange; COREv2, green) at latitudes between 26.5°N and 55°N. Shading indicates ±1σ uncertainty in the model-derived relationship. Black line shows observed AMOC strength from the RAPID array at 26.5°N. All time series are shown as 3-year running means and are expressed as anomalies relative to the entire period.

On the uncertainties of SST-based AMOC reconstructions

While air-sea heat fluxes are physically more directly linked to AMOC variability, they are admittedly difficult to observe accurately. SSTs, by contrast, are far easier to observe but are a much less direct proxy for AMOC strength, and I believe the uncertainties in SST-based reconstructions are substantially larger than acknowledged in the paper under discussion and in Caesar et al. (2018).

A first important concern relates to the model calibration underlying the SST-based approach. Caesar et al. (2018) calibrate the relationship between subpolar gyre SSTs and AMOC strength using a limited number of CMIP5 simulations, deriving a slope from the long-term trend. However, when more CMIP5 modes are used, not including these that exhibit a particular sensitivity to aerosol forcing, this relationship becomes substantially weaker and with a less steep slope (Menary et al., 2020, Fig. 4). Furthermore, Menary et al. (2020, Fig. S5) show that the Caesar et al. relationship only emerges clearly in first ensemble members and disappears entirely when ensemble means are used, which raises fundamental questions about whether the relationship is robust or is instead driven by internal variability aliasing. Similarly, Menary et al. (2020) show that there is no difference in the strength and slope of the relationship between CMIP5 and CMIP6 models when excluding models that are overly sensitive to aerosol forcing and that this relationship is weaker and the slope is less steep as the one shown in Caesar et al. (2018) based on a subset of CMIP5 simulations. Similarly, we also found in the CMIP6 models a weaker and not as steep relationship. This is not a minor correction. In our own analysis using CMIP6 models, applying this weaker relationship yields an inferred AMOC decline of only −0.2 ± 0.7 Sv (Terhaar et al., 2025. Fig. 1), which is indistinguishable from zero.

A second, arguably more serious concern is the calibration strategy itself that is used to reconstruct a timeseries of the AMOC. Caesar et al. (2018) derive the SST–AMOC relationship from the long-term trend in model simulations and then apply it to annual or decadal observed SST data. This is methodologically inconsistent: a calibration derived from multi-decadal trends cannot straightforwardly be applied at annual or decadal timescales where the relationship may be much weaker. This is not a theoretical concern; I have tested this explicitly. The relationship between the SST index and AMOC on annual and decadal timescales is indeed very weak ($r^2 = 0.16$ for annual averages and $r^2 = 0.22$ for overall model means; Fig. 2 here), far weaker than what the trend-based calibration would imply. As a direct consequence, the uncertainties on the annual and decadal reconstructions are very large, much larger than those presented by Caesar et al., as illustrated clearly in Fig. 3 here, which shows the annual and decadal reconstructions with uncertainty estimates derived from the model fit. These uncertainties are around 2.5 times larger than the uncertainties from the trend-based fit. Moreover, these uncertainties are ±1σ uncertainties from the fit alone, not ±2σ, and they already encompass zero trend. This means that even before considering any observational uncertainty in the SSTs themselves, the reconstructions cannot statistically distinguish an AMOC decline from no change.

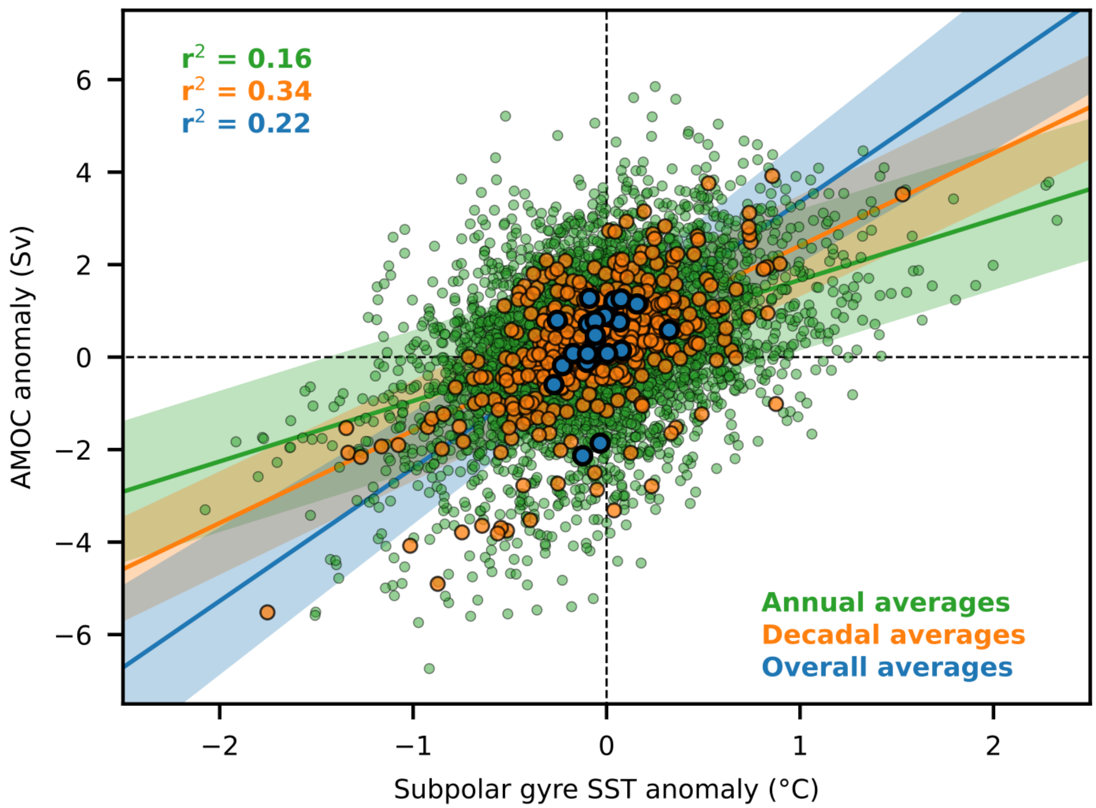

Fig. 2: Relationship between the subpolar gyre SST anomaly (°C) and AMOC anomaly (Sv) across CMIP6 model simulations at annual (green), decadal (orange), and overall mean (blue) timescales. Lines show linear fits for each averaging period with shading indicating the 95% confidence interval of the fit. Explained variance (r^2) is given for each timescale. An offset of 6 years is used between the AMOC and the SST index following Caesar et al. (2018) although the results do not change quantitatively without that offset. The weak relationships at annual and decadal timescales (r^2 = 0.16 and 0.34, respectively) highlight that a calibration derived from long-term model trends, as used in Caesar et al. (2018), is not appropriate for reconstructing annual or decadal AMOC variability from SST observations.

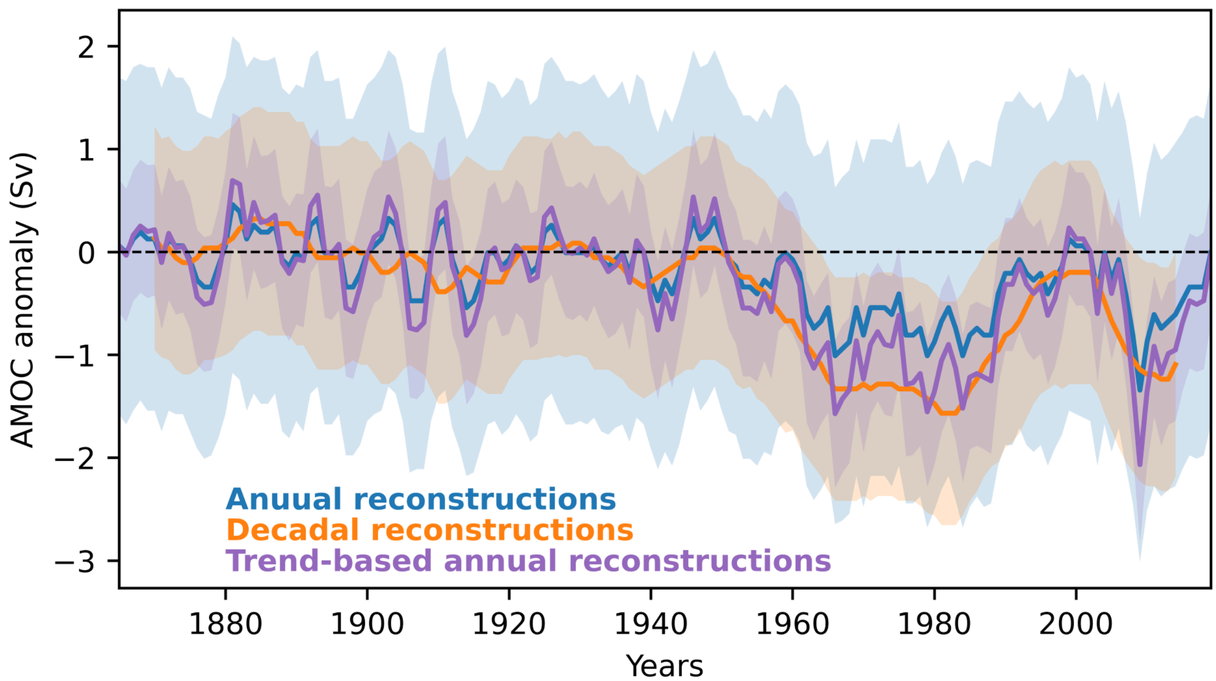

Fig. 3: AMOC anomaly (in Sv) reconstructed from observed HadSST subpolar gyre SSTs using calibrations derived from CMIP6 models. The purple line shows the trend-based annual reconstruction following the methodology of Caesar et al. (2018), the blue line shows the annual reconstruction calibrated on annual model values, and the orange line shows the decadal reconstruction calibrated on decadal model values. Shading indicates ±1σ uncertainty from the model fit. Note that observational uncertainties in the SST index itself are not included and would further increase the total uncertainty. The inferred AMOC anomalies are around half as large as those in Caesar et al. (2018) due to the less steep slope of the relationship shown for a larger ensemble of CMIP5 simulations (Menary et al., 2020) and for CMIP6 models (Menary et al., 2020; Terhaar et al., 2025).

This brings us to a third issue that is rarely adequately discussed: the observational uncertainty in the SST index itself. The Caesar et al. index is defined as subpolar gyre SSTs from November to May minus global mean SSTs from November to May. Both the first and second term in the equation therefore carry observational uncertainties, and both uncertainties, from the North Atlantic SSTs and from global ocean SSTs, contribute equally to the overall uncertainty of the index by Caesar et al. (2018). While the North Atlantic is indeed one of the better-observed ocean basins, wintertime coverage prior to the satellite era was far from complete, and the global mean SST is influenced heavily by large areas of the Southern Ocean, the tropical Pacific, and other regions where historical observational coverage was sparse and data quality was poor, particularly before 1958. These observational uncertainties in the SST index are entirely absent from the uncertainty quantification in Caesar et al. (2018) and in the paper under discussion, yet they would be expected to be very large given the data coverage and quality before the mid-20th century. This matters more than it might appear: a trend analysis of the SST index (see table at the end of the document) reveals that a statistically significant negative trend only emerges when the analysis starts in 1952 or earlier, that is, when it includes data from before the satellite era where observational coverage and quality are most uncertain. When starting from any year between the late 1950s and the mid-1980s, the trend is not only non-negative but actually positive, and even statistically significantly so for start years between approximately 1967 and 1973, indicating a warming of the subpolar North Atlantic relative to the global mean over this period rather than the cooling that would be expected from AMOC weakening. The trend only turns negative again from around 1987 onwards and reaches statistical significance only for start years between roughly the mid-1990s and mid-2000s, before becoming non-significant again and eventually turning positive for start years after around 2011. This behavior reflects the sensitivity of the index to decadal variability and the choice of start year rather than a robust long-term signal, and it does not support the notion of a persistent cold blob indicative of ongoing AMOC weakening. The sign and significance of the inferred trend are therefore highly sensitive to the choice of start year and to the inclusion of data from precisely the period where observational uncertainties are largest. The uncertainties shown in Fig. 3 already paint a sobering picture, and they represent only the model fit component. The full uncertainty, once observational SST uncertainty is properly accounted for, would be considerably larger still.

I therefore conclude that the SST-based AMOC reconstructions, while valuable as one line of evidence, carry uncertainties that are substantially underestimated in their current presentation. The best estimate does suggest a small decline, though considerably smaller than implied by Caesar et al. (2018) due to the weaker slope of the SST-AMOC relationship found in a larger CMIP5 ensemble (Menary et al., 2020) and in the CMIP6 ensemble (Menary et al., 2020; Terhaar et al., 2025). However, properly accounting for uncertainty in the model calibration, the timescale dependence of the SST-AMOC relationship, and the observational quality of the SST index itself leads to uncertainties so large that it is difficult to draw any statistically meaningful conclusion about the trend. The reconstruction is well within the range of a stable AMOC, and the best estimate of a small decline should therefore be interpreted with great caution. Strong conclusions about past AMOC changes can hardly be drawn from this proxy alone in its current form.

Conclusion

Reconstructing past AMOC variability from indirect proxies is inherently challenging, and all available approaches carry significant uncertainties. Air-sea heat fluxes are physically more directly linked to AMOC variability but are difficult to observe, while SSTs are easier to observe in the satellite period but represent a much more indirect proxy with a weak and timescale-dependent relationship to AMOC strength. In both cases, the uncertainties are large and, as I have argued here, substantially underestimated in the current presentation. This matters because conclusions about past and ongoing AMOC changes depend critically on how these uncertainties are treated. An oversimplification of the uncertainty structure, for example through a calibration strategy that is inconsistent with the timescales at which it is applied, can lead to conclusions that are not supported by the data. This is particularly important in the context of combining multiple lines of evidence through Bayesian inference: such approaches are only as reliable as the uncertainty estimates of each individual line of evidence, and overconfident uncertainty quantification in any one proxy will inevitably lead to overconfident conclusions about the AMOC as a whole. I therefore argue that uncertainties in all proxy-based AMOC reconstructions be treated rigorously and quantitatively, and that the significant methodological choices underlying each reconstruction be transparently acknowledged when drawing conclusions about the state and trajectory of the AMOC.

Trend analysis of SST index (start year to 2024)

Colour coding: dark green = statistically significant positive trend ($p < 0.05$); light green = positive trend ($p \geq 0.05$); light red = negative trend ($p \geq 0.05$); dark red = statistically significant negative trend ($p < 0.05$). Asterisk (*) denotes $p < 0.05$.

Start year	End year	N years	Trend (°C yr⁻¹)	Trend (°C decade⁻¹)	p-value
1940	2024	85	-0.0056	-0.0559	0.0006 *
1941	2024	84	-0.0055	-0.0550	0.0009 *
1942	2024	83	-0.0054	-0.0537	0.0015 *
1943	2024	82	-0.0053	-0.0525	0.0023 *
1944	2024	81	-0.0050	-0.0504	0.0041 *
1945	2024	80	-0.0046	-0.0460	0.0095 *
1946	2024	79	-0.0045	-0.0448	0.0135 *
1947	2024	78	-0.0047	-0.0467	0.0121 *
1948	2024	77	-0.0047	-0.0467	0.0144 *
1949	2024	76	-0.0048	-0.0479	0.0145 *
1950	2024	75	-0.0049	-0.0492	0.0144 *
1951	2024	74	-0.0045	-0.0453	0.0267 *
1952	2024	73	-0.0043	-0.0435	0.0381 *
1953	2024	72	-0.0035	-0.0346	0.0956
1954	2024	71	-0.0031	-0.0314	0.1393
1955	2024	70	-0.0024	-0.0239	0.2618
1956	2024	69	-0.0013	-0.0129	0.5352
1957	2024	68	-0.0009	-0.0095	0.6577
1958	2024	67	-0.0003	-0.0031	0.8845
1959	2024	66	-0.0001	-0.0007	0.9738
1960	2024	65	0.0001	0.0014	0.9529
1961	2024	64	0.0001	0.0013	0.9573
1962	2024	63	0.0006	0.0061	0.8026
1963	2024	62	0.0012	0.0117	0.6387
1964	2024	61	0.0016	0.0158	0.5371
1965	2024	60	0.0029	0.0291	0.2510
1966	2024	59	0.0036	0.0365	0.1592
1967	2024	58	0.0048	0.0482	0.0643
1968	2024	57	0.0057	0.0565	0.0339 *
1969	2024	56	0.0057	0.0567	0.0400 *

Start year	End year	N years	Trend (°C yr^{-1})	Trend (°C $decade^{-1}$)	p-value
1970	2024	55	0.0059	0.0594	0.0376 *
1971	2024	54	0.0064	0.0637	0.0316 *
1972	2024	53	0.0063	0.0630	0.0402 *
1973	2024	52	0.0052	0.0524	0.0913
1974	2024	51	0.0049	0.0490	0.1276
1975	2024	50	0.0047	0.0470	0.1594
1976	2024	49	0.0050	0.0503	0.1477
1977	2024	48	0.0044	0.0443	0.2176
1978	2024	47	0.0048	0.0485	0.1957
1979	2024	46	0.0053	0.0532	0.1731
1980	2024	45	0.0048	0.0484	0.2335
1981	2024	44	0.0056	0.0562	0.1843
1982	2024	43	0.0061	0.0609	0.1697
1983	2024	42	0.0049	0.0492	0.2819
1984	2024	41	0.0045	0.0449	0.3481
1985	2024	40	0.0030	0.0298	0.5457
1986	2024	39	0.0013	0.0126	0.8047
1987	2024	38	-0.0003	-0.0027	0.9593
1988	2024	37	-0.0013	-0.0127	0.8186
1989	2024	36	-0.0017	-0.0172	0.7688
1990	2024	35	-0.0048	-0.0476	0.4213
1991	2024	34	-0.0078	-0.0781	0.1958
1992	2024	33	-0.0094	-0.0940	0.1401
1993	2024	32	-0.0124	-0.1236	0.0610
1994	2024	31	-0.0168	-0.1684	0.0109 *
1995	2024	30	-0.0200	-0.1996	0.0038 *
1996	2024	29	-0.0208	-0.2084	0.0047 *
1997	2024	28	-0.0214	-0.2143	0.0066 *
1998	2024	27	-0.0232	-0.2320	0.0060 *
1999	2024	26	-0.0238	-0.2375	0.0089 *
2000	2024	25	-0.0248	-0.2484	0.0113 *
2001	2024	24	-0.0274	-0.2740	0.0097 *
2002	2024	23	-0.0293	-0.2928	0.0109 *
2003	2024	22	-0.0336	-0.3361	0.0069 *

Start year	End year	N years	Trend (°C yr⁻¹)	Trend (°C decade⁻¹)	p-value
2004	2024	21	-0.0346	-0.3460	0.0109 *
2005	2024	20	-0.0316	-0.3159	0.0301 *
2006	2024	19	-0.0263	-0.2630	0.0877
2007	2024	18	-0.0236	-0.2356	0.1636
2008	2024	17	-0.0145	-0.1451	0.4116
2009	2024	16	-0.0082	-0.0819	0.6737
2010	2024	15	-0.0026	-0.0260	0.9054
2011	2024	14	0.0166	0.1659	0.4492
2012	2024	13	0.0301	0.3005	0.2188
2013	2024	12	0.0598	0.5984	0.0081 *
2014	2024	11	0.0744	0.7439	0.0039 *
2015	2024	10	0.0684	0.6842	0.0188 *
2016	2024	9	0.0359	0.3588	0.0765
2017	2024	8	0.0487	0.4873	0.0520
2018	2024	7	0.0748	0.7483	0.0087 *
2019	2024	6	0.0644	0.6444	0.0545
2020	2024	5	0.0449	0.4491	0.2578
2021	2024	4	0.0454	0.4538	0.5016
2022	2024	3	0.1232	1.2316	0.3451