nature portfolio

Peer Review File

The Arctic has warmed nearly four times faster than the globe since 1979

Web links to the author's journal account have been redacted from the decision letters as indicated to maintain confidentiality.

7th Mar 22

Dear Dr Rantanen,

Your manuscript titled "The Arctic has warmed four times faster than the globe since 1980" has now been seen by 3 reviewers, and I include their comments at the end of this message. They find your work of interest, but some important points are raised. We are interested in the possibility of publishing your study in Communications Earth & Environment, but would like to consider your responses to these concerns and assess a revised manuscript before we make a final decision on publication.

We therefore invite you to revise and resubmit your manuscript, along with a point-by-point response that takes into account the points raised. Please highlight all changes in the manuscript text file. In the following, we list our editorial thresholds:

• Provide compelling quantitative insights of Arctic Amplification in the comparison between observation and models by considering different analysis periods and large model ensembles;
• Consider the uncertainties of the in-situ observations datasets used and provide a discussion in light of observational uncertainties;
• Ensure that differences between observations and model are estimated realistically.

We are committed to providing a fair and constructive peer-review process. Please don't hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the referees' comments (which should be in a separate document to any cover letter) and the completed checklist:
[link redacted]
** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to receive your revised paper within six weeks; please let us know if you aren't able to submit it within this time so that we can discuss how best to proceed. If we don't hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We understand that due to the current global situation, the time required for revision may be longer than usual. We would appreciate it if you could keep us informed about an estimated timescale for resubmission, to facilitate our planning. Of course, if you are unable to estimate, we are happy to accommodate necessary extensions nevertheless.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further. We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Best regards,

Viviane V. Menezes, PhD
Editorial Board Member
Communications Earth & Environment
orcid.org/0000-0002-4885-2056

Heike Langenberg, PhD
Chief Editor
Communications Earth & Environment

EDITORIAL POLICIES AND FORMATTING

We ask that you ensure your manuscript complies with our editorial policies. Please ensure that the following formatting requirements are met, and any checklist relevant to your research is completed and uploaded as a Related Manuscript file type with the revised article.

Editorial Policy: Policy requirements

Furthermore, please align your manuscript with our format requirements, which are summarized on the following checklist:
Communications Earth & Environment formatting checklist

and also in our style and formatting guide Communications Earth & Environment formatting guide .

*** DATA: Communications Earth & Environment endorses the principles of the Enabling FAIR data project (http://www.copdess.org/enabling-fair-data-project/). We ask authors to make the data that support their conclusions available in permanent, publically accessible data repositories. (Please contact the editor if you are unable to make your data available).

All Communications Earth & Environment manuscripts must include a section titled "Data Availability" at the end of the Methods section or main text (if no Methods). More information on this policy, is available at http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf.

In particular, the Data availability statement should include:
- Unique identifiers (such as DOIs and hyperlinks for datasets in public repositories)
- Accession codes where appropriate
- If applicable, a statement regarding data available with restrictions
- If a dataset has a Digital Object Identifier (DOI) as its unique identifier, we strongly encourage including this in the Reference list and citing the dataset in the Data Availability Statement.

DATA SOURCES: All new data associated with the paper should be placed in a persistent repository where they can be freely and enduringly accessed. We recommend submitting the data to discipline-specific, community-recognized repositories, where possible and a list of recommended repositories is provided at http://www.nature.com/sdata/policies/repositories.

If a community resource is unavailable, data can be submitted to generalist repositories such as figshare or Dryad Digital Repository. Please provide a unique identifier for the data (for example a DOI or a permanent URL) in the data availability statement, if possible. If the repository does not provide identifiers, we encourage authors to supply the search terms that will return the data. For data that have been obtained from publically available sources, please provide a URL and the specific data product name in the data availability statement. Data with a DOI should be further cited in the methods reference section.

Please refer to our data policies at http://www.nature.com/authors/policies/availability.html.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

This study examines the magnitude of Arctic amplification(AA) in recent observations and models. It is often reported that the Arctic is warming twice as fast as the rest of the world, however, the authors show that since 1980, the Arctic has warmed four times as fast as the global mean. They then show that this strong AA is very rare in climate models even when using large ensemble where internal variability is properly accounted for.

The headline result – that the Arctic is warming four times as fast as the globe – is a nice update to the data. However, the exact number is not that that useful because it will depend on the start date/trend length and on the definition of the 'Arctic'. The more interesting and important results are the comparisons to models which could point to deficiencies in the models which may have important implications. To my knowledge, there are not any studies that have done a proper quantitative comparison of AA between observations and models using both CMIP and single model large ensembles (which are needed because of the large internal variability in the Arctic). Therefore, I think that this study has the potential to be important. However. I do have some major concerns that need to be addressed. In particular, the authors are likely overstating the significance of the difference between models and observations.

Major issues:

1. The difference between observations and models are very likely to be overestimated in the statistical tests, and the p-values the authors calculated have little meaning. This is because the authors have chosen the length and start/end dates (1979-2018) of the observed trend because it was an extreme AA trend. These choices were not made a priori. Of course an extreme trend in the

observations will appear to be rare in the model as well.

The authors try to justify the start date and trend length, but their arguments are not convincing. They argue that this period is when warming is approximately linear, but then they contradict this point later when they say that the sustained period of warming started in 1971. In addition, the exact timing of when the trend may appear to be linear is going to depend a lot on internal variability cancelling/accelerating the trend (in addition the changes in GHG and aerosol forcing). They also mention the reliability of ERA5 before 1979, but the other datasets are in good agreement, and all will likely show weaker AA when starting trends before 1979. Also, there is no justification for ending the trend in 2018 when they have the 2019 data (and they could even update with 2020 and 2021 data).

The authors could do more comparisons between models and observations beyond the cherry-picked 40 year time span. The authors indeed show weaker observed AA when starting trends earlier (figure S1), but how do these compare with the models? If the discrepancy is in the forced response rather than due to internal variability, there should still be a discrepancy over longer periods. One thing that could be done is to make a plot similar to Fig 2, but showing the trends as function of start date with a fixed end date at present day.

2. Figures 2-4 are essentially illustrating the same point, so I am not sure why they all are needed. I like figure 2, but I don't see the point of figure 3. What new result/message is illustrated here that are not shown by figure 2? Figure 4 again makes the same point, but also that models that have more global warming have more Arctic warming, which seems like an obvious point that is not worth a whole figure. I think the authors were trying to reproduce a similar plot that was shown in Rosenblum and Eisenman (2017) for sea ice, but here with Arctic temperature. Rosenblum and Eisenman (2017) made a much more important point because there had been some studies comparing the sea ice trends in models and observations and claiming that they agreed because observation were within the model spread (but in reality were likely getting the right answer for the wrong reasons). I am not aware of any study making similar claims about Arctic warming, so I am not sure why this is relevant or interesting. It feels like the authors had an interesting result, but couldn't figure what else to fill the rest of the paper with, so added a few filler figures to increase the length.

The authors should either remove some of these figures, or more clearly state what the key messages are. The authors could also consider including additional analysis that would improve the paper and give more insight. Some examples they may want to consider: Examining the vertical structure of the AA or the seasonal differences. Are there discrepancies higher up in the mid-troposphere or only at the surface? Are the discrepancies seen in specific months/seasons or is it year round? The answers to these could begin to pinpoint the causes of any discrepancies.

Minor comments:
L16-17: "The difference.." Statistical significance here means that the observed trend is unlikely to occur in models beyond some arbitrary threshold, so it is essentially repeating the previous sentence and can be deleted.

L25: mowadays -> nowadays

L42-48 I was a little surprised that these reported much weaker AA than this study, especially since many of them are quite recent (and thus are unlikely to be outdated). I decided to give each of these references a quick skim to see why they might differ. What I found was that many of these did not do any calculation, they simply said these statements in the abstract or introduction, either without a reference or a with a reference to a much older study. This paragraph should be clarified to better represent what these studies did (or didn't do).

Methods: I think that the detailed methods are normally presented at the end of the paper for this journal, but I will leave that up to the editors to comment on.

L96-97: Why is only one member used for CMIP5, but all members are used for CMIP6?

L240-242: Why are the modelled AA40 trends starting before 1971 not used? The spread seems larger in this period, so this would change the results.

L254: Another possibility is that models underestimate the magnitude of multidecadal internal variability in the Arctic. There is evidence that models underestimate the multidecadal internal variability of the extratropical atmospheric circulation (e.g. O'Reilly et al. 2021 DOI: 10.1038/s43247-021-00268-7), which could impact the multidecadal variability of the Arctic temperature. It also seems plausible that if models underestimate the feedbacks in the Arctic that could result both an underestimation of the forced response and internal variability.

L268-277: Because these two models examined here are quite different, it might be useful to consider more large ensembles. There are many more that are available, such as the (1) CESM2 LENS (100 members): https://www.cesm.ucar.edu/projects/community-projects/LENS2/
(2) GFDL-SPEAR (only 30 members, but at a higher resolution than typical SMILEs and CMIP models): https://www.gfdl.noaa.gov/spear_large_ensembles/
(3) MIROC6 has a 50 member ensemble that is part of CMIP6. I am guessing that the authors did not use this because some members only go to 2039, but I doubt the one year will make a huge difference.

Reviewer #2 (Remarks to the Author):

Review of "The Arctic has warmed four times faster than the globe since 1980" by Rantanen et al.

General Comments:

The authors analyzed the Arctic-to-global-mean ratio (AA) of annual-mean warming trends from 1979-2018 (40-yr trends) in near-surface air temperature from ERA5 and 3 other observational datasets and compared it with those seen in 4 large ensembles of model simulations. They found an AA of ~4 in the four datasets, rather the commonly cited factor of 2, and that models underestimate this observation-based estimate of AA. The analyses appear to be comprehensive and solid, and the results are robust and important. However, I do have a few major concerns.

Main concerns:

1. ERA5 and all the observational temperature datasets have included very limited in-situ observations of surface air temperature (SAT) over the Arctic Ocean, especially over the polar ice cap during the winter, when SAT can be very different from SST, for which some marine observations

may exist and may have been used by ERA5 and the 3 global datasets. This could lead to overestimation of the SAT trends over the Arctic by these datasets. In any case, I think the authors need to recognize the possibility that apparent Arctic warming trends in the available datasets may an overestimation, or at least contain large uncertainties. Therefore, one cannot conclude that the models truly understand Arctic warming.

2. Even if the AA is underestimated by the models, it could be caused by the low global-warming rates in the recent decades, rather than due to underestimated Arctic warming. A quick comparison of Fig. 1a with model-simulated Arctic SAT anomaly series would help answer this question.

3. The estimated Arctic warming amount and thus the AA strongly depend on the time period, season, and location. For a short time period like 40 years, their estimates will contain large errors. Considering such error bars may improve the consistency between observations and model simulations.

Specific comments:

1. L18: “line” ◊ “in line”.

2. L25: “mowadays” ◊ “nowadays”

3. L42-45: Fig. 1 of Dai et al. (2019) is highly relevant to this study: It shows that the AA in ERA-I ranges from ~1.6 for July-August to ~9.0 for January during 1979-2016 and the CMIP5 models show much lower AA during this period. It’s well known that the AA is mainly a cold-season phenomenon. Thus, it’s better to discuss the AA explicitly for specific seasons or annual mean.

4. L49-54: I think the lack of reliable observations of surface air temperature over the Arctic region (especially over the Arctic Ocean) is the main source of uncertainty in our estimates of Arctic warming amount.

5. L71-72: Given that up to 50% of recent sea-ice loss may come from realization-dependent internal variability, it will be difficult for models to reproduce the observed sea-ice loss and the related warming. The effect of the realization-dependent internal variability should be mentioned here.

6. L81-82: The spatial interpolation does not mean they will produce better estimates of Arctic warming if there are no observations over most of the Arctic region. Numerically, one can fill in any data gaps but that does not mean the result will be reliable. Need to examine their actual in-situ data coverage in these datasets, not their interpolated data coverage. The 1200km interpolation radius may be suitable for low latitudes, but it may be inappropriate for the Arctic, as it will essentially fill in the data gaps over the Arctic Ocean by high-latitude land station data, and we know that near-surface air temperature changes over the ice-covered Arctic Ocean are very different from that over high-latitude land. Thus, the GISTEMP, BEST and CW likely reflect mainly the warming rates over the high-latitude land areas, not really the warming over the Arctic Ocean, which has few observations to be used in these datasets to my knowledge. This issue needs to be clearly addressed and emphasized in this study.

7. L87: Need to state the implications of this statement. For example, are these datasets representative of the true Arctic warming rates? Fig. S2 seems suggest that these datasets used

some marine observations over the Arctic Ocean to derive stronger warming over the Arctic Ocean than over the high-latitude land areas. A more detailed discussion is needed on what in-situ observations from the Arctic region were used in the global datasets.

The authors can use the ERA5 and CMIP6 model data to verify whether the high latitude land areas and the Arctic Ocean have similar warming rates. If yes, then the three global datasets may be useful for quantifying the Arctic warming. If no, then they can't be used to quantify warming over the Arctic Ocean, unless they did use observations over the Arctic Ocean in their analyses (Fig. S2 seems to suggest that is the case). To some extent, it's already been shown in previous studies (e.g., Fig. 2 of Dai eta l. 2019).

8. eq. 1: As AA refers to the enhanced Arctic warming relative to the global-mean warming rate; thus, AA can be defined as the difference or ratio of the Arctic to global-mean warming amount or warming rate. In eq. 1, the time interval should be the same, so that AA defined using the warming amount (dT_Arctic/dT_global) or warming rate should be the same. Clearly, a shorter time period will increase the contribution of realization-dependent internal variability on the warming amount, thus AA will depend strongly on the time interval used, as shown in Fig. S1.

9. L135: The dT/dtA is likely to be sensitive to the southern latitude as this determines how much land area is included in the calculation, as shown by Fig. S1. This point should be pointed out here. Please state the dataset used in Fig. S1.

10. L147: How were the 31 different 40-year periods in 1971-2040 generated? Do they contain some overlapping and thus some correlation among them (i.e., not totally independent)?

11. L172 and L180-184: Fig. S2 show large regional differences in the warming trends among the 4 datasets, despite the general agreement in the Arctic-mean time series. Some information regarding what observational data over the Arctic Ocean were used by each of the four datasets would be helpful in explaining their warming differences over the Arctic Ocean.

12. L204-205: I think that AA is around 2 was derived based on CMIP model projected warming rates, not based on recent warming rates that include effects of internal variability unrelated to GHG-induced warming. As shown previously (e.g., Dai et al. 2019), the AA is mainly a cold-season phenomenon. Thus, it would be helpful to show the AA as function of month, similar to Fig. 1 of Dai et al. (2019).

13. Fig. 2: What does the x-axis represent in this figure? Is the middle or ending year of the 40-year period used to compute the AA? I found some of the figure captions lack the key information needed to understand the results. It seems that you used all the years in the moving 40-yr window (this needs to be described in Method section).

14. L227-228: Fig. 2a only contains a relatively small sample runs, so it is possible that it may miss some possibilities, such as the observed realization.

15. L237: A 40-year period is short for quantifying externally-forced trends in individual realizations.

16. L253-254: Another possibility is that the four datasets overestimate the recent Arctic warming due to lack of observations of 2m air temperatures over the Arctic region (especially over the winter

polar ice cap, where the 2m air temperature can be much colder than nearby SST (a few SST observations may be used in the observational datasets).

17. L266-267: In Monto-Carlo simulations, tens of thousands of simulations were often used to generate the PDF for estimating the significance. Here, only 100-member runs were used and the 40-year periods from these runs were overlapped (thus not independent). Given this, I would caution the authors to make very strong statement on the likelihood of the recent warming rates, although I agree it seems very unlikely based on the PDFs from the limited number of model runs, especially for the MPI-GE that has a low ECS.

18. L278-279: Again, I think the authors need to consider the real possibility that the available datasets may overestimate Arctic warming since 1980 due to the lack of 2m air temperature observations over winter sea ice, rather than taking the face values of the 4 datasets as the truth and conclude that the models underestimate Arctic warming.

19: L294-302: Many studies showed that the recent global warming rates were significantly affected by the phase changes in IPO in the Pacific and AMO in the Atlantic. Because the models have difficulties in simulating the IPO and AMO in the first place, and the recent IPO/AMO phase combinations are highly realization dependent (i.e., it may take hundreds to thousands of realizations to capture it), I think it would be better to compare the observed and model-simulated Arctic warming rate or amount, rather than the Arctic/global dT ratio). This would avoid the effect of the errors in simulating the recent global warming rates. Also, estimates of trends in 40-year series are highly uncertain statistically (e.g., the end points may have a significant effect, etc.). This could be another source of the uncertainties in the comparison.

Reviewer #3 (Remarks to the Author):

Rantanen et al. (2022) discussed that the intensified Arctic warming in the past 40 years is much larger than was conventionally reported. In general, I found this study interesting and well-written. The results are also important, especially the implications for the potential underestimation of AA in state-of-the-art climate models. I support to publish this paper with some wording checks. Below are my minor comments:

To assess the internal variability, there is an observational large-ensemble datasets (https://www.cesm.ucar.edu/projects/community-projects/MMLEA/, McKinnon and Deser, 2018). It would be informative if the authors can estimate the internal variability of observations and compare it to those from SMILEs and spreads of CMIP5/6. The observational estimate could be added in Figures 3 and 4.

In the Discussion and conclusions section, the authors mentioned the effect of hiatus, which makes the global warming trend smaller and producing larger AA. I would like to know more about the authors' opinion on AA. Should we view AA as a global climate indicator or just a local one? We could get very large AA, assuming Arctic warming does not change but with very small global warming. And how should we interpret negative AA? If Arctic cools more than the globe does, one can still get positive AA.

Reference:

McKinnon, K.A. and Deser, C., 2018. Internal variability and regional climate trends in an observational large ensemble. Journal of Climate, 31(17), pp.6783-6802.

The responses to reviewers of "The Arctic has warmed four times faster than the globe since 1980"

Mika Rantanen, Alexey Yu. Karpechko, Antti Lipponen, Kalle Nordling, Otto Hyvärinen, Kimmo Ruosteenoja, Timo Vihma and Ari Laaksonen

We thank the reviewers for their constructive comments on our submitted manuscript. We are glad to see that our manuscript has been found of interest, although some important points have been raised.

Based on the concerns by the three reviewers, we have made a major revision to our work. In practice, we have recalculated almost all results. The main points of the revision can be listed as follows:

- The analysis extends now to 2021, so two years of more observational data have been added.
- The primary period of interest was changed from 1980-2019 to 1979-2021. This further strengthens our main results due to the longer time window considered.
- More comprehensive analysis of AA regarding various trend lengths and areal definitions of the Arctic are now presented in the main text.
- Seasonal variation of AA and its comparison to climate models are briefly discussed.
- The uncertainties related to the observed temperature trends in the Arctic are communicated explicitly in the Discussion section.

To better reflect the results of the revised study, we changed the title of the manuscript to "The Arctic has warmed nearly four times faster than the globe since 1979". In addition to all these points, your feedback has led to numerous smaller improvements of the manuscript. Thus, we kindly ask you to read it again.

The point-by-point replies to your comments are below, marked in black and our responses in blue.

Reviewer #1

This study examines the magnitude of Arctic amplification(AA) in recent observations and models. It is often reported that the Arctic is warming twice as fast as the rest of the world, however, the authors show that since 1980, the Arctic has warmed four times as fast as the global mean. They then show that this strong AA is very rare in climate models even when using large ensemble where internal variability is properly accounted for.

The headline result – that the Arctic is warming four times as fast as the globe – is a nice update to the data. However, the exact number is not that useful because it will depend on the start date/trend length and on the definition of the ‘Arctic’. The more interesting and important results are the comparisons to models which could point to deficiencies in the models which may have important implications. To my knowledge, there are not any studies that have done a proper quantitative comparison of AA between observations and models using both CMIP and single model large ensembles (which are needed because of the large internal variability in the Arctic). Therefore, I think that this study has the potential to be important. However. I do have some major concerns that need to be addressed. In particular, the authors are likely overstating the significance of the difference between models and observations.

Major issues:

1. The difference between observations and models are very likely to be overestimated in the statistical tests, and the p-values the authors calculated have little meaning. This is because the authors have chosen the length and start/end dates (1979-2018) of the observed trend because it was an extreme AA trend. These choices were not made a priori. Of course an extreme trend in the observations will appear to be rare in the model as well.

The authors try to justify the start date and trend length, but their arguments are not convincing. They argue that this period is when warming is approximately linear, but then they contradict this point later when they say that the sustained period of warming started in 1971. In addition, the exact timing of when the trend may appear to be linear is going to depend a lot on internal variability cancelling/accelerating the trend (in addition the changes in GHG and aerosol forcing). They also mention the reliability of ERA5 before 1979, but the other datasets are in good agreement, and all will likely show weaker AA when starting trends before 1979. Also, there is no justification for ending the trend in 2018 when they have the 2019 data (and they could even update with 2020 and 2021 data).

The authors could do more comparisons between models and observations beyond the cherry-picked 40 year time span. The authors indeed show weaker observed AA when starting trends earlier (figure S1), but how do these compare with the models? If the discrepancy is in the forced response rather than due to internal variability, there should still be a discrepancy over longer periods. One thing that could be done is to make a plot similar to Fig 2, but showing the trends as function of start date with a fixed end date at present day.

Thank you for these useful comments. We agree that analyzing the 40-year AA as occurred in 1979-2018 is not so relevant anymore in 2022 as there are now three full years of new data since 2018.

To get more robust and up-to-date comparison of AA between the observations and models, we extended the data up to 2021. We also changed the primary period of interest from 1980-2019 to 1979-2021. We argue that using 1979 as the starting year is the most reasonable choice due to the beginning of the satellite era in the observations. The new time period now spans 43 years, which is three years longer than in the original manuscript.

In order to analyse the sensitivity of the magnitude of AA to various trend lengths, we added two more figures to the manuscript: the first one, Fig. 2, shows the sensitivity of AA to the time window used in calculating the linear trends and the southern boundary of the Arctic. Fig. 2a was previously in the supplementary material, but we decided to bring it to the main text. Fig. 2b shows the discrepancy between observed AA and CMIP6 under a wide range of time periods and area definitions.

The second new figure (Fig. 4) shows the trends as a function of the starting year, as suggested by the referee. Naturally, discussion on these figures are provided in the main text, in Section 5.

We also replaced the Cowtan & Way dataset by HadCRUT5 (Morice et al., 2021) as Cowtan & Way was discontinued in June 2021.

Morice et al. (2021):
https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019JD032361

2. Figures 2-4 are essentially illustrating the same point, so I am not sure why they all are needed. I like figure 2, but I don't see the point of figure 3. What new result/message is illustrated here that are not shown by figure 2? Figure 4 again makes the same point, but also that models that have more global warming have more Arctic warming, which seems like an obvious point that is not worth a whole figure. I think the authors were trying to reproduce a similar plot that was shown in Rosenblum and Eisenman (2017) for sea ice, but here with Arctic temperature. Rosenblum and Eisenman (2017) made a much more important point because there had been some studies comparing the sea ice trends in models and observations and claiming that they agreed because observation were within the model spread (but in reality were likely getting the right answer for the wrong reasons). I am not aware of any study making similar claims about Arctic warming, so I am not sure why this is relevant or interesting. It feels like the authors had an interesting result, but couldn't figure what else to fill the rest of the paper with, so added a few filler figures to increase the length.

The authors should either remove some of these figures, or more clearly state what the key messages are. The authors could also consider including additional analysis that would improve the paper and give more insight. Some examples they may want to consider: Examining the vertical structure of the AA or the seasonal differences.

Are there discrepancies higher up in the mid-troposphere or only at the surface? Are the discrepancies seen in specific months/seasons or is it year round? The answers to these could begin to pinpoint the causes of any discrepancies.

Thank you for these comments. We agree that Figure 4 was less necessary, thus we decided to move it to the supplementary material. However, we kept Figures 2 and 3 in the main text (with changes from 40-year AA to 43-year AA). We would like to emphasize that in the PDF-histograms of Fig. 3, we have used all possible 43-year AA-ratios in the climate models starting from 1970 to 2040. Thus, Fig. 3 tells that even when taking into account the long time window (71 years), the probability of the observed AA is very small according to model simulations. This cannot be seen directly from Fig. 2. In the revised manuscript, Figures 2 and 3 are Figures 3 and 6, respectively.

The suggestion of analysing seasonal AA is relevant, and actually is something we had considered earlier. Therefore, in the revised manuscript, we added a new figure showing the monthly AA values in observations and CMIP6 climate models (Fig. 5). We discuss these results briefly at the end of Section 5. Comparison to CMIP5 models was added to supplementary material (Fig. S6).

We agree with the reviewer that examining the vertical structure of AA could be relevant for understanding the causes of the discrepancies between observed and simulated AA. However, downloading and analysing the pressure level data from both observations and climate models datasets would be a major effort that unfortunately goes beyond what is possible with our resources for this manuscript. Further, presenting results for AA at several altitudes would be challenging under the length restrictions of manuscripts for Communications Earth & Environment. We believe that the results presented in our manuscript do not depend on the analysis of the vertical structure, and that additional insights that can be obtained from such an analysis would deserve a separate study.

Minor comments:

L16-17: “The difference..” Statistical significance here means that the observed trend is unlikely to occur in models beyond some arbitrary threshold, so it is essentially repeating the previous sentence and can be deleted.

This sentence has been removed.

L25: mowadays -> nowadays

This typo has been fixed.

L42-48 I was a little surprised that these reported much weaker AA than this study, especially since many of them are quite recent (and thus are unlikely to be outdated). I decided to give each of these references a quick skim to see why they might differ. What I found was that many of these did not do any calculation, they simply said these statements in the abstract or introduction, either without a reference or a with a reference to a much older study. This paragraph should be clarified to better represent what these studies did (or didn't do).

We added a clarifying sentence here at lines 46-48: "However, the warming ratios reported in these and many other studies have usually been only referenced from older, possibly outdated, estimates and have not included recent observations. "

Methods: I think that the detailed methods are normally presented at the end of the paper for this journal, but I will leave that up to the editors to comment on.

In the submission guidelines it reads that "Manuscripts submitted to *Communications Earth & Environment* do not need to adhere to our formatting requirements at the point of initial submission; formatting requirements only apply at the time of acceptance."

Thus, the methods will be placed at the end of the manuscript provided that the manuscript is accepted for publication.

L96-97: Why is only one member used for CMIP5, but all members are used for CMIP6?

The primary reason for using only one member in CMIP5 is that we do not have the data covering all available ensemble members. However, for CMIP6, we performed a sensitivity analysis and noted that our key results were not highly dependent on whether we used one ensemble member or all the members (line 298 in the revised manuscript).

Furthermore, Notz and SIMIP Community (2020) found that CMIP6 models capture the sensitivity of Arctic sea ice to global warming better than CMIP5 models, which is in line with our results using only one member from CMIP5. For both of these reasons, we do not expect CMIP5 results to change significantly even if all ensemble members were in use.

Notz and SIMIP Community (2020):
https://agupubs.onlinelibrary.wiley.com/doi/full/10.1029/2019GL086749

L240-242: Why are the modelled AA40 trends starting before 1971 not used? The spread seems larger in this period, so this would change the results.

The definition of Arctic amplification makes only sense if there is clear global warming on which the Arctic warming is superimposed. We wanted to constrain our analysis to the time period which represents the current warming period. 1971 reflects approximately the point of time when the recent, sustained global warming started: the IPCC AR6 2nd chapter (p. 2-36) mentions that "most of the sustained warming has occurred after the 1970s".

However, to be more consistent with IPCC, and also because the trend length was extended from 40 years to 43 years, we extended the time window by one year from 1971 to 1970 in the revised manuscript.

L254: Another possibility is that models underestimate the magnitude of multidecadal internal variability in the Arctic. There is evidence that models underestimate the multidecadal internal variability of the extratropical atmospheric circulation (e.g. O'Reilly et al. 2021 DOI: 10.1038/s43247-021-00268-7), which could impact the multidecadal variability of the Arctic temperature. It also seems plausible that if models underestimate the feedbacks in the Arctic that could result both an underestimation of the forced response and internal variability.

Thank you for pointing out this interesting study. We briefly discuss this aspect in the Discussion section, including a citation to this paper.

L268-277: Because these two models examined here are quite different, it might be useful to consider more large ensembles. There are many more that are available, such as the (1) CESM2 LENS (100 members):
https://www.cesm.ucar.edu/projects/community-projects/LENS2/
(2) GFDL-SPEAR (only 30 members, but at a higher resolution than typical SMILEs and CMIP models):
https://www.gfdl.noaa.gov/spear_large_ensembles/
(3) MIROC6 has a 50 member ensemble that is part of CMIP6. I am guessing that the authors did not use this because some members only go to 2039, but I doubt the one year will make a huge difference.

Thank you for these suggestions. However, as far as we understand, none of these suggested large ensemble datasets is ideal for our needs:

1. CESM2 LENS includes only SSP3-7.0 for the future,
2. GFDL-SPEAR includes only SSP5-8.5 for the future,
3. MIROC6 is indeed reported as having 50 members for SSP2-4.5, but we had access only for three members.

Even though the impact of the emission scenario is probably very limited before 2040, for simplicity, we decided to stick with models which all have the same, SSP2-4.5/RCP4.5 scenario.

Reviewer #2

Review of “The Arctic has warmed four times faster than the globe since 1980” by Rantanen et al.

General Comments:

The authors analyzed the Arctic-to-global-mean ratio (AA) of annual-mean warming trends from 1979-2018 (40-yr trends) in near-surface air temperature from ERA5 and 3 other observational datasets and compared it with those seen in 4 large ensembles of model simulations. They found an AA of ~4 in the four datasets, rather the commonly cited factor of 2, and that models underestimate this observation-based estimate of AA. The analyses appear to be comprehensive and solid, and the results are robust and important. However, I do have a few major concerns.

Main concerns:

1. ERA5 and all the observational temperature datasets have included very limited in-situ observations of surface air temperature (SAT) over the Arctic Ocean, especially over the polar ice cap during the winter, when SAT can be very different from SST, for which some marine observations may exist and may have been used by ERA5 and the 3 global datasets. This could lead to overestimation of the SAT trends over the Arctic by these datasets. In any case, I think the authors need to recognize the possibility that apparent Arctic warming trends in the available datasets may an overestimation, or at least contain large uncertainties. Therefore, one cannot conclude that the models truly understand Arctic warming.

Thank you for this comment. We agree that the estimation of the actual warming rates over the Arctic is challenging due to sparseness of the observation network. In addition, it is well-known that reanalyses and climate models struggle to accurately simulate strong stable boundary layers (SBL), and thus they tend to have positive T2m-bias over the Arctic sea ice (e.g. Wang et al., 2019). However, when the ice- and snow-covered surfaces are replaced by open sea and land surfaces due to climate warming, the SBL-related bias decreases and thus the warming rate may in fact be underestimated. Thus, we see no reason or evidence why the warming trends would be overestimated; rather, it could be the opposite. For example, Davy and Esau (2014) show that CMIP5 models have generally negative bias in temperature trends in SBL situations (their Fig. 3b).

Furthermore, we would like to stress that we used three different in-situ datasets (Gistemp, Berkeley Earth, HadCRUT5) and one reanalysis dataset (ERA5). These four state-of-the-art datasets are used e.g. in the latest IPCC AR6 report for climate change monitoring. When looking at Fig. 1 of this letter (below), we see that the in-situ datasets agree well with each other and also with ERA5 in the Arctic (see also Fig. S3a of the manuscript for the trends).

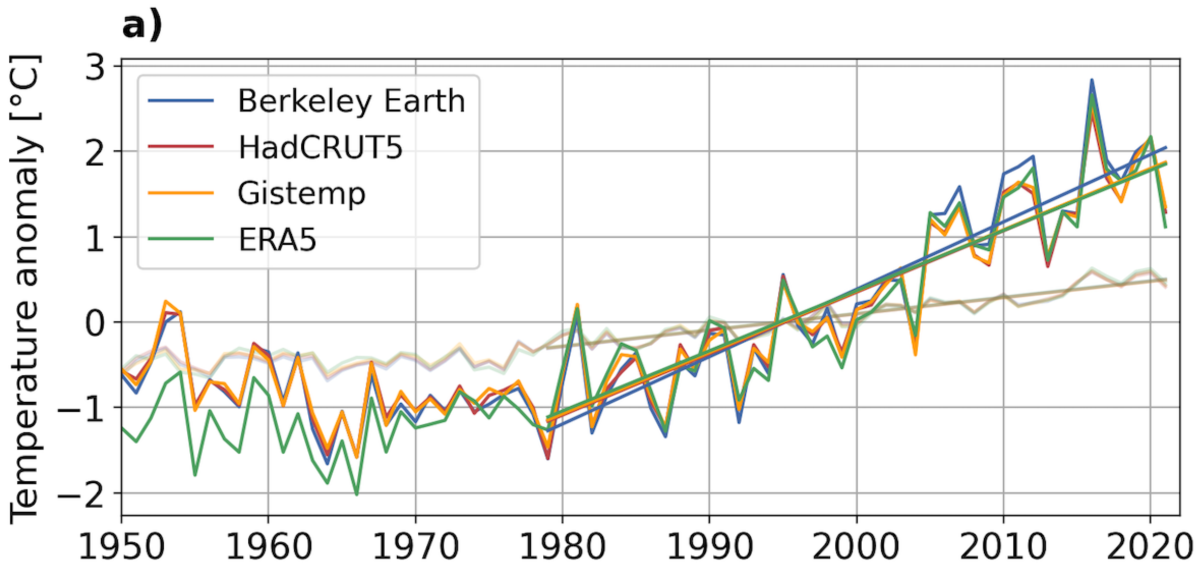

Fig. 1. Annual mean temperature anomalies in the Arctic (66.5° -90°N) (dark colors) and globally (light colors) during 1950-2021 derived from the various observational datasets. Temperature anomalies have been calculated relative to the 1981-2010 period. Shown are also the linear temperature trends for 1979-2021. The figure is identical with Fig. 1a in the revised manuscript.

Despite the good agreement of the area-averaged annual mean temperatures in the four data sets, we agree that there may be considerable uncertainty in the observational estimates of the Arctic warming trend. We also agree that this aspect was not adequately addressed in the original manuscript. Therefore, we added a new paragraph to the Discussion section in which we acknowledge this limitation of our study.

Wang et al. (2019): https://tc.copernicus.org/articles/13/1661/2019/
Davy & Esau (2014): https://doi.org/10.1088/1748-9326/9/11/114024

2. Even if the AA is underestimated by the models, it could be caused by the low global-warming rates in the recent decades, rather than due to underestimated Arctic warming. A quick comparison of Fig. 1a with model-simulated Arctic SAT anomaly series would help answer this question.

We agree. We compare the observed and simulated warming rates of 1979-2021 in the Arctic and globally in the supplementary figures S4 and S8. They illustrate that the underestimation of AA by climate models is due to both lower Arctic warming trend and higher global warming trend than in the observations. We discuss this aspect in the Discussion section at lines 353-368.

3. The estimated Arctic warming amount and thus the AA strongly depend on the time period, season, and location. For a short time period like 40 years, their estimates will contain large errors. Considering such error bars may improve the consistency between observations and model simulations.

This is true. We indeed carefully calculated the error bars of the observational annual AA estimates using the bootstrapping method (see Section 3 in the supplementary material). We also added a new figure to the revised manuscript (Fig. 2) which compares the observed AA using various time windows and area definitions for the Arctic. Furthermore, the seasonality of AA is now explicitly discussed in the revised manuscript (lines 266-286), along with a new Figure (Fig. 5 of the revised manuscript).

Specific comments:

1. L18: “line” “in line”.

Thank you for pointing out this typo.

2. L25: “mowadays” “nowadays”

Thank you for pointing out this typo.

3. L42-45: Fig. 1 of Dai et al. (2019) is highly relevant to this study: It shows that the AA in ERA-I ranges from ~1.6 for July-August to ~9.0 for January during 1979-2016 and the CMIP5 models show much lower AA during this period. It’s well known that the AA is mainly a cold-season phenomenon. Thus, it’s better to discuss the AA explicitly for specific seasons or annual mean.

Thanks for pointing this out. The seasonality of AA and the comparison to CMIP6 models are now explicitly discussed at the end of Section 5, including a new figure and a citation to Dai et al. (2019).

4. L49-54: I think the lack of reliable observations of surface air temperature over the Arctic region (especially over the Arctic Ocean) is the main source of uncertainty in our estimates of Arctic warming amount.

We agree. As stated already in the 1st response of this letter, we acknowledge the uncertainty related to the low number of observations in the Discussion section, at lines 353-368.

5. L71-72: Given that up to 50% of recent sea-ice loss may come from realization-dependent internal variability, it will be difficult for models to reproduce the

observed sea-ice loss and the related warming. The effect of the realization-dependent internal variability should be mentioned here.

Please note that the reason for using large ensembles in our study is exactly to take into account the contribution of internal variability, as well as forced component, to the trends. While any single model realization is not expected to reproduce observed internal variability, comparison of the large ensembles with observations allows us to estimate how unusual the observations are with respect to model simulations. We rephrased the sentence at line 69 to: “Because the sea ice loss is one of the main mechanisms causing AA, and given that up to 50 % of the recent loss may be due to realization-dependent internal variability, a relevant follow-up question is whether the climate models are able to reproduce the magnitude of the observed AA over the past 40 years or so.”

6. L81-82: The spatial interpolation does not mean they will produce better estimates of Arctic warming if there are no observations over most of the Arctic region. Numerically, one can fill in any data gaps but that does not mean the result will be reliable. Need to examine their actual in-situ data coverage in these datasets, not their interpolated data coverage. The 1200km interpolation radius may be suitable for low latitudes, but it may be inappropriate for the Arctic, as it will essentially fill in the data gaps over the Arctic Ocean by high-latitude land station data, and we know that near-surface air temperature changes over the ice-covered Arctic Ocean are very different from that over high-latitude land. Thus, the GISTEMP, BEST and CW likely reflect mainly the warming rates over the high-latitude land areas, not really the warming over the Arctic Ocean, which has few observations to be used in these datasets to my knowledge. This issue needs to be clearly addressed and emphasized in this study.

It is true that in the three in-situ datasets (Gistemp, Berkeley Earth, HadCRUT5), the temperatures over the ice-covered Arctic ocean are based on terrestrial stations. Thus, the temperatures in winter are mostly extrapolated from Greenland, Canada, Alaska, Scandinavia and Russia. However, in an ice-free ocean, the temperatures are extrapolated from the nearest SST measurements (which indeed may be scarce in the Arctic).

As seen from Fig. 2 of this response (below), all datasets clearly show that the warming in 1979-2021 has been the strongest over the Arctic Ocean. Thus, in this sense we disagree with the reviewer that these datasets would reflect mainly the warming rates over the high-latitude land areas. The warming maximum is located in the Barents Sea area, which is spatially collocated over the areas with the largest sea-ice loss.

In the Discussion section, we added a new paragraph in which we explicitly discuss the uncertainties of the observational datasets in the Arctic.

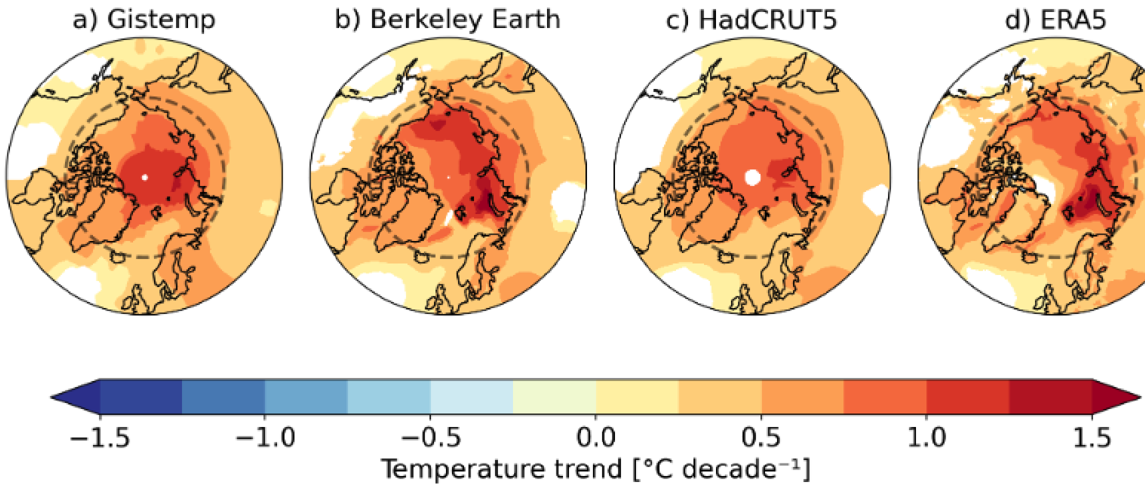

Fig. 2. Annual mean temperature trends for the period 1979-2021 derived from a) NASA GISTEMP v4, b) Berkeley Earth, c) HadCRUT5, and d) ERA5 reanalysis. Dashed line indicates the Arctic Circle (66.5°N latitude). Areas without a statistically significant change (determined using a two-sided Wald t-test, with P < 0.05 indicating a significant difference) are masked out. The figure is identical with Fig. S1 in the revised manuscript.

7. L87: Need to state the implications of this statement. For example, are these datasets representative of the true Arctic warming rates? Fig. S2 seems suggest that these datasets used some marine observations over the Arctic Ocean to derive stronger warming over the Arctic Ocean than over the high-latitude land areas. A more detailed discussion is needed on what in-situ observations from the Arctic region were used in the global datasets.

The authors can use the ERA5 and CMIP6 model data to verify whether the high latitude land areas and the Arctic Ocean have similar warming rates. If yes, then the three global datasets may be useful for quantifying the Arctic warming. If no, then they can’t be used to quantify warming over the Arctic Ocean, unless they did use observations over the Arctic Ocean in their analyses (Fig. S2 seems to suggest that is the case). To some extent, it’s already been shown in previous studies (e.g., Fig. 2 of Dai eta l. 2019).

We are not sure if we understood this comment correctly. The high latitude land areas and the Arctic Ocean do not have similar warming rates (Fig. 2 of this letter). This is mostly due to sea ice loss: these areas were mostly ice-covered in the 1980s or 1990s, but now they are open, at least for a longer period of the year. Since the sea surface temperature does not fall below the freezing point of the water and the surface air temperature on top of the sea ice can be much colder, the melting of the sea ice is causing a strong warming trend in these areas.

In the revised manuscript, we removed the sentence “Thus, temperatures in both terrestrial and marine Arctic are mostly based on terrestrial station observations.” because it can be misleading in the sense that terrestrial station observations are not used over the open ocean. We also mention that the observational datasets use sea ice dataset (HadISST2, Titchner and Rayner, 2014) to determine the presence or absence of sea ice in the grid cell.

Titchner and Rayner (2014): https://doi.org/10.1002/2013JD020316

8. eq. 1: As AA refers to the enhanced Arctic warming relative to the global-mean warming rate; thus, AA can be defined as the difference or ratio of the Arctic to global-mean warming amount or warming rate. In eq. 1, the time interval should be the same, so that AA defined using the warming amount (dT_Arctic/dT_global) or warming rate should be the same. Clearly, a shorter time period will increase the contribution of realization-dependent internal variability on the warming amount, thus AA will depend strongly on the time interval used, as shown in Fig. S1.

We agree. To communicate more clearly the effect of the time window to the magnitude of AA, Fig. S1 from the supplementary material was moved to the main text (Fig. 2a of the revised manuscript).

9. L135: The dT/dtA is likely to be sensitive to the southern latitude as this determines how much land area is included in the calculation, as shown by Fig. S1. This point should be pointed out here. Please state the dataset used in Fig. S1.

The effect of the southern boundary is explicitly discussed at line 215 onwards of the revised manuscript. In addition, the dataset used in Fig. 2a (Fig. S1 of the original manuscript) is the average across the four observational datasets, and is indicated now in the caption of Fig 2a in the revised manuscript.

10. L147: How were the 31 different 40-year periods in 1971-2040 generated? Do they contain some overlapping and thus some correlation among them (i.e., not totally independent)?

In the revised manuscript, we do the analysis using 43-year periods in 1970-2040. These 43-year periods are overlapping with each other and were defined as moving 43-year periods, e.g. 1970-2012, 1971-2013, ..., 1998-2040. This is now stated more clearly in the revised manuscript at line 148. Please also note that we have 29 different periods for each individual model realization, so that each histogram in Fig. 6 contains more than a thousand of samples.

11. L172 and L180-184: Fig. S2 shows large regional differences in the warming trends among the 4 datasets, despite the general agreement in the Arctic-mean time series. Some information regarding what observational data over the Arctic Ocean

were used by each of the four datasets would be helpful in explaining their warming differences over the Arctic Ocean.

We agree that there are regional differences in the warming trends which may be due to several reasons, such as differences in the background sea ice dataset or differences in the resolution of the datasets. For instance ERA5 has a peculiar lack of warming, or cooling trend north of Greenland (Simmons et al., 2021, p. 63)

In the revised manuscript, we emphasize the uncertainties in the warming trends, and briefly mention the inconsistent cooling trend in ERA5 north of Greenland, which is not visible in other datasets.

Simmons et al. (2021): https://www.ecmwf.int/node/19911

12. L204-205: I think that AA is around 2 was derived based on CMIP model projected warming rates, not based on recent warming rates that include effects of internal variability unrelated to GHG-induced warming. As shown previously (e.g., Dai et al. 2019), the AA is mainly a cold-season phenomenon. Thus, it would be helpful to show the AA as function of month, similar to Fig. 1 of Dai et al. (2019).

Thank you for this suggestion. The AA as a function of month has been added to Fig. 5 in the revised manuscript. The discussion on these results are provided at lines 266-268.

13. Fig. 2: What does the x-axis represent in this figure? Is the middle or ending year of the 40-year period used to compute the AA? I found some of the figure captions lack the key information needed to understand the results. It seems that you used all the years in the moving 40-yr window (this needs to be described in Method section).

The x-axis represents the ending year of the 43-year period. We state this more explicitly in the revised manuscript. We also went through other figure captions to improve their readability.

14. L227-228: Fig. 2a only contains a relatively small sample runs, so it is possible that it may miss some possibilities, such as the observed realization.

Yes, we agree. This is stated in the manuscript at line XX. We would also like to emphasize that we did a test with CMIP6 in which we used only one realization per model (line 298 in the revised manuscript). This did not significantly change the key results.

15. L237: A 40-year period is short for quantifying externally-forced trends in individual realizations.

We agree, and this is what we aim to express in the manuscript at line 246: “The large spread in CMIP6-simulated AA is in line with an earlier study (Ye and Messori, 2021) and highlights the effect of large internal variability for AA, even on a 43-year time scale.”

16. L253-254: Another possibility is that the four datasets overestimate the recent Arctic warming due to lack of observations of 2m air temperatures over the Arctic region (especially over the winter polar ice cap, where the 2m air temperature can be much colder than nearby SST (a few SST observations may be used in the observational datasets).

We would like to emphasize that SST observations are not used over the winter polar ice cap. Instead, those ice-covered regions are extrapolated from nearby land stations from Greenland, Canada, Alaska, Scandinavia and Russia. The datasets use satellite-derived information about sea ice coverage (via HadISST2 dataset). SST observations are used and extrapolated only for regions which are ice-free.

It is also important to recognize that while T2m temperatures in ERA5 may be warm-biased, the long-term trends may not be. In fact, we have not seen any evidence that the trends would be overestimated in the Arctic region, but rather it could be the opposite, as explained in the 1st response of this letter.

17. L266-267: In Monto-Carlo simulations, tens of thousands of simulations were often used to generate the PDF for estimating the significance. Here, only 100-member runs were used and the 40-year periods from these runs were overlapped (thus not independent). Given this, I would caution the authors to make very strong statement on the likelihood of the recent warming rates, although I agree it seems very unlikely based on the PDFs from the limited number of model runs, especially for the MPI-GE that has a low ECS.

We changed this statement to “MPI-GE does not capture the observed Arctic amplification…” and we believe that this statement is correct because none of the 100 members of the MPI-GE ensemble simulated as strong AA as observed in 1979-2021 even when the longer time gap (1970-2040) was allowed in the model simulations.

18. L278-279: Again, I think the authors need to consider the real possibility that the available datasets may overestimate Arctic warming since 1980 due to the lack of 2m air temperature observations over winter sea ice, rather than taking the face values of the 4 datasets as the truth and conclude that the models underestimate Arctic warming.

We agree that the observational estimates of Arctic warming have error bars, but we have not seen any evidence that they would systematically overestimate the Arctic warming trends. As a matter of fact, it may be rather the opposite, as explained in the 1st response of this document.

We stress the uncertainties more clearly in the Discussion section, acknowledging the possibility that the available datasets may have biases in the warming trends.

19: L294-302: Many studies showed that the recent global warming rates were significantly affected by the phase changes in IPO in the Pacific and AMO in the Atlantic. Because the models have difficulties in simulating the IPO and AMO in the first place, and the recent IPO/AMO phase combinations are highly realization dependent (i.e., it may take hundreds to thousands of realizations to capture it), I think it would be better to compare the observed and model-simulated Arctic warming rate or amount, rather than the Arctic/global dT ratio). This would avoid the effect of the errors in simulating the recent global warming rates. Also, estimates of trends in 40-year series are highly uncertain statistically (e.g., the end points may have a significant effect, etc.). This could be another source of the uncertainties in the comparison.

We indeed calculated the actual global and Arctic warming rates in the observations and model simulations. However, as the actual warming trends or their evaluation were not the main scope of our study, these figures are placed in the supplementary material. The underestimation of AA by climate models is generally due to both underestimating the Arctic warming and overestimating the global warming (Fig. S8 of the revised manuscript).

We also considered the error bars of the linear trends by considering statistical uncertainty. For error bars in the observed AA, we used a more sophisticated bootstrapping method (see Section 3 in the supplementary material).

Reviewer #3

Rantanen et al. (2022) discussed that the intensified Arctic warming in the past 40 years is much larger than was conventionally reported. In general, I found this study interesting and well-written. The results are also important, especially the implications for the potential underestimation of AA in state-of-the-art climate models. I support to publish this paper with some wording checks. Below are my minor comments:

To assess the internal variability, there is an observational large-ensemble datasets (https://www.cesm.ucar.edu/projects/community-projects/MMLEA/, McKinnon and Deser, 2018). It would be informative if the authors can estimate the internal variability of observations and compare it to those from SMILEs and spreads of CMIP5/6. The observational estimate could be added in Figures 3 and 4.

Thank you for these comments. The suggestion of using the above-mentioned observational dataset is potentially interesting. However, after reviewing the dataset and its reference paper (McKinnon and Deser, 2018), it appears that the dataset is unfortunately not suitable for our needs, mostly because a) the dataset is land-only and b) available only for 1921-2014. Comparing the internal variability from the Arctic land areas and only up to 2014 is not ideal given that the other datasets extend to present and cover both land and sea areas.

We acknowledge the uncertainty in the observations more clearly in the revised manuscript (lines 353-358).

In the Discussion and conclusions section, the authors mentioned the effect of hiatus, which makes the global warming trend smaller and producing larger AA. I would like to know more about the authors' opinion on AA. Should we view AA as a global climate indicator or just a local one? We could get very large AA, assuming Arctic warming does not change but with very small global warming. And how should we interpret negative AA? If Arctic cools more than the globe does, one can still get positive AA.

Thank you for this comment. Due to the definition of AA (a division of two trends), its calculation is sometimes problematic. This is especially true for periods characterized by weak global warming, when the denominator of the division approaches zero and AA can explode very high. However, physically the definition of AA only makes sense if there is a global warming on which Arctic warming is superimposed. For this reason, we made an attempt to constrain modeled AA values by neglecting those time periods when global temperature trend was not statistically significant. This arguably removed the most extreme AA values. However, there may still be better options.

In any case, the definition of AA as such, without any constraint, does not work for periods when global warming turns to cooling. For this reason, we believe that using AA as a global climate indicator may be problematic.

Another aspect is the internal variability. As found in our study, the internal variability may have contributed considerably to the magnitude of AA. Hence, if one wants to quantify AA of anthropogenic climate warming, it should be calculated over a reasonably long time period.
Negative AA characterizes the time periods when the Arctic cools but the globe as a whole warms. This is a possible scenario given the large multi-decadal variability in the Arctic. In situations when the Arctic cools more than the globe does, it is true that one can still get positive AA. However, we think that these situations are not realistic at least on a multi-decadal scale unless climate change is heavily mitigated.

Reference:

McKinnon, K.A. and Deser, C., 2018. Internal variability and regional climate trends in an observational large ensemble. Journal of Climate, 31(17), pp.6783-6802.

21st Apr 22

Dear Dr Rantanen,

Your revised manuscript titled "The Arctic has warmed nearly four times faster than the globe since 1979" has now been seen by the original three reviewers, and I include their comments at the end of this message. They continue to find your work of interest, but a few important points are still raised.

We therefore invite you to revise and resubmit your manuscript once again, along with a point-by-point response that takes into account the remaining issues. Specifically, we will need you to:

1) appropriately and prominently discuss in your analysis and interpretation sections the influence of the start date of your data analysis on your conclusions

2) either make a compelling case that land warming contamination is negligible for your data set or discuss the associated caveats and uncertainties transparently

Please highlight all changes in the manuscript text file.

We are committed to providing a fair and constructive peer-review process. Please don't hesitate to contact us if you wish to discuss the revision in more detail.

Please use the following link to submit your revised manuscript, point-by-point response to the referees’ comments (which should be in a separate document to any cover letter) and the completed checklist:
[link redacted]
** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to receive your revised paper within six weeks; please let us know if you aren’t able to submit it within this time so that we can discuss how best to proceed. If we don’t hear from you, and the revision process takes significantly longer, we may close your file. In this event, we will still be happy to reconsider your paper at a later date, as long as nothing similar has been accepted for publication at Communications Earth & Environment or published elsewhere in the meantime.

We understand that due to the current global situation, the time required for revision may be longer than usual. We would appreciate it if you could keep us informed about an estimated timescale for resubmission, to facilitate our planning. Of course, if you are unable to estimate, we are happy to accommodate necessary extensions nevertheless.

Please do not hesitate to contact me if you have any questions or would like to discuss these revisions further. We look forward to seeing the revised manuscript and thank you for the opportunity to review your work.

Best regards,

Viviane V. Menezes, PhD
Editorial Board Member
Communications Earth & Environment
orcid.org/0000-0002-4885-2056

Heike Langenberg, PhD
Chief Editor
Communications Earth & Environment

EDITORIAL POLICIES AND FORMATTING

We ask that you ensure your manuscript complies with our editorial policies. Please ensure that the following formatting requirements are met, and any checklist relevant to your research is completed and uploaded as a Related Manuscript file type with the revised article.

Editorial Policy: Policy requirements

Furthermore, please align your manuscript with our format requirements, which are summarized on the following checklist:
Communications Earth & Environment formatting checklist

and also in our style and formatting guide Communications Earth & Environment formatting guide .

*** DATA: Communications Earth & Environment endorses the principles of the Enabling FAIR data project (http://www.copdess.org/enabling-fair-data-project/). We ask authors to make the data that support their conclusions available in permanent, publically accessible data repositories. (Please contact the editor if you are unable to make your data available).

All Communications Earth & Environment manuscripts must include a section titled "Data Availability" at the end of the Methods section or main text (if no Methods). More information on this policy, is available at http://www.nature.com/authors/policies/data/data-availability-statements-data-citations.pdf.

In particular, the Data availability statement should include:
- Unique identifiers (such as DOIs and hyperlinks for datasets in public repositories)
- Accession codes where appropriate
- If applicable, a statement regarding data available with restrictions
- If a dataset has a Digital Object Identifier (DOI) as its unique identifier, we strongly encourage including this in the Reference list and citing the dataset in the Data Availability Statement.

DATA SOURCES: All new data associated with the paper should be placed in a persistent repository where they can be freely and enduringly accessed. We recommend submitting the data to discipline-specific, community-recognized repositories, where possible and a list of recommended repositories is provided at http://www.nature.com/sdata/policies/repositories.

If a community resource is unavailable, data can be submitted to generalist repositories such as figshare or Dryad Digital Repository. Please provide a unique identifier for the data (for example a DOI or a permanent URL) in the data availability statement, if possible. If the repository does not provide identifiers, we encourage authors to supply the search terms that will return the data. For data that have been obtained from publically available sources, please provide a URL and the specific data product name in the data availability statement. Data with a DOI should be further cited in the methods reference section.

Please refer to our data policies at http://www.nature.com/authors/policies/availability.html.

REVIEWER COMMENTS:

Reviewer #1 (Remarks to the Author):

The authors have addressed most of my comments and have improved the manuscript. My second major comment about the redundancy and lack of clear message from some of the figures has been adequately addressed. I like the new figure that shows the seasonal cycle of Arctic amplification. My first major comment about the overestimation of the discrepancy between models and observations because of cherry-picking has been partially addressed. The authors now focus on the last 43 years instead of a cherry-picked 40 year period and they have included some analysis showing the sensitivity to start year. However the results of this analysis are not reflected in the interpretation and conclusions.

I still think the authors are likely overestimating the discrepancy between models and observations described in section 6 by only choosing 1979 as the start date. Figure 4 shows that choosing the start date around this time leads to the largest discrepancy compared to any other start date. Because of this, internal variability likely contributes to the magnitude trend. If the authors chose 1950 as the start date, the magnitude of the observed Arctic amplification is very similar to the ensemble means from the models and the probabilities and conclusions in this section would be very different. Also, this longer trend is more likely to reflect the forced response compared to the 43 year trend that the authors chose to use.

The justification for the choice of focusing on 1979 is not well reasoned. While there is more uncertainty in the observations when going back further in time, the different datasets show similar Arctic warming prior to 1979. I also do not understand the argument stated on L129-131, about how the definition of AA only makes sense if there is global warming. This applies if there is actually no global warming (i.e. the denominator in eq. 1 is 0), but there clearly is global warming if trends are started before 1979, so I don't see the relevance. The authors even state that the period of sustained warming started in about 1970 (L150), so why not pick 1970?

At the very minimum, the authors need to acknowledge that the probabilities/p-values and conclusions in section 6 are strongly dependent on the choice of start year, and that it was not chosen a priori. They also need to make it clear that the longer trends (which are more likely to reflect a forced response) show closer agreement/higher p-values. Some of this is evident from the previous figures (Fig2b, Fig 4), but it is neglected in the analysis and interpretation in section 6.

Other comments:

L19/302: I think that 'extremely unlikely event' is a bit of an overstatement for the reasons described above. When considering all possible start years, it probably not that unlikely that some of the trends will appear as unlikely as they do here.

L108-109: Is this really why they were chosen? From the previous responses, it seems like these two models were the only two large ensembles available with this emissions scenario. I also do not see why climate sensitivity would be most relevant quantity to consider when picking the models.

L316-325: There are many other differences between these models, so I don't see how the different behaviour in CanESM5 can be attributed to the higher climate sensitivity based on only the analysis performed here.

Reviewer #2 (Remarks to the Author):

I thank the authors for their efforts to address my concerns. I'm generally satisfied with their responses, although a few comments still come to my mind and they are listed below.

1. The good agreement among the 3 global datasets does not necessarily demonstrate that they are reliable. In fact, they may contain similar biases due to their use of the similar data sources. For example, if they interpolated land warming onto the polar ice cap, then they could have the same warming biases if the ice cap and land areas had different warming rates since 1980. I think this is a major issue that the authors need to investigate a bit more. For example, as I suggested previously, they could use the ERA5 or CMIP6 data to investigate whether the warming rates in 2m air temperature is comparable over high-latitude land areas (with station data) and over the polar ice cap (which seem to be different based on their maps of trends, but a more detailed analysis with high-latitude station locations and interpolated ice cap warming may help). If not, then such an interpolation using land observations onto ice cap is problematic and could lead to biases in Arctic mean warming rates.

2. Another issue is related to the use of SST observations for quantifying warming rates in 2m air temperature over the open water surfaces. The SST and T2m over the Arctic, especially in the cold season, are very different and their trends may also differ. This issue also need to be investigated, e.g., using the ERA5 or CMIP6 SST and T2m data over the open water surfaces.

Reviewer #3 (Remarks to the Author):

Please find attached pdf file for my quick comments.

I thank the authors for the revision and answering my questions. I do not have further comments. However, the observational large ensemble indeed provides the global temperature fields (at least on NCAR casper repository). I quickly plotted out a snapshot for the DJF mean surface air temperature profile in Figure R1. There are four separate files for DJF, MAM, JJA, and SON means, one can sum them up and take average to retrieve the annual mean values. Could the authors check the dataset again?

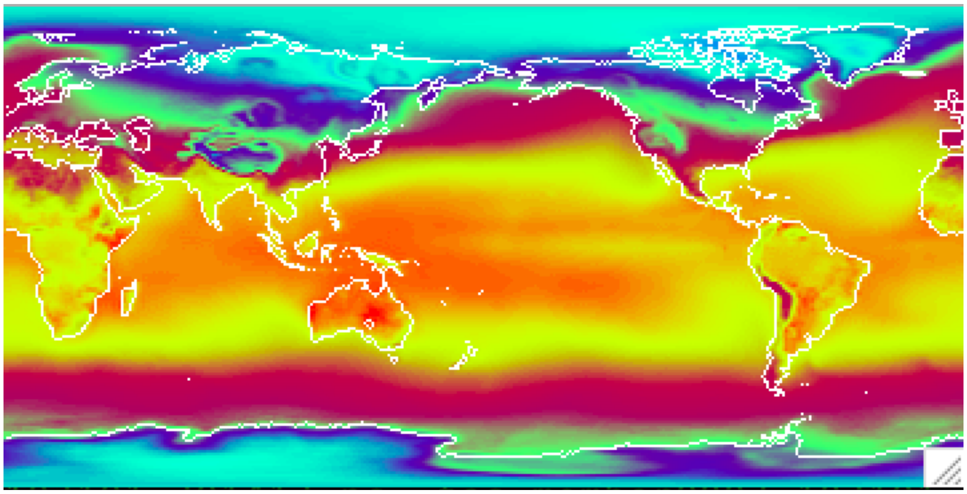

Figure R1.

The responses to reviewers of “The Arctic has warmed nearly four times faster than the globe since 1979”

Mika Rantanen, Alexey Yu. Karpechko, Antti Lipponen, Kalle Nordling, Otto Hyvärinen, Kimmo Ruosteenoja, Timo Vihma and Ari Laaksonen

We are glad that our major revision has largely satisfied you. We thank you for your additional comments and hope that our responses will clarify the rest of unclear matters.

In addition to the revision we made according to your comments, we noticed an error in Figure 4a: the CMIP5-simulated AA values were mistakenly calculated using RCP8.5 scenario, while they should have used RCP4.5 scenario. In the revised manuscript, a correct figure is used. This did not change the conclusions of the figure.

The point-by-point replies to your comments are below, marked in black and our responses in blue.

Reviewer #1 (Remarks to the Author):

The authors have addressed most of my comments and have improved the manuscript. My second major comment about the redundancy and lack of clear message from some of the figures has been adequately addressed. I like the new figure that shows the seasonal cycle of Arctic amplification. My first major comment about the overestimation of the discrepancy between models and observations because of cherry-picking has been partially addressed. The authors now focus on the last 43 years instead of a cherry-picked 40 year period and they have included some analysis showing the sensitivity to start year. However the results of this analysis are not reflected in the interpretation and conclusions.

I still think the authors are likely overestimating the discrepancy between models and observations described in section 6 by only choosing 1979 as the start date. Figure 4 shows that choosing the start date around this time leads to the largest discrepancy compared to any other start date. Because of this, internal variability likely contributes to the magnitude trend. If the authors chose 1950 as the start date, the magnitude of the observed Arctic amplification is very similar to the ensemble means from the models and the probabilities and conclusions in this section would be very different. Also, this longer trend is more likely to reflect the forced response compared to the 43 year trend that the authors chose to use.

We are glad that our revision has mostly satisfied you. We agree that the analysis showing the sensitivity to the starting year of the trends was not reflected adequately in Section 6. In the revised manuscript, we communicate more clearly that the model-data discrepancies are sensitive to the starting year of the trends. We made small changes to the Abstract, Section 5, 6, and 7.

The justification for the choice of focusing on 1979 is not well reasoned. While there is more uncertainty in the observations when going back further in time, the different datasets show similar Arctic warming prior to 1979. I also do not understand the argument stated on L129-131, about how the definition of AA only makes sense if there is global warming. This applies if there is actually no global warming (i.e. the denominator in eq. 1 is 0), but there clearly is global warming if trends are started before 1979, so I don't see the relevance. The authors even state that the period of sustained warming started in about 1970 (L150), so why not pick 1970?

Thank you for this comment. We respectfully disagree on the similarity of the different datasets prior to 1979. As seen from Fig. RL1 below, ERA5 diverges from the three other in-situ datasets when going back from 1979. Especially in the pre-1970 period, ERA5 is clearly colder (down to -0.75°C) than the other three datasets. In the revised manuscript, we state more clearly that the reason why we focus on the last 43 years is due to (i) linearity of the Arctic warming since 1979, (ii) less uncertainty due to availability of remote-sensing observations and (iii) the divergent behavior of ERA5 when going back from 1979 (Section 3.3).

One can also note that during 1950-1979, the temperatures in the Arctic were slightly cooling (except in ERA5, Fig. RL1). Thus, the linear trends over 1950-2021 do not correctly reflect the dynamics of the Arctic warming due to non-linear behavior of the temperatures during this period.

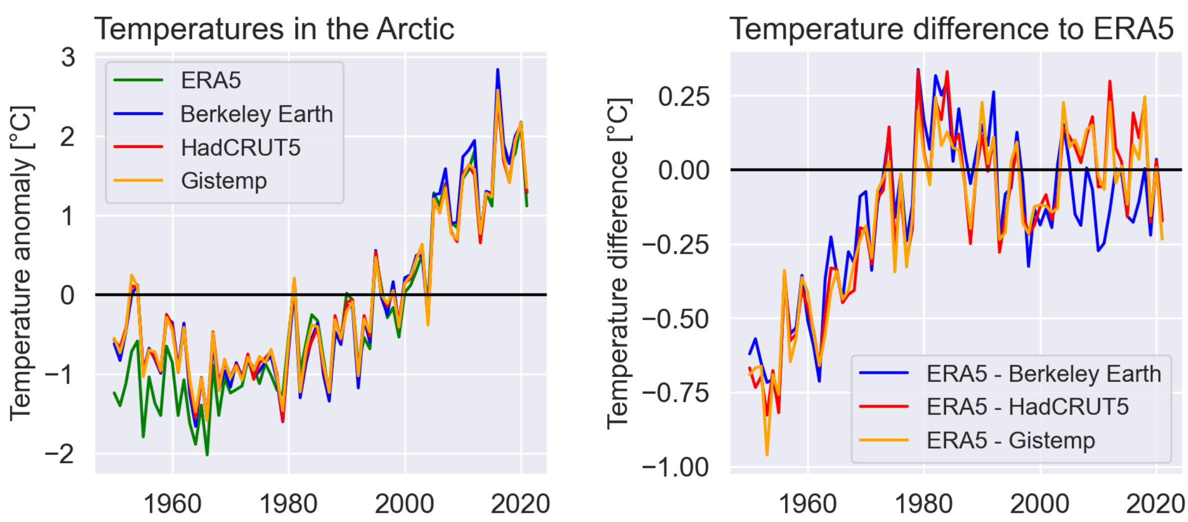

Figure RL1. Annual mean temperatures in the Arctic (>66.5°N, left) and their differences to ERA5 (right).

In addition to the above-mentioned difference in the datasets, there are reasons to focus more on the recent four decades or so. There is little observed global warming from 1950s to 1970s due to cooling effects of atmospheric aerosols (e.g. Hegerl et al. 2019). So, while choosing 1950 as a start date would indeed bring observed and simulated AA closer to each other, it would not necessarily make a better comparison because of larger uncertainty in observed temperatures before 1979 and non-linear behavior of temperatures since 1950. The reviewer is right that internal variability has contributed to the observed AA but we compare observations with 380 individual model realizations so that the role of internal variability is taken into consideration.

The argument on the definition of AA making physical sense only during global warming is primarily related to short trends (< 30 years). In these time scales the simulated global temperature trends may be close to zero due to internal variability, and thus calculating AA ratio for these time periods does not make physical sense. We agree that the constraint has no relevance when the trends are calculated from 1979 (or earlier) because on longer time scales the global warming trend is almost always positive.

Hegerl et al. (2019): https://iopscience.iop.org/article/10.1088/1748-9326/ab4557

At the very minimum, the authors need to acknowledge that the probabilities/p-values and conclusions in section 6 are strongly dependent on the choice of start year, and that it was not chosen a priori. They also need to make it clear that the longer trends (which are more likely to reflect a forced response) show closer agreement/higher p-values. Some of this is evident from the previous figures (Fig2b, Fig 4), but it is neglected in the analysis and interpretation in section 6.

We agree. In the revised manuscript, we acknowledge the sensitivity of the probabilities to the starting year of the trends in the end of Section 6. We also mention that the model-data discrepancy is smaller when considering londer trends, although longer trends do not well represent the observed temperature evolution due to the non-linearities in the warming.

Other comments:

L19/302: I think that ‘extremely unlikely event’ is a bit of an overstatement for the reasons described above. When considering all possible start years, it probably not that unlikely that some of the trends will appear as unlikely as they do here.

We indicate now the time period 1979-2021 explicitly in the Abstract. We also added the following sentence to the Abstract: “The observed and simulated amplification ratios are more consistent with each other if calculated over a longer period; however the comparison is obscured by observational uncertainties before 1979”.

Thus, we make it clear that the observed AA ratio is extremely unlikely only when calculated over 1979-2021. Likewise, on line 302, we added the time range 1979-2021 to emphasize that the observed AA ratio is extremely unlikely only when considering the given time period.

L108-109: Is this really why they were chosen? From the previous responses, it seems like these two models were the only two large ensembles available with this emissions scenario. I also do not see why climate sensitivity would be most relevant quantity to consider when picking the models.

Yes, thanks for pointing this out. The emission scenario was indeed one of the reasons why MPI-GE and CanESM5 were chosen. In the revised manuscript, we indicate this at line 126.

L316-325: There are many other differences between these models, so I don't see how the different behaviour in CanESM5 can be attributed to the higher climate sensitivity based on only the analysis performed here.

We agree that the climate sensitivity alone does not necessarily explain the different behavior of AA43 in CanESM5 compared to the other three model ensembles (Fig. 3 of the manuscript). For example, positive model biases in snow and sea ice can cause a long-lasting sea ice feedback in the model, and thus stronger AA later in the 21st century.

We changed the words "Presumably for this reason" to "In addition" on line 359.

Reviewer #2 (Remarks to the Author):

I thank the authors for their efforts to address my concerns. I'm generally satisfied with their responses, although a few comments still come to my mind and they are listed below.

1. The good agreement among the 3 global datasets does not necessarily demonstrate that they are reliable. In fact, they may contain similar biases due to their use of the similar data sources. For example, if they interpolated land warming onto the polar ice cap, then they could have the same warming biases if the ice cap and land areas had different warming rates since 1980. I think this is a major issue that the authors need to investigate a bit more. For example, as I suggested previously, they could use the ERA5 or CMIP6 data to investigate whether the warming rates in 2m air temperature is comparable over high-latitude land areas (with station data) and over the polar ice cap (which seem to be different based on their maps of trends, but a more detailed analysis with high-latitude station locations and interpolated ice cap warming may help). If not, then such an interpolation using

land observations onto ice cap is problematic and could lead to biases in Arctic mean warming rates.

Thank you for this comment. Verifying the warming trends over the polar ice cap is indeed challenging due to missing observations. However, there are stations in the archipelago of the Arctic Ocean which are strongly influenced by the presence (or absence) of sea ice. A validation of the temperature trends was done in a recent study (Cai et al., 2021), which compared ERA5 with the station data for 1979-2014. According to the study, "ERA5 generally captures well the temporal and spatial characteristics of the near-surface mean temperature in the Arctic, which makes it suitable to use as a reference in the model evaluations".

To assess the accuracy of the four datasets applied in our study (GISTEMP, HadCRUT5, Berkeley Earth, ERA5), we conducted a similar validation against the Global Historical Climatology Network monthly (GHCN-M) stations. We selected all the stations located north of 66.5°N that had at least 39 years of data over the 43-year period of 1979-2021. In total, these criteria resulted in 87 stations (Fig. RL2 below). We calculated the warming trends of each station, and compared them with the gridded data. These results are shown below (Fig. RL2). Furthermore, the validation against the individual datasets are shown at the end of this letter.

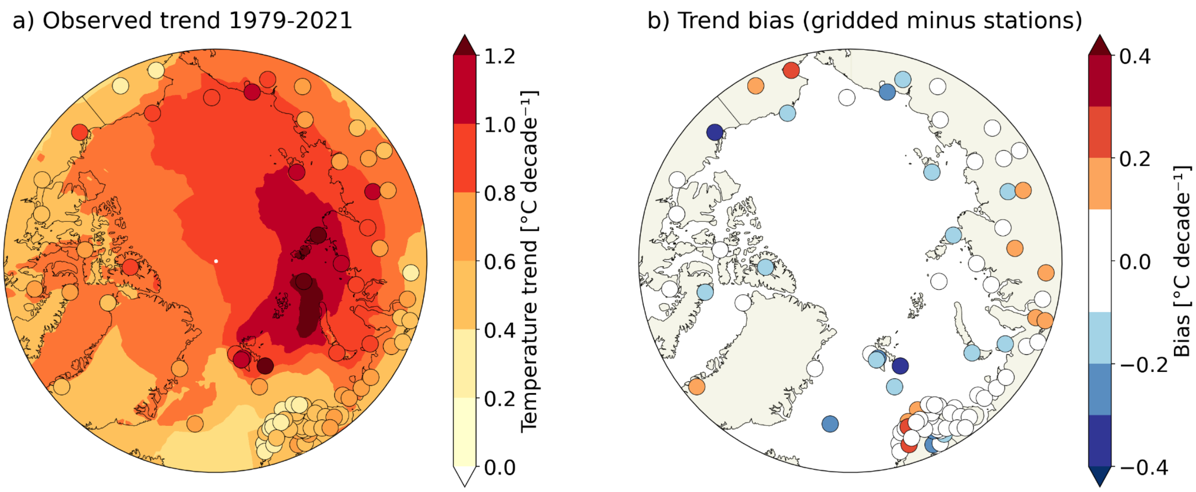

Figure RL2. The warming trend in the Arctic (> 66.5°N) over the 1979-2021 period (left), and the bias in the trend (right). The shading in a) depicts the observational average across the four datasets (Gistemp, HadCRUT5, Berkeley Earth, ERA5), and the colored circles indicate the station observations. The bias in b) is defined as gridded data minus station observations.

It can be seen that the warming rates in the stations mainly match with gridded data, and there is no systematic bias in one direction or another. If anything, some of the high Arctic stations (e.g. in Svalbard) show a slightly weaker warming trend (cold bias, Fig. RL2b) than the gridded dataset. Of course, there are no observations from the central Arctic Ocean, so the comparison is not exhaustive. However, based on our validation and the published paper (Cai et al., 2021) we argue that ERA5 and the other three observational datasets are suitable to use as a basis of our study.

In the revised manuscript, we wrote a new paragraph to Section 3.1 where we present the above-mentioned validation. The figure was added to the Supplementary material.

Cai et al., (2021):
https://journals.ametsoc.org/view/journals/clim/34/12/JCLI-D-20-0791.1.xml

2. Another issue is related to the use of SST observations for quantifying warming rates in 2m air temperature over the open water surfaces. The SST and T2m over the Arctic, especially in the cold season, are very different and their trends may also differ. This issue also need to be investigated, e.g., using the ERA5 or CMIP6 SST and T2m data over the open water surfaces.

We agree that comparing in-situ observational datasets with climate models is not really apples-to-apples comparison because the in-situ datasets report a blend of land near-surface air temperature and sea surface temperature (BST), whereas the model output is the near-surface air temperature (SAT). Cowtan et al. (2015) reported that in climate models the global warming trend calculated from BST over the 1975–2014 period is about 7 % lower than the trend derived from SAT. Therefore, the absolute warming trends in the models are likely biased high compared to observations.

However, we argue that the blending effect is not significantly affecting our results, because

1) We mainly compare the warming trend ratios between the models and observations, not the absolute warming trends.
2) We also use ERA5 which allows like-for-like comparison with the models. In our results, the absolute warming trends in ERA5 are not high compared to the three in-situ datasets, despite ERA5 using T2m instead of SST (see Fig. S3).

Nevertheless, we added a mention to the revised paper where we acknowledge the possibility of small biases in the three in-situ datasets when comparing the direct warming trends to the climate models. This disclaimer can be found at the end of Section 3.2 of the revised manuscript.

Cowtan et al. (2015):
https://agupubs.onlinelibrary.wiley.com/doi/10.1002/2015GL064888

Reviewer #3 (Remarks to the Author):

Please find attached pdf file for my quick comments.

I thank the authors for the revision and answering my questions. I do not have further comments. However, the observational large ensemble indeed provides the global temperature fields (at least on NCAR casper repository). I quickly plotted out a snapshot for the DJF mean surface air temperature profile in Figure R1. There are four separate files for DJF, MAM, JJA, and SON means, one can sum them up and take average to retrieve the annual mean values. Could the authors check the dataset again?

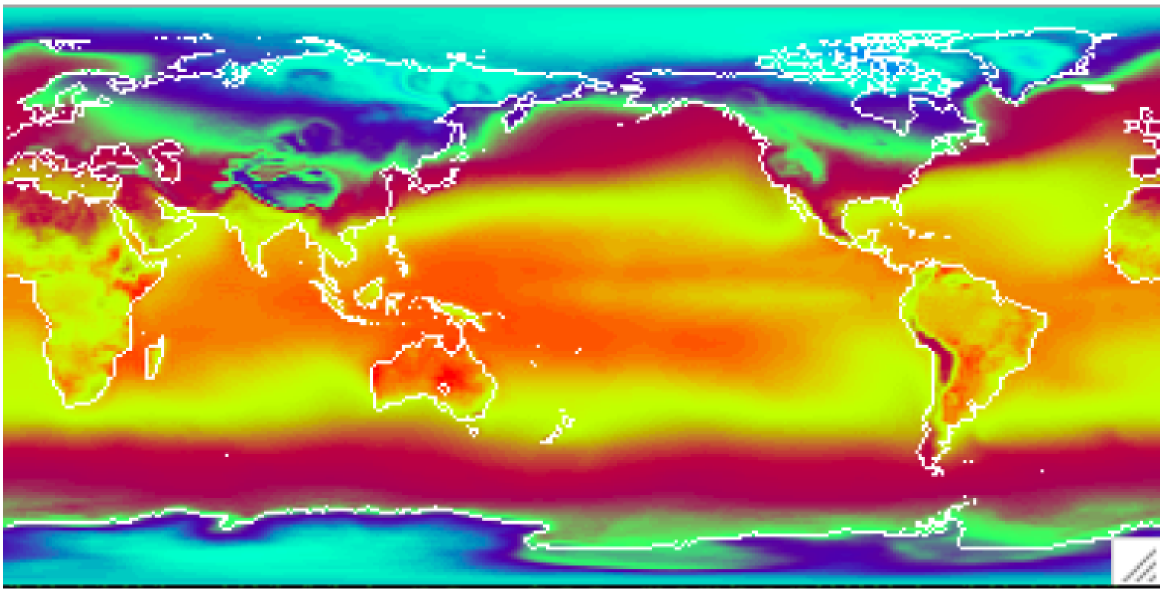

Figure R1.

Thank you for this comment. We appreciate your efforts to investigate the dataset. We checked the dataset again, and noticed the same: at least DJF fields are indeed global although the documentation mentions that the dataset is land-only. Nevertheless, the dataset does not extend to present (year 2021), which makes it incompatible to combine with our analysis. Furthermore, it would be a huge effort to download and analyze the 1000-member ensemble to the current manuscript. Therefore, we decided not to include that dataset in this analysis but we keep it in mind for potential future research.

Supplementary figures related to comments by Reviewer 2

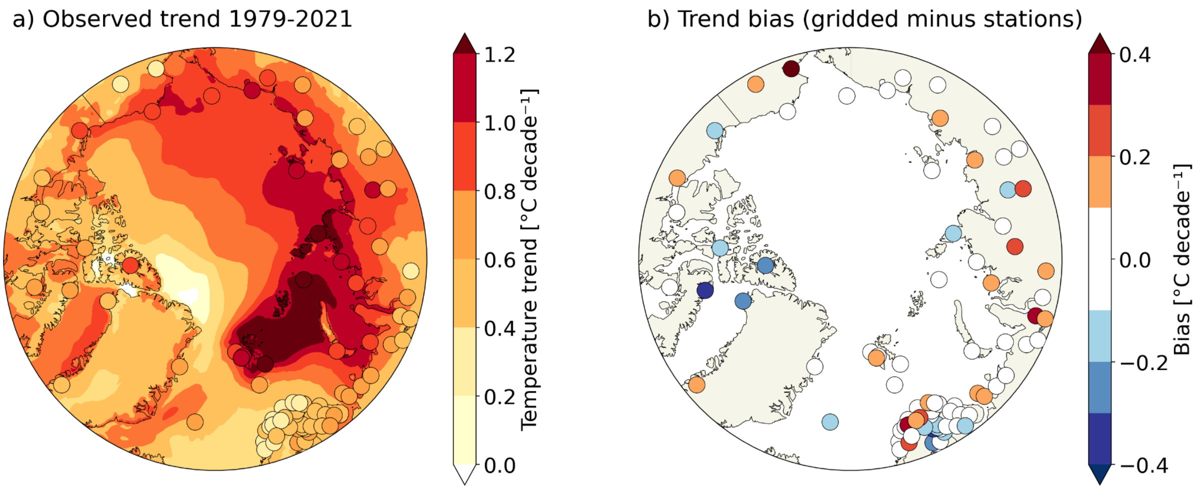

Fig. RL3. Same as Fig. RL2, but for ERA5 reanalysis.

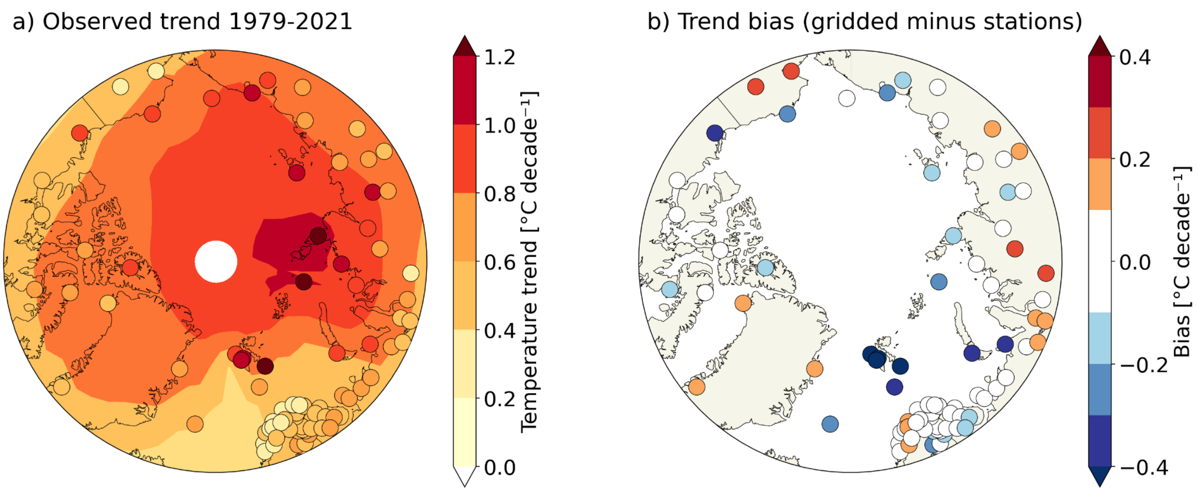

Fig. RL4. Same as Fig. RL2, but for the HadCRUT5 dataset.

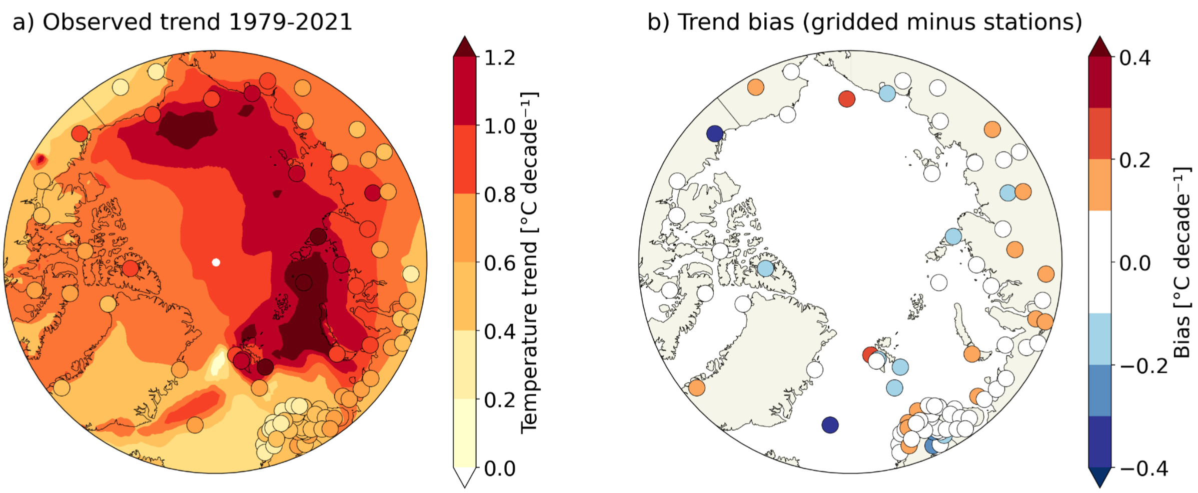

Fig. RL5. Same as Fig. RL2, but for the Berkeley Earth dataset.

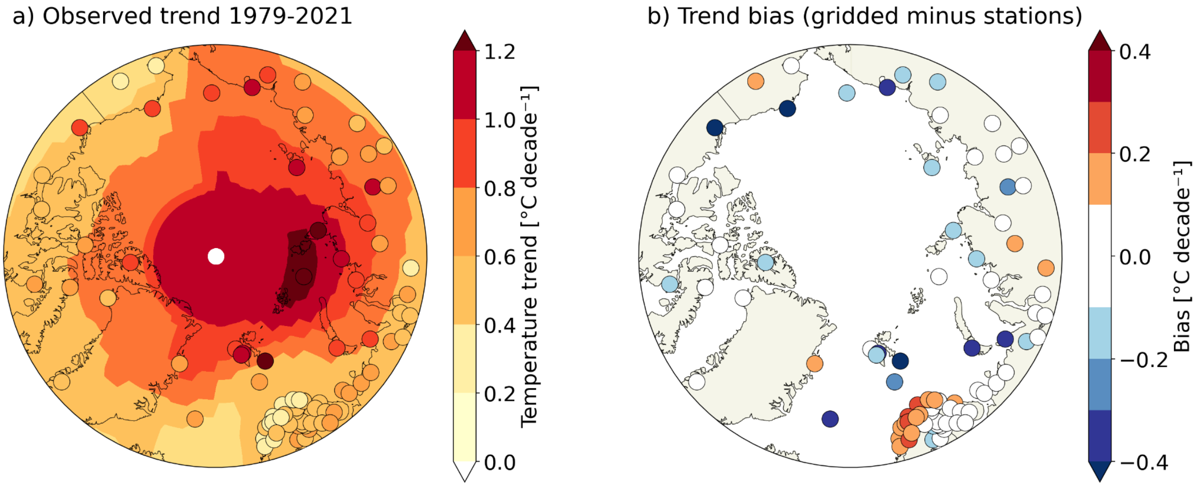

Fig. RL6. Same as Fig. RL2, but for the Gistemp dataset.

13th Jun 22

Dear Dr Rantanen,

Your manuscript titled "The Arctic has warmed nearly four times faster than the globe since 1979" has now been seen by our reviewers, whose comments appear below. In light of their advice I am delighted to say that we are happy, in principle, provided you can fully address the remaining concerns. If so, we will publish a suitably revised version in Communications Earth & Environment under the open access CC BY license (Creative Commons Attribution v4.0 International License).

We therefore invite you to revise your paper one last time to address the remaining concerns of our reviewers. Specifically, please:

1) Explain in the introduction why the analysis of the observational datasets is limited to the period 1979-2021, i.e., why the starting year is 1979.

2) Ensure consistent periods are used throughout the text. When periods need to be different (e.g., for the calculation of anomalies), explain very clearly the reasons.

3) Discuss the sensitivity of the study conclusions to the choice of period of investigation

4) Add more information about the limitations of observational-based datasets (e.g., whether there is land contamination) and discuss clearly the sensitivity of the study conclusions to these limitations

At the same time we ask that you edit your manuscript to comply with our format requirements and to maximise the accessibility and therefore the impact of your work.

EDITORIAL REQUESTS:

Please review our specific editorial comments and requests regarding your manuscript in the attached "Editorial Requests Table". Please outline your response to each request in the right hand column. Please upload the completed table with your manuscript files.

If you have any questions or concerns about any of our requests, please do not hesitate to contact me.

SUBMISSION INFORMATION:

In order to accept your paper, we require the files listed at the end of the Editorial Requests Table; the list of required files is also available at https://www.nature.com/documents/commsj-file-checklist.pdf .

OPEN ACCESS:

Communications Earth & Environment is a fully open access journal. Articles are made freely accessible on publication under a <a href="http://creativecommons.org/licenses/by/4.0"

target="_blank"> CC BY license (Creative Commons Attribution 4.0 International License). This license allows maximum dissemination and re-use of open access materials and is preferred by many research funding bodies.

For further information about article processing charges, open access funding, and advice and support from Nature Research, please visit https://www.nature.com/commsenv/article-processing-charges

At acceptance, you will be provided with instructions for completing this CC BY license on behalf of all authors. This grants us the necessary permissions to publish your paper. Additionally, you will be asked to declare that all required third party permissions have been obtained, and to provide billing information in order to pay the article-processing charge (APC).

Please use the following link to submit the above items:
[link redacted]

** This url links to your confidential home page and associated information about manuscripts you may have submitted or be reviewing for us. If you wish to forward this email to co-authors, please delete the link to your homepage first **

We hope to hear from you within two weeks; please let us know if you need more time.

Best regards,

Viviane V. Menezes, PhD
Editorial Board Member
Communications Earth & Environment
orcid.org/0000-0002-4885-2056

Heike Langenberg, PhD
Chief Editor
Communications Earth & Environment

On Twitter: @CommsEarth

REVIEWERS' COMMENTS:

Reviewer #1 (Remarks to the Author):

I still disagree with the authors regarding the justification to focus on the start date of 1979, which I explain in more detail below. However, despite these disagreements, I am still recommending publication because the authors have done a reasonable job mentioning these issues and the caveats with their analysis.

My point about the similarity of the datasets was that the long-term trends are similar enough that the observations and models will show better agreement over longer-term trends regardless of the dataset used. There still could be systematic biases that effect all datasets, but they are would have to be very large to change this result.

I also still disagree about the argument for wanting the trends to be linear for the comparisons to make sense. Some of the non-linearity could be from forcing (e.g. aerosols), but some of it could be from internal variability superimposed on top of the long-term trend. This makes the analysis susceptible to cherry-picking, because the choice of start date will be influenced by internal variability. Even when compared to the large ensembles where internal variability is taken into account, the cherry-picking will still result in p-values/probabilities that have little meaning.

Even if we ignore these issues, the authors literally state that 1970 reflects the time that the sustained global warming started (L174), so why not pick 1970? Their arguments could make sense for not picking 1950, but I don’t see why they justify picking specifically 1979.

Reviewer #2 (Remarks to the Author):

The authors made some efforts to address two of my remaining concerns on the potential biases that may be induced by spatial extrapolations from land station data onto the polar ice cap in the observational datasets, and the use of SST data for SAT trends. I'm a bit disappointed with their response: I thought I outlined some simple ways for them to address these concerns. The authors either did not fully understand my concerns and suggestions, or simply chose not to perform the simple analyses that I suggested. For example, my first concern is whether the SAT trend over the polar ice cap is comparable to nearby stations (shown in Fig. R2) and suggested them to compare the SAT trends over those two regions in CMIP6 model and/or ERA5 data. If they are different, then the observational datasets may have biases in their estimates of the SAT trend over the polar ice cap because they extrapolated nearby station data onto the polar cap. Their Fig. RL2 and their response do not really address this issue.

In response to my second concern on the potential difference between SST and SAT trends over the Arctic open waters, they cited a previous study that noticed 7% difference, but argued that this difference may have little impact because they used trend ratio, instead of trend difference. I can't understand this argument, as the mean bias will affect both the ratio and difference. Their second argument with comparison to ERA5 SAT trend is helpful. Still, they could follow my suggestion to perform some simple comparison between the SST and SAT trends over Arctic open waters using ERA5 and/or CMIP6 model data, for which they should have both the SST and SAT data.

Reviewer #3 (Remarks to the Author):

I do not have further comments. And I am looking forward to the future works the authors mentioned in the replies.

The responses to reviewers of "The Arctic has warmed nearly four times faster than the globe since 1979"

Mika Rantanen, Alexey Yu. Karpechko, Antti Lipponen, Kalle Nordling, Otto Hyvärinen, Kimmo Ruosteenoja, Timo Vihma and Ari Laaksonen

We thank the reviewers for your final comments on our manuscript. We hope that the rest of the answers will clarify the remaining matters.

The point-by-point replies to your comments are below, marked in black and our responses in blue.

Reviewer #1 (Remarks to the Author):

I still disagree with the authors regarding the justification to focus on the start date of 1979, which I explain in more detail below. However, despite these disagreements, I am still recommending publication because the authors have done a reasonable job mentioning these issues and the caveats with their analysis.

My point about the similarity of the datasets was that the long-term trends are similar enough that the observations and models will show better agreement over longer-term trends regardless of the dataset used. There still could be systematic biases that effect all datasets, but they are would have to be very large to change this result.

I also still disagree about the argument for wanting the trends to be linear for the comparisons to make sense. Some of the non-linearity could be from forcing (e.g. aerosols), but some of it could be from internal variability superimposed on top of the long-term trend. This makes the analysis susceptible to cherry-picking, because the choice of start date will be influenced by internal variability. Even when compared to the large ensembles where internal variability is taken into account, the cherry-picking will still result in p-values/probabilities that have little meaning.

Even if we ignore these issues, the authors literally state that 1970 reflects the time that the sustained global warming started (L174), so why not pick 1970? Their arguments could make sense for not picking 1950, but I don't see why they justify picking specifically 1979.

Thank you for these comments. First, we agree that the model-observations difference is indeed smaller when applied for longer periods, such as 1970-2021 or 1950-2021. We hope that we have emphasized this enough in our manuscript.

However, we still argue that focusing on the time period starting in 1979 is scientifically justified. In addition to the reasons listed in our manuscript, almost all studies of Arctic sea ice and its declining trend use 1979 as a starting year (e.g. https://doi.org/10.1038/s41561-018-0256-8 or https://doi.org/10.1029/2019GL086749). As the sea ice loss is one of the major components causing Arctic amplification, we find it's important that our main analysis is comparable in time with other studies. Thus, we do not see it as cherry-picking to focus mainly on the 1979-2021 time period, especially as some of our figures show the situation for different starting years, and we acknowledge the sensitivity of our results to different starting years in the main text as well.

We indeed state that the period of sustained global warming started in around 1970. However, we do have much more certainty about the magnitude of the warming since 1979. In addition, there is a reason why almost all reanalyses start from 1979, because we have reliable satellite observations since 1979 that can constrain the rate of Arctic warming. Even though the use of 1979 results in low p-values and large discrepancy between observations and models, we find it an interesting result which is important to document.

We had to choose some single year as the lower limit of the time window for our AA calculations, because the internal variation in the models is not in phase with the real world. The year had to be before 1979 in order to capture a large sample of simulated 43-year AA ratios from the models. Since IPCC mentions that sustained global warming began in the 1970s, we simply chose 1970 to allow 10 years more data for the models.

Reviewer #2 (Remarks to the Author):

The authors made some efforts to address two of my remaining concerns on the potential biases that may be induced by spatial extrapolations from land station data onto the polar ice cap in the observational datasets, and the use of SST data for SAT trends. I'm a bit disappointed with their response: I thought I outlined some simple ways for them to address these concerns. The authors either did not fully understand my concerns and suggestions, or simply chose not to perform the simple analyses that I suggested. For example, my first concern is whether the SAT trend over the polar ice cap is comparable to nearby stations (shown in Fig. R2) and suggested them to compare the SAT trends over those two regions in CMIP6 model and/or ERA5 data. If they are different, then the observational datasets may have biases in their estimates of the SAT trend over the polar ice cap because they extrapolated nearby station data onto the polar cap. Their Fig. RL2 and their response do not really address this issue.

Thank you for this comment. We are sorry that we have partly misunderstood your previous comment. We understood that the concern was whether there are systematic biases in the Arctic warming trends of the gridded datasets. However, we are still a bit unsure of what you mean by *two regions*. We cannot expect that the trends between ocean and land are comparable because the loss of sea ice causes strong amplification over the ocean which is not necessarily present over land.

Over the polar ice cap, where there are no in-situ observations, the only option is to extrapolate the temperatures from the nearest stations, which are located in terrestrial land areas and islands. The GHCN-Monthly dataset which was used for validation is perhaps the most extensive in-situ dataset which we have to validate the warming trends in the gridded datasets across the whole Arctic.

We performed the validation against 1979-2021 temperature trends for all four observational datasets: Berkeley Earth, HadCRUT4, GISTEMP, and ERA5. These validation results are shown below in Figures RL1-RL4. For CMIP6 models, as suggested by the referee, we did not make the validation as they are not observational-based.

As one can see from Fig. RL1b, the warming trends in ERA5 is mostly within ±0.1°C/decade range of the stations. In the Canadian Archipelago the warming trend in ERA5 is somewhat weaker than in the stations (negative bias, Fig. RL1b). Anyways, there is no *systematic* bias in ERA5 to one direction or another.

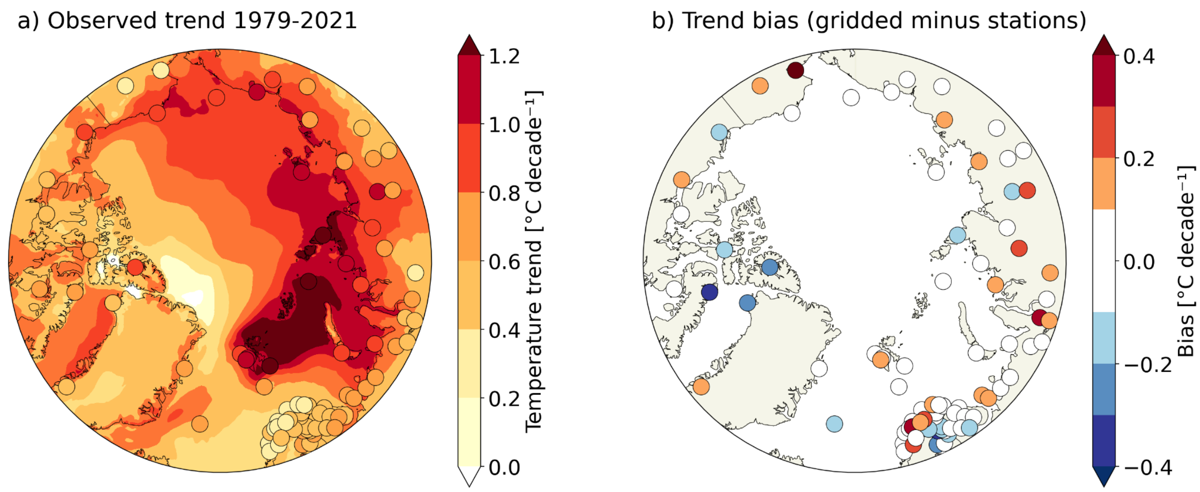

Figure RL1. The warming trend in the Arctic (> 66.5°N) over the 1979-2021 period (left), and the bias in the trend (right). The shading in a) depicts the trend in ERA5 data, and the colored circles indicate the station observations. The bias in b) is defined as ERA5 data minus station observations.

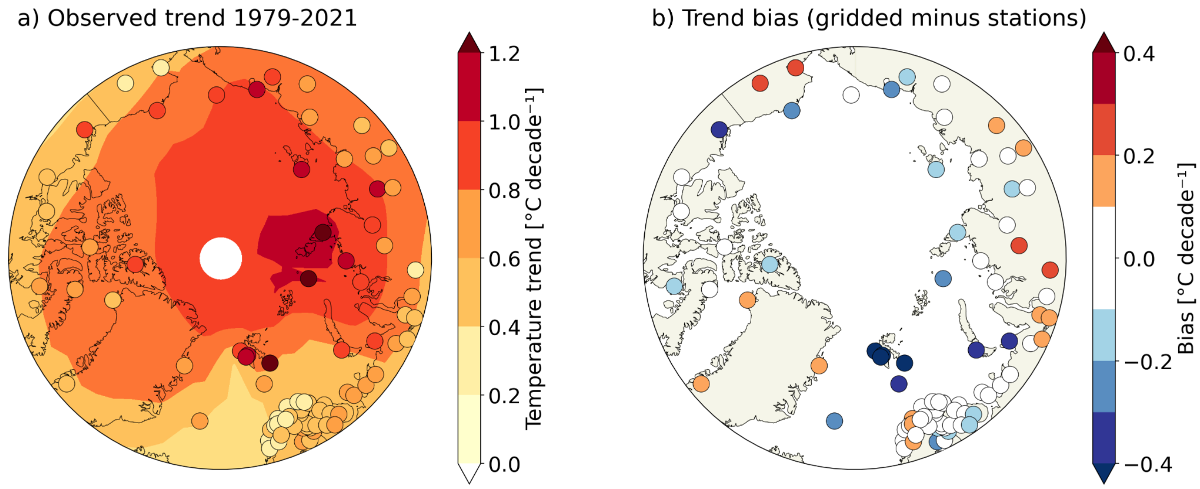

Fig. RL2. Same as Fig. RL1, but for the HadCRUT5 dataset.

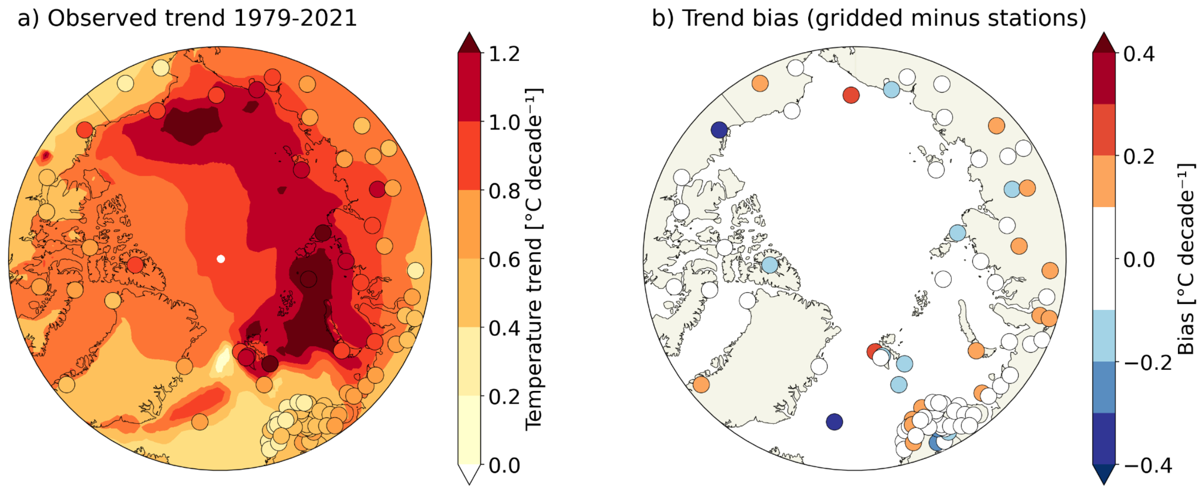

Fig. RL3. Same as Fig. RL1, but for the Berkeley Earth dataset.

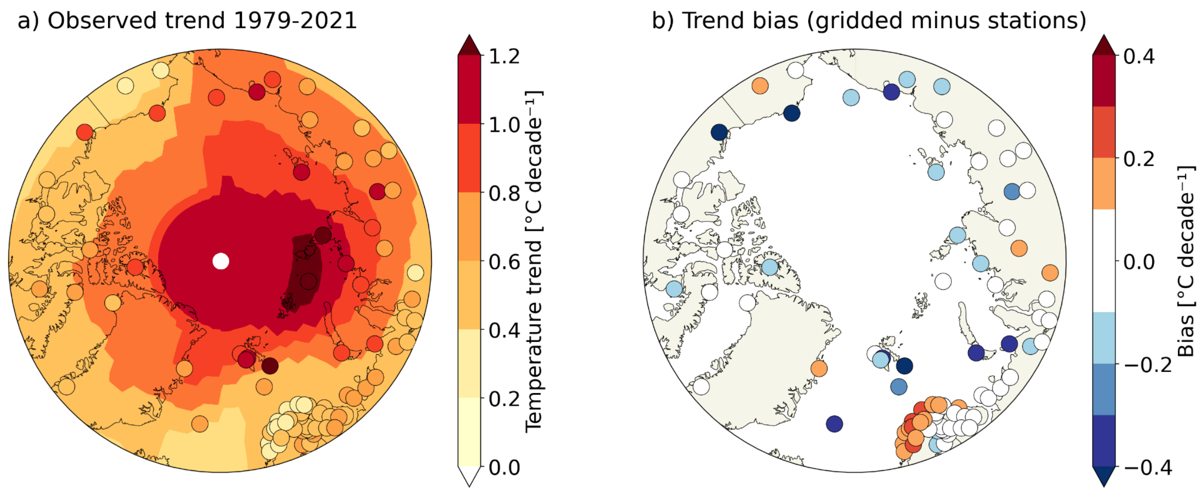

Fig. RL4. Same as Fig. RL1, but for the Gistemp dataset.

To conclude, we note that the observational datasets, including ERA5, do not have markedly large positive biases in their warming trends which could significantly make our results an overestimate. Some of the datasets, such as HadCRUT5 (Fig. RL2) or GISTEMP (Fig. RL4) have mostly negative biases in the Arctic, meaning that their warming trend is in fact underestimated i.e. smaller than in the stations. Thus, we conclude that our estimate of observed four-fold amplification ratio is not at least overestimated; in fact it could be partly underestimated due to the cold biases in HadCRUT5 and GISTEMP.

In response to my second concern on the potential difference between SST and SAT trends over the Arctic open waters, they cited a previous study that noticed 7% difference, but argued that this difference may have little impact because they used trend ratio, instead of trend difference. I can't understand this argument, as the mean bias will affect both the ratio and difference. Their second argument with comparison to ERA5 SAT trend is helpful. Still, they could follow my suggestion to perform some simple comparison between the SST and SAT trends over Arctic open waters using ERA5 and/or CMIP6 model data, for which they should have both the SST and SAT data.

We thank you also for this comment. What we meant by using trend ratio is that the ratio of two trends can remain the same even if both of them would be increased by 7 % (i.e. multiplied by 1.07). Nevertheless, we recognize that the 7 % bias was for the global trend (i.e. the denominator of AA), and for the Arctic trend it may not be the same. Thus, a small bias can still be present in the simulated AA ratio. This is acknowledged in the manuscript in the Section 3.2

Comparing SST and SAT trends between the models and observations would indeed help to quantify the magnitude of this bias. However, we feel that this goes somewhat outside the scope of our paper, especially as it would involve downloading the SST data from CMIP6 models, a time-consuming task which we haven't done so far.

In addition, we would like to stress here that we performed all of our analyses with ERA5 data, which is an observational dataset. ERA5 provides a fair comparison to climate models because both use 2-m temperatures over the whole globe. Still, our key results were not different compared to the in-situ datasets.

We were unsure of what you mean by comparing the SST and SAT trends over Arctic open waters using ERA5 and/or CMIP6 model data. The fraction of open water in the Arctic area is relatively small, and thus, the trends calculated only over those areas do not tell the whole truth. In our opinion, it would be more relevant to calculate the global mean temperature trend and consequently Arctic amplification ratio in CMIP6 models using SST instead of near-surface temperature; however using ERA5 as observations offers a possibility not to perform this option for the reasons mentioned above.