Author's Responses to RC1's comments on *"Brief communication: Reanalyses underperform in cold regions, raising concerns for climate services and research"*

Bin Cao[1], Stephan Gruber[2]

[1]State Key Laboratory of Tibetan Plateau Earth System Environment and Resources (TPESER), National Tibetan Plateau Data Center (TPDC), Institute of Tibetan Plateau Research, Chinese Academy of Sciences, Beijing, China
[2]Department of Geography & Environmental Studies, Carleton University, Ottawa, ON, Canada

Correspondence: Bin Cao (bin.cao@itpcas.ac.cn)

The authors would like to thank the reviewer for their constructive feedback, and the thorough assessment of the manuscript. Below, we provide a point-by-point response to each comment. Reviewer comments are given in black and responses in blue. Additionally, we have included details of how we intend to address these changes in a revised submission.

Cao and Gruber investigate the performance of five modern reanalyses (JRA-3Q, ERA5, MERRA-2, JRA-55 and NCEP2) over cold regions, with a focus on air temperature and snow water equivalent (SWE). They show that the ensemble spread in mean annual air temperature (MAAT) in reanalyses is up 90% greater over cold regions (defined as regions with a MAAT $< 0°C$) relative to regions with a MAAT $\geq 0°C$. The study explicitly shows the relationship between station density and ensemble spread for both SWE and MAAT and is able to show that the reduced reanalysis performance is at least partially related to the low station density over cold regions.

The bulk of the conclusions come from Figure 1, which show the average ensemble spread in MAAT (Panel A) and SWE (Panel B) binned by MAAT. The station density in each MAAT bin, and the proportion of the grid cells within the MAAT bin covered by ice sheets and glaciers, snow cover, permafrost, and seasonally frozen ground is also shown. Herrington et al. (2024) showed a similar plot for soil temperature, that identified the average reanalysis spread in soil temperatures binned by MAAT, against sample size, though it didn't explicitly consider station density, or the proportion of the grid cells covered by the cryosphere - so this is a novel analysis (along with the focus on SWE and MAAT).

While Figure 1 clearly shows a clear correlation between ensemble spread, station density, and the presence of cryospheric elements, there is no attempt to separate the contributions of low station density from those related to inadequate representation of cold region processes in reanalyses. As a reader it raises questions as to what the relative contributions of station density, and inadequate process representation to the ensemble spread in MAAT and SWE are?

To me, the novel science is in quantifying what proportion of the uncertainty can be attributed to station density, and what proportion is related to inadequate representation of cold region processes, which have been thoroughly discussed in the literature (e.g. Broxton et al., 2016; Cao et al., 2020, 2022; Hu et al., 2019; Mortimer et al., 2020; Wang et al., 2019). Thus, I recommend that the authors extend their analysis to explicitly quantify what proportion of the uncertainties or spread in MAAT and SWE can be attributed to the low station density in cold regions, and what proportion can be attributed to the inadequate process representation in the products. This will greatly enhance the contribution of the paper to the literature and provide the community with useful and quantifiable estimates of the uncertainty attribution.
Response (given in AC1 at `https://doi.org/10.5194/egusphere-2025-575-AC1`): This is an important point that has also been raised by colleagues who commented on an earlier version of the manuscript. We have made a deliberate decision here: A detailed and conclusive analysis of the causes for the large spread in cold regions will likely be an involved process requiring a broad range of knowledge, skills, and perspectives that differ from ours, and that will take time to bring together in a research project. We made a choice to expose our finding as a Brief Communication quickly to motivate and accelerate this research.

Specific Comments

- P2, L31: Why was NCEP2 investigated over NCEP CFSR/CFSv2, for example? NCEP CFSR/CFSv2 is available at a much higher resolution than NCEP2, and is available over the period of analysis (1991-2020).
  Response (given in AC1 at `https://doi.org/10.5194/egusphere-2025-575-AC1`): In selecting reanalyses to include, we have opted to not include CFSR/CFSv2 because it mixes two differing simulation and assimilation systems.

- P2, L36-L37: What do the authors mean by the statement "the decades 1991-2020 were used, likely a period of high

quality for reanalyses." Is there a particular standard by which the authors determined this? Some clarification would be helpful.
Response: In the revision, we will change this part to:

"The three most recent decades 1991–2020, which had improved satellite observation and data assimilation (Hersbach et al., 2019), were used."

In addition, we will 1) add a brief description regarding to the temporal changes of spread (see Sec. 3.1 Lower agreement among reanalyses in cold regions); and 2) revise Figure 2 by adding the temporal change of $MAAT_s$ and relative $maxSWE_s$.

"The temporal analyses revealed that the spread in MAAT and maxSWE was generally reduced since 1980, with the increased assimilation of satellite datasets (Fig. 2E and F, Hersbach et al., 2019). For example, the $MAAT_s$ in 2010s was reduced by 0.23 °C compared to that in 1960s for the 4DVar reanalyses. But a persistent spread found since 1980 despite improved observations may indicate process-representation issues in the numerical weather prediction models.

- P2, L38: What is the CDS? It appears that CDS is in brackets, but the acronym is not defined. I presume this may be the Climate Data Store?
  Response: The citation will be revised as below.

  Copernicus Climate Change Service, Climate Data Store, (2021): Global land surface atmospheric variables from 1755 to 2020 from comprehensive in-situ observations. Copernicus Climate Change Service (C3S) Climate Data Store (CDS). (Accessed on 21-04-2025), 10.24381/cds.cf5f3bac

- P4, L101: The all 5-reanalysis value for MAAT spread was (1.5°C, 0.5°C-3.0°C) - is this statistically different from the value for the 4DVar reanalyses reported here?
  Response: No. The MAAT (1.5, 0.5–3.0°C *vs.* 1.3, 0.3–2.9°C) as well as SWE (105, 51–206% *vs.* 101%, 56–186%) spread based on all 5-reanalysis and 4DVar reanalyses are generally close.

  "Compared to all five reanalyses, the 4DVar reanalyses show a reduced spread in MAAT (1.3, 0.3–2.9 °C) and its trend (0.13, 0.04–0.24 °C dec^{-1}), as expected from a consistent and more advanced assimilation method. However, the average ensemble spread for $MAAT_s$ *in cold regions is still up to 45% greater than that of other regions. The relative* $maxSWE_s$ *among 4DVar reanalyses is about 101% (56–186%), and is comparable to that derived from all five reanalyses."*

- P10, Figure 1: What do the dashed lines represent in Figure 1? I don't see a dashed line in the figure legend?
  Response: The dashed lines represent trend for MAAT and SWE. We will revise the figure and caption to clarify (see Fig. 1).

- P11, Figure 2: It may be helpful to have a "difference" panel between the All 5 reanalyses and the three 4DVar reanalyses to highlight the regions with the largest differences, particularly for SWE, since it is a little harder to notice the differences in Panel B.
  Response: We agree the "difference" between the all 5 reanalyses and the three 4DVar reanalyses would provide additional helpful information. We treat 4DVar reanalyses separately *"As better performance is expected from the newer 4DVAR reanalyses"* rather than demonstrating 4D. Our results, however, indicated the reanalyses significantly underperformed in cold regions regardless the assimilation method. In addition, 3DVar and 4DVar reanalyses generally show very close performance. To keep the manuscript "brief", the difference are given in Supporting Information (SI) as Figure S2 (see below).

References

Broxton, P. D., Zeng, X., and Dawson, N.: Why Do Global Reanalyses and Land Data Assimilation Products Underestimate Snow Water Equivalent?, J. Hydrometeorol., 17, 2743–2761, https://doi.org/10.1175/JHM-D-16-0056.1, 2016.

Cao, B., Gruber, S., Zheng, D., and Li, X.: The ERA5-Land soil temperature bias in permafrost regions, The Cryosphere, 14, 2581–2595, https://doi.org/10.5194/tc-14-2581-2020, 2020.

Cao, B., Arduini, G., and Zsoter, E.: Brief communication: Improving ERA5-Land soil temperature in permafrost regions using an optimized multi-layer snow scheme, The Cryosphere, 16, 2701–2708, https://doi.org/10.5194/tc-16-2701-2022, 2022.

Herrington, T. C., Fletcher, C. G., and Kropp, H.: Validation of Pan-Arctic Soil Temperatures in Modern Reanalysis and Data Assimilation Systems, The Cryosphere, 18, 1835–1861, https://doi.org/10.5194/tc-18-1835-2024, 2024.

Hersbach, H., Bell, W., Berrisford, P., HorÃ¡nyi, A., J, M.-S., Nicolas, J., Radu, R., Schepers, D., Simmons, A., Soci, C., Dee, D., 2019. Global reanalysis: goodbye ERA-Interim, hello ERA5. ECMWF Newsletter. https://doi.org/10.21957/vf291hehd7.

Hu, G., Zhao, L., Li, R., Wu, X., Wu, T., Xie, C., Zhu, X., and Su, Y.: Variations in soil temperature from 1980 to 2015 in permafrost regions on the Qinghai-Tibetan Plateau based on observed and reanalysis products, Geoderma, 337, 893–905, https://doi.org/10.1016/j.geoderma.2018.10.044, 2019.

Mortimer, C., Mudryk, L., Derksen, C., Luojus, K., Brown, R., Kelly, R., and Tedesco, M.: Evaluation of long-term Northern Hemisphere snow water equivalent products, The Cryosphere, 14, 1579–1594, https://doi.org/10.5194/tc-14-1579-2020, 2020.

Wang, C., Graham, R. M., Wang, K., Gerland, S., and Granskog, M. A.: Comparison of ERA5 and ERA-Interim near-surface air temperature, snowfall and precipitation over Arctic sea ice: effects on sea ice thermodynamics and evolution, The Cryosphere, 13, 1661–1679, https://doi.org/10.5194/tc-13-1661-2019, 2019.

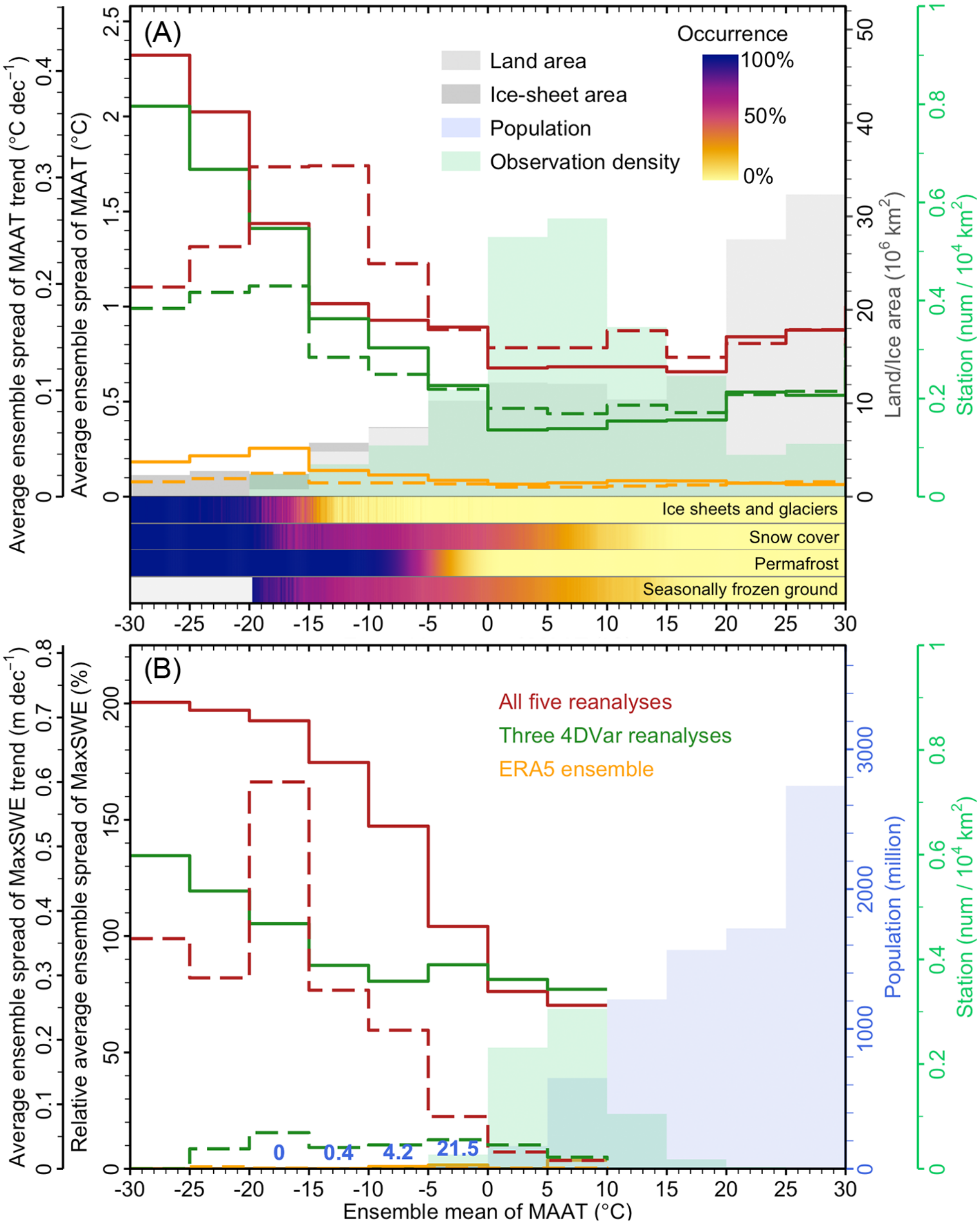

Figure 1: The 1991–2020 average ensemble spread of (A) mean annual air temperature (MAAT) and (B) relative maximum snow water equivalent (MaxSWE) among different reanalyses. The red (3DVar and 4DVar) and green (4DVar only) lines represent ensembles of differing numerical weather prediction models and assimilation systems, whereas the yellow line (ERA5) represents uncertainty in observations and physical parameterizations in a single modelling and assimilation system. **The solid lines represent the mean state and dashed lines indicate the trend (left vertical axis).** Land area and population are shown for context **(right vertical axis)**. Values are summarized in intervals of 5 °C for the ensemble mean of MAAT. The occurrence of cryosphere elements, estimated as the probability of occurrence during the analysis period, is scaled per MAAT bin of 0.1 °C (see Methods). Only reanalysis cells with a significant ($P < 0.05$) trends are used for the analysis of change. Blue numbers express low population counts in million. The peak in the trend of MaxSWE observed for MAAT class from −15 °C to −20 °C is caused by increased uncertainty in ice-free areas of Greenland and Antarctic.

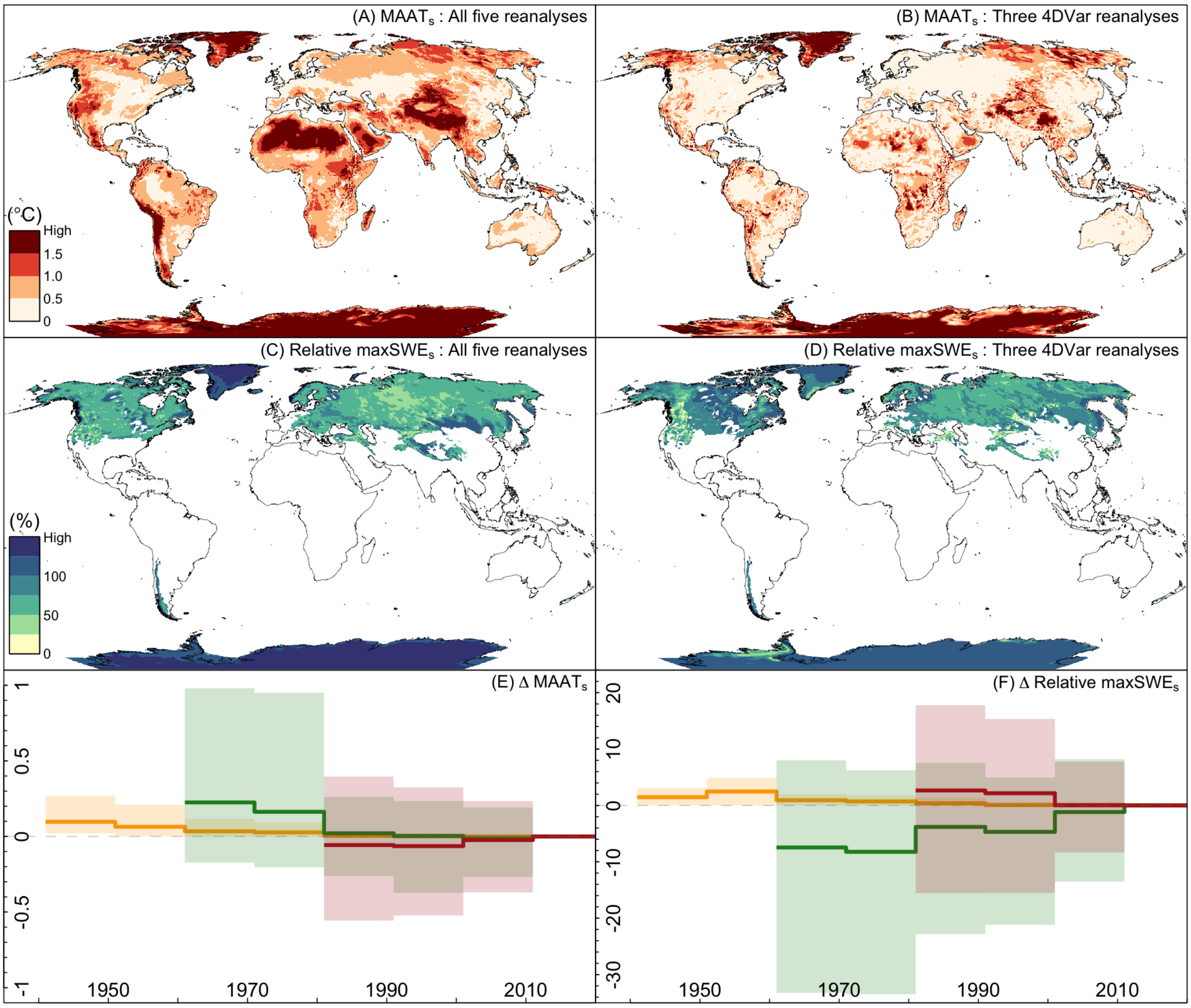

Figure 2: The 1991–2020 average ensemble spread of mean annual air temperature (MAAT$_s$) and relative spread of maximum snow water equivalent (maxSWE$_s$). Only areas with a mean maxSWE$_s$ greater than 0.0125 m (0.05 m snow height at a snow density is 250 kg m^{-3}) are shown. Snow water equivalent is not available for the two continental ice sheets in MERRA-2, and therefore, MERRA-2 is not included in these regions. The overall temporal changes for (E) MAAT$_s$ and (F) relative maxSWE$_s$ was derived with the reference period of 2011–2020, and a positive value means the spread is reduced relative to the referenced period. In E and F, the soil lines represent the mean state and shaded areas indicate 10th to 90th percentile. The spread difference between all five reanalyses and three 4DVar reanalyses is given in Fig. S2.

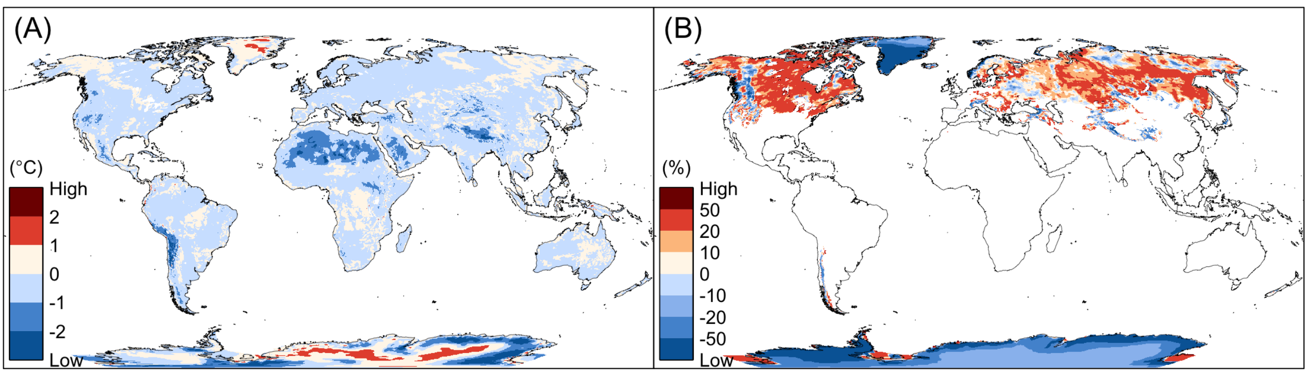

Figure S2: The difference between all five reanalyses and three 4DVar reanalyses for (A) mean annual air temperature (B) and relative maximum snow water equivalent (maxSWE). Blue areas indicate in improvement (smaller spread) in the 4DVar results relative to all five reanalyses, Red areas indicate a deterioration (greater spread) in the 4DVar results relative to all five reanalyses.