Supporting Information for "Illuminating snow droughts: The future of Western United States snowpack in the SPEAR large ensemble"

Julian Schmitt[1], Kai-Chih Tseng[234], Mimi Hughes[5], Nathaniel C. Johnson[2]

[1]California Institute of Technology

[2]Geophysical Fluid Dynamics Laboratory

[3]Princeton University

[4]Department of Atmospheric Sciences, National Taiwan University, 10617, Taipei, Taiwan

[5]Earth System Research Laboratories/ Physical Sciences Laboratory

Contents of this file

1. Text S1
2. Figures S1 to S4
3. Table S1

Introduction

We include additional text and figures to complement the discussion in the main text.

Text S1: Historical Changes in Temperature and Precipitation Extremes

To assess SPEAR's ability to represent historical temperature and precipitation extremes, we applied the methodology used in section 2.4 of the main text for snow drought classification to temperature and precipitation. We first aggregated SPEAR and Livneh to monthly time scales, taking the maximum high and minimum low daily temperature for each month. We do so because we assume these metrics are more likely to capture heat waves and cold extremes than an average of daily high and low temperatures. For example, a severe multi-day heat wave in January has the potential to melt snowpack quickly but might fail to show up on a 30-day average. To measure changes in meteorological drought conditions, we utilize the D2+ severity notation for dry extremes, and introduce W2+ for wet extremes. For temperatures, we introduce H2+ to indicate heat extremes and C2+ to indicate cold extremes; refer to Figure S2 for notation. As our analysis focuses on both monthly maximum and minimum temperatures, we append the subscripts "max" and "min" to distinguish between these conditions, respectively. Therefore, $H2+_{\mathrm{min}}$ represents the change in frequency of warm extremes for monthly minimum temperatures. To assess whether changes are significant across the historical time period, we evaluate a 95% confidence interval for the SPEAR ensemble mean, assuming the underlying changes were distributed normally. If the interval does not contain zero change, then the forced component is significant in SPEAR. We present these calculations in Figure S3 which assesses changes to monthly meteorological drought (D2+), warm temperature extremes ($H2+_{\mathrm{max}}$), and cold temperature extremes ($C2+_{\mathrm{min}}$ and $H2+_{\mathrm{min}}$). Together, these panels provide a method to validate the SPEAR ensemble against changes in extreme temperature and precipitation observations.

Across the historical period (1930-2011), we found that while SPEAR's changes in D2+ meteorological drought were not significant, several measures of temperature extremes were. Across the five HUC2 regions, we find that H2+$_{max}$ extreme heat increased on average between 59% and 73%, C2+$_{min}$ extreme cold decreased between 18-21%, and H2+$_{min}$ extreme heat increased between 41% and 60% (Figure S3(b, c, d)). These changes indicate significant warming in both maximum and minimum temperatures and both of these trends are expected to negatively impact WUS snowpack. In both SPEAR and Livneh, extreme heat events have increased in frequency while extreme cold events have decreased on average. When we assess agreement between SPEAR and Livneh, we find that all but one Livneh observation falls within the SPEAR ensemble range, suggesting SPEAR is able to accurately reproduce changes in precipitation and temperature extremes across the historical period. Examining which trends are significant in SPEAR, we find all temperature trends to be significant while only the increase in D2+ meteorological drought in the LC region is significant. The LC saw an average increase in meteorological drought of 16% in SPEAR, while in Livneh the increase in the LC was 48%. Amongst the five HUC2 regions, the LC region increase was also the most extreme increase among Livneh meteorological drought observations (Figure S3(a)). Together, these observations may indicate the LC is drying more rapidly than other regions. When examining temperature trends, the PNW stood out as it experienced the smallest changes in extreme temperatures and was the only region to observe a decrease in meteorological drought. We assume these underlying colder, wetter conditions across the latter half of the historical period explains the decrease in D2+ SD frequency over the historical period in the PNW seen in Figure 3 and perhaps the deviation in the early 2000s of Livneh D2+ SD frequency from the

SPEAR ensemble in Figure 5. While PNW falls further from the SPEAR ensemble mean, the changes are still within the SPEAR ensemble range and thus may be attributable to internal climate variability. The strong agreement between changes to historical meteorological and temperature conditions in SPEAR and Livneh further lends confidence to SPEAR's ability to capture historical trends across the WUS.

References

Delworth, T. L., Cooke, W. F., Adcroft, A., Bushuk, M., Chen, J.-H., Dunne, K. A., ... Zhao, M. (2020). Spear: The next generation gfdl modeling system for seasonal to multidecadal prediction and projection. *Journal of Advances in Modeling Earth Systems*, *12*(3), e2019MS001895. Retrieved from `https://agupubs.onlinelibrary.wiley.com/doi/abs/10.1029/2019MS001895` (e2019MS001895 2019MS001895) doi: https://doi.org/10.1029/2019MS001895

Huning, L. S., & AghaKouchak, A. (2020). Global snow drought hot spots and characteristics. *Proceedings of the National Academy of Sciences*, *117*(33), 19753-19759. Retrieved from `https://www.pnas.org/doi/abs/10.1073/pnas.1915921117` doi: 10.1073/pnas.1915921117

Siirila-Woodburn, E. R., Rhoades, A. M., Hatchett, B. J., Huning, L. S., Szinai, J., Tague, C., ... Kaatz, L. (2021). A low-to-no snow future and its impacts on water resources in the western united states. *Nature Reviews Earth & Environment*, *2*(11), 800–819. Retrieved from `https://doi.org/10.1038/s43017-021-00219-y` doi: 10.1038/s43017-021-00219-y

Svoboda, M., LeComte, D., Hayes, M., Heim, R., Gleason, K., Angel, J., ... Stephens, S. (2002). The drought monitor. *Bulletin of the American Meteorological Society*, *83*(8), 1181 - 1190. Retrieved from `https://journals.ametsoc.org/view/journals/bams/83/8/1520-0477-83_8_1181.xml` doi: 10.1175/1520-0477-83.8.1181

Walton, D., & Hall, A. (2018). An assessment of high-resolution gridded temperature datasets over california. *Journal of Climate*, *31*(10), 3789–3810. Retrieved 2022-11-18, from `https://journals.ametsoc.org/view/journals/clim/31/10/jcli-d`

-17-0410.1.xml (Publisher: American Meteorological Society Section: Journal of Climate) doi: 10.1175/JCLI-D-17-0410.1

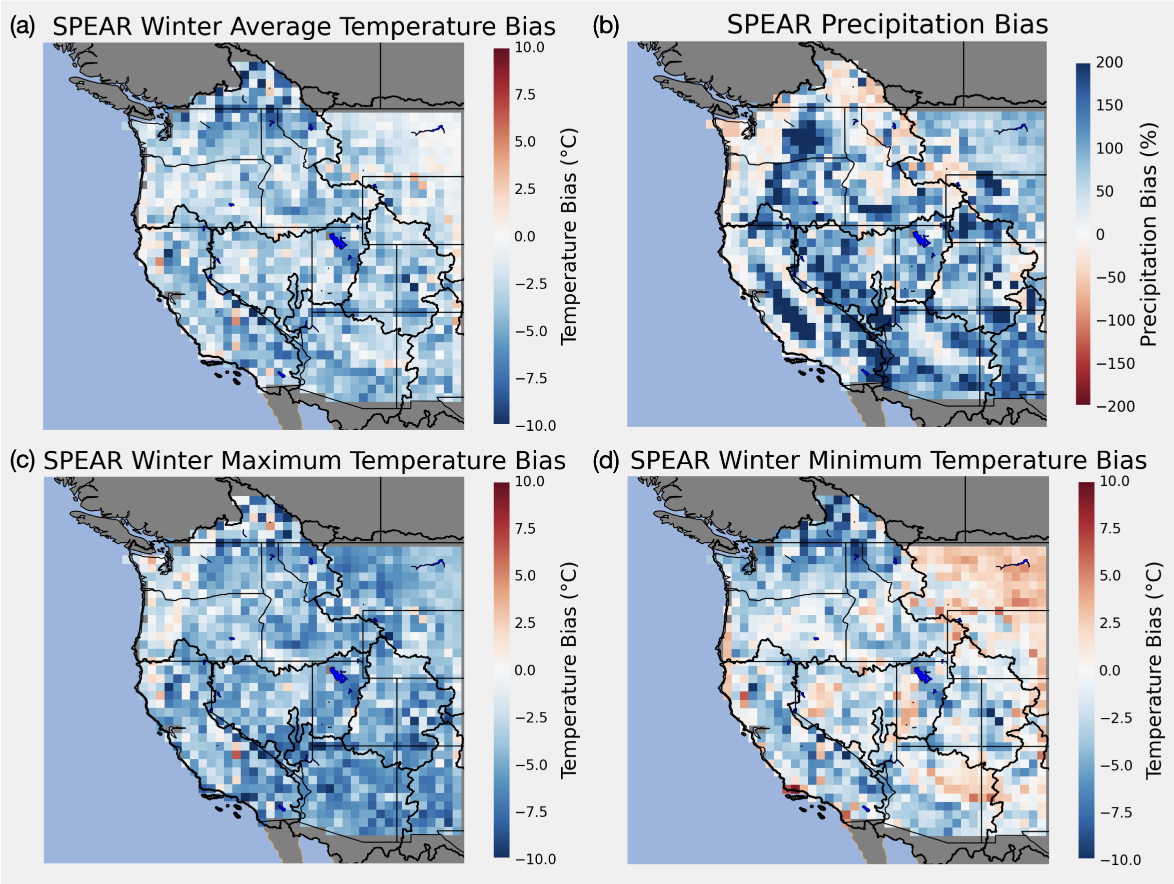

Figure S1. 90-Year winter temperature and precipitation biases including (a) average temperature bias, (b) precipitation bias, (c) maximum temperature bias, and (d) minimum temperature bias. The winter average temperature bias was computed by taking the difference of the average maximum and average minimum temperatures of SPEAR and Livneh between October 1st and April 31st. Overall SPEAR has a slight cold and wet bias across much of the Western United States. The wet bias is consistent with Delworth et al. (2020). By examining the maximum and minimum temperature biases we see that SPEAR has a significant cold bias for maximum temperatures across the entire WUS, while it has a systematic cold bias for minimum temperatures over mountainous regions and slight warm bias over the rest of the WUS. We expect that some of the bias can be explained by the differences in model resolution: SPEAR is on a 1/2° grid while Livneh is on a 1/16° grid. We also note that Livneh has a particularly high lapse rate of 6.5°C/1000m which may contribute some additional bias (Walton & Hall, 2018).

Classification of Extremes by Z-Score				
Drought Severity	**Temperature Severity**	**Description**	**Z-Score**	**Probability of at least as Extreme**
D4	C4	Exceptional	Z ≤ -2.0	0.023
D3	C3	Extreme	-2.0 < Z ≤ -1.6	0.055
D2	C2	Severe	-1.6 < Z ≤ -1.3	0.097
D1	C1	Moderate	-1.3 < Z ≤ -0.8	0.21
D0	C0	Abnormal	-0.8 < Z ≤ -0.5	0.31
NN	NN	Near Normal	-0.5 < Z < 0.5	--
W0	H0	Abnormal	0.5 ≤ Z < 0.8	0.31
W1	H1	Moderate	0.8 ≤ Z < 1.3	0.21
W2	H2	Severe	1.3 ≤ Z < 1.6	0.097
W3	H3	Extreme	1.6 ≤ Z 2.0	0.055
W4	H4	Exceptional	2 ≤ Z	0.023

Figure S2. List of drought and temperature classification abbreviations, a text description for each, the corresponding Z-Score, and the probability of an event being at least as extreme. This probability captures how likely a random historical month is to be classified in that category or one that is more extreme. This table uses identical Z-Score ranges to Huning and AghaKouchak (2020) and attempts to mimic frequencies of hydrological drought given by the US Drought Monitor (Svoboda et al., 2002).

Figure S3. Changes in wintertime (Oct-Apr) historical precipitation and temperature extremes in (a) D2+ meteorological drought and (b) H2+$_{\mathrm{max}}$, (c) C2+$_{\mathrm{min}}$, and (d) H2+$_{\mathrm{min}}$ temperatures. The shaded histogram depicts the SPEAR ensemble distribution, with ensemble mean and confidence interval marked with vertical black dashed and solid lines, respectively. The observed value from the Livneh dataset is marked as a vertical line shaded red in (a) and blue in (b-d). A vertical dotted zero trend line is included for reference.

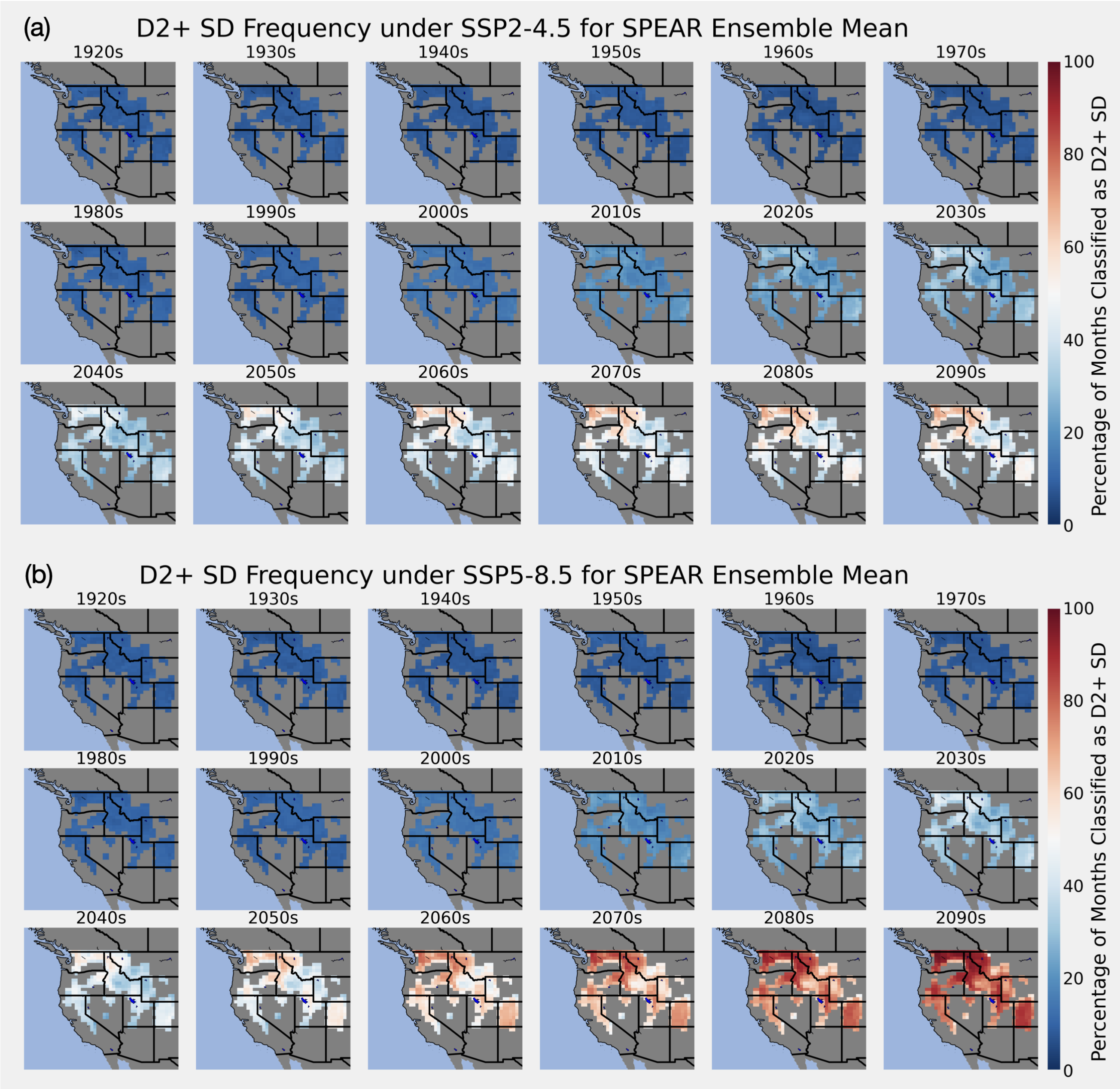

Figure S4. Panel plots for all 18 study decades between 1920 and 2100 for the SSP2-4.5 and SSP5-8.5 D2+ SD classification frequencies for the SPEAR ensemble mean. This figure emphasizes just how dramatic SPEAR projects the increase in D2+ SD occurrence to be, conditioned on the emissions scenarios, as the historical variability of the 20th century is barely distinguishable when placed on the same color scale as changes in the 21st century.

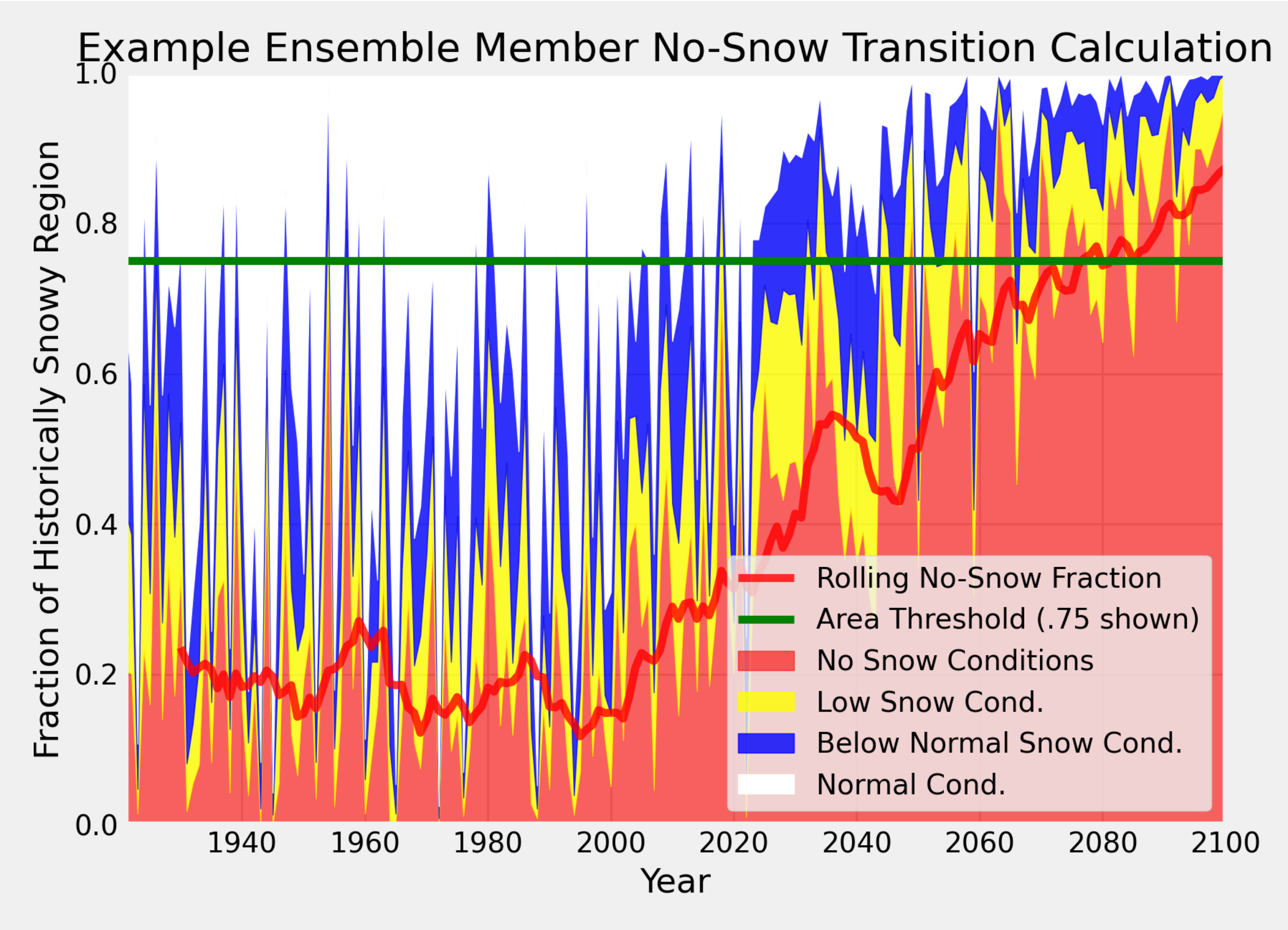

Figure S5. An illustration of how the no-snow transition is calculated as a function of the area threshold. This figure shows the fraction of the historically-snowy region experiencing no-snow (red), low-snow (yellow), below average-snow (blue), and near-normal or above-average snow (white) following the categories used by Siirila-Woodburn et al. (2021). The dark red curve represents a 10-year moving average of the yearly no-snow values (in red), while the green horizontal line indicates the chosen area threshold, in this case $\mathcal{A} = .75$. For this particular region in one ensemble member, we see that the red curve crosses the green line for the last time in 2082. Thus, this ensemble member records a no-snow transition time of 2082 for the given threshold.