# Supporting Information for "Changes in United States summer temperatures revealed by explainable neural networks"

Zachary M. Labe[1], Nathaniel C. Johnson[2], and Thomas L. Delworth[2]

[1]Atmospheric and Oceanic Sciences Program, Princeton University, NJ, USA

[2]NOAA/OAR/Geophysical Fluid Dynamics Laboratory, Princeton, NJ, USA
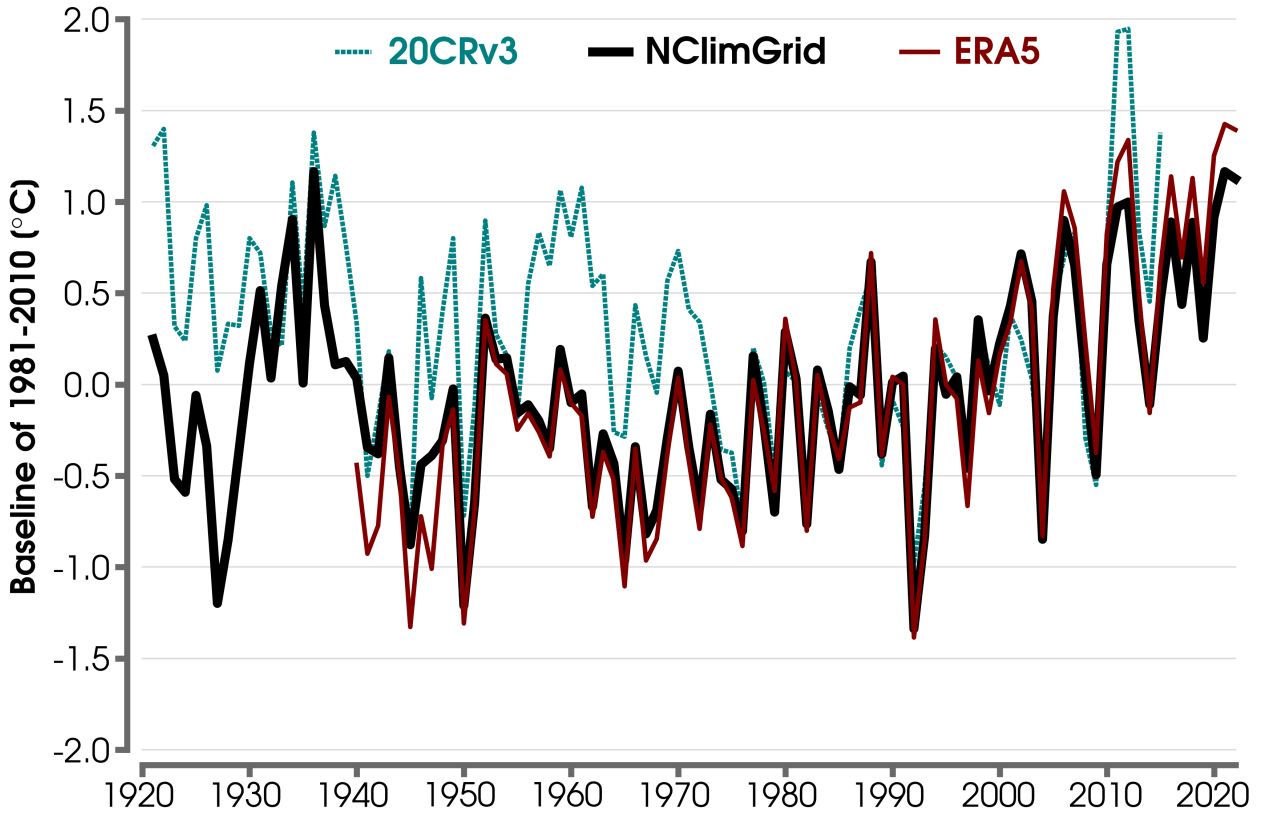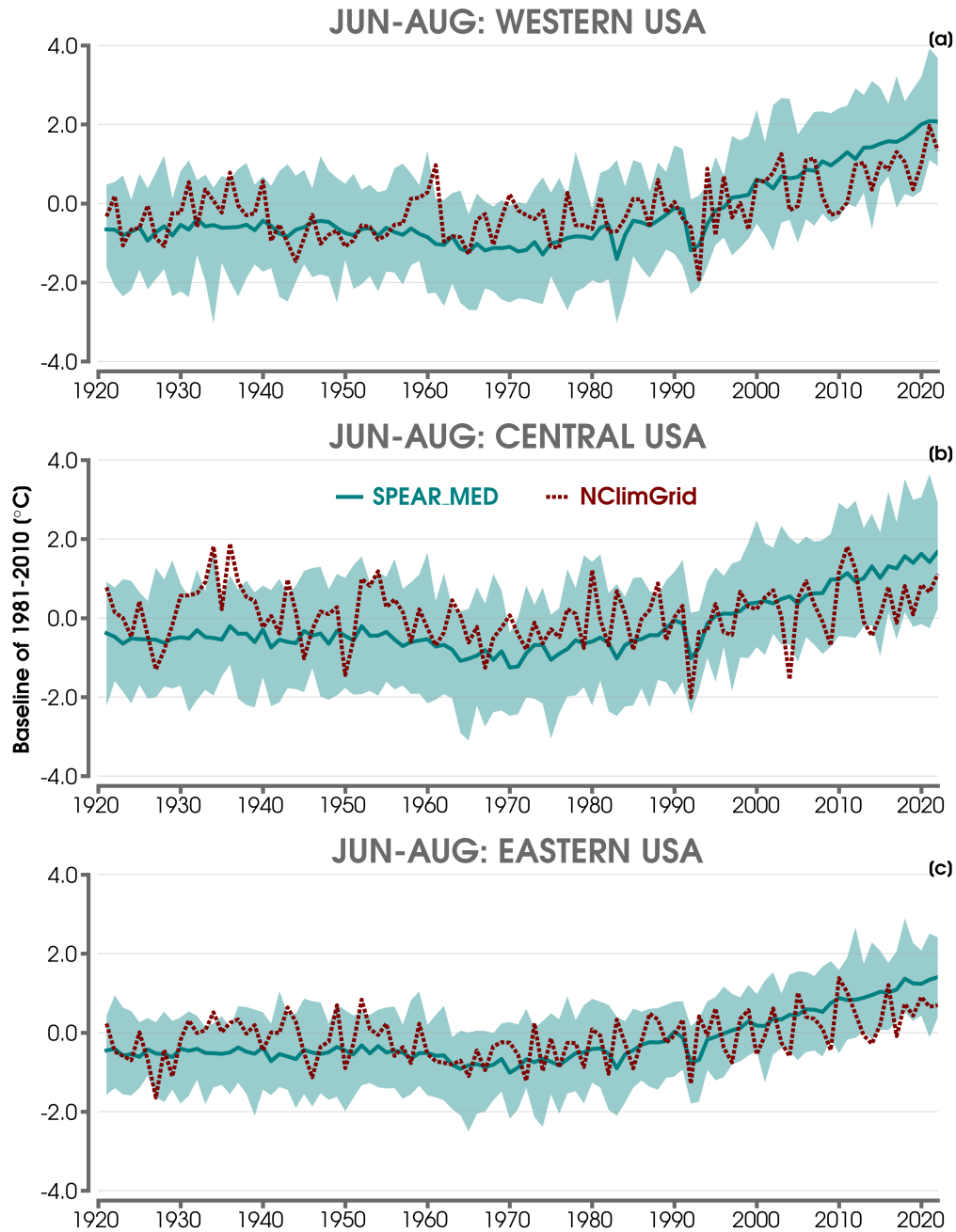
## Contents of this file

Corresponding author: Zachary M. Labe (zachary.labe@noaa.gov)

July 11, 2023, 9:23am

**Table S.1.** Summary of additional climate model large ensembles used in this study. The artificial neural network (ANN) grid notation refers to the size of the input maps for the neural network, which are interpolated to either the same atmospheric horizontal resolution as SPEAR_MED (i.e., MED; 0.5°) or SPEAR_LO (i.e., LOW; 1.0°) (see Section 3).

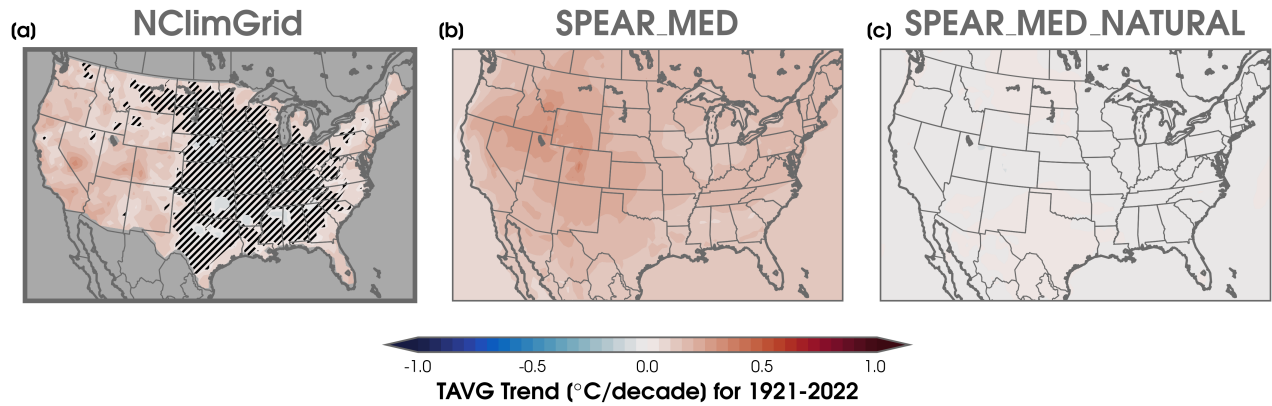| Name | CMIP5/6 Forcing | Years | # Members | Horizontal Resolution (Atmosphere / Ocean) | ANN Grid | Reference |
|---|---|---|---|---|---|---|
| CESM1-LE | Historical to 2005, RCP 8.5 | 1920-2100 | 40 | $\sim 1.25° \times 0.9°$/nominal 1.0° | LOW | Kay et al. (2015) |
| CESM2-LE | Historical to 2014, SSP3-7.0 | 1850-2100 | 100 | $\sim 1.25° \times 0.9°$/nominal 1.0° | LOW | Rodgers et al. (2021) |
| FLOR | Historical to 2005, RCP 8.5 | 1921-2100 | 30 | nominal 0.5°/nominal 1.0° | MED | Vecchi et al. (2014) |
| MIROC6-LE | Historical to 2014, SSP5-8.5 | 1850-2100 | 50 | $\sim 1.4° \times 1.4°$/nominal 1.0° | LOW | Shiogama et al. (2023) |
| SPEAR_LO | Historical to 2014, SSP5-8.5 | 1850-2100 | 30 | nominal 1.0°/nominal 1.0° | LOW | Delworth et al. (2020) |

**Figure S1.** Average June to August (JJA) mean temperature (TAVG) anomalies for the contiguous United States (CONUS) using the National Oceanic and Atmospheric Administration (NOAA)/Cooperative Institute for Research in Environmental Sciences (CIRES)/Department of Energy (DOE) Twentieth Century Reanalysis (20CR) version 3 (20CRv3; dashed blue line) from 1921 to 2015 (Slivinski et al., 2019, 2021), the European Centre for Medium-Range Weather Forecasts (ECMWF) fifth generation reanalysis (ERA5; solid red line) from 1940 to 2022 (Hersbach et al., 2020), and the NOAA Monthly U.S. Climate Gridded Dataset (NClimGrid; solid black line) from 1921 to 2022 (Vose et al., 2014). All anomalies are computed from their 1981-2010 climatologies.
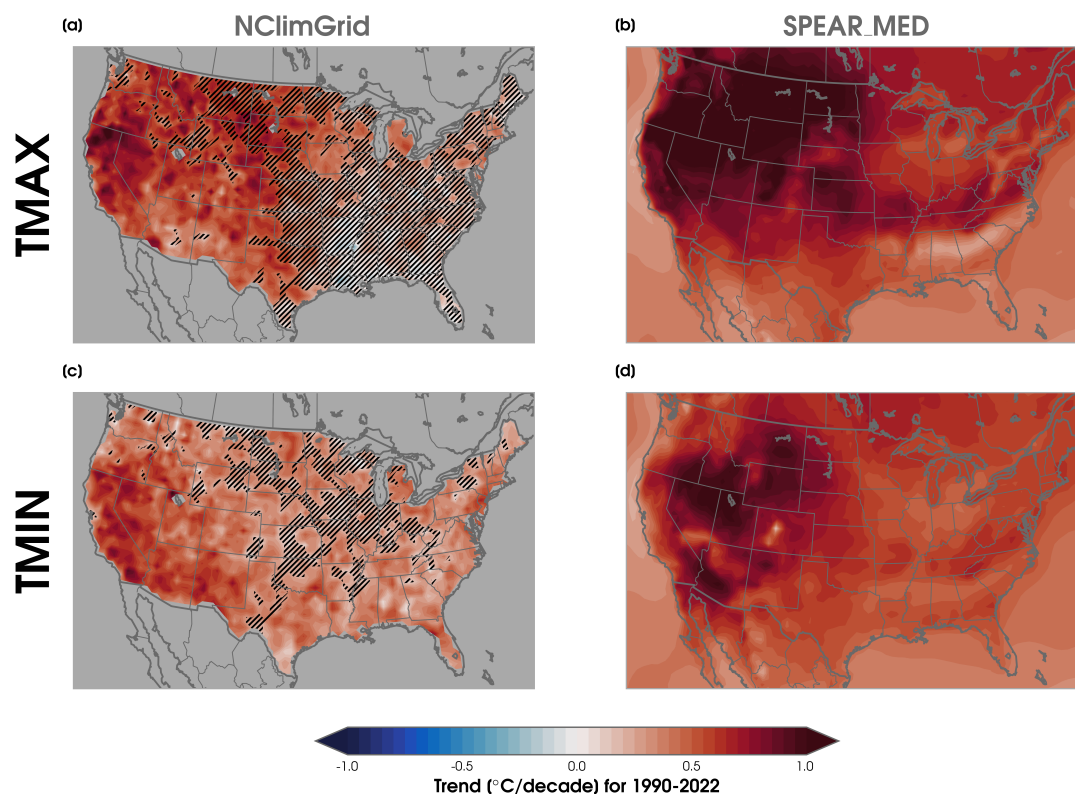
**Figure S2.**    (a) Time series of mean JJA TAVG anomalies over the Western USA from 1921 to 2022 for the ensemble mean of SPEAR_MED (dark green line) compared to observations from NClimGrid (dashed red line).  The spread across SPEAR_MED ensemble members is shown with the light green shading.  Anomalies are computed for each dataset with respect to their own 1981-2010 climatological mean. (b) As in (a), but for the Central USA. (c) As in (a), but for the Eastern USA. See Figure 2 for the regional outlines.
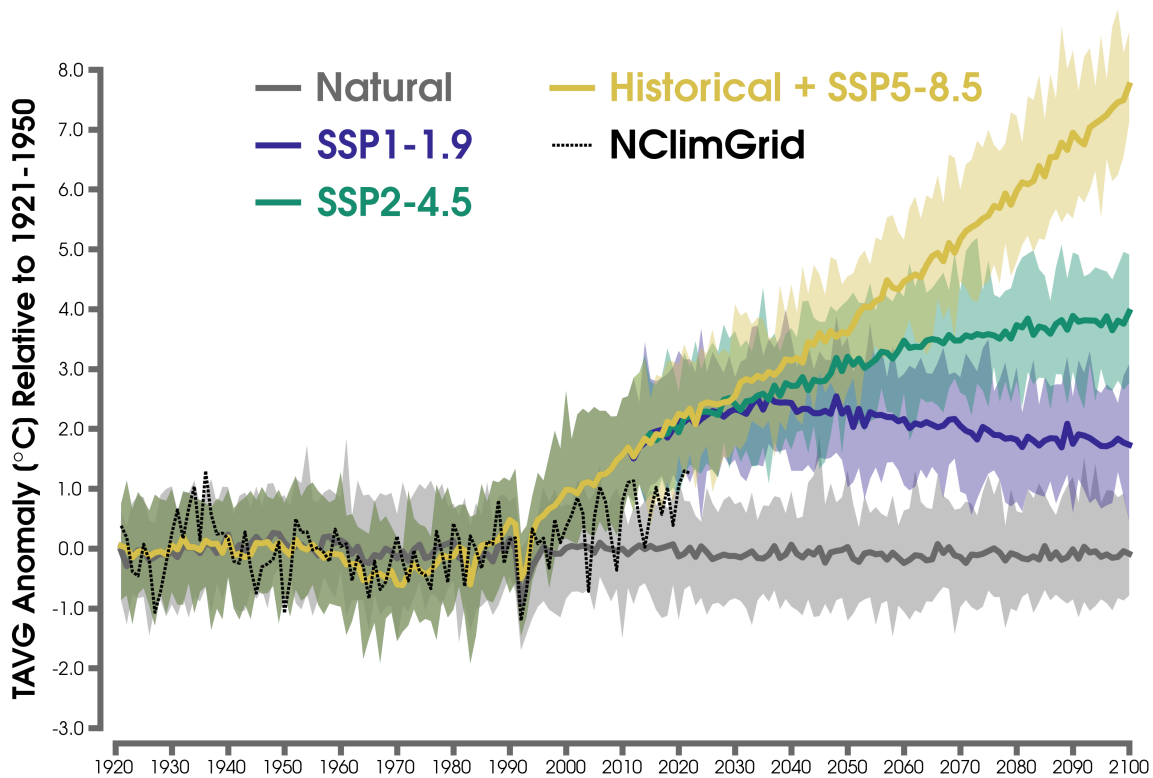
**Figure S3.** (a) Linear least squares trends of average JJA TAVG from 1921 to 2022 for NClimGrid, (b) the ensemble mean from SPEAR_MED, and (c) the ensemble mean from SPEAR_MED_NATURAL (c, f). For maps of NClimGrid, black hatch marks indicate TAVG trends that are not statistically significant following a Mann-Kendall test (Bevan & Kendall, 1971; Mann, 1945) for the 95% confidence level.
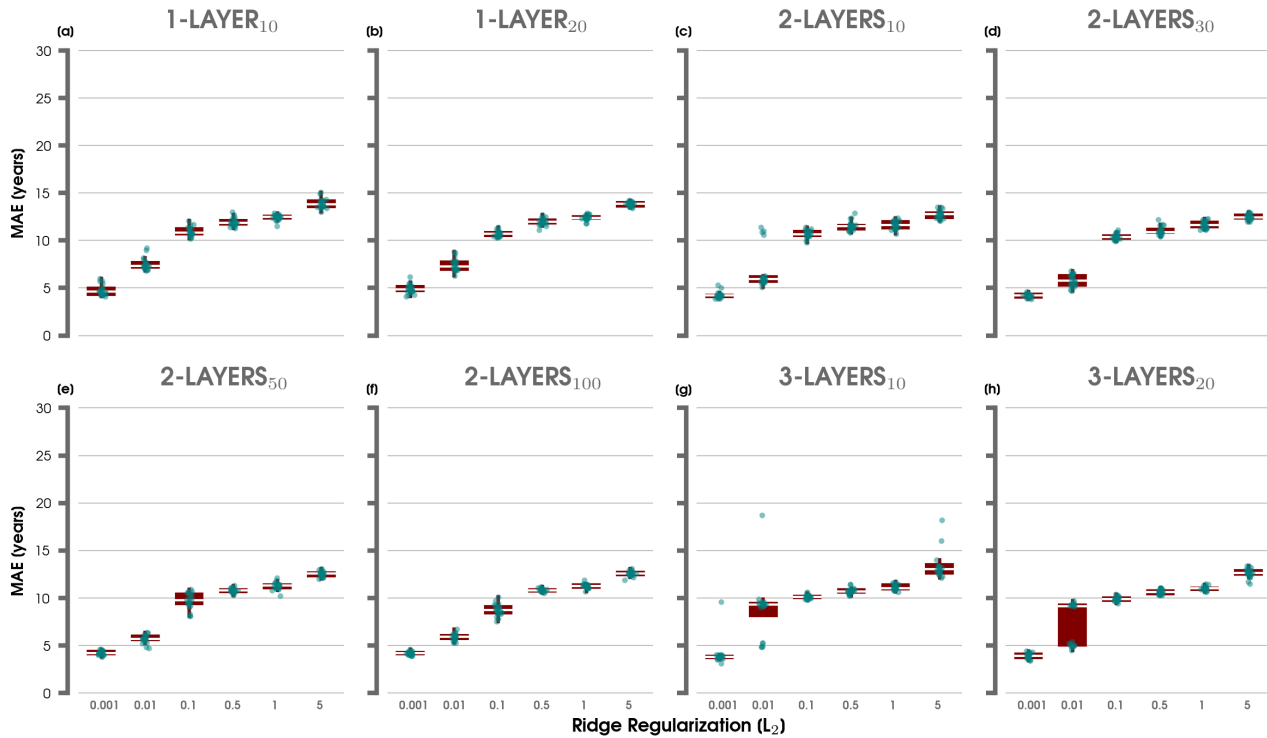
**Figure S4.**    (a) Linear least squares trends of average JJA TMAX from 1990 to 2022 for NClimGrid. (b) As in (a), but for the ensemble mean from SPEAR_MED. (c-d) As in (a-b), but for TMIN. Black hatch marks indicate TAVG trends that are not statistically significant following a Mann-Kendall test for the 95% confidence level for NClimGrid data.
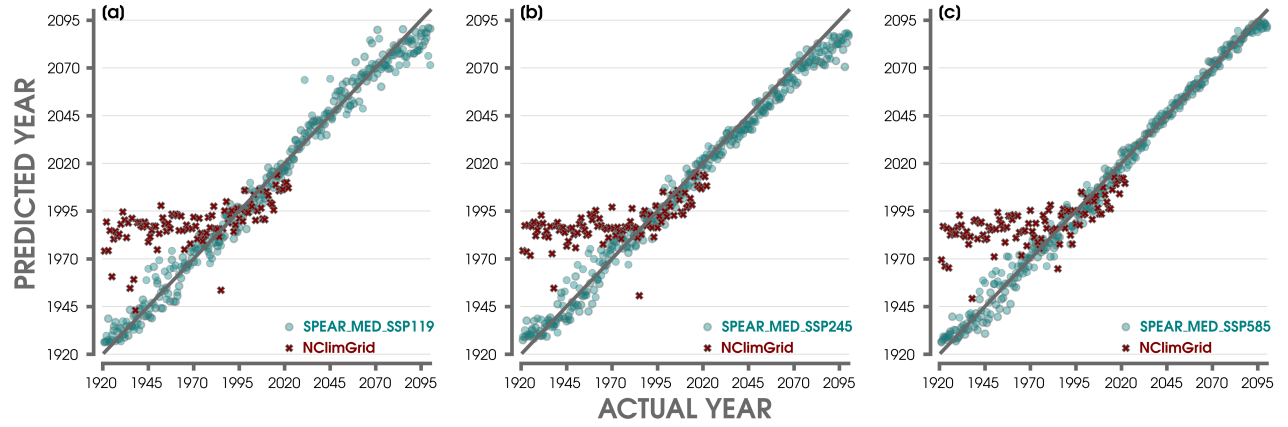
**Figure S5.** Time series of mean JJA TAVG anomalies over the CONUS from 1921-2022 for NClimGrid observations (dashed black line) compared to the ensemble mean of SPEAR_MED_NATURAL from 1921 to 2100 (solid gray line), the historical-only simulation of SPEAR_MED from 1921 to 2014 (solid yellow line), the future ensemble mean of SPEAR_MED following the SSP1-1.9 climate scenario from 2015 to 2100 (solid purple line), the future ensemble mean of SPEAR_MED following the SSP2-4.5 climate scenario from 2015 to 2100 (solid green line), and the future ensemble mean of SPEAR_MED following the SSP5-8.5 climate scenario from 2015 to 2100 (solid yellow line). The spread across SPEAR_MED ensemble members is shown with the lighter shading for each respective simulation. Anomalies are computed with respect to their 1921-1950 climatological time means.
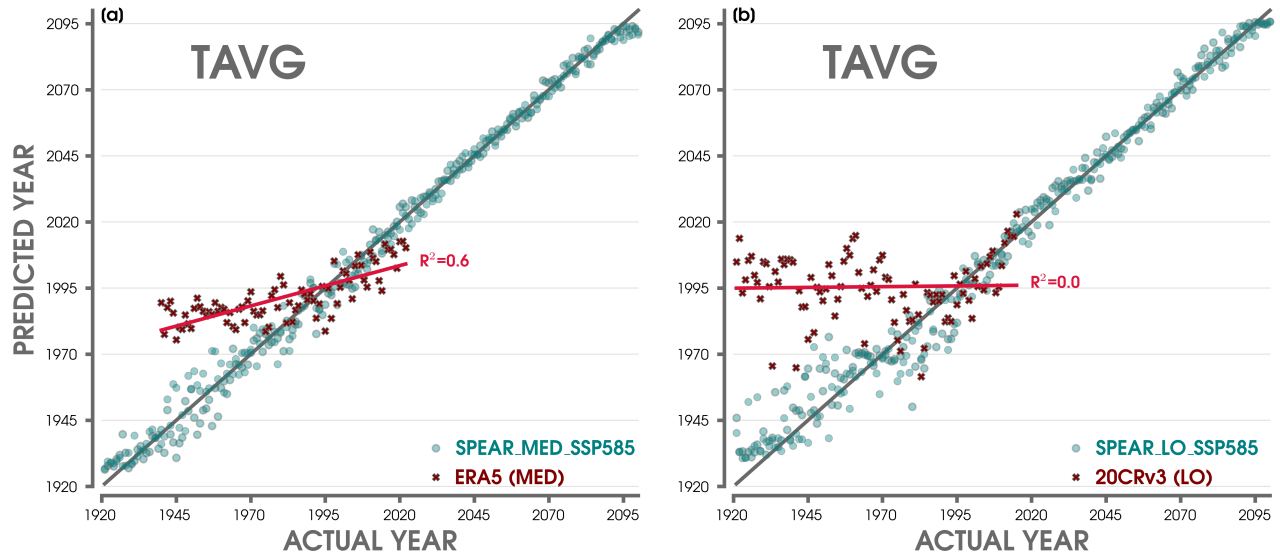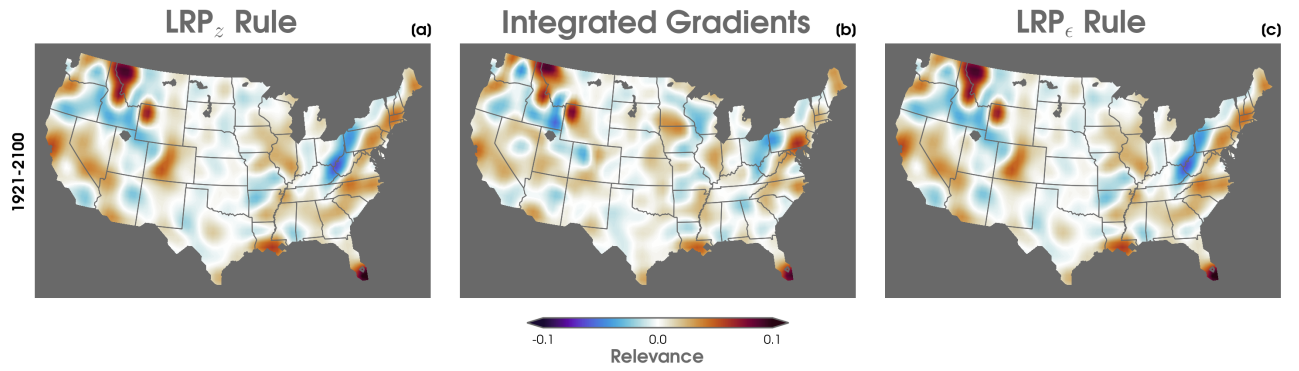
**Figure S6.** (a) Distribution of Mean Absolute Error (MAE) scores for predictions based on inputs of CONUS maps using different $L_2$ regularization values (0.001, 0.01, 0.1, 0.5, 1, 5) for an ANN architecture with 1 hidden layer and 10 nodes. The MAE is computed for validation data from the SPEAR_MED ensemble members over the 1921 to 2100 period. Each distribution of scores (blue points) is constructed from 20 ANN iterations (different combinations of training, testing, and validation ensemble members and random initialization seeds). The median score is shown with a thin white horizontal line. (b) As in (a), but for an ANN architecture with 1 hidden layer and 20 nodes. (c) As in (a), but for an ANN architecture with 2 hidden layers and 10 nodes each. (d) As in (a), but for an ANN architecture with 2 hidden layers and 30 nodes each. (e) As in (a), but for an ANN architecture with 2 hidden layers and 50 nodes each. (f) As in (a), but for an ANN architecture with 2 hidden layers and 100 nodes each. (g) As in (a), but for an ANN architecture with 3 hidden layers and 10 nodes each. (h) As in (a), but for an ANN architecture with 3 hidden layers and 20 nodes each.
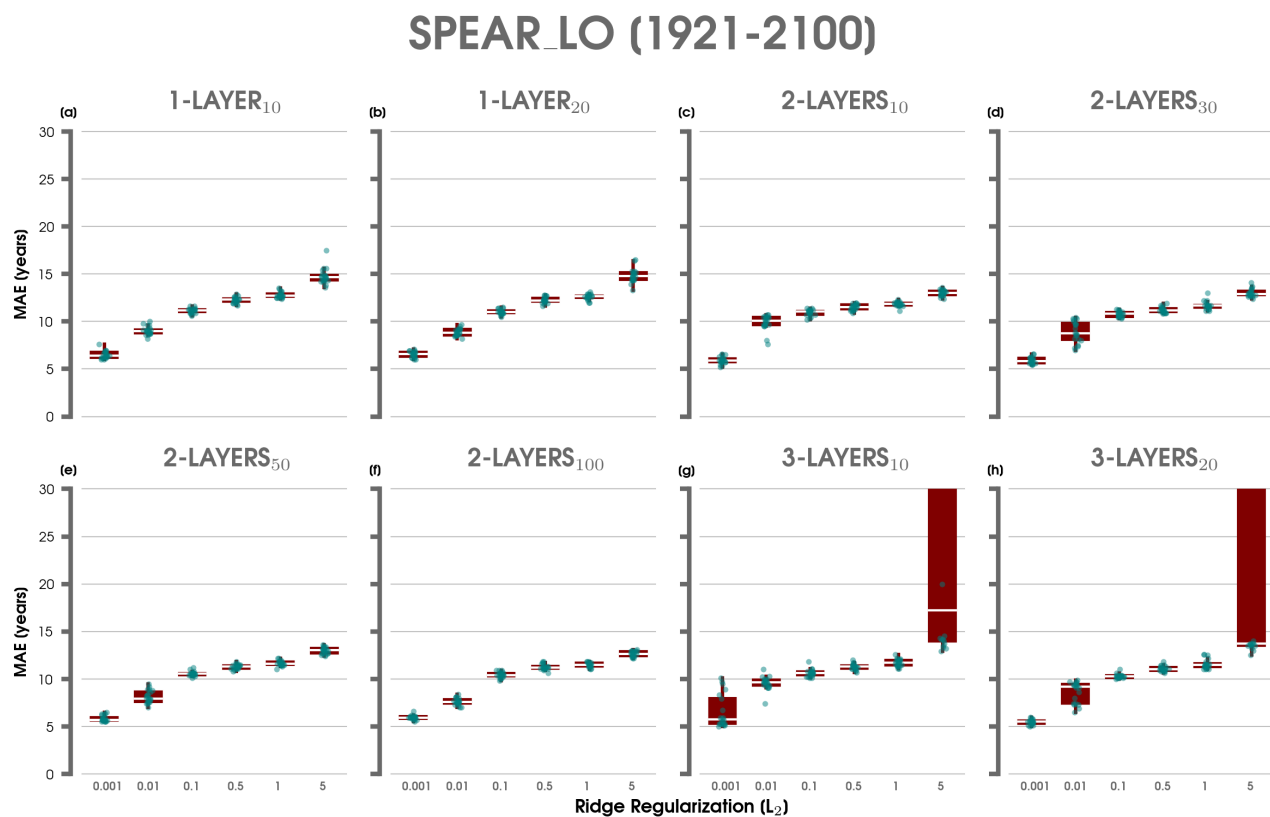
**Figure S7.** (a) ANN predictions of SPEAR_MED testing ensemble members for input maps of TAVG averaged over JJA with the future forcing of SPEAR_MED following the SSP1-1.9 pathway. The actual year is denoted on the x-axis and the predicted year on the y-axis. Red markers are shown for ANN predictions after inputting maps from NClimGrid. A perfect prediction (1:1 slope) is annotated behind all ANN predictions with a solid gray line. (b) As in (a), but for the future forcing of SPEAR_MED following the SSP2-4.5 pathway. (c) As in (a), but for the future forcing of SPEAR_MED following the SSP5-8.5 pathway.

**Figure S8.**   (a) ANN predictions of SPEAR_MED testing ensemble members for input maps of TAVG averaged over JJA. The actual year is denoted on the x-axis and the predicted year on the y-axis. Red markers are shown for ANN predictions after inputting maps from ERA5. A perfect prediction (1:1 slope) is annotated with a solid gray line, and a red line is displayed showing the linear least squares fit through the observational predictions along with its corresponding $R^2$ value. (b) As in (a), but for an ANN trained and tested on SPEAR_LO. Red markers are shown for ANN predictions after inputting maps from 20CRv3.
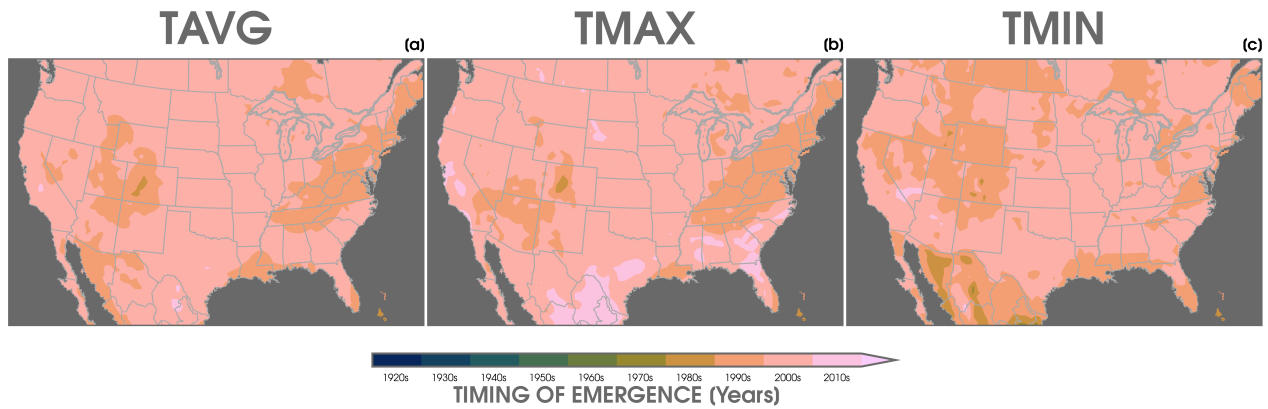
**Figure S9.** (a) Relevance heatmaps using the layerwise relevance propagation (LRP) z-rule (LRP$_z$), (b) Integrated Gradients method, and (c) LRP epsilon-rule (LRP$_\epsilon$) for testing ensemble members from SPEAR_MED composited over 1921 to 2100. The composited heatmaps are smoothed using a Gaussian filter to improve interpretability. Positive relevance indicates regions that pushed the ANN to make its predicted year. Negative relevance suggests areas that tried to push the ANN away from its predicted year.
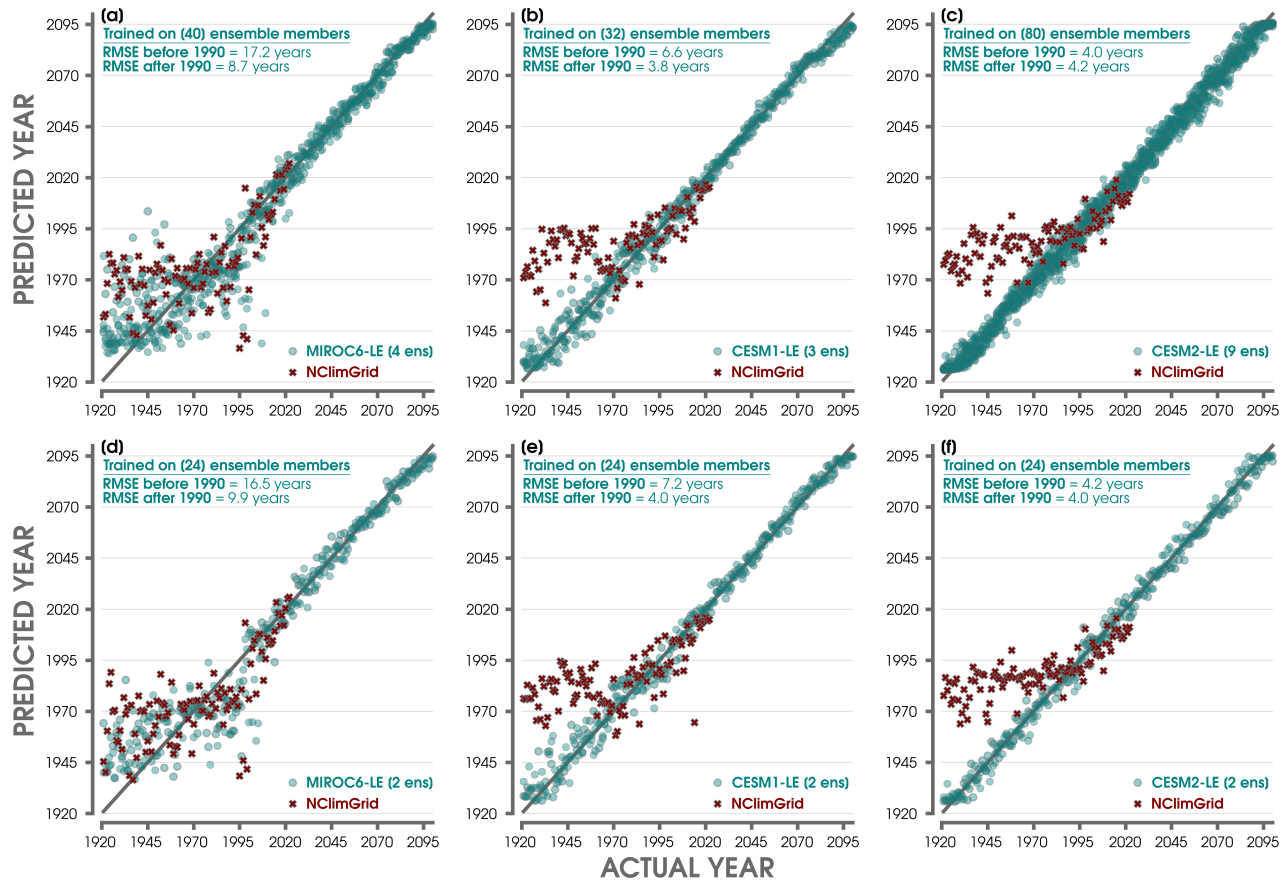
July 11, 2023, 9:23am

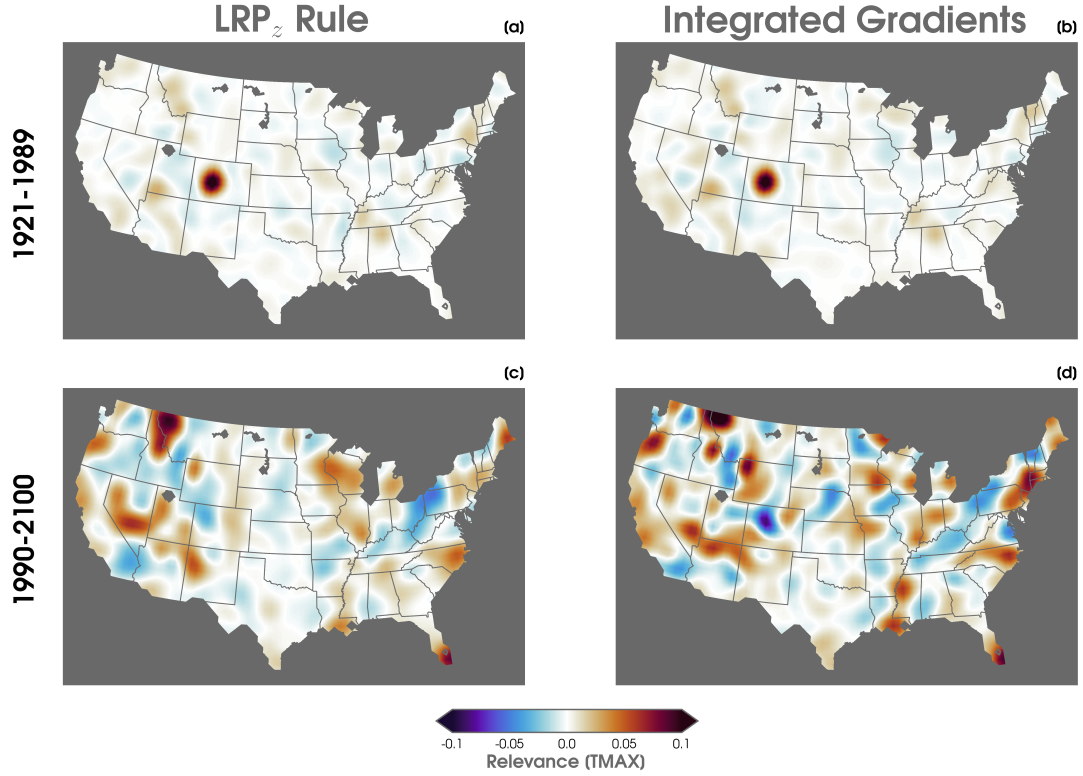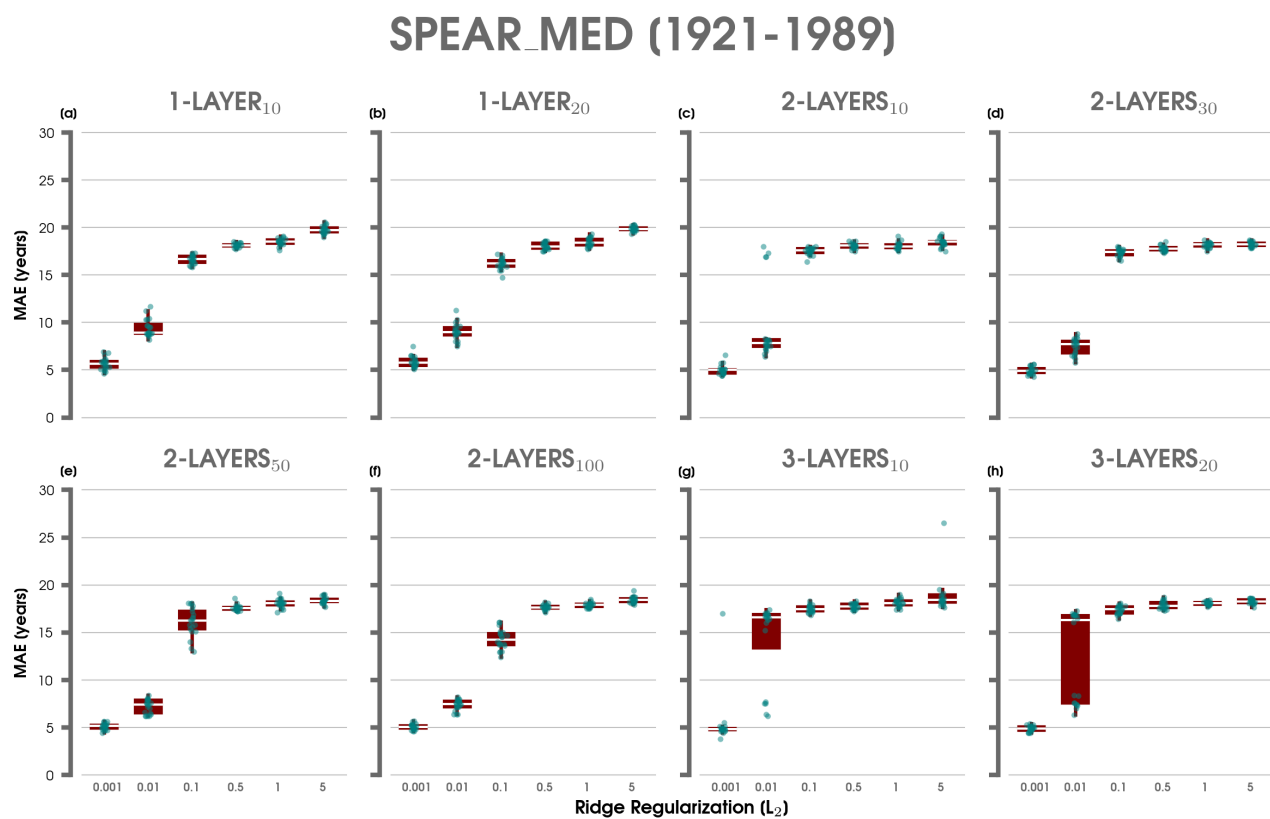**Figure S10.**  As in Figure S6, but for SPEAR_LO

**Figure S11.**   (a) Average JJA timing of emergence (ToE) maps, which are defined here as the first year the 10-year running-mean TAVG temperature exceeds and stays above the mean 1921-1950 period by more than two standard deviations (e.g., Lehner et al., 2017) for each ensemble member in SPEAR_MED (historical + SSP5-8.5). (b) As in (a), but for TMAX. (c) As in (a), but for TMIN.
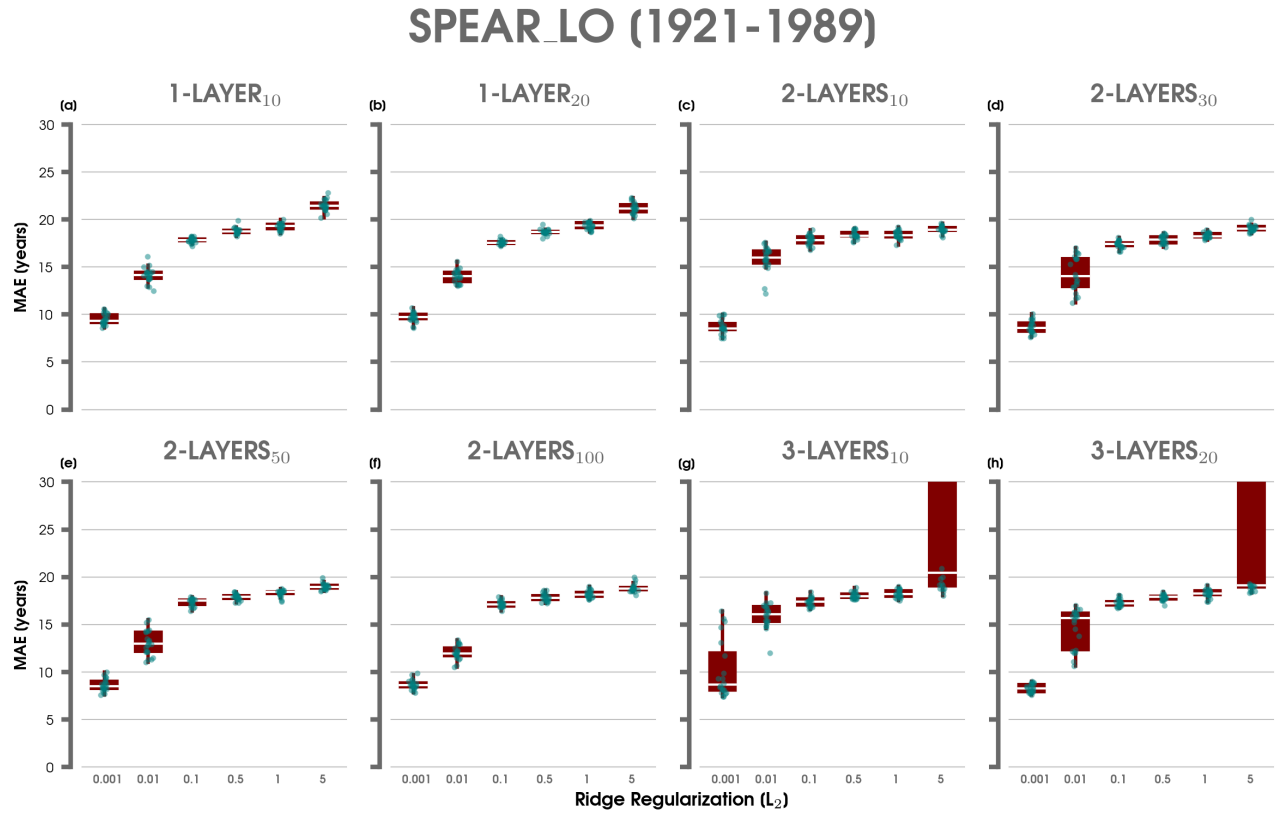
**Figure S12.**    (a) Predictions of MIROC6-LE's four testing ensemble members by an ANN trained on 40 MIROC6-LE ensemble members after inputting CONUS maps of TAVG averaged over JJA. The actual year is denoted on the x-axis and the predicted year on the y-axis. Red markers are shown for ANN predictions after inputting maps from NClimGrid. A perfect prediction (1:1 slope) is annotated behind all of the ANN predictions with a solid gray line. The Root Mean Squared Error (RMSE) for the MIROC6-LE testing ensemble members is included for predictions over the actual years of before and after the year 1990. (b) As in (a), but for predictions by three testing ensemble members for an ANN trained on 32 CESM1-LE ensemble members. (c) As in (a), but for predictions by nine testing ensemble members for an ANN trained on 80 CESM2-LE ensemble members. (d-f) As in (a-b), but for ANNs trained on only 24 ensemble members from each large ensemble. Testing predictions are displayed for two ensemble members each.
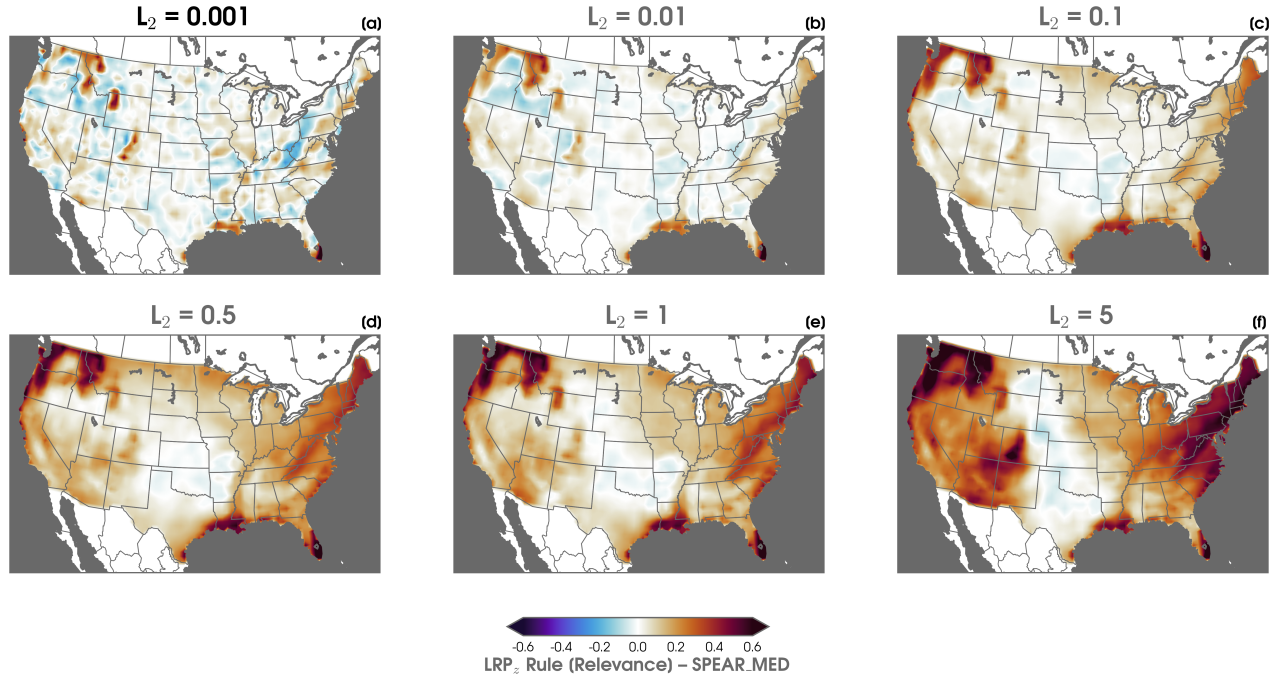
July 11, 2023, 9:23am

**Figure S13.** (a) Relevance heatmaps using the layerwise relevance propagation z-rule ($LRP_z$) and (b) Integrated Gradients method for the testing ensemble members from SPEAR_MED composited over 1921 to 1989 for TMAX. The composited heatmaps are smoothed using a Gaussian filter to improve interpretability. Positive relevance indicates regions that pushed the ANN to make its predicted year. Negative relevance suggests areas that tried to push the ANN away from its predicted year. (c-d) As in (a-b), but for composites over 1990-2100.
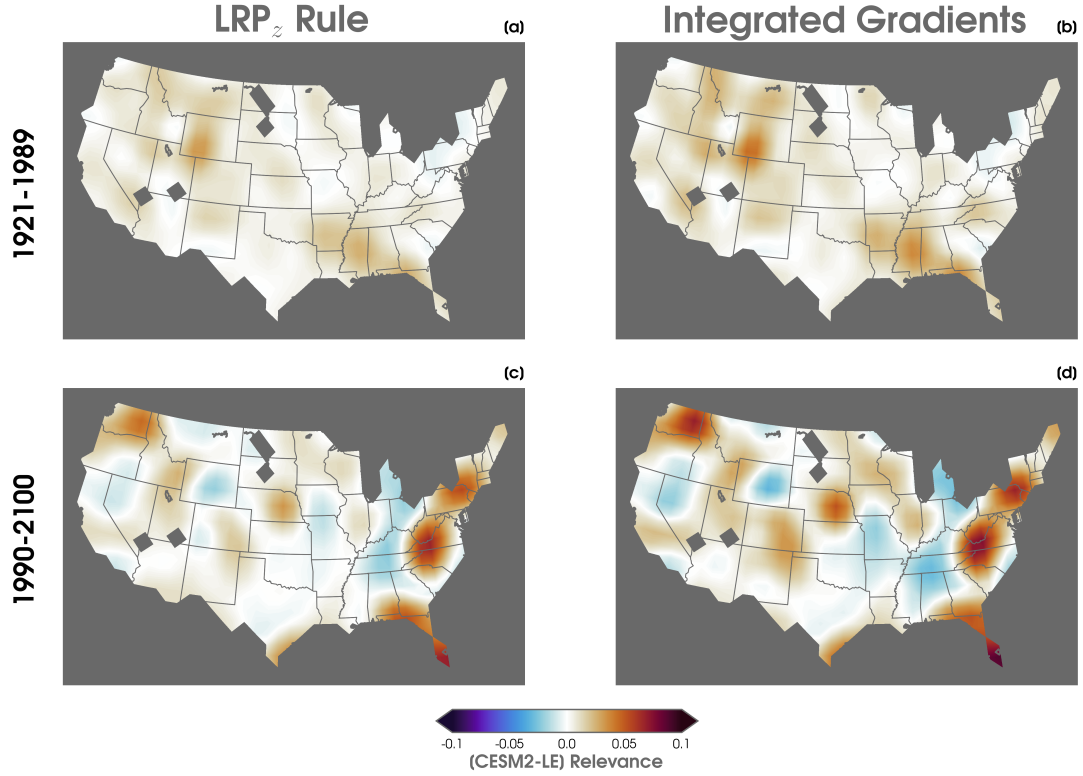
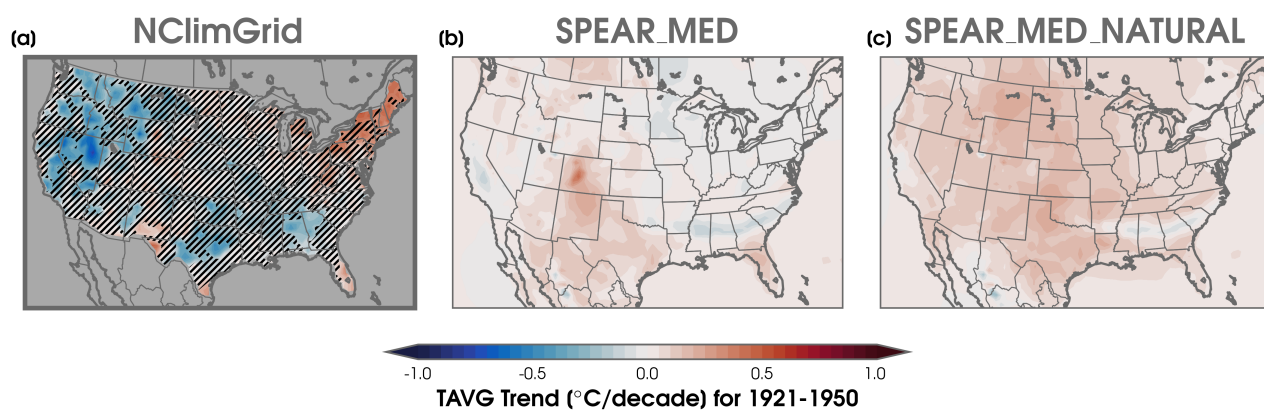**Figure S14.**    As in Figure S6, but for SPEAR_MED with the MAE calculated only over the 1921 to 1989 period.

**Figure S15.** As in Figure S6, but for SPEAR_LO with the MAE calculated only over the 1921 to 1989 period.

**Figure S16.** (a-f) Composite of $LRP_z$ heatmaps of TAVG using the testing ensemble members from SPEAR_MED, which are averaged over 1921 to 2100 for ANNs with different $L_2$ regularization values (0.001, 0.01, 0.1, 0.5, 1, 5). The $L_2$ value used for the main results of the paper is labeled in bold font. Positive relevance indicates regions that pushed the ANN to make its predicted year. Negative relevance suggests areas that tried to push the ANN away from its predicted year.

**Figure S17.** (a) Relevance heatmaps of TAVG using the layerwise relevance propagation z-rule (LRP$_z$) and (b) Integrated Gradients method for nine testing ensemble members from CESM2-LE (i.e., an ANN trained on 80 members from CESM2-LE) composited over 1921 to 1989. The composited heatmaps are smoothed using a Gaussian filter to improve interpretability. Positive relevance indicates regions that pushed the ANN to make its predicted year. Negative relevance suggests areas that tried to push the ANN away from its predicted year. (c-d) As in (a-b), but for composites over 1991-2100.
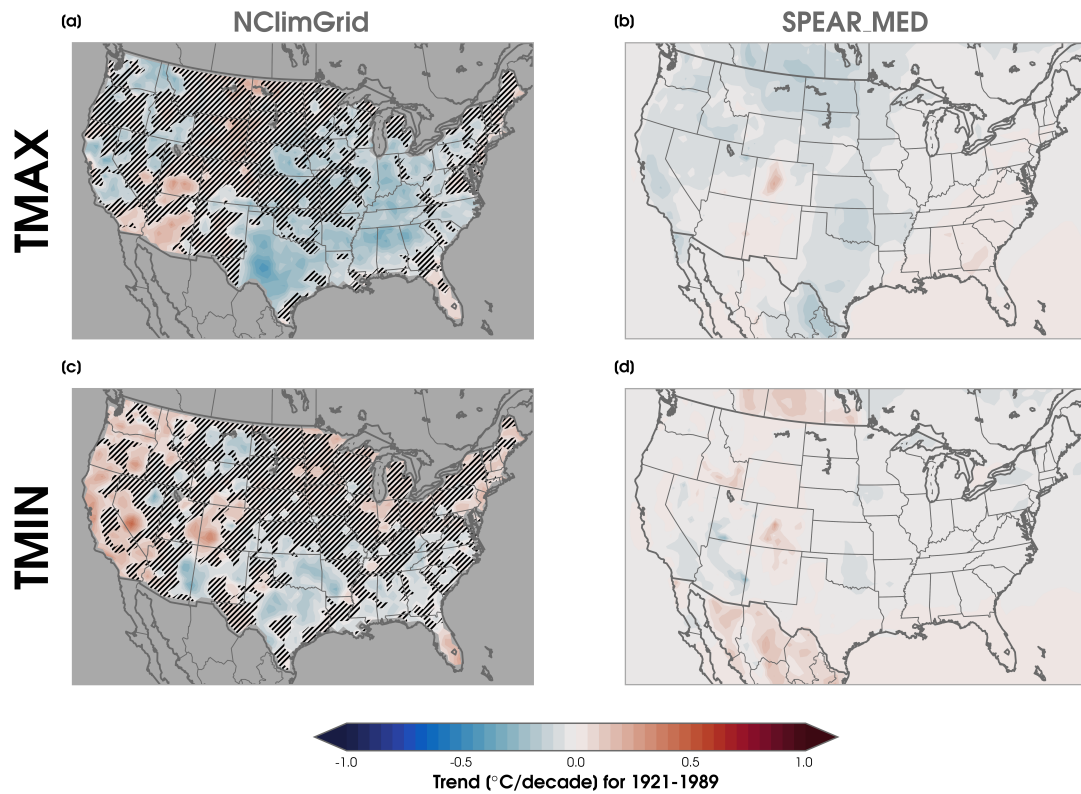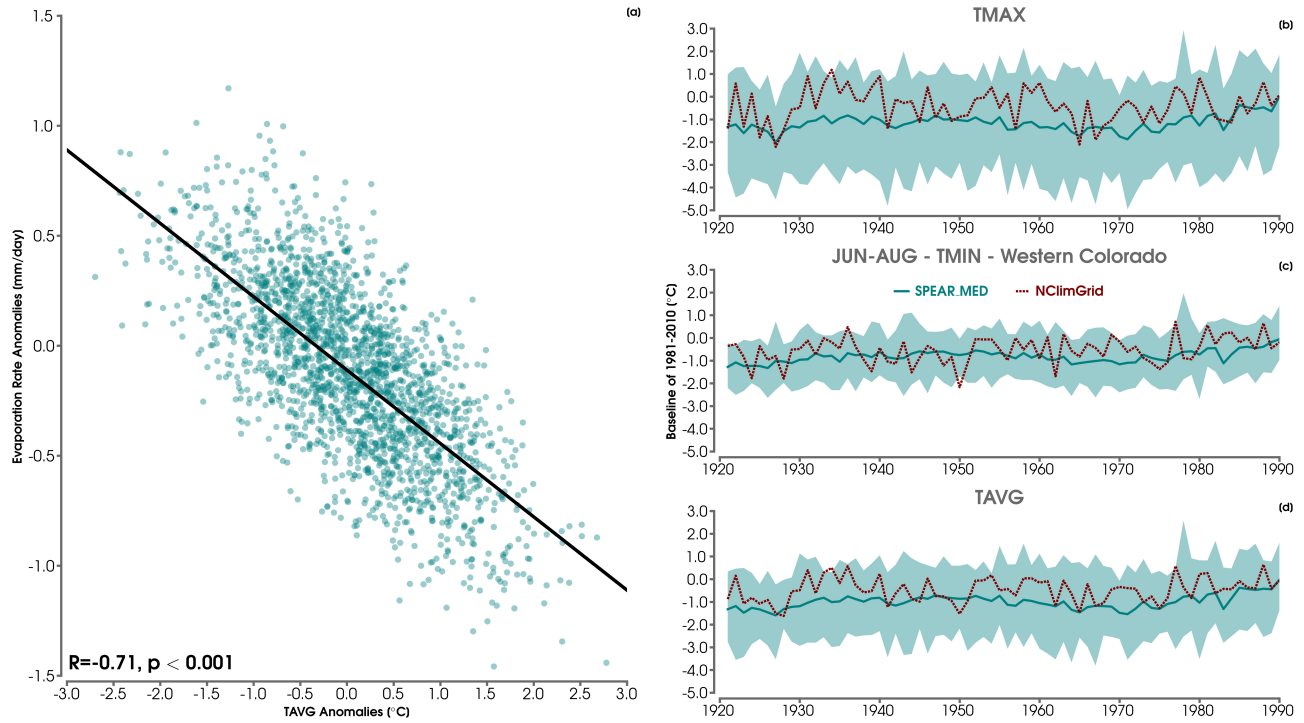
**Figure S18.**    As in Figure S3, but for trends over 1921 to 1950.
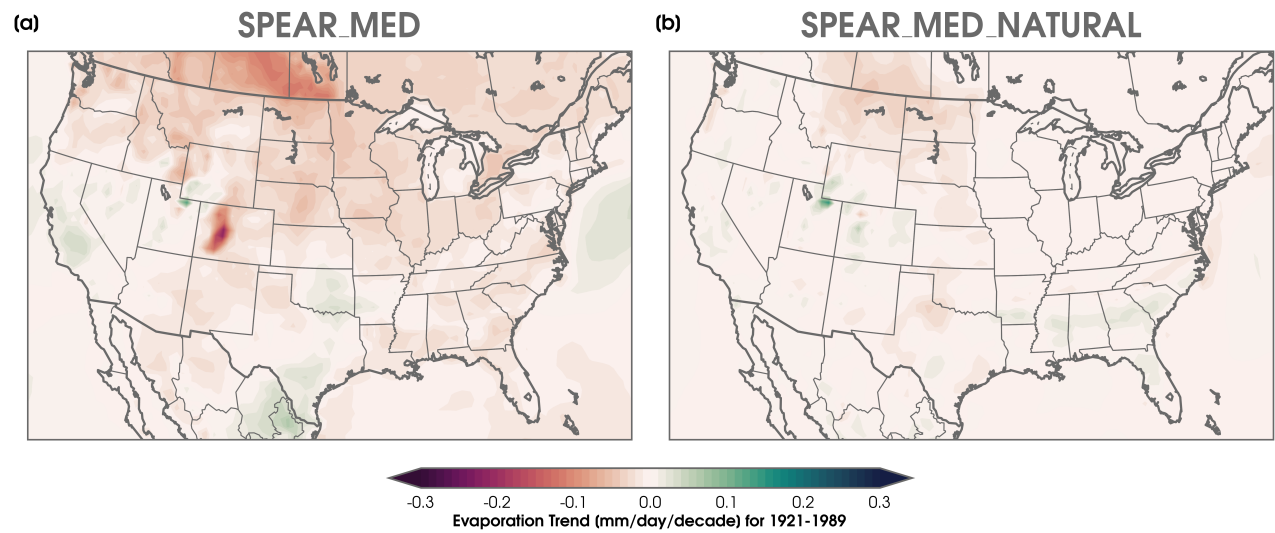
**Figure S19.** As in Figure S4, but for trends over 1921 to 1989.

**Figure S20.**     (a) Scatter plot of mean JJA TAVG anomalies versus mean JJA evaporation

rate anomalies averaged over western Colorado (see yellow box in Figure 2) for 1921-1989.  A

solid black line is displayed showing the linear least squares regression along with its correspond-

ing Pearson correlation coefficient and $p$-value.  Anomalies are computed with respect to their

1921-1950 climatological mean.  (b) Time series of mean JJA TMAX anomalies over western

Colorado from 1921 to 1990 for the ensemble mean of SPEAR_MED (dark green line) compared

to observations from NClimGrid (dashed red line).  The spread across SPEAR_MED ensemble

members is shown with the light green shading.  Anomalies are computed for each dataset with

respect to their own 1981-2010 climatological mean. (c) As in (b), but for the mean JJA TMIN.

(d) As in (b), but for the mean JJA TAVG.

**Figure S21.** (a) Linear least squares trends of average JJA evaporation rate from 1921 to 1989 for the ensemble mean from SPEAR_MED and (b) the ensemble mean from SPEAR_MED_NATURAL.

## References

Bevan, J. M., & Kendall, M. G. (1971). Rank correlation methods. *The Statistician*. doi: 10.2307/2986801

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

Hersbach, H., Bell, B., Berrisford, P., Hirahara, S., Horányi, A., Muñoz-Sabater, J., . . . Thépaut, J.-N. (2020, 5). The era5 global reanalysis. *Quarterly Journal of the Royal Meteorological Society*. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3803` doi: 10.1002/qj.3803

Kay, J. E., Deser, C., Phillips, A., Mai, A., Hannay, C., Strand, G., . . . Vertenstein, M. (2015, 8). The community earth system model (cesm) large ensemble project: A community resource for studying climate change in the presence of internal climate variability. *Bulletin of the American Meteorological Society*, *96*, 1333-1349. Retrieved from `http://journals.ametsoc.org/doi/10.1175/BAMS-D-13-00255.1` doi: 10.1175/BAMS-D-13-00255.1

Lehner, F., Deser, C., & Terray, L. (2017). Toward a new estimate of "time of emergence" of anthropogenic warming: Insights from dynamical adjustment and a large initial-condition model ensemble. *Journal of Climate*, *30*. doi: 10.1175/JCLI-D-16-0792.1

Mann, H. B. (1945). Nonparametric tests against trend. *Econometrica*. doi: 10.2307/1907187

Rodgers, K. B., Lee, S.-S., Rosenbloom, N., Timmermann, A., Danabasoglu, G., Deser, C., . . . Yeager, S. G. (2021, 12). Ubiquity of human-induced changes in climate variability.

*Earth System Dynamics*, *12*, 1393-1411. Retrieved from `https://esd.copernicus.org/articles/12/1393/2021/` doi: 10.5194/ESD-12-1393-2021

Shiogama, H., Tatebe, H., Hayashi, M., Abe, M., Arai, M., Koyama, H., ... Watanabe, M. (2023). Miroc6 large ensemble (miroc6-le): experimental design and initial analyses. *Earth Syst. Dynam. Discuss.*. Retrieved from `https://doi.org/10.5194/esd-2023-12` doi: 10.5194/esd-2023-12

Slivinski, L. C., Compo, G. P., Sardeshmukh, P. D., Whitaker, J. S., McColl, C., Allan, R. J., ... Wyszynski, P. (2021, 2). An evaluation of the performance of the twentieth century reanalysis version 3. *Journal of Climate*, *34*, 1417-1438. Retrieved from `https://journals.ametsoc.org/view/journals/clim/34/4/JCLI-D-20-0505.1.xml` doi: 10.1175/JCLI-D-20-0505.1

Slivinski, L. C., Compo, G. P., Whitaker, J. S., Sardeshmukh, P. D., Giese, B. S., McColl, C., ... Wyszyński, P. (2019, 10). Towards a more reliable historical reanalysis: Improvements for version 3 of the twentieth century reanalysis system. *Quarterly Journal of the Royal Meteorological Society*, *145*, 2876-2908. Retrieved from `https://onlinelibrary.wiley.com/doi/abs/10.1002/qj.3598` doi: 10.1002/qj.3598

Vecchi, G. A., Delworth, T., Gudgel, R., Kapnick, S., Rosati, A., Wittenberg, A. T., ... Zhang, S. (2014, 11). On the seasonal forecasting of regional tropical cyclone activity. *Journal of Climate*, *27*, 7994-8016. Retrieved from `https://journals.ametsoc.org/view/journals/clim/27/21/jcli-d-14-00158.1.xml` doi: 10.1175/JCLI-D-14-00158.1

Vose, R. S., Applequist, S., Squires, M., Durre, I., Menne, C. J., Williams, C. N., ... Arndt, D. (2014, 5). Improved historical temperature and precipitation time series for u.s. climate divisions. *Journal of Applied Meteorology and Climatology*, *53*, 1232-1251.

55    Retrieved from `https://journals.ametsoc.org/view/journals/apme/53/5/jamc-d-13`

56    `-0248.1.xml`  doi: 10.1175/JAMC-D-13-0248.1