PNAS
www.pnas.org

Supplementary Information for

Increased drought severity tracks warming in the United States' largest river basin

Justin T. Martin[1], Gregory T. Pederson[1], Connie A. Woodhouse[2,3], Edward R. Cook[4], Gregory J. McCabe[5], Kevin J. Anchukaitis[2,3], Erika K. Wise[6], Patrick Erger[7], Larry Dolan[8], Marketa McGuire[9], Subhrendu Gangopadhyay[9], Katherine Chase[10], Jeremy Littell[11], Stephen Gray[11], Scott St. George[12], Jonathan Friedman[13], Dave Sauchyn[14], Jeannine St. Jacques[15], and John King[16]

[1]U.S. Geological Survey, Northern Rocky Mountain Science Center, Bozeman, MT, USA
[2]School of Geography and Development, University of Arizona, Tucson, AZ, USA
[3]Laboratory of Tree-Ring Research, University of Arizona, Tucson, AZ, USA
[4]Lamont-Doherty Earth Observatory, Palisades, New York, USA
[5]U.S. Geological Survey, Water Resources Division, Denver, Colorado, USA
[6]Department of Geography, University of North Carolina, Chapel Hill, NC, USA
[7]U.S. Bureau of Reclamation, Great Plains Regional Office, Billings, MT, USA
[8] Montana Department of Natural Resources and Conservation, Helena, MT, USA
[9] U.S. Bureau of Reclamation, Technical Service Center, Denver, CO, USA
[10]U.S. Geological Survey, Wyoming-Montana Water Science Center, Helena, MT, USA
[11]U.S. Geological Survey, Alaska Climate Adaptation Science Center, Anchorage AK, USA
[12]Department of Geography, Environment and Society, University of Minnesota, Minneapolis, MN, USA
[13]U.S. Geological Survey, Fort Collins Science Center, Ft. Collins, CO, USA
[14]Department of Geography and Environmental Studies, University of Regina, Regina, Saskatchewan, Canada
[15]Department of Geography, Planning and Environment, Concordia University, Montreal, Quebec, Canada
[16]Lone Pine Research, Bozeman, MT, USA

Corresponding author: Justin Martin (justinmartin@usgs.gov)

This PDF file includes:

Supplementary text S1-S4
Figures S1 to S14
Tables S1 and S2
SI References

Supplemental Information Text

S1. Cross-validation of modeled temperature

The nested Composite-Plus-Scale (CPS) method was used to reconstruct regional mean runoff-season (Mar-Aug) temperatures from a network of high-elevation, temperature sensitive ring-width and maximum density (MXD) tree-ring data (Fig. S1 and S2). In this approach, all of the available tree-ring data are normalized and averaged to create a mean composite proxy series. This single series is then scaled to the mean and standard deviation of the instrumental data over the calibration period, and reconstruction skill is evaluated based on the residuals. We used a calibration period of 1930 to 1976 (bounded by the last year of the common period covered by all the available tree-ring chronologies) and a validation period of 1901 to 1930 (bounded by the first year of available PRISM temperature data). In a nested CPS approach, the reconstruction is extended back in time by recalculating the composite and rescaling it to the instrumental temperature data iteratively each time there is a change in sample size. This allows the reconstruction to extend further back in time while also accounting for changes in sample size, reconstruction accuracy, and the resulting change in reconstruction calibration and validation skill.

The reconstruction of regional runoff-season temperatures is skillful back to approximately 1500 CE (Fig. S2). Importantly, both the Reduction of Error (RE) and Coefficient of Efficiency (CE) statistics remain positive through this period. Prior to 1500, the loss of proximal MXD chronologies to our study region results in a temperature reconstruction without sufficient skill for further statistical analyses. The reconstruction, however, is still plotted alongside the streamflow and drought deficits prior to 1500 since the sign of decadal temperature anomalies likely retains enough useful information to illustrate whether a drought event occurred under generally cool or warm conditions.

S2. Assessing the sensitivity of comparing droughts over time to methodological choices

Drought deficits are presented here as a calculation based on standardized streamflow values (*z*-scores) but can be calculated on observed streamflow values as well (units of cubic feet per second [CFS]). In generating the UMRB composite record, taking the average of the former may yield a different result from the latter in the case where flow time series of very different magnitudes are averaged. Calculating drought deficits on observed streamflow values more closely approximates the mass dynamics of a drought estimate that is akin to the total flow volume measured at a single terminal gaging location at the outlet of the basin, which is non-existent in the UMRB. To determine how this affects our estimate of basin-wide deficits, we also quantified deficits based on the UMRB composite record generated by taking the simple average of the streamflow records in CFS units. In this case, the temporal dynamics of individual large rivers can be expected to have a

greater effect on the basin wide average than individual small rivers. Figure S3 depicts this deficit record calculated on streamflow values in CFS. In this alternative approach, the decadal deficits of the Turn-of-the-Century Drought remain substantially more severe than those of other droughts in the record.

Additionally, to contextualize the Turn-of-the-Century drought with paleoclimatic and paleohydrologic records, it was necessary to splice the basin-wide composite streamflow reconstruction (which ends in 1998) to the observed record that runs through 2010. This splice occurred at the year 1930 in order to include the dry years of the 1930s as reflected in the actual gage records. Although the streamflow reconstruction is mean and variance scaled to its respective observational record over the common period, splicing the instrumental and reconstructed records inevitably introduces bias at the transition of record types in that the reconstruction itself will fail to perfectly capture the variation in the observed hydrology. Thus, we attempted to assess potential bias that existed in the reconstruction relative to the observational record.

In this bias assessment, we first quantified the error in the prediction of low flow values in the streamflow reconstruction versus the observed record over the common period for each gage in the composite record. We did so by comparing reconstructed and observed flows from years where flow values were more than 1 s (standard deviation) below the mean flow level. The mean low-flow level in the reconstruction matched that of the observed record very closely over the calibration period such that the mean low-flow level for the reconstruction was 0.006 s lower than the observed flow. To determine the effect this very small underestimation of observed low flows had on the full-length drought deficit record, we explicitly aligned the reconstructed and observed flow records based on their respective low-flow means. This places a greater emphasis on the accuracy of the lowest reconstructed flows in comparing drought years between reconstructed and observed records. Even under this scenario, where the reconstruction mean is matched to the observed over only the driest years of the calibration period, the peak severity of Turn-of-the-Century Drought still appears unprecedented over the last 1200 years (Fig S4).

Also, because we focused on decadal-scale variability in drought severity, we filtered higher frequency variability (< 10 years) from the long-term streamflow record using the cubic smoothing spline of streamflow anomalies with a spline length of 10 years and a frequency response of 0.5. To avoid the potential for end-effects to exaggerate the severity of the Turn-of-the-Century Drought, we padded the end of the time-series with flow values set to the maximum observed flow in the 1200 year record for a number of years equal to half the length of the spline. Thus, the severity of the Turn-of-the-century drought is almost certainly underestimated due to the application of a maximum flow value padding.

Since the choice of filtering method affects both the definition and magnitude of drought events, we explored how calculating drought severity based on splines of

lengths from five to fifteen years affected the comparison of droughts over time. Figure S5 shows the drought severity records for each of these spline lengths, demonstrating that Turn-of-the-Century Drought is more severe than other droughts on record when assessed over periods of various lengths from five to twelve years. At spline lengths of 13 years and beyond, it is likely that the 13th century drought represents a drought event of greater persistent severity than the Turn-of-the-Century Drought, as is to be expected from an event over three times as long. However, because fitting longer splines to the streamflow record requires more maximum flow value padding of the recent years, the underestimation of drought severity during the Turn-of-the-Century Drought also increases when assessed via longer splines. Future assessments of changing drought dynamics through the 21st century in the UMRB will require updated naturalized streamflow records.

Finally, to test another method that integrates information on both drought severity and duration, we also calculated the intensity of each drought in the 1200-year record as the cumulative deficit / length of the drought. We used the same drought definition (two or more consecutive negative-anomaly years in the 10-year spline of streamflow) but calculated the intensity of each drought event directly from the un-smoothed annual flow anomalies. Again, the Turn-of-the-Century drought was ranked as significantly more intense than any other drought event on record (Fig. S6).

S3. Assessing the sensitivity of relationships between temperature, precipitation, and streamflow to uncertainties in gridded climate data

Numerous gridded estimates of precipiation and temperature for the western United States are now available for various time periods since the late 1800s. It is well known that all gridded climate datasets are inherently uncertain owing to issues that may include measurement errors in underlying data (1, 2), inhomogeneities in individual or network records (3–5), and methodological choices specific to the development of each dataset (6). Additionally, the sparsity of current and espescially earlier observations from which gridded datasets are derived is an inherent limitation in all gridded products (7).

While a certain lack of independence between all climate datasets exists, we attempted to determine the effect of our choice of climate data (i.e. PRISM) on the findings of this study to the degree possible by carrying out the hydroclimatic analyses reported herein on homoginized climate datasets of equivalent length to PRISM (8), and on a modified version of the PRISM precipitation dataset that extends back to 1950. Modifications made to the orgininal PRISM dataset are from Luce et al. (9), and were developed to account for potential biases that may, or may not, be present in the precipitation dataset related to reduced orgraphic precipitation due to a reduction in 700 hectopascal (hpa) zonal (U700) windspeeds. These efforts are described in detail below.

In the construction of the PRISM gridded precipitation dataset, the 4x4 km grid cells are first interpolated from observations (i.e met stations) using a local regression function informed by climatological normals for the 1981-2010 period (10). This local regression approach accounts for the influence of lapse rates (and other variables) due to topography on the interpolation of monthly climate values between distant obervation locations. Based on the work by Luce et al., 2013 (9) covering the Pacific Northwest region, we recognize the possibility that the use of 1981-2010 climatology may not accurately reflect the relationship between topography and precipitation prior to the early 1980s in light of evidence suggesting that winter zonal 700 hpa westerly winds over portions of the region have declined since the mid-20th century. This is because during winter the speed of lower tropospheric winds is strongly correlated with precipitation in mountainous regions conducive to orographic enhancement of precipitation. Thus decreasing westerly winds in recent decades may have decreased mountain precipitation, and this trend may not be accurately reflected in gridded precipitation records due to the lack of high-elevation weather stations in the region. However, the likelihood that this situation exists in the UMRB is highly uncertain as the work of Luce et. al., 2013 focused primarily on the Pacific Northwest, not the UMRB. Additonally, hydrologically significant moisture influxes into the the Upper Missouri over both the cool and warm seasons occur from a greater diversity of source regions and stormtracks (11)(e.g. wintertime northwesterly flow patterns, springtime moisture advection off the Gulf of Mexico via the low-level jet and cutoff lows) than the dominant zonal flow pattern responsible for the majority of the moisture delivery across the the Columbia Basin (i.e. the Pacific Nortwest). Climatically important sources of moisture aside, since Luce et al. (9) do show correlations between U700 winds and precipitation in some less rain-shadowed catchments in the far western portion of the UMRB it is worth invetigating the potential influence such a procces could have on our results.

To understand how an underestimation of mountain precipitation in the UMRB prior to the early 1980s could affect our analyses, we used the standardized regression slopes for the effect of U700 windspeed on precipitation from Luce et al. (9) to calculate the ammount of precipitation that could hypothetically be missing from PRISM grids in each year prior to 1981. We were able to perform this analysis from 1950 forward based on the reported decrease in average U700 winds of 0.19 m/s/decade. We estimated a standardized regression slope of $\sim$0.12 fraction precipitation per m/s for the reported stations that fall within the UMRB. We calculated the hypothetical underestimation of precipitation in each year as the total annual precipitation times 0.019, times the number of years before 1981 times 0.12, times the mean of 1981-2010 precipitation, then added that quantity back to the total precipitation for each year (Fig. S7). Because we carried out this exercise on HU 8 watershed-aggregated PRISM grids, all grid cells in the UMRB headwater basins were effectively adjusted wet in the earlier decades (not just those mountain grids with the most significant potential for orographic enhancement). The adjustment to PRISM precipitation likely over-corrects for any actual

underestimation of precipitation that may be present in PRISM data prior to 1981. This essentially biases the precipitation of the 1981 – 2010 period dry relative to earlier periods, creating a situation in which precipitation appears to have more of an increasing influence on streamflow generation than it does if PRISM precipitation is left unaltered.

Even in this rather extreme hypothetical situation, the influence of temperature on UMRB streamflow during the drought periods of the late 1980s and early 2000s has been more negative and of greater magnitude than during previous decades (Fig. S8). As expected, the primary influence of adusting PRISM wet in earlier decades is to assign slightly more importance to the effect of precipitation on recent droughts and slightly less importance on recent pluvials. Nonetheless, a clear shift to an increasingly negative forcing of temperature on streamflow is evident in most sub-basins since the 1980s, and is coincident with very dry conditions in the late 1980s and early 2000s (Fig. S8).

Adjusting PRISM precipitation data for hypothetical precipitation biases also affects our estimates of UMRB runoff efficiency over time. Although doing so shortens the available period of record for the analysis to 1950 – 2010, figure S9 shows that even with higher precipitation estimates prior to 1981, runoff efficiency remains significantly lower during the period from 1984 – 2010 than during the preceding three decades (p<0.001).

An additional uncertainty in PRISM data is the known existence of inhomogeneities in some observational networks and records underlying many gridded products including the PRISM dataset (3, 4). Some level of inhomogeneity over time in portions of the available station data network is likely due to climate-independent changes in station site conditions, instrumentation, and location. These issues can be difficult or impossible to quantify, however, several gridded climate datasets now exist that employ pairwise homogenization techniques to limit the influence of potential inhomogeneities in an effort to make them more suitable for trend and long-term analyses (8, 12, 13). The nCLIMDIV dataset by Vose et al. (10) is directly comparable to the PRISM dataset in terms of temporal coverage (1895 to 2010) and employs such a homogenization technique. Thus, while the nCLIMDIV and PRISM datasets are not independent of each other owing to shared underlying data and certain methodologies, comparison of the two allows an understanding of the potential effects of inhomogeneities in PRISM data on trends reflected by that dataset. Figure S10 shows the comparison of the UMRB aggregated runoff-season temperature and water-year precipitation anomalies from the PRISM and nCLIMDIV datasets. Though the PRISM dataset shows slightly more warming and wetting over the period of record, the differences in trend are not statistically significant, with 95% confidence bounds encapsulating each other at all points in time. Strong fidelity between interannual to multidecadal anomalies in basin-wide precipitation and temperature is also evident with correlations between the two datasets of r = 0.96 and r = 0.98 respectively, with p < 0.001 in both cases.

We also assessed the influence of precipitation and temperature on streamflow using the nCLIMDIV dataset, as was done using PRISM (Fig. S11). Compared with PRISM, the nCLIMDIV data highlight a slightly greater influence of temperature on the low flows of the Dustbowl drought, in particular the driest years of that drought. The nCLIMDIV data also highlight the importance of precipitation in driving the lowest flows of the Turn-of-the-Century Drought, while demonstrating the sustained shift to negative temperature forcing of flow since 1984 and the importance of temperature in sustaining the Turn-of-the-Century drought through the second half of the 2000s, when UMRB precipitation was actually above its long-term trend line (Figs. S7, S10, & S11).

Finally, period comparisons of runoff efficiency based on PRISM were repeated using the nCLIMDIV dataset and are shown in Figure S12. In all cases, the distributions of temperature and precipitation are significantly different across time with later years of the 1984-2010 period and the Turn-of-the-Century Drought displaying higher temperatures and/or lower runoff efficiency than earlier periods ($p < 0.029$, all cases).

S4. Assessing the role of temperature, precipitation and evapotranspiration in driving trends in UMRB streamflow using a monthly water balance model

A simplified model for the generation of streamflow (Q) in the UMRB can be expressed as Q=P-ET, where P reflects precipitation and ET reflects evapotranspiration. Using PRISM precipitation anomalies and natural flow anomalies, we described decreasing UMRB runoff efficiency (RE) in recent decades (Fig. 5 a-c) where RE is expressed as Q-P. Given this simplified framework, decreasing RE implies increasing ET relative to P and establishes those increases in ET as a likely contributor to recent droughts in the basin. However, increases in ET implied by decreasing RE cannot be independently verified by analyses based on streamflow and precipitation data alone. In order to provide a process-based assessment of ET over the UMRB we used a well-verified monthly water balance model (MWBM) (14) to estimate ET and Q from UMRB precipitation and temperature data for comparison with those results based only on PRISM precipitation and natural streamflow data.

All MWBM records considered here are derived from aggregated precipitation and temperature data for the 36 HU 8 level watersheds listed in table S2, averaged first by sub-basin and water-year, and then by water-year for the entire basin. This reflects the same averaging approach used to generate the basin-wide estimate of natural streamflow from the 17 gages depicted in Figure 1. Figure S13 shows the model input data (precipitation and temperature) along with model output (ET and runoff). For comparison with MWBM runoff, runoff modeled by multiple linear regression (MLR) using water-year temperature, precipitation, and snow water equivalent (SWE) estimates are shown along with the basin-wide natural flow estimate. Both MWBM and MLR estimates were generated from raw PRISM data as well as U700 adjusted PRISM precipitation data (see supplemental text S3 for details

on U700 adjustment). For this reason, the comparisons shown in figure S13 are limited to the timeframe from 1950-2010, over which estimates of U700 wind speeds are available.

Since we expected increases in ET to be associated with decreases in RE as described above, we also directly examined the relationship between the two variables over time (Fig. 5, d and e; Fig S14). Here again, both raw PRISM and U700 adjusted PRISM precipitation data were used. To make this analysis comparable in time to that shown in Fig 5, a-c of the main text, we made two different comparisons of ET and RE in the URMB over time which are depicted in Figure S14. The first compares water-year values during the Dust Bowl drought to those during the Turn-of-the-Century Drought (Fig 5, d; Fig S14, a). The second compares the water-year values from 1906-1983 to those from 1984-2010 (Fig 5, e and Fig S14, b). Here it should be noted that because the analysis extends prior to 1950, U700 adjustment of PRISM precipitation was uncertain from 1906-1949. For these years, we used the level of adjustment used in 1950 as a constant level applied to the beginning of the record, which implies no further trend in U700 wind speeds back in time prior to 1950.

Several important observations in climate, modeled runoff, and natural streamflow trends are notable in Figure S13. First, a negative trend in natural streamflow (panels f,l) is evident over the 1950-2010 period, and runoff modeled by MLR and the MWBM also exhibit negative trends regardless of whether raw or U700-adjusted PRISM precipitation data are used (panels d,e,j,k). Over the same period, precipitation exhibits a slight positive trend (raw PRISM) (panel a), or no trend (U700 adjusted PRISM) (panel g), although all basin-wide trends in precipitation are insignificant at the 95% confidence level. Temperature and ET, on the other hand, exhibit significant positive trends over the period (panels b,c,h,i). Because runoff within the MWBM is roughly estimated as Q=P-ET (see McCabe and Wolock, 2011 for details), negative trends in runoff associated with a positive trend, or lack of trend in precipitation implies increased ET over time as estimated by the model.

Increasing ET in the UMRB is shown to be negatively associated with RE (Fig 5, d and e), and this is true for the U700 adjusted PRISM dataset as well (Fig. S14). Given the relationships discussed above, it follows that the lower runoff efficiencies observed in natural streamflow since 1984, and especially during the Turn-of-the-Century Drought, should be associated with increased ET relative to earlier periods. Kernel density estimates for the distributions of ET and RE based on MWBM output (Fig 5 d and e margins; Fig. S14 margins) reaffirm the tendency towards increased ET and decreased RE over the 20th and early 21st century implied by the estimates of RE based on precipitation and natural flow data alone (Figure 5, a-c). When assessed at the level of the individual HU 8 watersheds of the UMRB (Table S1), modeled runoff efficiency was significantly lower and modeled ET was significantly higher during the Turn-of-the-Century Drought relative to the Dust Bowl ($p<0.026$ all cases), and since 1984 relative to the 1906-1983 period ($p<0.001$ all cases).

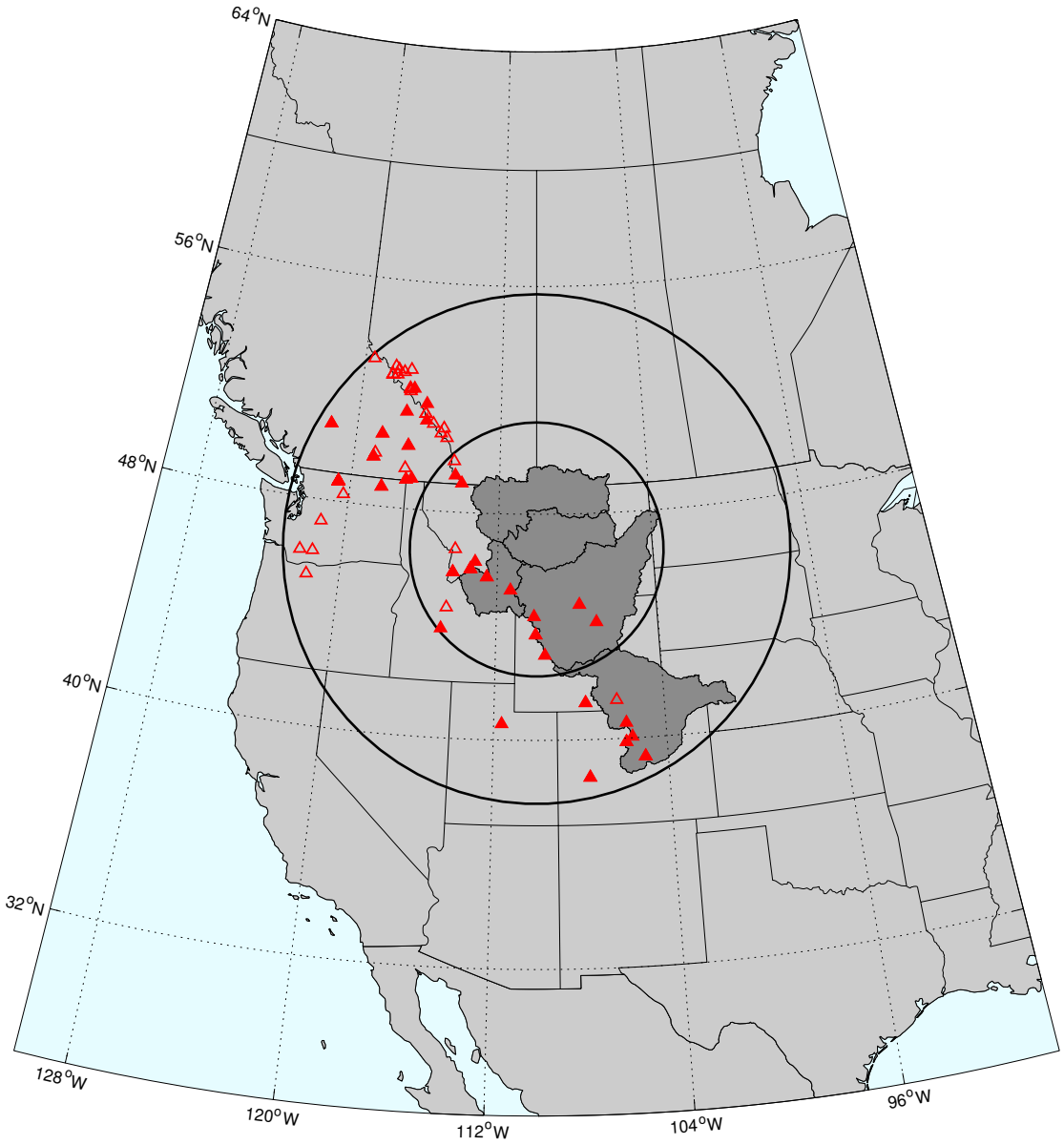

Fig. S1. Temperature reconstruction tree-ring records. The locations of high-elevation, ring-width and maximum latewood density (MXD) chronologies that are within the 1000 km UMRB search radius (outer circle), and are significantly ($p < 0.05$) positively correlated with runoff-season temperature. 500 km radius (inner circle) shown for reference. Filled symbols indicate chronologies that passed record screening and entered into the temperature reconstruction.

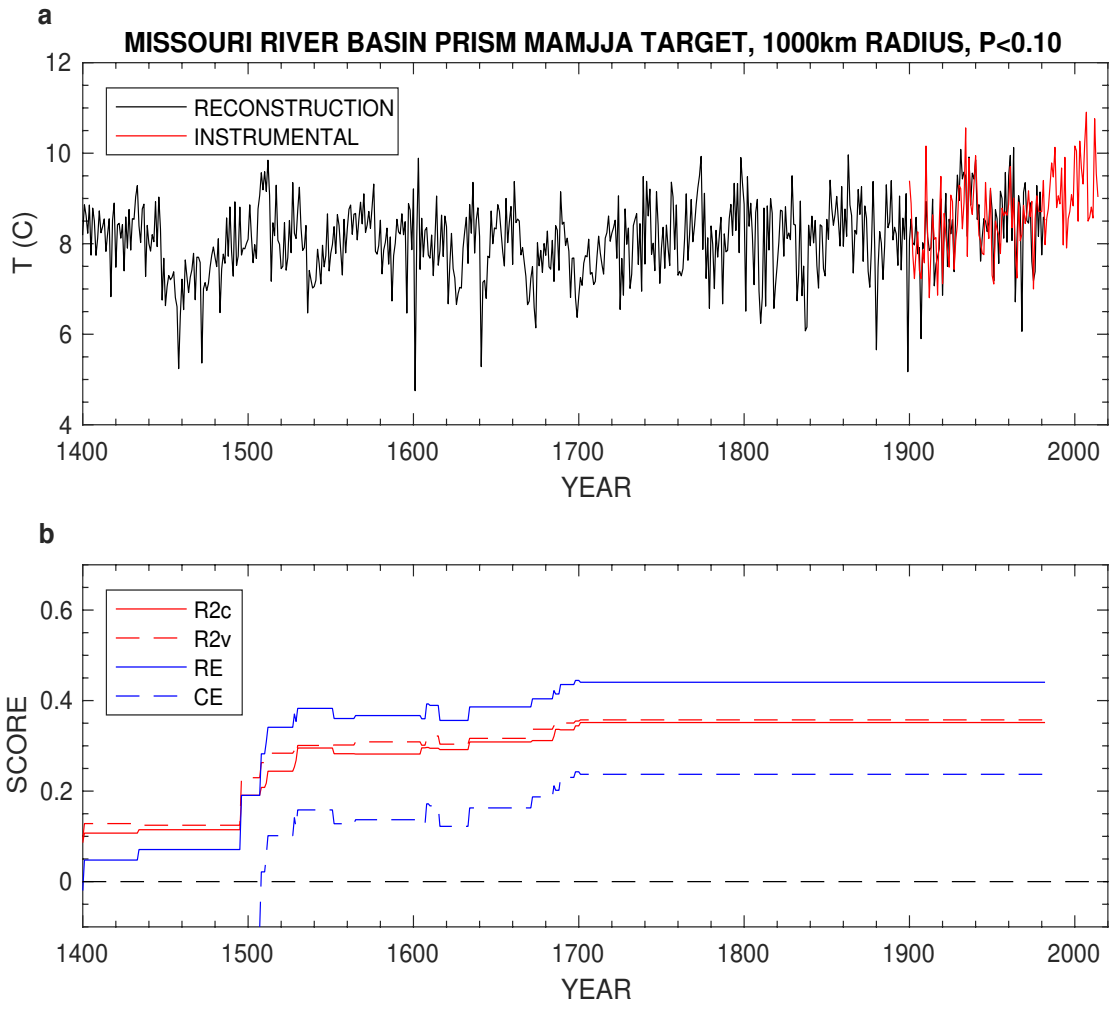

Fig. S2. UMRB runoff-season temperature reconstruction. Cross-validation statistics for the UMRB runoff-season (Mar-Aug) temperature reconstruction showing **(a)** the reconstruction (black) and target temperature record (red), and **(b)** validation R^2 (R2v), calibration R^2 (R2c), reduction of error statistic (RE), and coefficient of efficiency (CE).

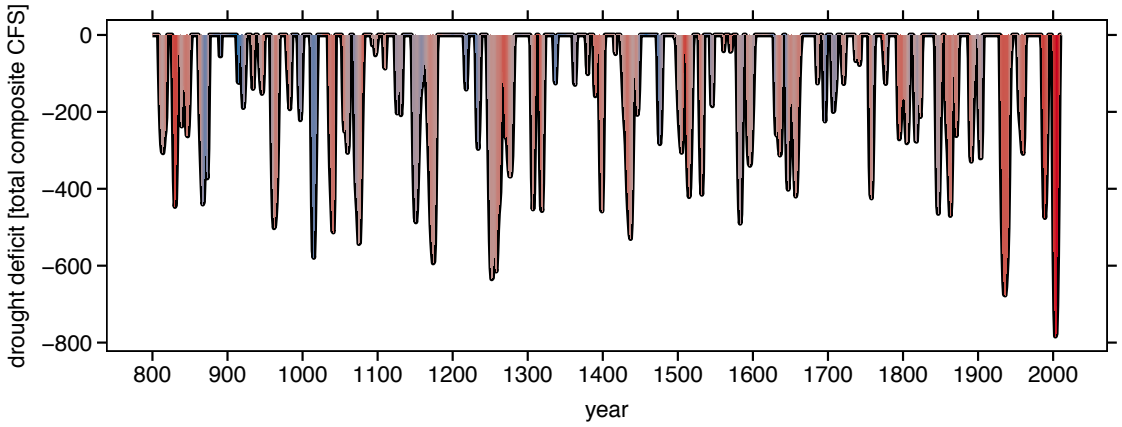

Fig. S3. UMRB un-standardized droughts. The basin-wide drought deficit reconstruction generated by averaging un-standardized streamflow values for the 17 gages in the UMRB composite record.

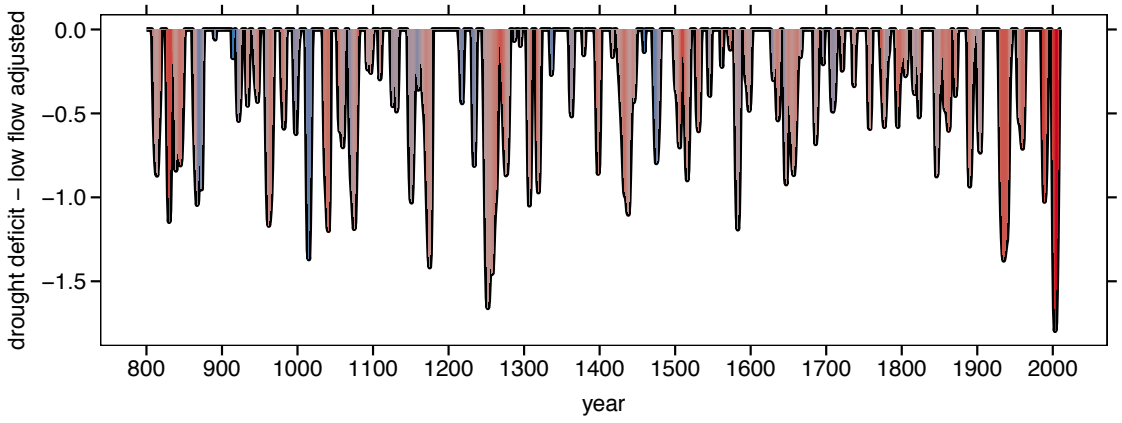

Fig. S4. UMRB droughts from low-flow aligned observed and reconstructed records. The basin-wide drought deficit reconstruction generated from the UMRB composite record in which reconstructed and observed flows are aligned based on their respective means over only the driest years of the calibration period.

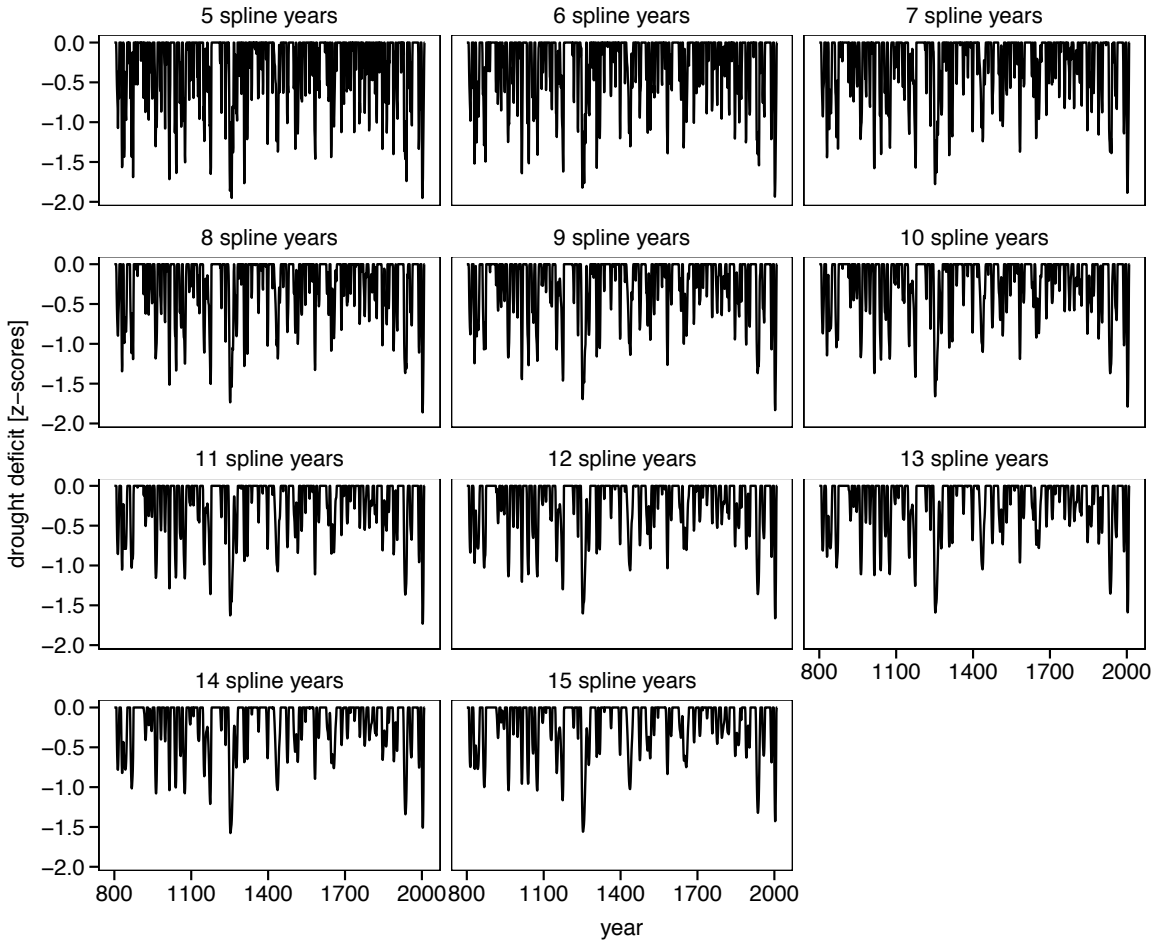

Fig. S5. Five to Fifteen-year UMRB drought splines. The basin-wide drought deficit reconstructions generated from splines lengths ranging from 5 to 15 years.

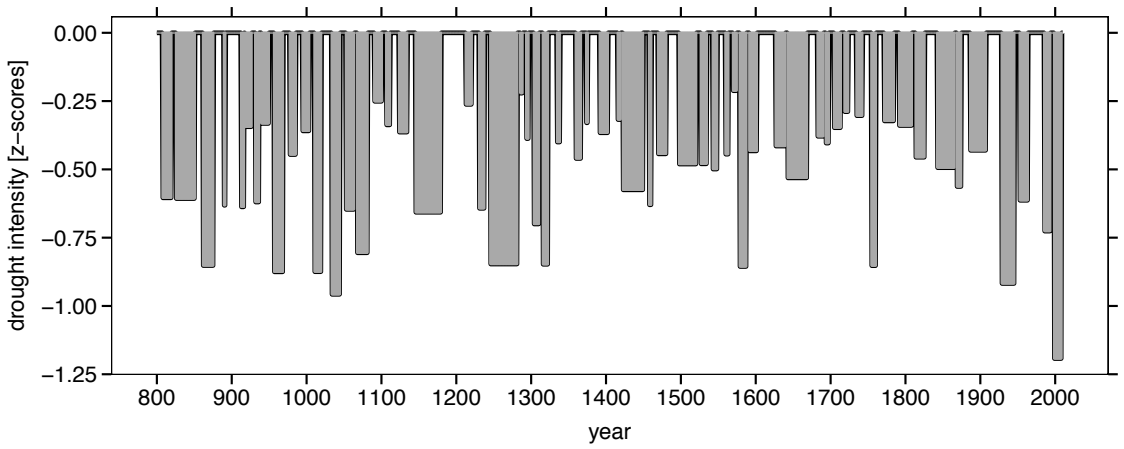

Fig. S6. UMRB drought intensity. The basin-wide decadal drought intensity record generated from droughts defined by the 10-year spline of streamflow and calculated as the cumulative deficit of the drought divided by the duration of the drought in units of z-scores.

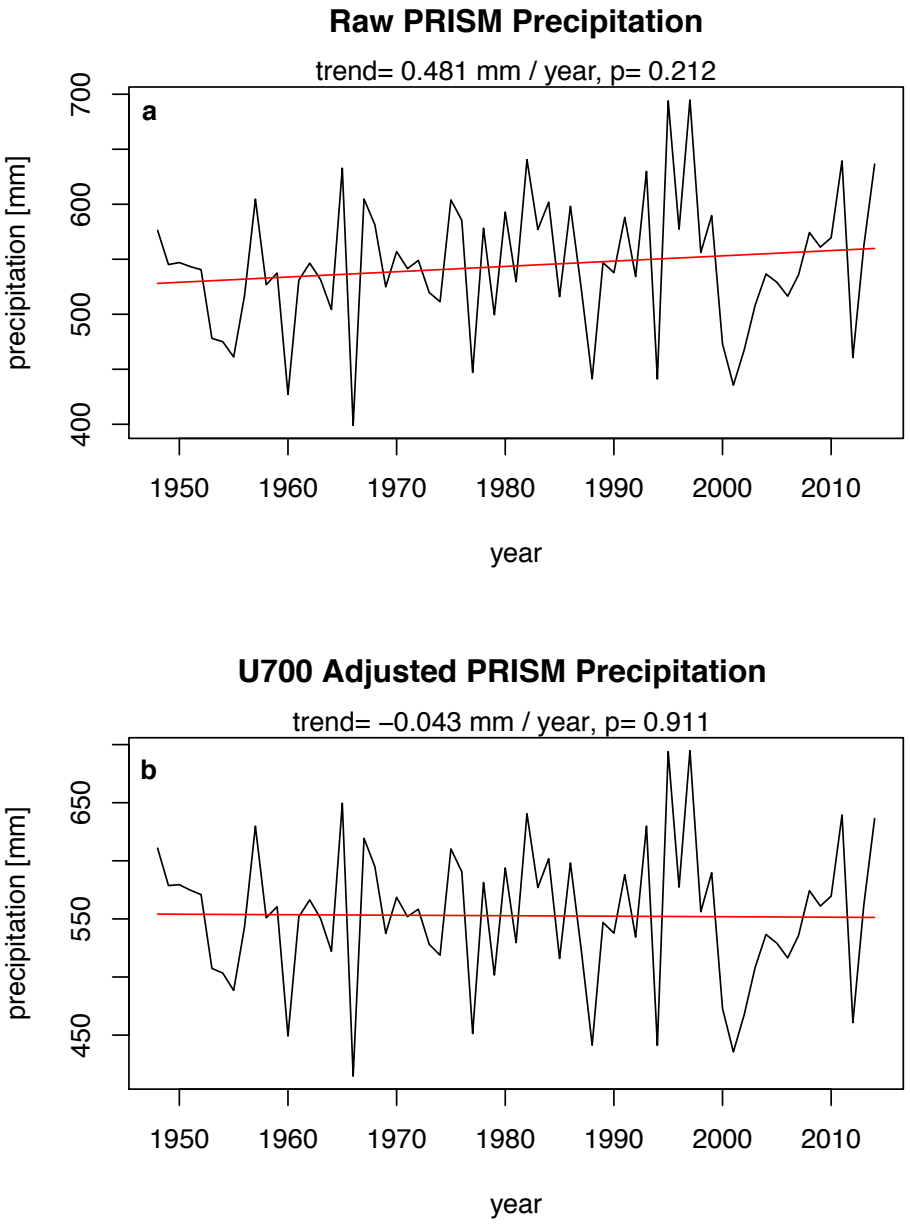

Fig. S7. Raw and U700-adjusted precipitation data. (**a**) Average total annual precipitation for PRISM grids within the UMRB derived from raw PRISM data, and (**b**) PRISM data adjusted for a hypothetical decrease in orographic enhancement of precipitation during the later 20th century.

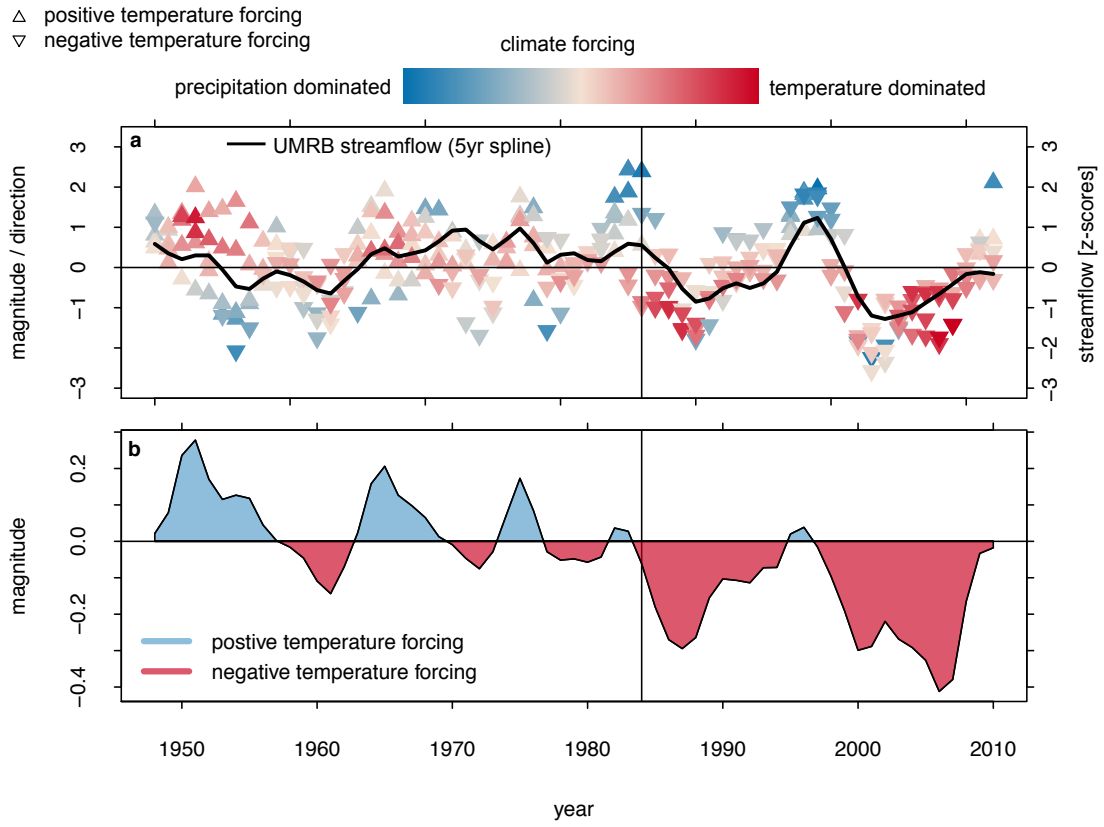

Fig. S8. Climate forcing of streamflow derived from U700-adjusted PRISM precipitation data. (a) The relative forcing of precipitation and temperature (arrows) on basin-wide mean annual streamflow (black line) in the UMRB. Colored arrows show the individual relative forcing estimates for each sub-region of the UMRB for each year. Colors denote which climate variable was more dominant in the combined forcing of streamflow relative to the long-term average influence of each variable. The y-axis shows the relative magnitude and direction of that combined forcing. The direction of the arrows shows the direction of forcing of the temperature component of that combined forcing (supporting or suppressing streamflow). All data shown are derived from the 5-year cubic smoothing splines of streamflow and climate data. **(b)** The relative forcing of temperature on streamflow determined as the temperature anomaly times its multiple-regression coefficient for predicting streamflow along with precipitation. The black line denotes the mean temperature forcing of temperature on streamflow for all sub-regions of the UMRB.

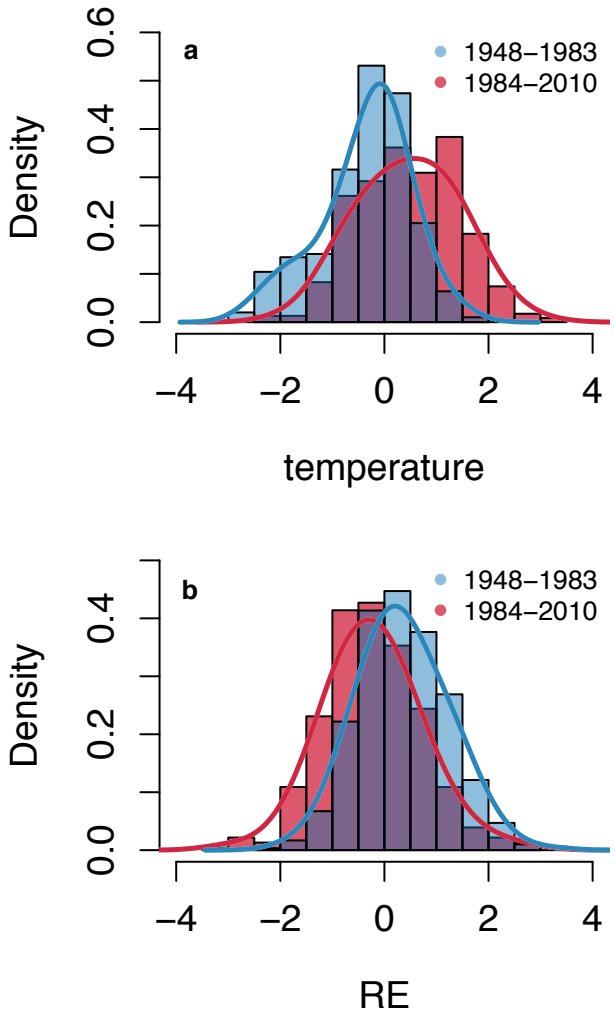

Fig. S9. UMRB basin-wide temperature and runoff efficiency derived from U700-adjusted PRISM precipitation data. (a) Distributions of temperature and **(b)** runoff efficiency from 1900-1983 (blue) and 1984-2010 (red). Lines show the kernel density estimates of the distributions.

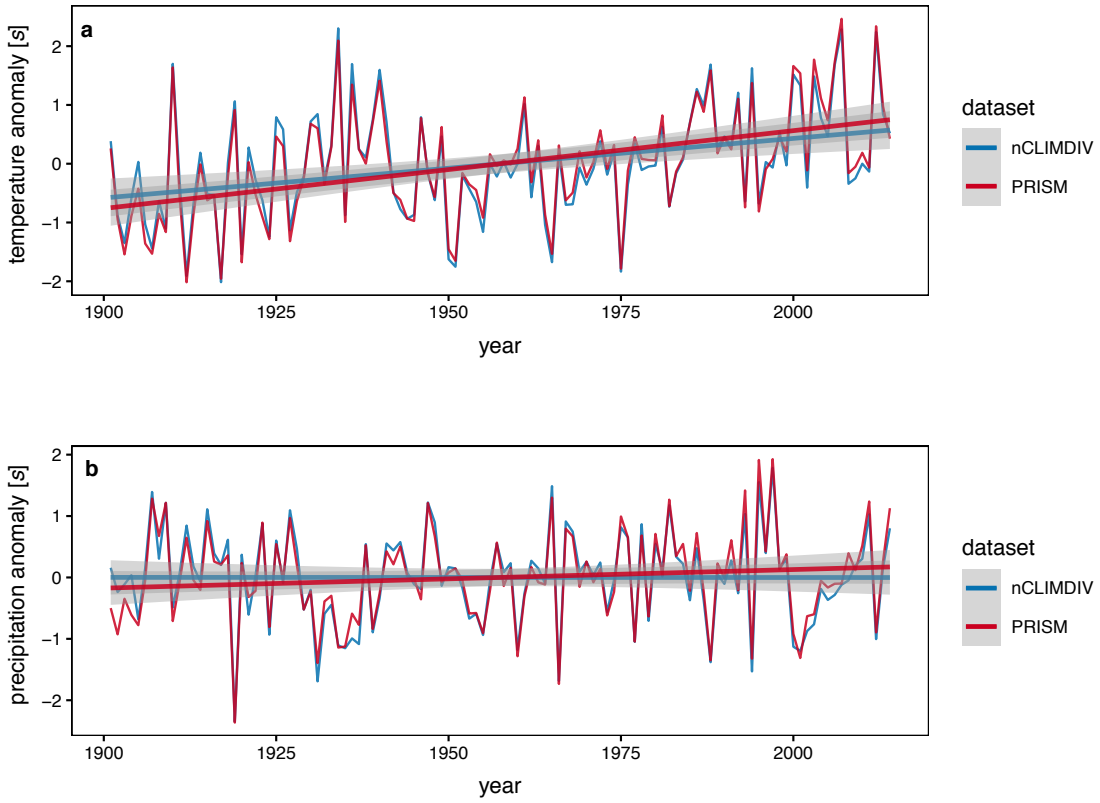

Fig. S10. UMRB basin-wide climate records. Comparison of (**a**) basin-wide average March-August temperature and (**b**) total annual precipitation anomalies in the UMRB derived from the PRISM and nCLIMDIV datasets. Least-squares regression trend lines are shown bracketed by their 95% confidence intervals (gray shading).

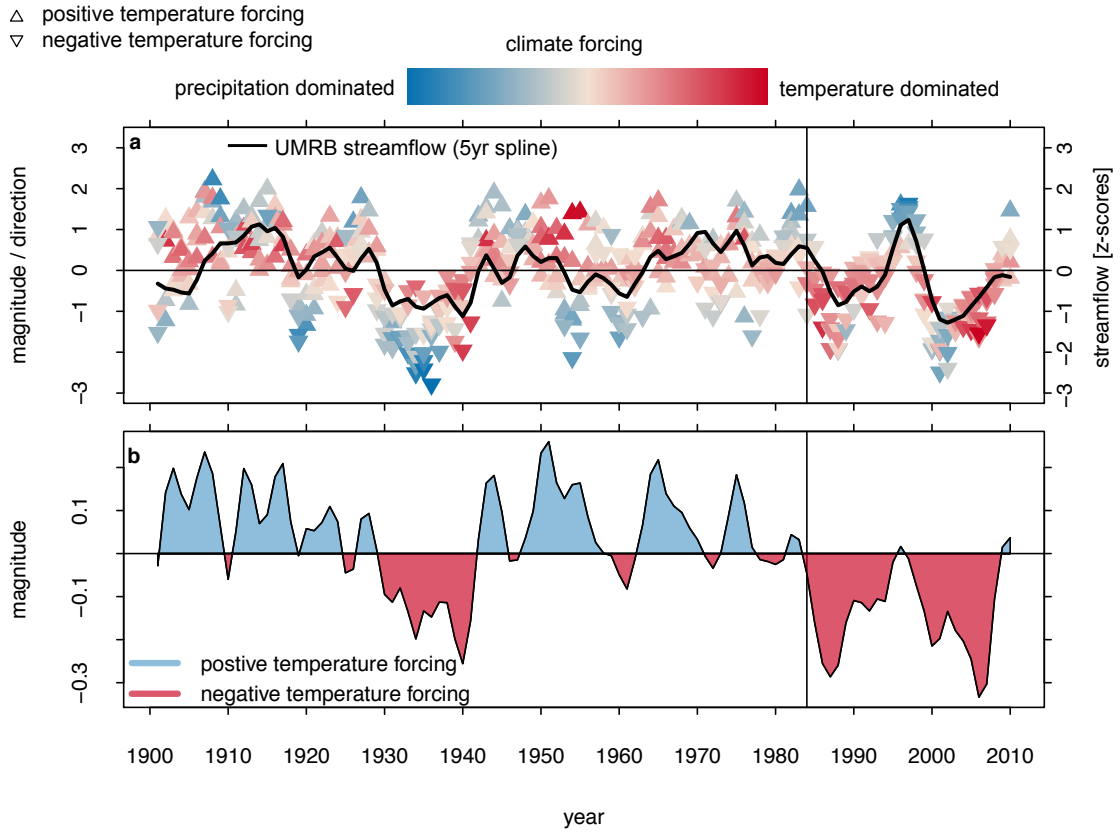

Fig. S11. Climate forcing of streamflow derived from the nCLIMDIV dataset. (a) The relative forcing of precipitation and temperature (arrows) on basin-wide mean annual streamflow (black line) in the UMRB. Colored arrows show the individual relative forcing estimates for each sub-region of the UMRB for each year. Colors denote which climate variable was more dominant in the combined forcing of streamflow relative to the long-term average influence of each variable. The y-axis shows the relative magnitude and direction of that combined forcing. The direction of the arrows shows the direction of forcing of the temperature component of that combined forcing (supporting or suppressing streamflow). All data shown are derived from the 5-year cubic smoothing splines of streamflow and climate data. **(b)** The relative forcing of temperature on streamflow determined as the temperature anomaly times its multiple-regression coefficient for predicting streamflow along with precipitation. The black line denotes the mean temperature forcing of temperature on streamflow for all sub-regions of the UMRB.

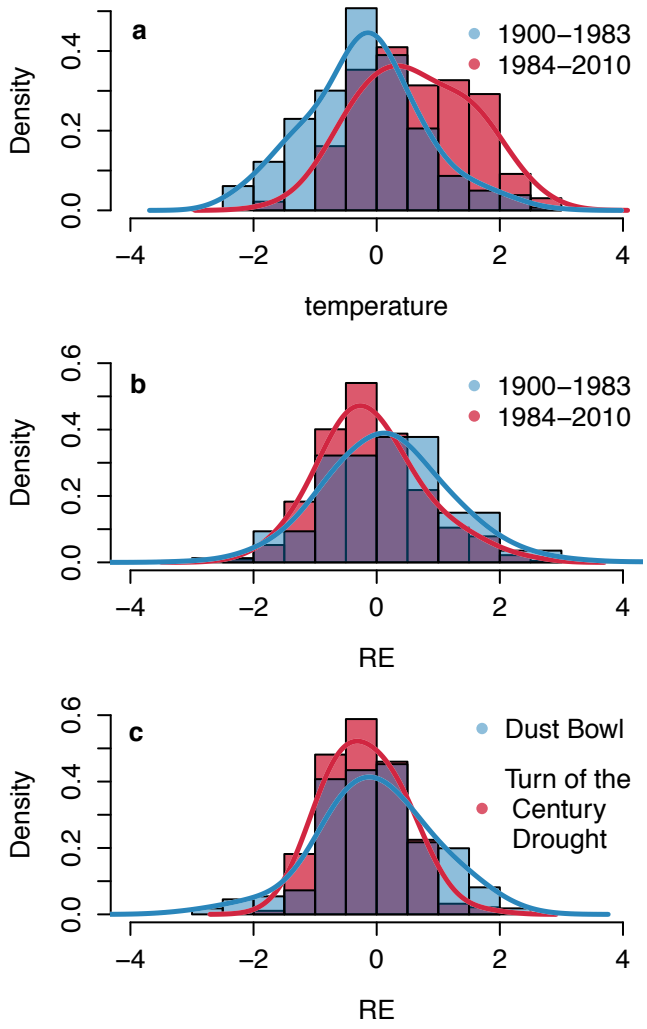

Fig. S12. UMRB basin-wide temperature and runoff efficiency derived from the nCLIMDIV dataset. Distributions of **(a)** temperature and **(b)** runoff efficiency from 1900-1983 (blue) and 1984-2010 (red). Panel **c** shows the distributions of runoff efficiency during the years of the Dust Bowl drought (blue) and Turn-of-the-Century Drought (red). Lines show the kernel density estimates of the distributions.

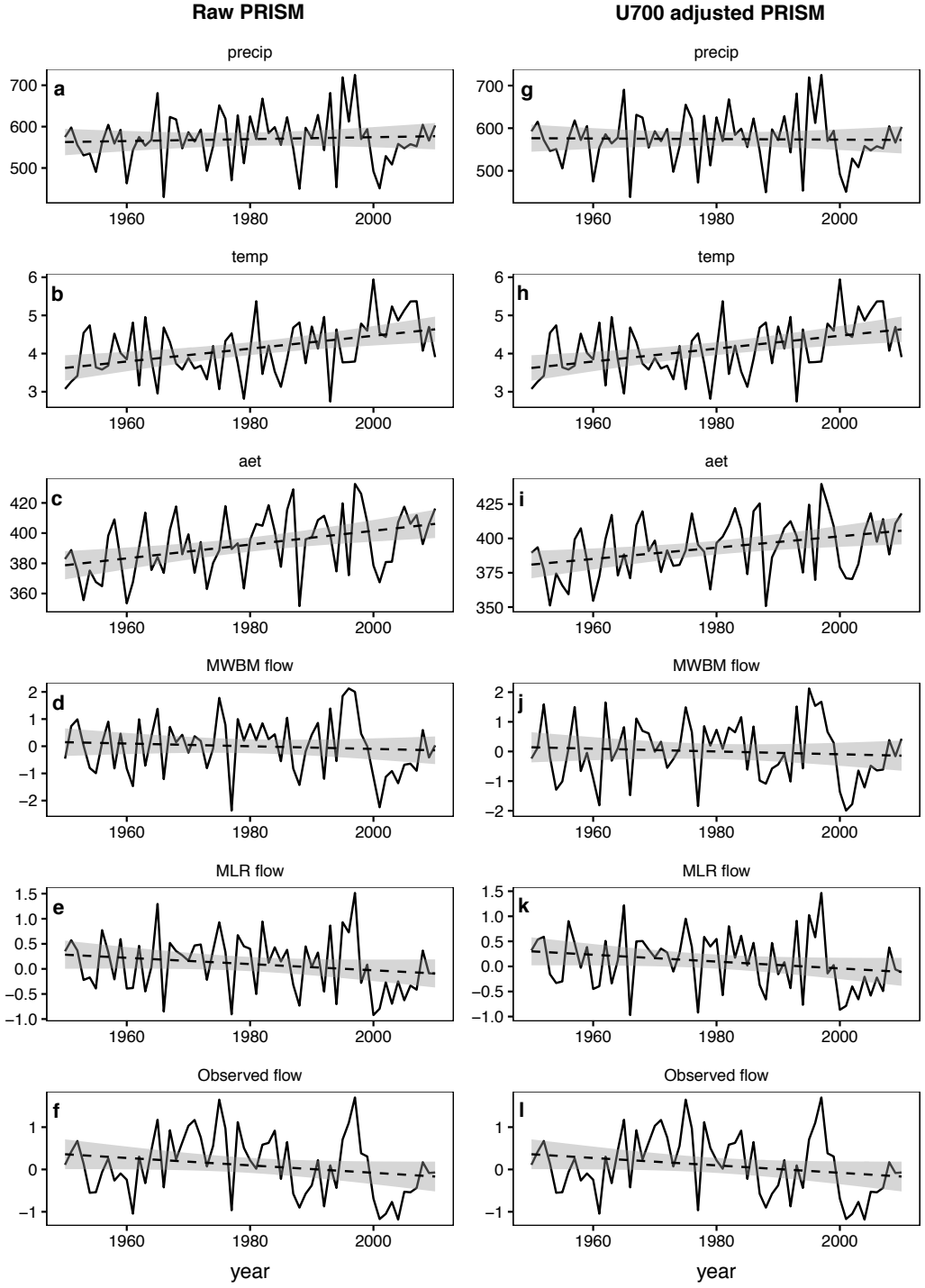

Fig. S13. UMRB basin-wide climate and streamflow records. (**a,g**) Water-year total precipitation [mm], (**b,h**) average temperature [°C], (**c,i**) evapotranspiration [mm], (**d,j**) monthly water balance model runoff [*z*], (**e,f**) linear regression modeled runoff [*z*], and (**f,l**) naturalized streamflow [*z*]. All records except naturalized flows are derived from precipitation and temperature data for the 36 HU 8 level watersheds listed in table S2, averaged by sub-basin and water-year, then by water-year. Data records in the left column use raw PRISM data while records in the right column use U700 adjusted PRISM precipitation data

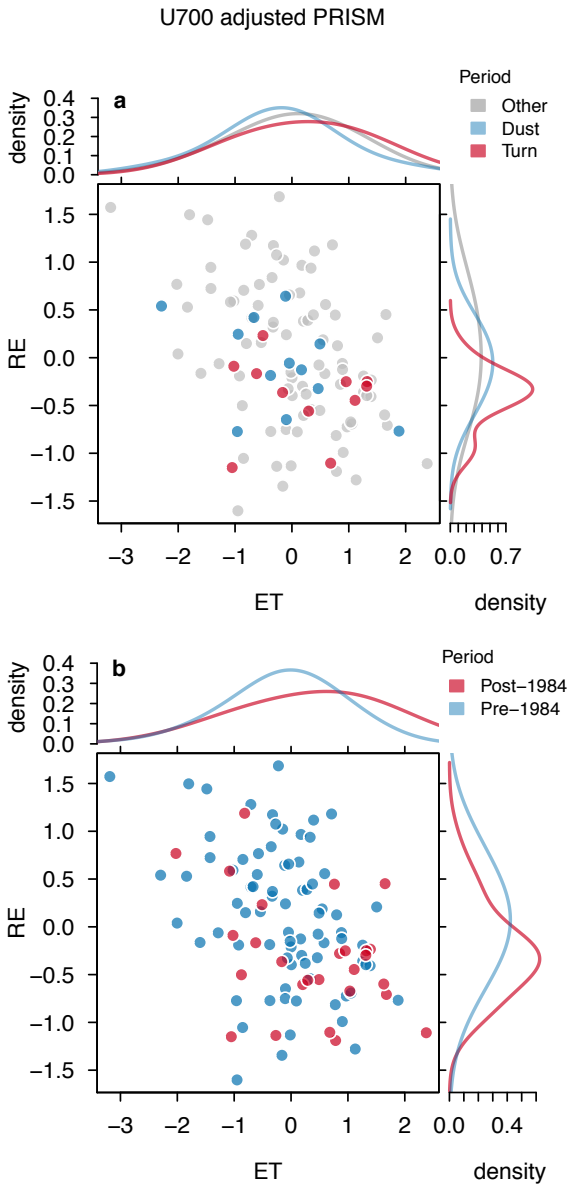

Fig. S14. U700-adjusted, UMRB basin-wide evapotranspiration (ET) versus runoff efficiency (RE). All values shown are derived from U700-adjusted PRISM precipitation and raw PRISM temperature data for the 36 HU 8 level watersheds listed in table S2, averaged by sub-basin and water-year, then by water-year (supplemental text S4). Statistics for the group comparisons ((**a**) Dust Bowl (Dust) vs. Turn-of-the-Century (Turn) Drought; and, (**b**) pre- vs. post-1984) are based on non-aggregated values for each watershed and each water-year within the groups being compared and are reported in supplementary table S1.

comparison	data	variable	effect size (sd)	t	p	n
Post'84-Pre'84	Raw PRISM	ET	0.34	9.14	<0.001	3780
Post'84-Pre'84	Raw PRISM	RE	-0.08	-5.82	<0.001	3780
Post'84-Pre'84	U700-adjusted PRISM	ET	0.27	7.06	<0.001	3780
Post'84-Pre'84	U700-adjusted PRISM	RE	-0.12	-6.58	<0.001	3780
Turn-Dust	Raw PRISM	ET	0.29	4.20	<0.001	828
Turn-Dust	Raw PRISM	RE	-0.05	-2.23	0.026	828
Turn-Dust	U700-adjusted PRISM	ET	0.17	2.34	0.019	828
Turn-Dust	U700-adjusted PRISM	RE	-0.06	-2.31	0.021	828

ET=evapotranspiration
RE=runoff efficiency

Table S1. Statistics for the group comparisons shown in Fig. S14 ((a&c) Dust Bowl (Dust) vs. Turn-of-the-Century (Turn) Drought; and, (b&d) pre- vs. post-1984) are based on non-aggregated values for each watershed and each water-year within the groups being compared.

Cluster	HUC
Missouri Headwaters	10020001
Missouri Headwaters	10020007
Missouri Headwaters	10020008
Missouri Headwaters	10020003
Missouri Headwaters	10020002
Missouri Headwaters	10020005
Missouri Mainstem	10030101
Missouri Mainstem	10030103
Missouri Mainstem	10030105
Missouri Mainstem	10040103
Missouri Mainstem	10040201
Yellowstone	10070001
Yellowstone	10080012
Yellowstone	10080013
Yellowstone	10080009
Yellowstone	10070006
Yellowstone	10070005
Yellowstone	10070002
Yellowstone	10080010
Yellowstone	10080001
Yellowstone	10080002
Yellowstone	10080003
Yellowstone	10080008
Northern Tributaries	10030104
Northern Tributaries	10030201
Northern Tributaries	17010207
Platte	10190001
Platte	10190002
Platte	14010002
Platte	14010001
Platte	10180001
Platte	10190007
Platte	10190006
Platte	10190005
Platte	10190004
Platte	10180002

Table S2. HUC 8 watersheds over which climate data were aggregated to serve as average sub-basin climate records.

SI References

1. Yang D, et al. (1998) Accuracy of NWS 8″ standard nonrecording precipitation gauge: Results and application of WMO intercomparison. *J Atmos Ocean Technol* 15(1):54–68.
2. Clark MP, Slater AG (2006) Probabilistic quantitative precipitation estimation in complex terrrain. *J Hydrometeorol* 7(1):3–22.
3. Groisman PY, Easterling DR (1994) Variability and trends of total precipitation and snowfall over the United States and Canada. *J Clim* 7(1):184–205.
4. Oyler JW, Dobrowski SZ, Ballantyne AP, Klene AE, Running SW (2015) Artificial amplification of warming trends across the mountains of the western United States. *Geophys Res Lett* 42:153–161.
5. McAfee SA, McCabe GJ, Gray ST, Pederson GT (2018) Changing station coverage impacts temperature trends in the Upper Colorado River basin. *Int J Climatol* 39(3):1517–1538.
6. Newman AJ, Clark MP, Longman RJ, Giambelluca TW (2019) Methodological intercomparisons of station-based gridded meteorological products: Utility, limitations, and paths forward. *J Hydrometeorol* 20(3):531–547.
7. Schaefer G, Paetzold R (2000) SNOTEL (SNOwpack TELemetry) and SCAN (soil climate analysis network). *Automated Weather Stations for Applications in Agriculture and Water Resources Management: Current Use and Future Perspectives*, p 30.
8. Vose RS, et al. (2014) Improved historical temperature and precipitation time series for U.S. climate divisions. *J Appl Meteorol Climatol* 53(5):1232–1251.
9. Luce CH, Abatzoglou J, Holden ZA (2013) The Missing Mountain Water: Slower Westerlies Decrease Orographic Enhancement in the Pacific Northwest USA. 342(6164):1360–1364.
10. Doggett M, Daly C, Smith J (2014) Challenges in Mapping Daily Temperature and Precipitation across the Conterminous United States. *AMS100* (Denver, CO).
11. Wise EK, Woodhouse CA, McCabe GJ, Pederson GT, St-Jacques JM (2018) Hydroclimatology of the Missouri River Basin. *J Hydrometeorol* 19(1):161–182.
12. Deems J, Hamlet AF (2010) *Historical Meteorological Driving Data Set* Available at: http://warm.atmos.washington.edu/2860/report/.
13. Oyler JW, Ballantyne A, Jencso K, Sweet M, Running SW (2015) Creating a topoclimatic daily air temperature dataset for the conterminous United States using homogenized station data and remotely sensed land skin temperature. *Int J Climatol* 35(9):2258–2279.
14. McCabe GJ, Wolock DM (2011) Century-scale variability in global annual runoff examined using a water balance model. *Int J Climatol* 31(12):1739–1748.