This document describes files output by TRISH for reconstruction method “Analog,” a two-stage method for reconstruction of a hydrologic or climatic time series, *y*, from a network of tree ring chronologies, X. For each year of the tree-ring record, some observed *y* for a year of the instrumental period with a similar multivariate pattern of tree-ring data is assigned as the analog reconstruction.

The first stage of the reconstruction method is to convert individual tree-ring chronologies to single-site reconstructions (SSRs) of *y* by distributed-lag stepwise multiple linear regression. Assume a time series matrix X whose columns, *x*, are standard or residual tree-ring chronologies. Each *x* is converted into a separate SSR of *y* by stepwise regression (Weisberg 1985) of *y* on *x* lagged -2 to +2 years from *y*. Such a regression, called “distributed lag” regression, has a long history of application in dendroclimatology (e.g., Stockton and Fritts 1973; Fritts 1976; Stockton and Meko 1983; Meko and Graybill 1995; Cook et al. 1999). The SSR model is cross-validated by leave-9-out cross-validation Meko (1997) and split-sample validation (Snee 1977) at each step as variables are entered in order of maximum reduction of the residual variance. Entry into the model is stopped if the next predictor to enter would result in decreased skill, as indicated by a drop in cross-validation RE (Fritts et al. 1990). This approach to stopping entry of predictors in stepwise regression is a cross-validation cutoff rule (Wilks 2019). Regardless of the change in cross-validation RE, entry of variable is not allowed to proceed beyond the step at which the model reaches maximum adjusted R^2 (Myers 1990). In summary, entry of variables into the SSR model is guided by a cross-validation cutoff rule and entry is limited by the maximum adjusted R^2.

TRISH can be run with either standard or residual chronologies (Cook and Kairiukstis 1990) as input. An option for whitening site chronologies within TRISH allows you to get a reasonable estimate of how residual chronologies would perform without actually uploading a matrix of residual chronologies to TRISH. From TRISH screens you can choose to prewhiten chronologies with an order-*p* autoregressive model (AR(*p*)) where *p* is 1 2 or 3. Prewhitening at the site chronology level (Meko et al. 1993) is not the same as prewhitening at the individual-core level, as done in computing residual chronologies, but the resulting chronologies are generally very similar.

TRISH also allows you to optionally specify that lag-0 only is allowed in the pool of potential predictors. Such specification amounts to ignoring lag effects and may be useful for sensitivity analyses – i.e., models allow lags vs not allowing lags.

The second stage of reconstruction is multi-site reconstruction, or MSR, whose objective is to combine the individual SSRs into a final reconstruction. In preparation for MSR, the SSRs are screened for strength and temporal stability of the signal for *y*, and the screened SSRs are converted to orthogonal variables by principal components analysis (PCA; Wilks 2019). The screened network includes only those SSRs with the following properties: 1) a statistically significant (p<0.05) overall-F of calibration, 2) positive skill of cross-validation, as reflected in a reduction-of-error (RE) statistic (Fritts et al. 1990) greater than zero, 3) positive skill of split-sample validation, as reflected in RE>0 for both halves of split-sample validation, and 4) a physically logical lag structure given the expected causative relationship of climate and tree rings. This last constraint means that the the SSR is rejected if the lags imply that the current year’s *y* is predictable from just past years’ *x*. Logically, tree-growth should not be able to respond to a climate fluctuation before that fluctuation occurs. You can make the screening less stringent by optionally removing the last two constraints (3 and 4) through input settings in TRISH.

As already mentoned, the SSRs are converted to orthogonal variables, or principal component (PC) scores. The PCA is done on the covariance matrix rather than the correlation matrix of the screened SSRs. I prefer the covariance matrix because the individual variances of the SSRs are meaningful to the analysis – they reflect the strength of signal for *y*, which varies from chronology to chronology, or from one SSR to another. Namely, the variance of an SSR computed over its overlap with *y* is proportional to the variance of *y* explained by the SSR model.

The time series matrix of screened SSRs input to PCA covers all years in common to those SSRs plus any additional years until the last year of the most recently ending SSR. Any SSR with an earlier ending year than that last year is extended statistically by a quantile-analog procedure described later (see “**Figure03-SSR3**).

The last step of the MSR modeling is the selection of analog years of observed *y* to use as the reconstructed *y* for each year of the tree-ring record. Before that step, the PCs are screened to eliminate any PC whose scores are not significantly correlated (default α=0.05; no adjustment for autocorrelation) with *y*. The user can override the default α within TRISH.

Analog selection of the observed y most appropriate for some reconstruction year is done by a nearest-neighbor algorithm applied to the vector of PC scores retained after the screening just decribed. Consider the following periods

- A: full time coverage by PC scores
- B: subset of years of A before start of *y*
- C: subset of years of A overlapping *y*

The PCA is done on the SSRs for period A. For any year, t_0, to be reconstructed in B, the Euclidian distance of the vector of PC scores from the vectors of scores in each year of C is computed. The year, t_c, in C with the shortest Euclidian distance is defined as the analog year for year t_0. The observed *y* for year t_c is then assigned as the reconstruction for year t_0.

The procedure just described has to be modified to get the reconstruction for period C because the method would by design yield a perfect reconstruction of *y* for those years. The question then is how can we judge the “calibratin-perkiod accuracy” of reconstruction by this method? That assessment requires time series of observed and reconstructed *y* for the common period. The solution is to use the second-nearest-neighbor pattern of PC scores for the reconstructed values in period C. The reconstruction for period C is therefore based on the second-nearest-neighbor PC scores. The observed and reconstructed *y* for period C can then be compared to assess accuracy of reconstruction, but the comparison likely underestimates the true skill of the analog method when applied to period B.

A much more sophisticated analog reconstruction approach is taken for reconstruction of Colorado River flows by Gangopadhyay et al. (2009). There, separate PCA are done on periods B and C, and Monte-Carlo resampling is applied to arrive at cross-validation reconstructed *y* for period A. That more sophisticated approach was deemed to time consuming due to computation demands for the online TRISH tool.

Reconstruction in TRISH by the analog method yields various files in an output directory:

1. **Fifteen figure files**. Figures 1-4 are for the SSRs. Figures 5-6 are for the PCA. Figures 7-15 are for the MSR.
2. **Six tables of statistics**. Tables 1-2 are for the SSRs. Tables 3-4 are for the PCA (loadings; correlations of y with PCs. Tables 5-6 are for the MSR.
3. **Three files with tab-separated time series.** One file has the full-length PC scores (time series) of the screened SSRs. A second file lists the analog years for each reconstruction year. A third file lists observed y, reconstructed y, and a 50% confidence interval for reconstructed y. Files are described in more detail below.

FIGURE FILES (all .png)

1. **Figure01-SSR1. . Summary of chronology screening by single-site reconstruction (SSR)**. Bar chart at left shows the number of chronologies at various stages of screening. Box plots at right summarize distributions of adjusted R-squared of SSR models for the subset of chronologies suitable for modeling. Screening tally as follows: 1) the N1 chronologies in the source tree-ring network, 2) the N2≤N1 chronologies in the user-drawn map polygon, 3) the N3 ≤N2 chronologies with sufficient time coverage for SSR modeling, and 4) the N4≤N3 chronologies passing screening for strength and temporal stability of hydrologic signal.

   Screening consists of several stages, starting with a user-specified network of N1 chronologies. First is the reduction to N2, which includes only those chronologies both in the user-drawn map polygon and with full time coverage over some specified interval. Second is reduction to N3, which includes only those chronologies whose time coverage has sufficient time coverage for some specified calibration interval. Third is reduction to N4, which includes only those chronologies passing SSR screening, described previously.

   Only N3 chronologies are SSR-modeled, and only N4 of those pass SSR screening. Because the F-level of adjusted R-squared is one of the screening criteria, the box plots usually indicate that the N4 screened chronologies have higher adjusted R-squared than the large set of N3 chronologies fit to SSR models.
2. **Figure02-SSR2 Bar charts summaring lags in SSR models**. At left is a histogram of the number of models that include lags t-2 to t+2 relative to *y* for the N3 chronologies subjected to SSR modeling. At right is a similar histogram for just the N4 chronologies passing screening.

   The histograms are useful for a quick assessment of whether lags are important for the chronologies modeled. It is not unusual for some models to include positive lags but not lag-0, meaning a delayed response of growth to climate.
3. **Figure03-SSR3**. **Drop in maximum SSR signal-strength with loss of chronologies toward present**. Depending on the input specifications, SSR models may have been fit to a specific uniform calibration period, or to whatever overlap period of y is available for each chronology (variable calibration period). The maximum R-squared (left y-axis) applies to that calibration period. Chronologies generally have different ending years, so that the nearer to the present the fewer the available chronologies (or SSRs), because each chronology is modeled by an SSR).

   The red line shows the drop in available chronologies from an initial year when all N4 SSRs are available, to the last year with any chronologies are available. The black line, and right y-axis, shows the corresponding maximum adjusted R-squared of the SSR models that extend to the given year. For example, if the SSR with the strongest signal (highest R-squared) also happens to be the chronology with the most recent time coverage, the the black line will be at a constant over the time period of dropping sample size.

   The plot is intended as interactive feedback to allow you to pick a reasonable ending year for the calibration period of SSR models, and equivalently, of the final MLR model. You would not want the calibration period to extend past the last year of tree-ring data with a strong SSR model. For example, if all screened SSRs were complete through 1998, and if chronologies dropped out such that only one was available in 2017, you could specify an ending year for calibrating the MSR model anywhere in the interval 1998-2017. If the chronology with the strongest signal (highest R-squared of SSR) had the end year 2017, it might make sense to to specify the end year for MSR calibration as 2017. But if there was a big drop in maximum R-squared after some year (say, 2004), you might want to specify 2004 as the end year for calibration of the MSR.

   The currently specified last year of the default calibration period is annotated on the figure as "calibration end year." Calibration of models is not allowed to include any *y* data after that year. The actual calibration periods of SSR models might not extend to the specified year, depending on the ending date of chronologies. "NA" as "calibration end year" means that you have allowed the data itself to set the maximum possible calibration period depending on the overlap of the chronology with *y*. If you happen to set the calibration end year to "NA," the actual calibration period for all SSR models will extend to the year of the most recently collected tree-ring chronology. Calibration periods of individual SSR models are listed in the tables in output files Table1-SSR1.txt and Table2_SSR1.txt.

How is it possible to specify and ending year of the calibration period later than the end of one of the chronologies. This is possible because before MLR modeling, TRISH extends all of the SSRs by a quantile analog method to the year of the most recently ending SSR. Extension is done by combination of correlation analysis and quantile ranking, as follows, assuming **A** is the SSR needing extended in year *k*. (1) A common period of length at least 50 years is identified for which all SSRs are complete – no missing data. (2) The correlation matrix (Spearman) of SSRs is computed for the common period, and correlations are sorted from largest to smallest. (3) The SSR that is most highly correlated with **A**, has data in year *k*, and has at least 100 year of overlap of complete data with **A** is defined as the predictor SSR, or series **B**. (4) The quantile of the value of **B** in year *k* is computed in the overlap period of **A** and **B**. (5) The same quantile of **A** in that overlap period is defined as the estimate of **A** in year *k*. Steps 3-5 are repeated until all SSR series have been extended to the ending year of the most recently ending SSR.

To avoid using extended SSRs in the reconstruction, specify at the TRISH input window an ending year of calibration that is the same as the last year of coverage by all of the screened SSRs before extension. This year can be identified from Figure 3.

4. **Figure04-SSR4**. **Scatterplot of the mean of the N4 SSRs of *y* against observed y**. The plot also has a fitted least-squares straight line, a locally-weighted (loess) fit to the scatter of points, and separate loess fits to the positive and negative residuals from the loess-fit to the points. The correlation coefficient is annotated on the plot. This figure plot gives a quick view of the strength of signal for *y* in the mean of the screened SSRs and whether the relationship of observed *y* with the mean of the SSRs is linear, curvilinear, restricted to one particular part of the range of *y*, or driven by outliers.
5. **Figure05-PCA1**. **Scree plot of eigenvalues of PCs on the screened subset of SSRs**. The percentage of variance and cumulative percentage of variance of SSRs accounted for by the first few (up to 7) PCs is annotated on the plot. A horizontal dashed line marks the "average" eigenvalue. For uses wanting to specify how many of the "most important" PCs to retain for the pool of potential predictor for the MSR model, those PCs above the dashed line are consistent with an "eigenvalue of 1" cutoff if the PCA had been done on the correlation matrix (It was done on the covariance matrix.)
6. **Figure06-PCA2**. **Heat map of of PC loadings**. Heat map includes all PCs of the screened SSRs. The heat map summarizes the modes of variability of the tree-ring chronologies after they have been converted to SSRs by lagged regression. Each column gives the loadings of a PC on the individual chronologies. These loading are also listed in one of the output tables. Loading of all the same sign indicate a mode of variability in common to to all chronologies.
7. **Figure07-Calibration1**. **Bar plot of correlations and autocorrelations**. Blue bars are Pearson correlation of the predictand, *y*, with the PCs of the screened SSRs. Magenta bars are lag-1 autocorrelation of the PCs. A 95% confidence band on the correlations is plotted by default. (User can change this to 90% or 99% within TRISH.) The statistical significance of the computed correlations of *y* with PCs is not adjusted for autocorrelation in the time series. The

autocorrelation of *y* is shown with a horizontal magenta line. Difference in the height of this line and the heights of the magenta bars shows whether the PC scores of the SSRs are more or less autocorrelated than the hydrologic variable to be reconstructed. It is important to recognize that the lag-1 autocorrelation can reflect long-memory (e.g., trend) as well as short-memory processes.

The correlation plotted as a blue bar directly gives the strength of linear correlation of *y* with a specific PC of the SSRs. Futher interpretation in terms of direction of influence of individual chronologies is complicated, and depends on several factors, including the sign of the PC loading of the SSR, and the signs of the regression coefficients in the lagged regression model that converts the chronology to the SSR. A more direct way of assessing importance of individual chronologies to *y* is from the coefficients and statistics of the SSR models (**Table1-SSR1.txt**, see below)

The correlation bars in Figure 7 also directly indicate which PCs are used to identify analog years. Only those PCs significantly correlated with y are used in this identification. The default threshold significance level is 0.05, but this can be changed to any of {0.01 0.05 0.10} at a TRISH screen.

8. **Figure08-Calibration2**. **Summary of the analog MSR model**.. A scatterplot of reconstructed against observed y, with Pearson correlation coefficient annotated, is at upper left. A time series plot showing tracking of observed y by the reconstruction is at the bottom. Information on the periods for principal components analysis (PCA) and and analog years for the MSR model are annotated at upper right. There are no conventional "regression" statistics because the analog reconstruction is done without calibrating a regression model. Because this reconstruction is not done by regression, the squared correlation coefficient cannot be be interpreted as equivalent to decimal fraction of "explained variance" of *y*. Unlike regression, the analog method does not partition variance so that total variance of *y* is equal to the sum of variance of the predictions and variance of the error (observed-reconstructed *y*).
9. **Figure09-Calibration3**. Histograms and autocorrelation functions (acf's) of observed and reconstructed predictand for period of calibration. The histograms at left allow comparison of distributions of observed and reconstructed *y*. These histograms have the same x-axis scale to make it easier to assess differences in the distributions. The analog method, unlike regression, does not necessarily give a reconstruction whose variance is reduced from that of of observed *y*.

   The acf's allow comparison of the autocorrelation properties of observed and reconstructed *y* over a common period. The acf reflects persistence of above-mean or below-mean values, and so is related to duration of wet and dry periods. Greatly different acf's could indicate that the reconstruction overestimates or underestimates such duration.
10. **Figure10-AnalysisResiduals1. Analysis of residuals for normality and for constancy of variance**. Unlike regression, the analog method of reconstruction makes no assumptions on properties of the residuals. But the diagnostic plots and statistics of the model residuals (observed minus reconstructed y) are still useful for interpreting the quality of the reconstruction, and for placing a confidence interval on the reconstructed *y*. TRISH output therefore includes some of the plots used in conventional analysis of residuals for regression (e.g., Draper and Smith 1981; Myers 1990).

    Normality can be checked visually by the histogram of residuals, at left, and statistically by the annotated results of the Lilliefors Test (Conover 1980). The p-value for Lilliefors Test is annotated at top of the histogram.

    Constancy of variance is checked visually with the scatterplot of residuals against predicted values (right). The scatter plot ideally shows no noticeable pattern (e.g., fanning out, curvature). "Heteroscedastic" means that there is a dependence of variance of residuals on the fitted values of the regression. Visually, one common example of heteroscedasticity residuals that fan out, or become more spread, toward higher fitted values. If residuals are heteroscedastic, a confidence interval that assumes variance of residuals is the same for the full range of reconstructed *y* would be problematic.

11. **Figure11-AnalysisResiduals2**. **Analysis of residuals for trend.** Regression residuals are plotted as a time series and the the time plot is tested for trend using the non-parametric Mann-Kendall trend test (Wilks 2019). The plot includes a non-parametric best-fit line,following Haan (2002), and is annotated with results of the Mann-Kendall test and with the estimates of the fitted parameters of the line as well as a variance inflation factor (VIF). The VIF reflects possible adjustment of the significance of trend for autocorrelation in the residuals – that is, autocorrelation in addition to that due to trend. The autocorrelation adjustment follows Wilks (2019), and its application is indicated by a VIF>1.0. If VIF=1.0, the residuals after removal of linear trend have no significant lag-1 positive autocorrelation (one-sided test, α=0.05).

    Residuals ideally have no trend. Trend might indicate that some factor, with trend, other than the reconstruction predictand, is influencing the tree-ring index. There could be other causes (e.g., nonclimatic growth trend still in tree-ring index after standardization).

12. **Figure12-AnalysisResiduals3. Analysis of residuals for autocorrelation**. The acf of residuals is plotted with a 95% confidence interval for lags k=0 to k=m years, where m is the minimum of 20 and N/4, and N is the length of the time series of residuals. Ideally, the acf is within its 95% confidence interval for all lags k>0. A Durbin-Watson (DW) test is not applied because that test is specific to regression residuals (Myers 1990). The analog method does not use regression.

    The resampling of analog years without explicit consideration of persistence is likely to prevent the analog reconstruction from reproducing autocorrelation in the observed *y*.

13. **Figure13-Validation1**. **Time plots and statistics summarizing tracking of the observed predictand by the analog reconstruction and by a "best possible" reconstruction**. The analog reconstruction for any year outside the overlap with observed *y* (common period) is the observed *y* for the year whose row of PC scores is closest (Euclidean distance) the scores in the reconstruction year. If this same rule were followed for the common period, the reconstruction would by definition be perfect, because the nearest-neighbor is the same as the year to be reconstructed. For the common period, a different rule is therefore followed to get values of the analog reconstruction: the instrumental-period year whose scores are second-most similar is assigned as the analog year. The maximum possible accuracy of reconstruction for the common period is therefore if the analog years happen to those closest in value to the observed *y*. To assess how such a best possible reconstruction preforms compared with the observed reconstruction, those closest values are plotted (green asterisk) on the time series plots of observed and reconstructed *y*.

    Statistic listed to the right of the time series plot include the root-mean-square error of reconstruction (RMSE), reduction-of-error (RE) and Pearson correlation coefficient (r). The RMSE is the square root of the mean squared error of reconstruction. RMSE is computed by 1) computing the time series of errors, E, defined as the difference of observed and reconstructed y, 2) squaring elements of E, 3) summing the squared E over the common period, 4) dividing that sum by the length of *y*, and 5) taking the square root of that quotient. RMSE can be used for a confidence interval around the reconstructions. For example,if E is assumed to be normally distributed, a 95% confidence interval is ±1.96 RMSE around the reconstruction. The correlation, *r*, of observed with reconstructed y measures the agreement between standardized departures from their means of reconstructed *y* and observed *y*. It is important to recognize that r is not affected by shifting the two time series (changing their means) or scaling them (changing their variance).

The RE statistic is analogous to the conventional reduction-of-error statistic in dendroclimatology (Fritts et al. 1990), and measures skill of reconstruction relative to an “null” reconstruction consisting of the observed mean of *y* in every year. It is possible for the correlation coefficient, *r*, to be positive and RE negative. A negative RE indicates is no better – in terms of lower sum-of-squares of errors – that just assuming the long-term mean of *y* as the reconstruction in every year. than using the analog-reconstructed y.

14. **Figure14-Reconstruction1**. Full-length reconstruciton with confidence interval. The full-length reconstruction of y is plotted with its 50% confidence interval computed as yhat±0.67449 RMSE, where yhat is reconstructed annual y. Note that 0.6449 is the 0.75 quantile of the standard normal distribution and RMSE is the root-mean-square error of the analog MSR model. This RMSE is a “validation” RMSE because the analog model has not been tuned, as in regression, to produce minimum square error in the instrumental period. The confidence interval therefore assumes normally distributed errors.

    The last few years of the reconstruction will have amplified uncertainty if based on statistically extended SSRs (see description of Fig. 3, above). For that reason, you may want to truncate the reconstruction to end with the last year that all of the screened SSRs (before extension) have data. That way all years of the reconstruction are base on observed tree-ring chronologies rather than possibly being distorted because some tree-ring chronologies (their SSRs) are extended from other chronologies.
15. **Figure15-Reconstruction2**. **Autocorrelation functions (acf) and boxplots of reconstructed y for the calibration period of the MSR model and for earlier years**. These plots allow a quick graphical assessment of the calibration period *y* in a long-term perspective. The comparisons here are “apples vs apples” in that reconstructed *y* is compared with reconstructed *y* for different time intervals.

TABLES OF STATISTICS (all tab-sep .txt)

When viewing tables on screen with text editor, use monospaced font; otherwise columns will not line up properly with headings.

1. **Table1-SSR1.txt.** Summary statistics of SSR models for all chronologies fit with reconstruction models.
   1. **N_1**: sequential number in this table
   2. **N_2**: corresponding site number in original tree-ring network
   3. S**ite**: unique alphanumeric identifier for tree-ring chronology
   4. **Goc**: first year of model calibration period
   5. **Endc**: last year of model calibration period
   6. **Model**: code indicating lags in model and order that they entered stepwise. The five slots correspond to lags -2, -1, 0, +1, +2 years relative to the predictand year. For example, code [0 2 1 0 0] means lag 0 entered first, lag t-1 entered second, and no other lags are in the model.
   7. **Sign**: code that goes along with "Model" and tells the sign of the SSR regression coefficients on the the chronology lagged -2, -1, 0, +1, +2 years relative to the predictand year. Codes "P", "N" ad "0" indicate positive coefficient, negative coefficient, and not in model. For example, [00P0P] means positive coefficent on lag 0, and positive coefficient on lag t+2.
   8. **R2a**: adjusted R-squared of model; this is the regression R-squared adjusted downward as a penalty for number of predictors in the model.
   9. **pF**: p-value of the overall F of regression ($pF<0.05$ indicates significant model at 0.05 level)
   10. **REcv**: Reduction of error (RE) statistic from leave-9-out cross-validation
   11. **REa**: Split sample RE for fitting model to first half of record and validating on second half
   12. **REb**: Split sample RE for fitting model to second half of record and validating on first half
   13. **Refit**: Logical (TRUE or FALSE) variable indicating whether the model was re-fit with expanded calibration period after exploratory stepwise regression.

       First the stepwise regression allows lags -2 to +2 in the model, which could restrict the calibration period if the tree-ring data happen to end in the same year or in the year after the end of y. For example, if y and the chronology end in 2019, the calibration period cannot have an ending year later than 2017, because of the possible need for lags +1 and +2 on the tree-ring series. The stepwise process might result in a model that does not include lags +1 or +2. If so, the model is re-fit, resulting in the maximum possible length of calibration period given the lags in the model and time coverage of y and the tree-ring chronology. "Refit" indicates whether the model was re-fit. .
   14. **Gor**: First year of reconstruction.
   15. **Endr**. Last year of reconstruction.
   16. **Reject**: Logical (TRUE or FALSE) variable indicating whether the chronology and its SSR are rejected from further use in the later step of multi-site reconstruction. Rejection occurs if any of the following are true:
       1. $pF\geq0.05$
       2. $REcv\leq0$
       3. $REa\leq0$ or $REb\leq0$

   4. Illogical causal model: the final model has y predicted from x at negative lags only from y. This is physically unreasonable, because past years' tree-ring values alone should not be able to detect current year's climate.

2. **Table2-SSR2.txt.** Summary statistics of SSR models for just those chronologies passing screening for hydrologic signal. These are the chronologies for which "Reject" is FALSE in Table 1. The columns are the same as those of Table 1.
3. **Table3-PCA1.txt.** Loadings of PCs of the screened subset of SSRs.
   These loadings are the PCA weights of each PC on the SSRs derived from individual tree-ring chronologies. Each SSR is associated with a specific tree-ring chronology, as indcated by the column "SiteID." The site number, in the user database supplied to TRISH, is listed in column "Site#." The percentage of variance of SSRs accounted for by the PCs applies to the full overlap of the SSRs, not just the period in common with the hydrologic variable. Note that the PCA was done on that full period of overlap. Accordingly, PC scores for the full overlap are by design not intercorrelated, but they may be intercorrelated during the shorter overlap with the hydrologic variable.
4. **Table4-PCA2.txt.** Correlation of y with PCs of the screened subset of SSRs.
   The analysis is done on the calibration period of the MSR model. Thresholds (95%) are shown for significance disregarding (Thresh1) and considering (Thresh2) lag-1 autocorrelation in y and the PCs. The signs of the correlations of y with the PCs is related to whether high or low tree-ring index for a particular chronology contributes to a positive or negative score of a PC, but the interpetation is complicated. The interpretation must consider the signs of the regression coefficients in the distibuted-lag regression model the converts a chronology to and SSR, and on the loading of the resulting SSR in the PC (Figure 6).
5. **Table5-Calibration1.txt.** Calibration statistics of the analog multi-site reconstruction (MSR) model. The "calibration" period is the overlap of observed y with the PCs of the single-site reconstructions (SSRs).
   1. **YearGo:** first year of calibration period
   2. **YearStop:** last year of calibration period
   3. **Npool**: number of PCs before correlation screening against *y*
   4. **alphaR**: α-level for screening of the Npool PCs before selecting analog years. For example, alphaR=0.05 indicates that Pearson correlation was required to be significant at the 95% level for a two-tailed test of the null hypothesis of zero correlation. The correlations and threshold level of correlation are dispayed graphically in Figure 7.
   5. **Npredictors**: number of PCs after screening. The analog years are selected by the Euclidean distance of the scores of these PCs scores for any reconstruction year from the scores of the PCs in the calibration period. See table footnote for which PCs these are.
   6. **RMSE:** root-mean-square error, computed from the difference of observed and analog-reconstructed y – the "errors".
   7. **RE**: reduction-of-error statistics, defined as RE=1-(SSE1/SSE2), where SSE1 is the sum-of-squares of reconstruction errors and SSE2 is the sum-of-squares of differences of the calibration-period mean of y and the reconstucted *y*.
   8. ***r***: Pearson correlation coefficient of observed with reconstructed y for the calibration period
6. **Table6-AnalysisResiduals1.txt**. Normality and trend
   1. **YearGo:** Start ear of calibration period
   2. **YearStop**: End ear of calibration period
   3. **pNormal**: p-value of Lilliefors test for normality (pNormal<0.05: reject H0 that residuals from normal distribution)

4. **TrendSlope**: slope coefficient of non-parametricfit of trend of residuals
5. **pTrend:** Slope of trend line, with p-value from Mann-Kendall test

TIME SERIES OUTPUT

1. **PCscoresTimeSeries.** Scores of all PCs of the screened SSRs. Note that depending on screening of PCs for correlation with y, not all of these PCs may have been used to identify analog years (see Table 5)
2. **AnalogYearsTimeSeries.txt.** Listing of the reconstructed y, the analog year providing it, and whether this analog year is the nearest neighbor or second-nearest neighbor from the calibration period. Nearest neighbor is used for all years not in the overlap period (calibration period) of *y* and the PCs of SSRs. Second-nearest neighbor is used for the calibration period.
    1. **Year:** year of data
    2. **<code for predictand – e.g., RO>**: reconstructed values of predictand
    3. **Analog Year***:* the year whose observed y supplies the reconstruction.
    4. **Neighbor** (1 or 2): whether analog year is nearest (1) or second-nearest (2) neighbor.
3. **ReconstructedWithConfidenceIntervalTimeSeries.txt**. Listing of time series of observed predictand, reconstruction, and confidence interval on reconstruction (5 columns). Call the observed predictand y and the reconstruction yhat. Confidence interval is estimated assuming that reconstution errors are normally distributed with a standard deviation equal to RMSEcv.
    1. **Year:** year of data
    2. **y**: observed predictand
    3. **yhat:** the reconstruction
    4. **Lower 50% CI:** true (unknown) value of predictand has 50% chance of being lower than this threshold of the confidence interval (CI)
    5. **Upper 50% CI:** true (unknown) value of predictand has 50% chance of being higher than this threshold of the confidence interval (CI)

REFERENCES

Breusch, T. S., & Pagan, A. R. (1979). A simple test for heteroscedasticity and random coefficient variation. Econometrica, 47 (5), 1287–1294.

Conover, W. (1980). Practical nonparametric statistics (Second ed.). New York: John Wiley & Sons. (493 pp)

Cook, E. R., & Kairiukstis, L. A. (Eds.). (1990). *Methods of dendrochronology: Applications in the environmental sciences*. Dordrecht: Kluwer Academic Publishers. (394 pp)

Cook, E. R., Meko, D. M., Stahle, D. W., & Cleaveland, M. K. (1999). Drought reconstructions for the continental United States. J. Climate, 12 , 1145-1162.

Draper, N. R., & Smith, H. (1981). Applied regression analysis. New York: John Wiley Sons. (709 pp)

Fritts, H. C. (1976). Tree rings and climate. London: Academic Press. (567 pp)

Fritts, H. C., Guiot, J., & Gordon, G. A. (1990). Verification. In E. R. Cook & L. A. Kairiukstis (Eds.), Methods of dendrochronology: Applications in the environmental sciences (pp. 178–185). Kluwer Academic Publishers.

Gangopadhyay, S., Harding, B. L., Rajagopalan, B., Lukas, J. J., & Fulp, T. J. (2009). A non-parametric approach for paleo reconstruction of annual streamflow ensembles. Water Resour. Res., 45 (W06417), doi:10.1029/2008WR00720.

Haan, C. T. (2002). Statistical methods in hydrology (Second ed.). Iowa State University Press. (496 pp)

Hutton, P. H., Meko, D. M., & Roy, S. B. (2021). Supporting restoration decisions through integration of tree-ring and modeling data: reconstructing flow and salinity in the San Francisco Estuary over the past millennium. Water , 13 (15), 2139. ('Received 2 June 2021, revised: 21 July 2021, accepted 27 July 2021, published 3 August 2021; this article belongs to the Special Issue Decision Support Tools for Water Quality Management)) doi: 10.3390/w13152139

Meko, D. (1997). Dendroclimatic reconstruction with time varying predictor subsets of tree indices. J. Climate, 10 (4), 687–696.

Meko, D. M., Cook, E. R., Stahle, D. W., Stockton, C. W., & Hughes, M. K. (1993). Spatial patterns of tree-growth anomalies in the United States and southeastern Canada. J. Climate, 6 , 1773–1786.

Meko, D., & Graybill, D. A. (1995). Tree-ring reconstruction of Upper Gila River discharge. J. Am. Water Resour. Assoc., 31 (4), 605–616.

Meko, D. M., Woodhouse, C. A., Baisan, C. H., Knight, T., Lukas, J. J., Hughes, M. K., & Salzer, M. W.

(2007). Medieval drought in the Upper Colorado River Basin. Geophys. Res. Lett., 34 (L10705), 10.1029/2007GL029988.

Myers, R. H. (1990). Classical and modern regression with applications. Pacific Grove, California: Doxbury. (488 pp)

Stockton, C. W., & Fritts, H. C. (1973). Long-term reconstruction of water level changes for Lake Athabasca by analysis of tree rings. J. Am. Water Resour. Assoc., 9 (5), 1006–1027. doi: 10.1111/j.1752-1688.1973.tb05826.x

Stockton, C. W., & Meko, D. M. (1983). Drought recurrence in the Great Plains as reconstructed from long-term tree-ring records. J. Clim. Appl. Meteor., 22 , 17-29.

Weisberg, S. (1985). Applied linear regression, 2nd ed. New York: John Wiley. (324 pp)

Wilks, D. S. (2019). Statistical methods in the atmospheric sciences (Fourth ed.). Cambridge, MA: Elsevier. (818 pp)

Woodhouse, C. A., Meko, D. M., & Bigio, E. R. (2020). A long view of Southern California water supply: perfect droughts revisited. J. Am. Water Resour. Assoc., 56 (2), 212–229. doi: https://doi.org/10.1111/1752-1688.12822