

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).

TRISH allows variation on the SSR model. From TRISH screens you can choose to prewhiten chronologies and order- p autoregressive model (AR(p)) where p is 1 2 or 3. Prewhitening at the site chronology level (Meko et al. 1993) is slightly different than the prewhitening at the individual-core level used to generate so-called “residual” chronologies, but experience has shown that the resulting site chronologies are very similar. This option eliminates the need for you upload separate networks of standard and residual chronologies to TRISH to check whether a better hydrologic signal might be obtained from a network of residual rather than standard chronologies. TRISH also allows you to optionally dispense with lags in the SSR models.

The second stage of the reconstruction method in TRISH multi-site reconstruction, or MSR. The goal of MSR is to combine the individual SSRs into a final reconstruction. Prior to MSR, the SSRs are screened for strength of 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 causative relationship of climate and tree rings. This last constraint means that the SSR is rejected if the lags imply that the current year's y is predictable from a model including just past years' x (e.g., this year's precipitation predictable from a model including just past years' tree-ring index).

The PCA is done on the covariance matrix rather than the correlation matrix of the full-length common period of SSRs. The covariance matrix is used for the PCA because the individual variances of the SSRs are meaningful to the analysis – they reflect the strength of signal for y of the individual chronologies. Optionally, all SSRs may be extended to the ending year of the SSR ending most recently to allow the PCA and subsequent MSR reconstruction to come as close to the present as possible. Such extension might be desirable if the SSRs have various ending years, (e.g., 2018 to 2022). The optional extension is done by a quantile method with the following steps for any SSR -- the “key” SSR -- requiring extension: 1) find a common period, A , covered by all SSRs, and compute the correlation matrix of the key SSR with all others for that period; 2) find the highest correlated (Spearman r) SSR that also has data in the year, t_0 , that the key SSR needs replacement of a missing value; 3) for the common period of the two SSRs (key and predictor) find the quantile of the predictor SSR for year t_0 in the common period of the two SSRs; and 4) replace the missing value in the key SSR in year t_0 with its same quantile for the common period. In the special case of the predictor SSR for year t_0 outside its range for the common period, the maximum or minimum (whichever is more appropriate) of the key SSR for the common period replaces the missing value. A diagnostic figure (Figure 3) is provided by TRISH to help the user decide whether to apply this optional extension of SSRs. For

example, one would not want to extend if the only SSRs remaining in the last few years (predictors must come from these) are SSRs with weak signals for y .

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 $\alpha=0.05$; 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 described. Consider the following periods

- A: full coverage of PCs
- 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 “accuracy” of reconstruction by this method? That assessment requires time series of observed and reconstructed y for the common period. Our solution is to use the second-nearest neighbor pattern of PC scores for the reconstruction 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 years in period B.

A much more 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.

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 has the calibration-period data of predictand and predictors for the MSR model. 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 N_1 chronologies in the source tree-ring network, 2) the $N_2 \leq N_1$ chronologies in the user-drawn map polygon, 3) the $N_3 \leq N_2$ chronologies with sufficient time coverage for SSR modeling, and 4) the $N_4 \leq N_3$ 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.

The following criteria for SSR screening must all be satisfied for a chronology and its SSR to be accepted for later use in multi-site reconstruction: 1) significant calibration signal, as defined regression overall F having $p < 0.05$, 2) cross-validation reduction of error (RE_{cv}) greater than zero, and 3) split-sample validation RE_{cv} > 0 for both halves of the data (validating on second half the model fit to first half, and validation on first half of model fit to second half).

2. **Figure02-SSR2 Bar charts summarizing 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.

A SSR regression model is a stepwise regression of the predictand, y , on the tree-ring chronology lagged -2 to $+2$ years from y . Variables are entered forward stepwise according to maximum reduction in residual variance. Entry is stopped at the earliest step one of the following conditions are found: 1) adjusted R-squared of calibration reaches a maximum, or 2) cross-validation RE_{cv} reaches its first local maximum.

3. **Figure03-SSR3. Drop in maximum SSR signal-strength with loss of chronologies toward present.** X-axis is number of chronologies, from N4 at left end to 1 at right end. Left y-axis is number of chronologies, or SSRs, available in each year. Right y-axis is the maximum adjusted R-squared of the SSR models for those chronologies. Depending on year of collection, chronologies may drop out of the network before the most recent year available for any of the screened SSRs. Reconstruction method LR1 statistically extends all screened SSRs to the ending year of the most recently ending SSR to allow a maximum possible length of calibration period for the MSR. This may not be a good strategy if the SSRs with the strongest signals drop out before that most recent ending year.

TRISH gives an option for specifying the end year of the calibration period, and this plot will help guide you by indicating loss of signal with data extension. 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 R² of SSR) had the end year 2017, it might make sense to specify the end year for MSR calibration as 2017. But if there was a big drop in maximum R² after some year (say, 2004), you might want to specify 2014 as the end year for calibration of the MSR.

You may wonder how the SSRs are extended statistically to the year of most recent coverage by any SSR. This is done by combination of correlation analysis and quantile ranking. For any year that needs to be estimated for some SSR, the first step is to find which other SSR has data for year in question and also has the highest bivariate Pearson correlation with the SSR needing estimation (correlation based on full available overlap of pairs of SSRs). One SSR would then be the predictor for the missing value of the other SSR. If the year to be estimated is 2017, say, the overlap period of the two SSRs would be examined and the quantile matching the 2017 value of the predictor SSR determined. The same quantile for the predictand in that overlap would be assigned as the SSR for the predictand series in 2017.

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 set of screened SSRs and whether the of observed y with the mean of the SSR 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 all PCs on the screened SSRs. This heat map summarizes modes of variability of tree-ring chronologies after converting them to SSRs with lagged regression.
7. **Figure07-Calibration1.** Bar plot of correlations and autocorrelations. Blue bars have length proportional to correlation of the predictand, y , and the PCs of the screened SSRs. Magenta bars are lag-1 autocorrelations of the PCs. A 95% confidence band on the correlations is plotted by default, but the user can change this to 90% or 99% within TRISH. 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.
8. **Figure08-Calibration2.** Summary of 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 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 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 observed y . The acf's similarly allow comparison of the autocorrelation properties of observed and reconstructed y over their common period.
10. **Figure10-AnalysisResiduals1.** Analysis of residuals for normality and 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 are tested for trend using the non-parametric Mann-Kendall trend test (Wilks 2019). The plot includes a non-parametric best-fit line as described by 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 reflect possible adjustment of significance of trend for autocorrelation in the residuals, beyond the autocorrelation represented by the 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, $\alpha = 0.05$).
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.
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 perfects, 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 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 reconstruction with confidence interval. The full-length reconstruction of y is plotted with its 50% confidence interval computed as $\hat{y} \pm 0.67449$ RMSE, where \hat{y} 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.
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 whether the reconstruction is more or less autocorrelated in the instrumental period than in previous years, and whether the distribution (e.g., spread and level) of reconstructed y is greater or less in the instrumental period than in earlier years.

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. **N₁**: sequential number in this table
 2. **N₂**: corresponding site number in original tree-ring network
 3. **Site**: 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. **R2a**: adjusted R-squared of model; this is the regression R-squared adjusted downward as a penalty for number of predictors in the model.
 8. **pF**: p-value of the overall F of regression (pF<0.05 indicates significant model at 0.05 level)
 9. **REcv**: Reduction of error (RE) statistic from leave-9-out cross-validation
 10. **REa**: Split sample RE for fitting model to first half of record and validating on second half
 11. **REb**: Split sample RE for fitting model to second half of record and validating on first half
 12. **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. .
 13. **Gor**: First year of reconstruction.
 14. **Endr**: Last year of reconstruction.
 15. **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. $pF \geq 0.05$
 2. $REcv \leq 0$
 3. $REa \leq 0$ or $REb \leq 0$
 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 indicated 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 interpretation is complicated. The interpretation must consider the signs of the regression coefficients in the distributed-lag regression model that converts a chronology to an 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, $\alpha R=0.05$ 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 displayed 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 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 reconstructed 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 year of calibration period
 2. **YearStop:** End year of calibration period
 3. **pNormal:** p-value of Lilliefors test for normality ($pNormal < 0.05$: reject H_0 that residuals from normal distribution)
 4. **TrendSlope:** slope coefficient of non-parametric fit 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 \hat{y} . Confidence interval is estimated assuming that reconstruction errors are normally distributed with a standard deviation equal to RMSEcv.
 1. **Year:** year of data
 2. **y :** observed predictand
 3. **\hat{y} :** 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., 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>