Standardization II:

Standardization

1. Objective
2. Alternative approaches
3. Detrending
4. Averaging
5. Common signal
6. Variance stabilization
7. Adjusting for persistence

“Common Signal”

1. Measure of strength of co-variation in core indices from different trees
2. Defined relative to growth of the trees only – not to any particular driving factor (e.g. precipitation)

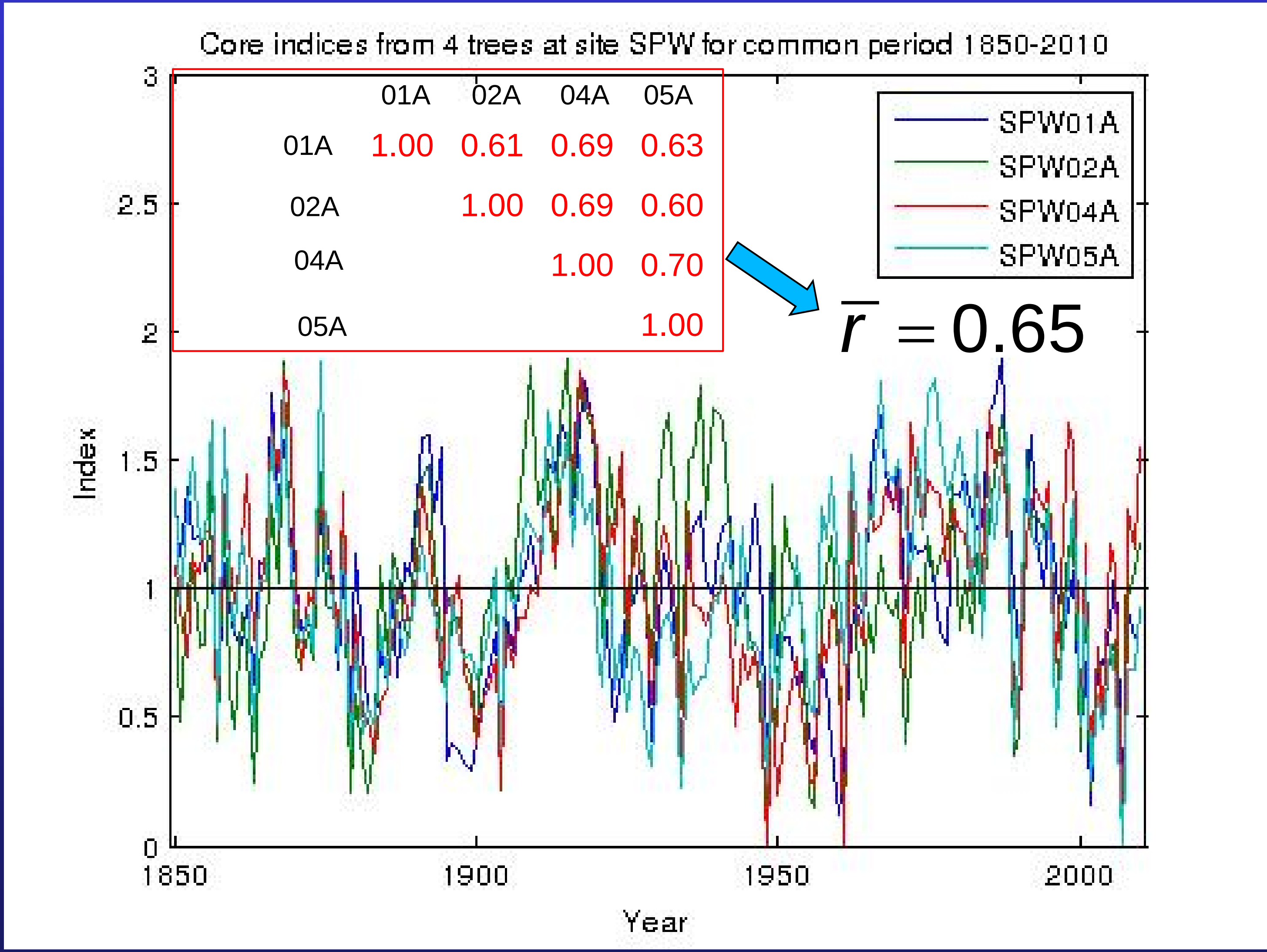
4. Assessment of Common Signal

The common growth signal in the site chronology will be either weak or strong, depending on how well the core indices from different trees show the same patterns of variation

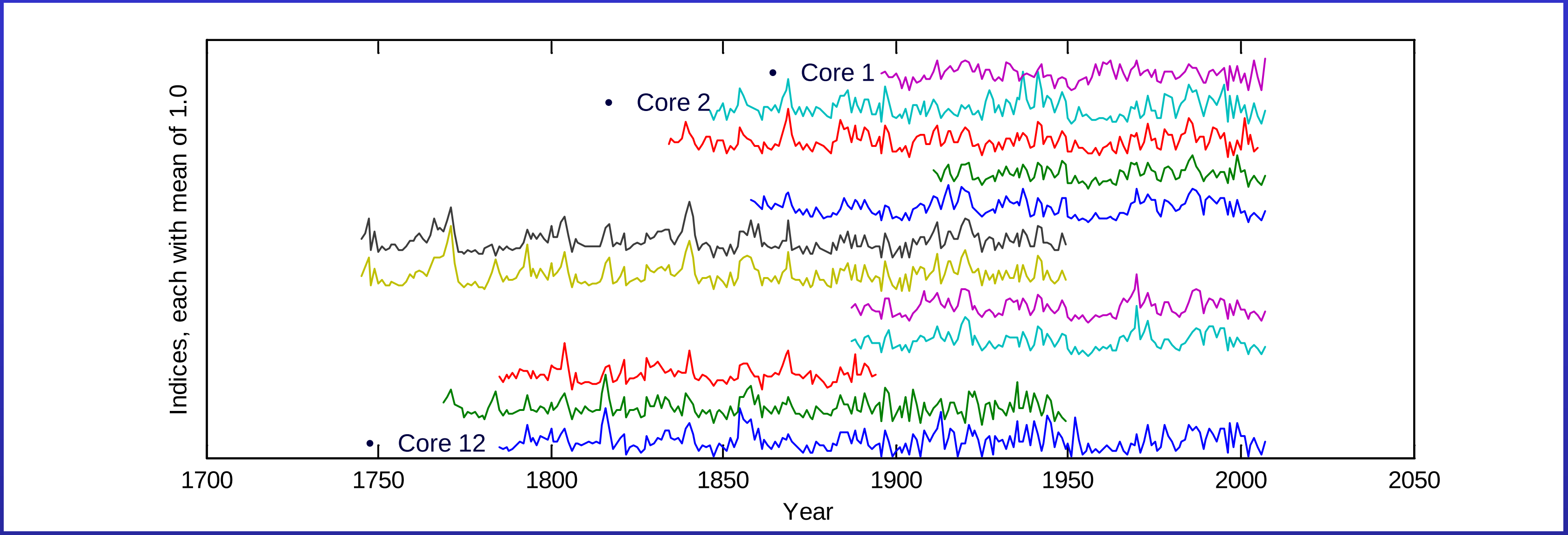
$$\bar{r}$$

"r-bar", the average inter-series correlation of core indices is a statistical measure of the common signal

r-bar for set of cores from SPW



Sample Depth and the Common Signal



1. Higher inter-series correlation → smaller sample needed to summarize population signal
2. Sample drops too low → sample inadequate for inferring climate
3. Two important tree-ring statistics to help assess what is an “acceptable” sample size:
    1. EPS, expressed population statistic → how well the sample represents the unknown population (if we had sampled an infinite number of trees)
    2. SSS, subsample signal strength → how well an *m*-core chronology represents the *n*-tree chronology, where *m* is the number of cores in some year, and *n* is the maximum number of cores in any year

Expressed Population Signal (EPS)

$$\text{EPS} = \frac{nr}{1 + (n-1)r}$$

n = number of trees

r = mean between-tree correlation

- measures how well an *n*-tree chronology represents the population (an ∞-tree chronology)
- Increases with *r* and *n*
- If just one tree, EPS=*r*
- One rule of thumb is that sample size should be large enough so that EPS≥0.85 before the chronology is used for climatic inference
- Equation assumes one core per tree; program ARSTAN modifies the equation to handle multiple cores per tree
- Derived from theory of an average of correlated random variables (Wigley et al. 1984); derivation strictly applies to series that are not autocorrelated

Subsample Signal Strength (SSS)

$$SSS_n = \frac{EPS_n}{EPS_N}$$

n = number of trees in given year

N = maximum number of trees in any year

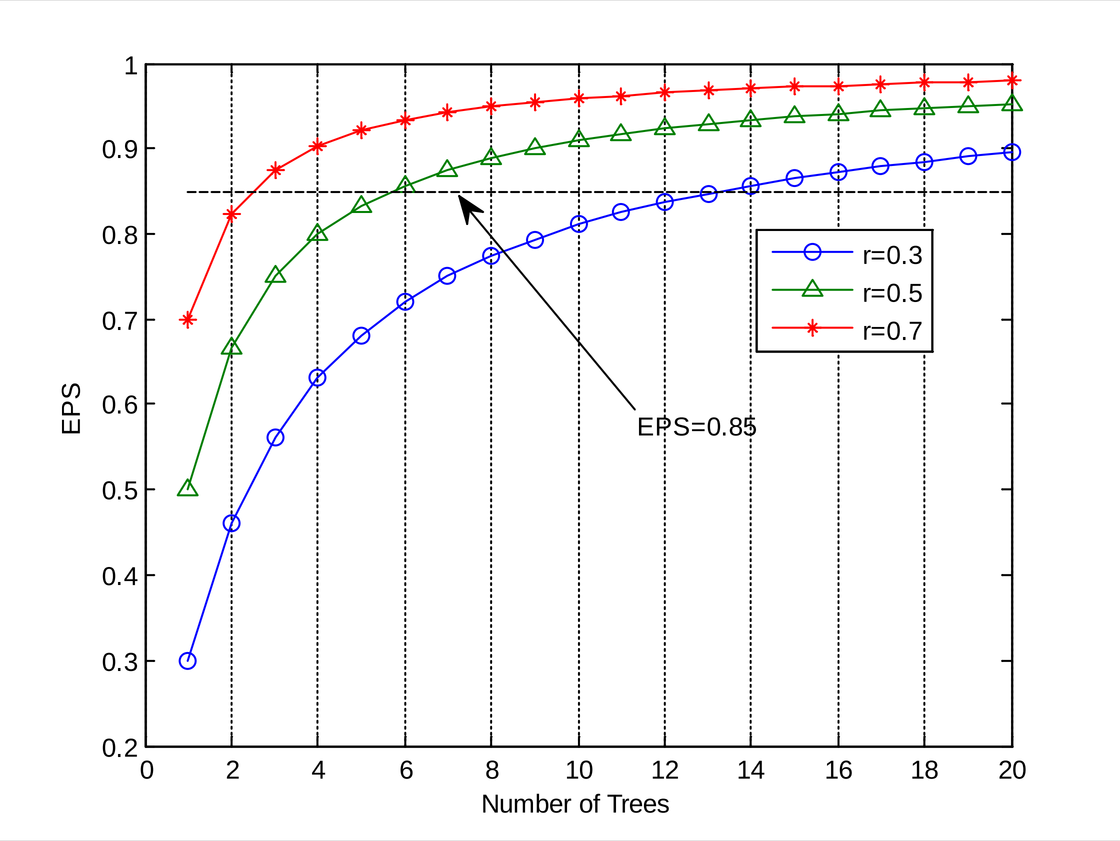
- Measures how well an n-tree chronology approximates an N-tree chronology
- Of interest because the N-tree chronology is used to calibrate the reconstruction model, while the n-tree chronology is the predictor in earlier centuries
- A rule of thumb is that the reconstruction should not be extended to the period for which SSScarried back further than the period for which SSS<0.85
- If EPS≥0.85, it is also true that also guarantees that SSS≥0.85

EPS graph

$$\text{EPS} = \frac{nr}{1 + (n-1)r}$$

- The equation for EPS (above) can be used to build a graph
- The graph tells how many trees are needed to reach a “critical” EPS for different levels of r-bar

EPS as function of number of trees



- If r=0.3, need 14 trees for suitable EPS, while if r=0.7 need only 3 trees
- Slope of curve gets more gradual the larger the number of trees: biggest gain with additional sampling happens when sample size is small
- For example, huge benefit in going from 2 to 4 trees, but small benefit in going from 18 to 20 trees

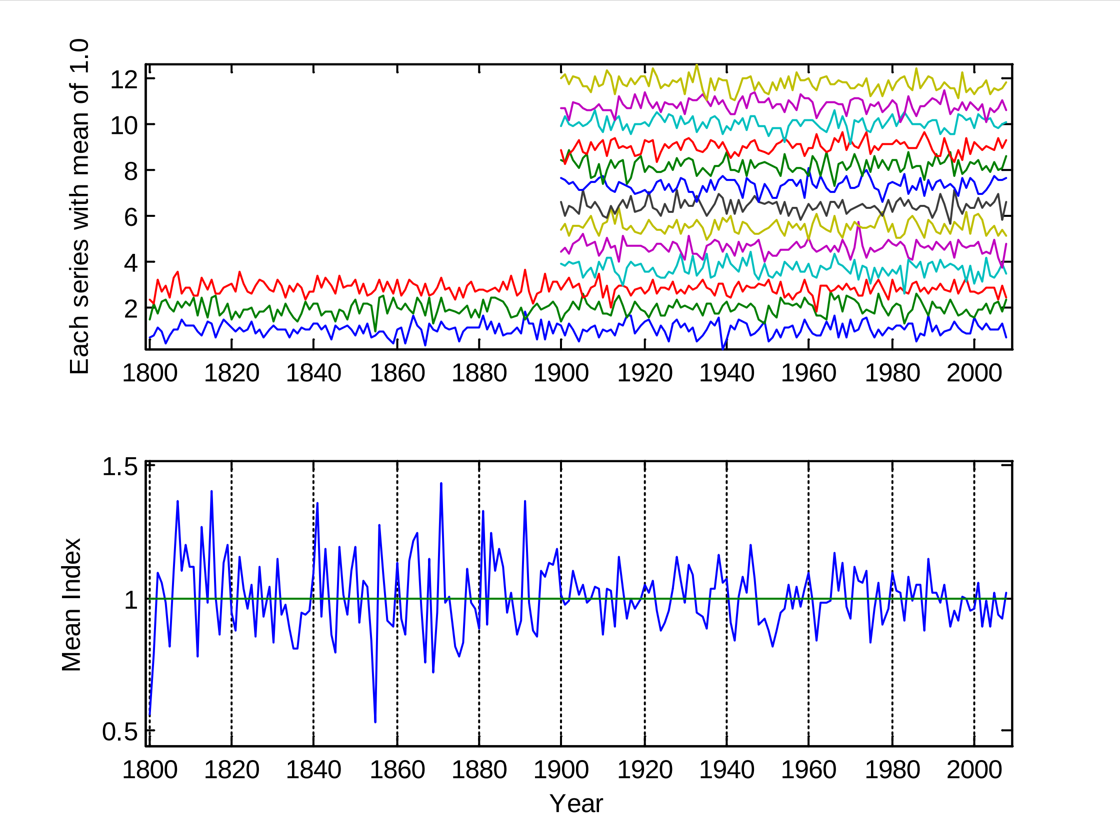
Standardization

1. Objective
2. Detrending
3. Averaging
4. Common signal
5. Variance stabilization
6. Adjusting for persistence

5. Variance Stabilization

- This is an adjustment of the variance of the site chronology for temporal changes in sample size (number of trees, cores)
- Bigger sample size →average over cores is better at smoothing out, or removing the ring-width variations specific to only one or a few trees, and at capturing the growth signal common to all trees
- The variance of the site chronology generally decreases as the sample size increases. This trend in variance should be removed because it is due to changes in sample size (number of trees) and not to climate

Variance Inflation in Simulated Mean Chronology



- Top: 10 simulated core indices, random normal series with mean 1.0 and standard deviation 0.3. Series shifted vertically for plotting
- Bottom: mean chronology as arithmetic mean of the available series
- Sample size changes abruptly from 3 to 10 at year 1900. At same time, variance decreases
- Pattern reflects relationship of varianceof a sample mean to the individual variances of samples (ratio=1/n)

Variance Stabilization

$$y_t = x_t \sqrt{n'_t}$$

x_t site chronology, as departures from mean, before adjustment

n' effective independent sample size

$$n'_t = \frac{n_t}{1 + (n_t - 1)r}$$

n_t number of cores in year t

r mean between-tree correlation

See Osborn et al. (1997)

Standardization

1. Objective
2. Detrending
3. Averaging
4. Common signal
5. Variance stabilization
6. Adjusting for persistence

6. Adjusting for persistence

Persistence, or autocorrelation, is sometime removed from tree-ring indices before use of those indices in climate reconstruction.

The reasons are

1. Standard tree-ring chronologies are sometimes much more highly autocorrelated than the climate variable being reconstructed
2. Autocorrelation in core indices can vary greatly from tree-to-tree, suggesting this autocorrelation has a biological source
3. There are many plausible explanations for autocorrelation being biological in origin

The “residual” index

Two versions of the site chronology are the “standard” and “residual” index

The standard index is based on core indices computed as the ratio or ring-width to fitted trend line (standard core indices)

The residual index is based on autoregressive (AR) residuals of those standard core indices. Steps:

1. Multivariate autoregressive modeling of all standard core indices to select a “pooled” autoregressive order and estimated model coefficients.
2. AR modeling of each standard core index by the pooled AR modeling; define the residuals as the residual core index
3. Average over cores to get the residual site index

Computation of a Residual Core Index

$$e_t = x_t - \hat{a}_1 x_{t-1} - \hat{a}_2 x_{t-2}$$

$x_t =$ standard core index

$\hat{a}_1, \hat{a}_1 =$ estimated autoregressive parameters

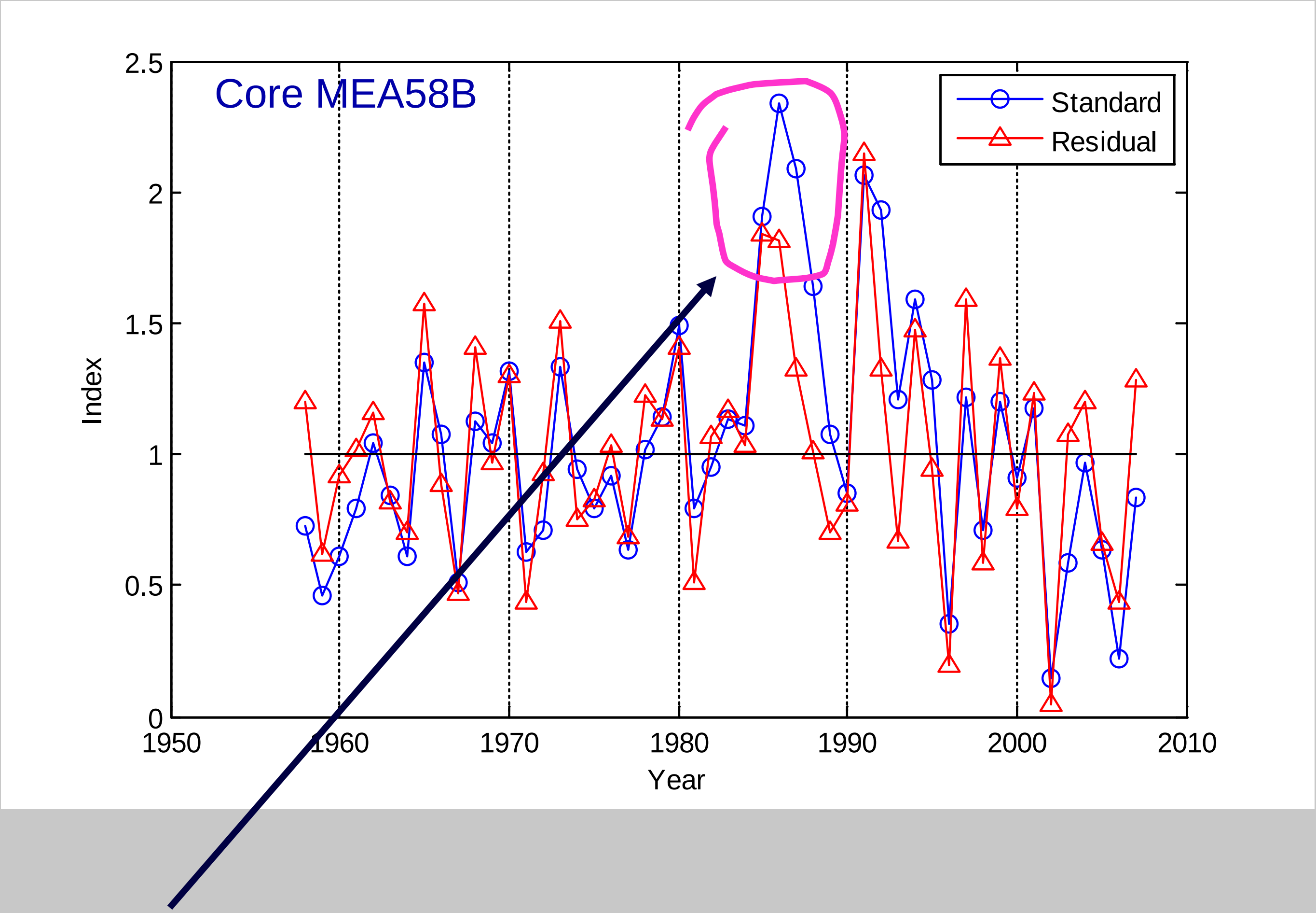
$\mathrm{e}_t =$ residual core index

The equations above assume that the standard core index has had its mean subtracted, so that the mean is zero; and that the autoregressive model is order 2

*The "ARSTAN" version of site chronology

- This is a third version (standard, residual, ARSTAN) of site chronology made by program ARSTAN.
- It is produced from the residual chronology by restoring the persistence using the coefficients of the multivariate autoregressive, or "pooled" model
- The idea, developed by Ed Cook in his dissertation, is that persistence in common among trees could be a large-scale environmental signal.
- The ARSTAN chronology is rarely used in climate reconstructions. You can forget about it.

Example: a standard and residual core index



Departures from mean for residual index are less persistent than departures for standard index