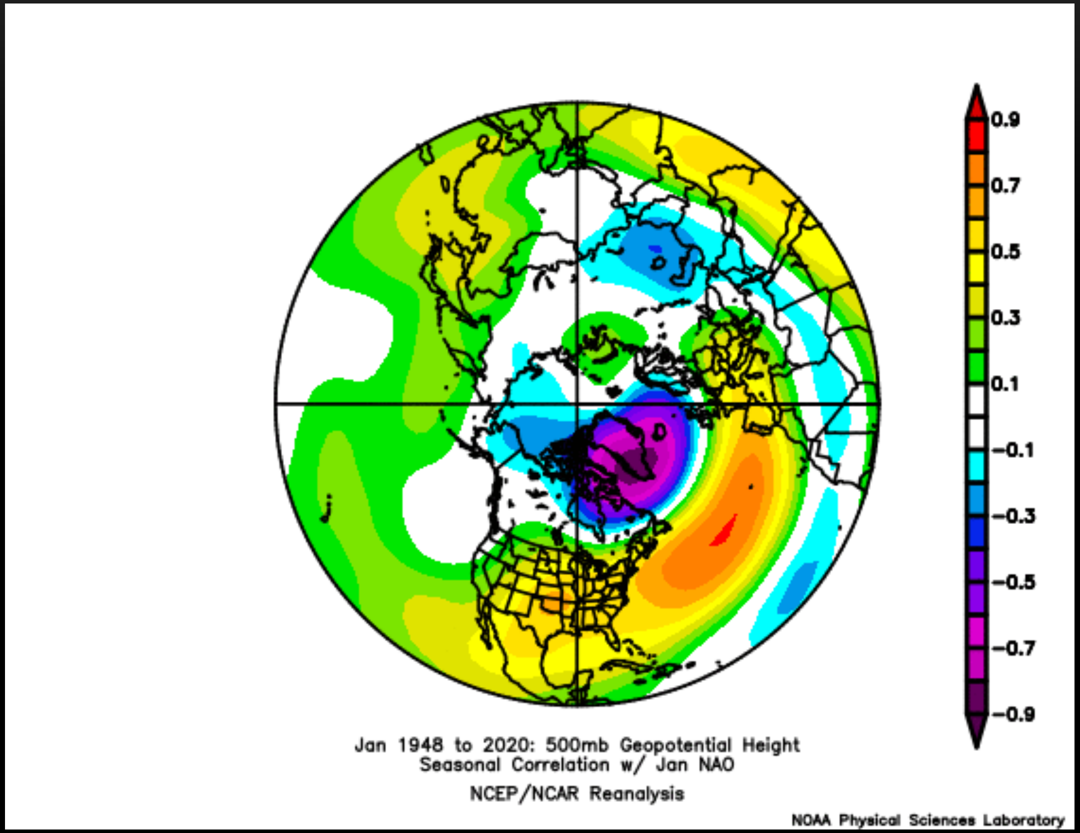
Statistical Techniques for Identifying Climate Modes of Variability

Climate Modes of Variability

- The climate system is characterized by recurrent patterns of variability, such as the ENSO.
- The identification of climate modes is useful because low-frequency modes of variability are important sources of predictability for medium- to long-range prediction.
- We will provide an overview of some statistical methods here
    - Correlation maps
    - Composite analysis
    - Empirical Orthogonal Functions (EOFs)
    - Singular Value Decomposition (SVD)
    - Cluster Analysis

Correlation Maps

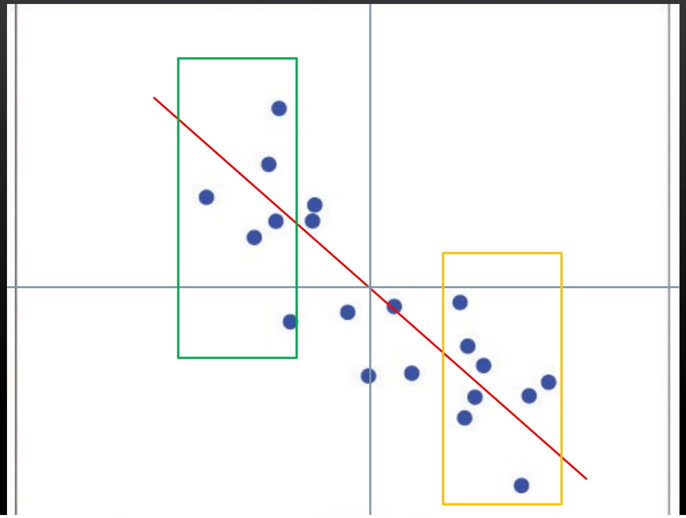
- Correlation maps can effectively reveal the linear relationship between a reference index and a field variable.
- See the example on the right.
- If you use the time series of H500 at the center of the positive node, instead of the NAO index, as the reference time series, you will get the so-called one-point correlation map.
- Caution:
  - Correlation does not imply causality
  - One-point correlation map is sensitive to the reference point. Instead of choosing the time series at a single grid point, the areal average over a certain domain may produce more robust results.

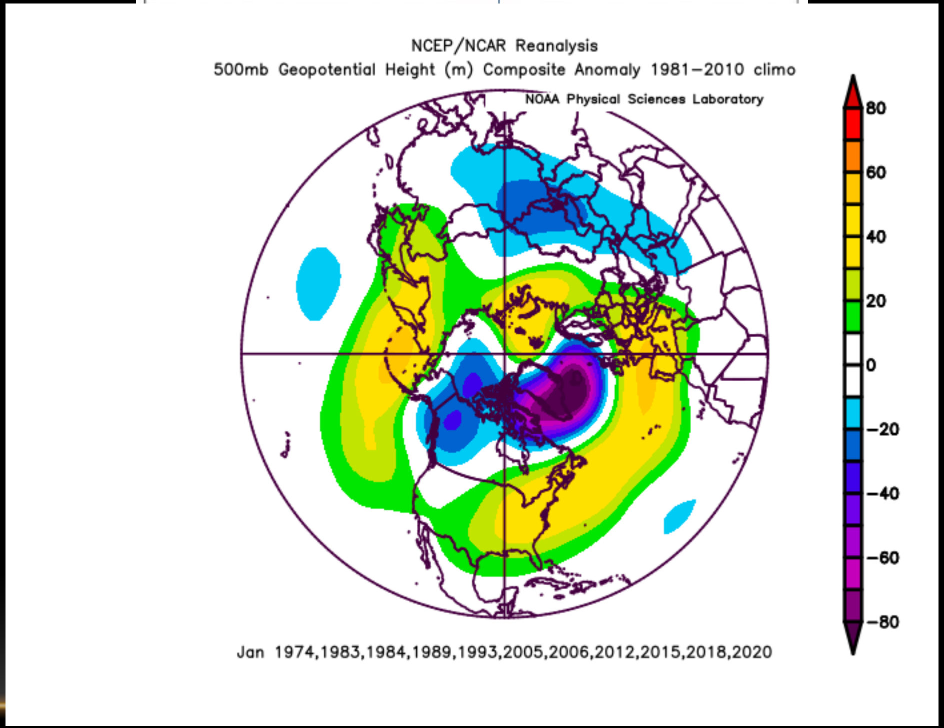


The example on the right shows the correlation between H500 and the NAO index in Jan from 1948-2020. Produced from https://psl.noaa.gov/data/correlation/

Composite Analysis

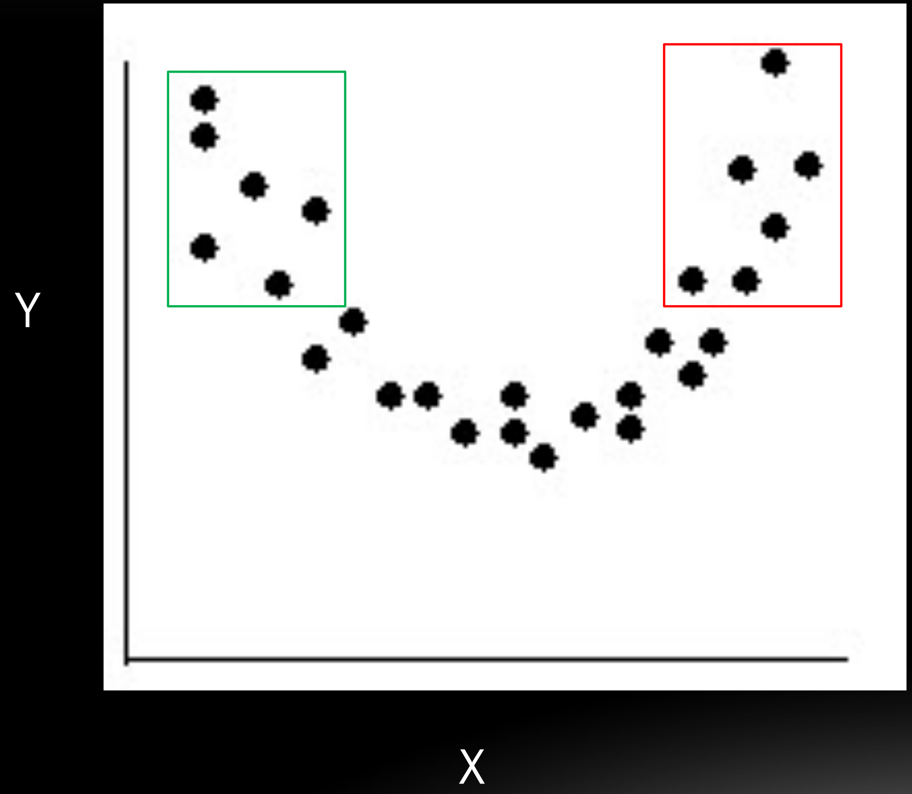
- Climate modes of variability can be determined through composite analysis. In composite analysis, you will average patterns with similar features. The average can be carried out based on a reference index.
- The average is expected to smooth out sample fluctuations and retain recurrent climate patterns. Significance tests can help assess the consistency among members.
- Different from correlation analysis, composite analysis is mainly determined by extreme values.
    - In the example shown in the upper right, the correlation between y and x is derived using all data points. In contrast, only points with x exceeding a certain threshold go into the composite analysis.
- The H500 composite pattern based on the NAO index is very similar to the correlation map we saw in the last slide.





Composite anomalies of H500 for the years with the normalized NAO index anomaly exceeding 1.0 in Jan from 1948-2020. Produced from https://psl.noaa.gov/cgi-bin/data/composites/printpage.pl

Do correlation and composite maps provide the same information?



- What would you get if you calculate the Pearson correlation between x and y?
- What would you get if you calculate the composites of Y for extreme values of X?

Composite analysis can be used to reveal the nonlinear relationship between x and y.

Composite Analysis (cont'd)

- The figure on the right shows the composite anomalies of 500-hPa geopotential height and SST during June-August for Ei Nino and La Nina years. The two patterns are apparently asymmetric.
- Composite analysis can be used to reveal nonlinearity of a climate mode.

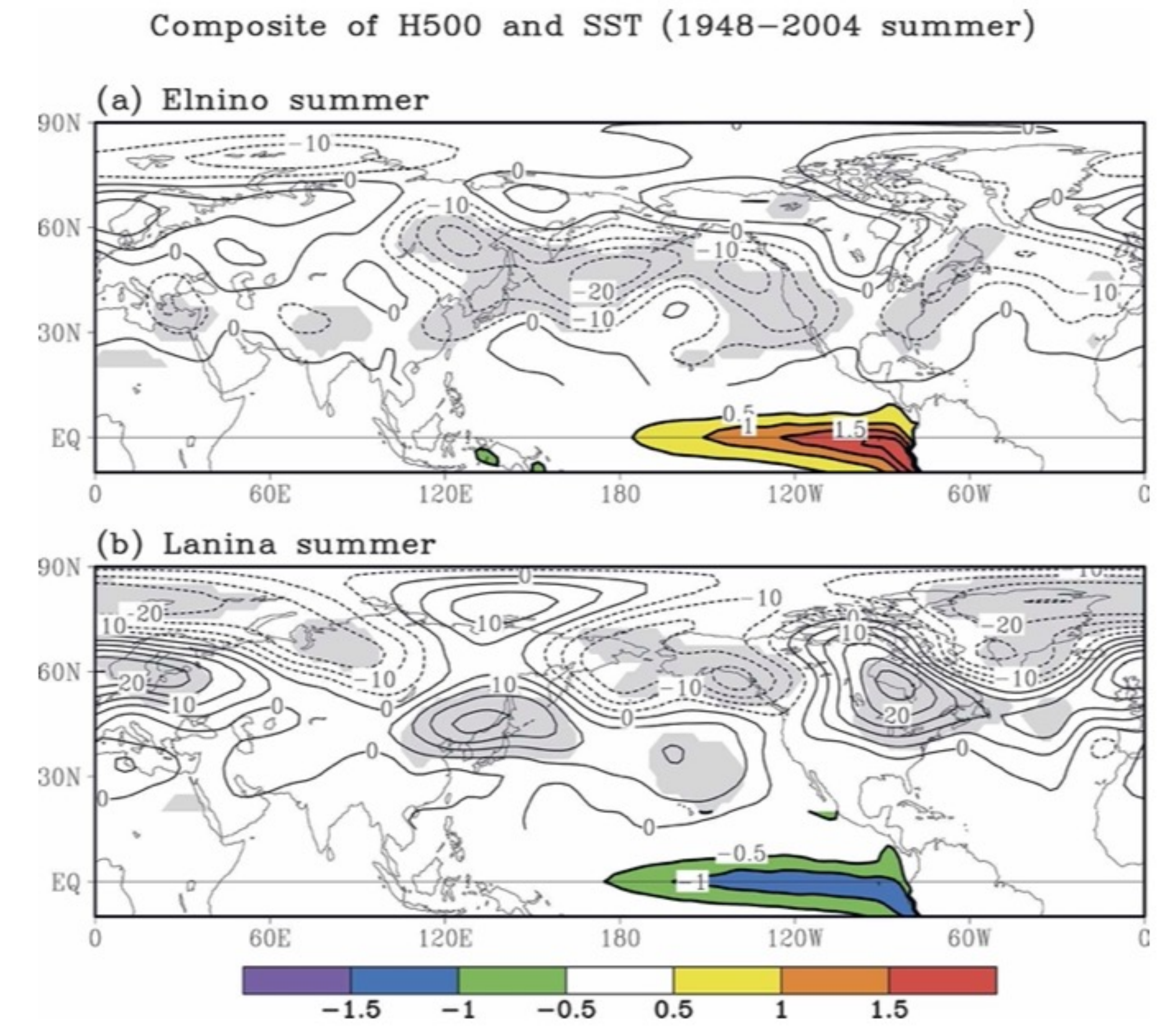


FIG. 1. Composites of 500-hPa height (15°–90°N) and tropical SSTAs (color shading; K) for (a) El Niño and (b) La Niña events during 1948–2004. Gray shading indicates height anomalies above the 95% confidence level. SSTAs below the 95% confidence level are not shown.

Significance Testing: correlation analysis

- To determine if a correlation coefficient, r, is statistically significant you can perform a t-test, which involves calculating a t-score and a corresponding p-value.

$$t = r\sqrt{(n-2) / (1-r^2)}$$

where n is the degree of freedom.

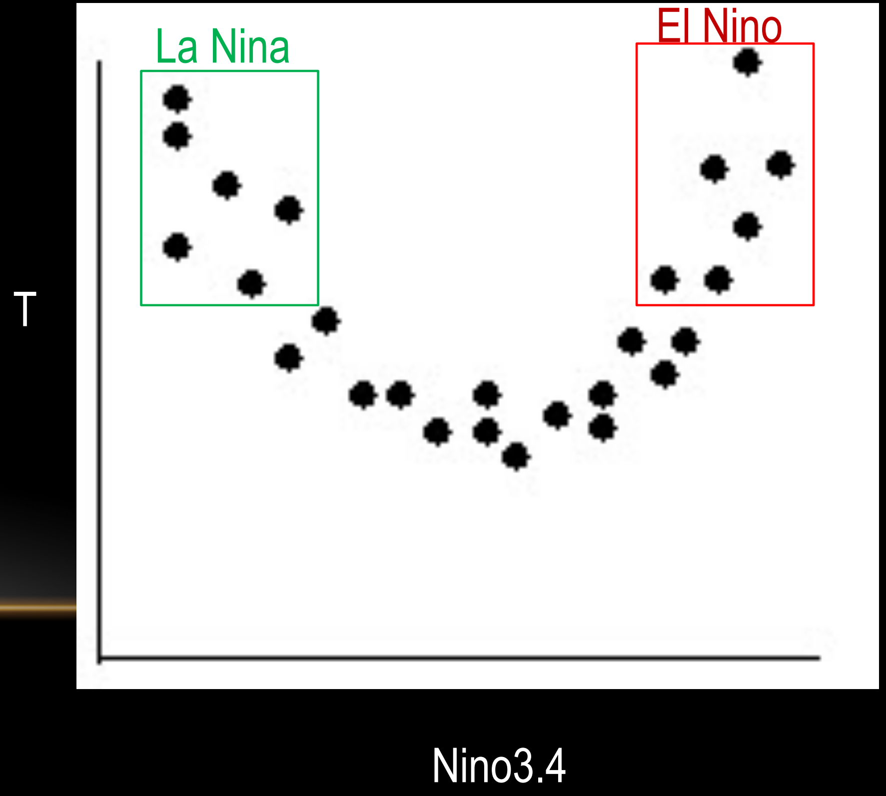
- For data with strong serial correlation, n can be much smaller than the sample size.
  - Chang and Wang (2020) applied 5-year running mean to their time series in the multi-year prediction of Atlantic tropical cyclone activity. Although the time series covered more than 50 years (1954-2015), the effective degree of freedom of the 5-yr-mean time series is between 8 and 10.

Significance Testing: composite analysis

- One sample vs. two sample t-test:
  - One Sample: You are studying one group of data, and the null hypothesis is that an observed sample mean is equal to a specified value. For example, you can test whether the composite mean anomalies of T for strong El Nino years are significantly different from zero.
  - Two Sample: You are studying two groups of data, and the null hypothesis is that the difference of the two sample means is zero. For example, you can test whether the composite mean anomalies of T for strong El Nino years are significantly different from the composite mean anomalies of T for strong La Nina years.
- One-sided (one-tailed) vs. two-sided (two-tailed) t-test
  - One-sided: you want to determine if there is a difference between groups in a specific direction. For example, you can test whether the composite mean anomalies of T for strong El Nino years are significantly larger than zero.
  - Two-sided: you want to determine if there is a difference between groups. For example, you can test whether the composite mean anomalies of T for strong El Nino years are significantly different from zero (either larger or smaller than zero).

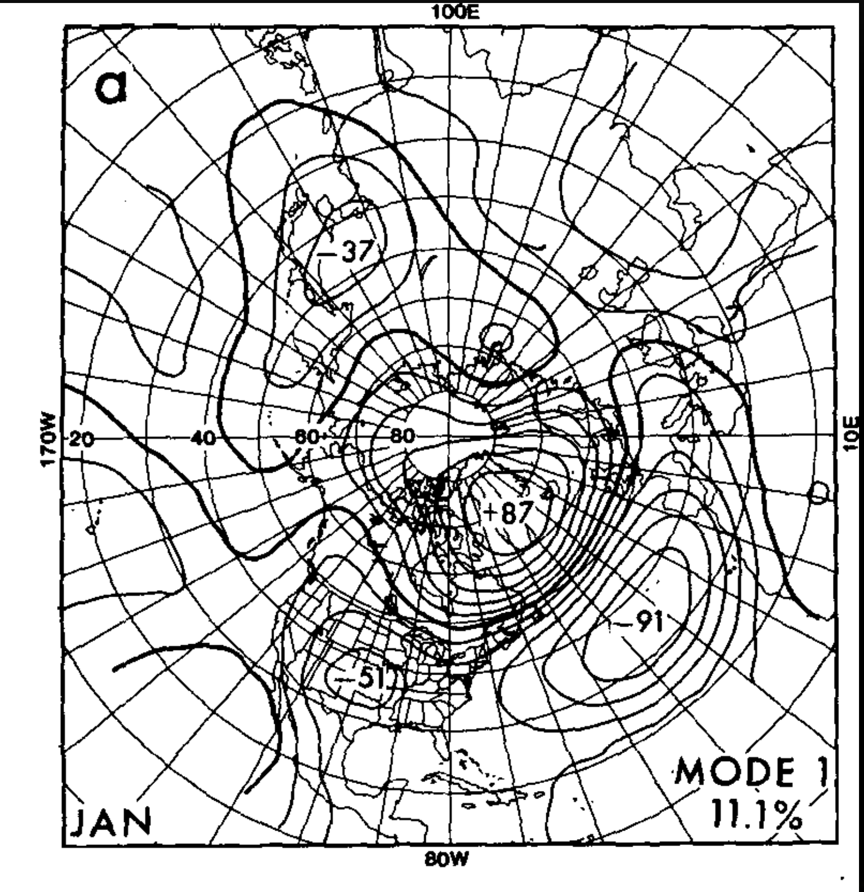
student's t test is often used for significance testing in composite analysis, but that's where it often gets confusing.

A hypothesized example, showing the scatterplot between Nino3.4 and T anomalies



Empirical Orthogonal Functions (EOFs)

- The EOF analysis helps to extract the dominant modes of variability for the time series of a field variable.
- Barnston (1987) showed that the rotated EOF can effectively identify the NAO pattern. It explains 11.1% of the interannual variance of 700-hPa geopotential height in Jan.
- Weaknesses:
  - Sensitive to the domain of analysis
  - Symmetry about the mean state (no sign preference)
  - Higher order modes may not be physically meaningful due to the requirement of orthogonality



The NAO pattern as represented by the leading EOF model of H700. (Barnston 1987 © American Meteorological Society. Used with permission)

Singular Value Decomposition (SVD)

- The singular value decomposition of the covariance matrix between the two fields can identify important coupled modes of variability between the two fields.
  - The two field variables may not have the same spatial dimension.
  - The leading SVD pair explains a larger fractional covariance than the other individual pairs.
- The figure on the right shows the leading SVD pair between the US precipitation and 500-hPa geopotential height.
- The bottom panel shows the corresponding time series of the SVD modes. This pair explains 32% of the covariance between the two fields.

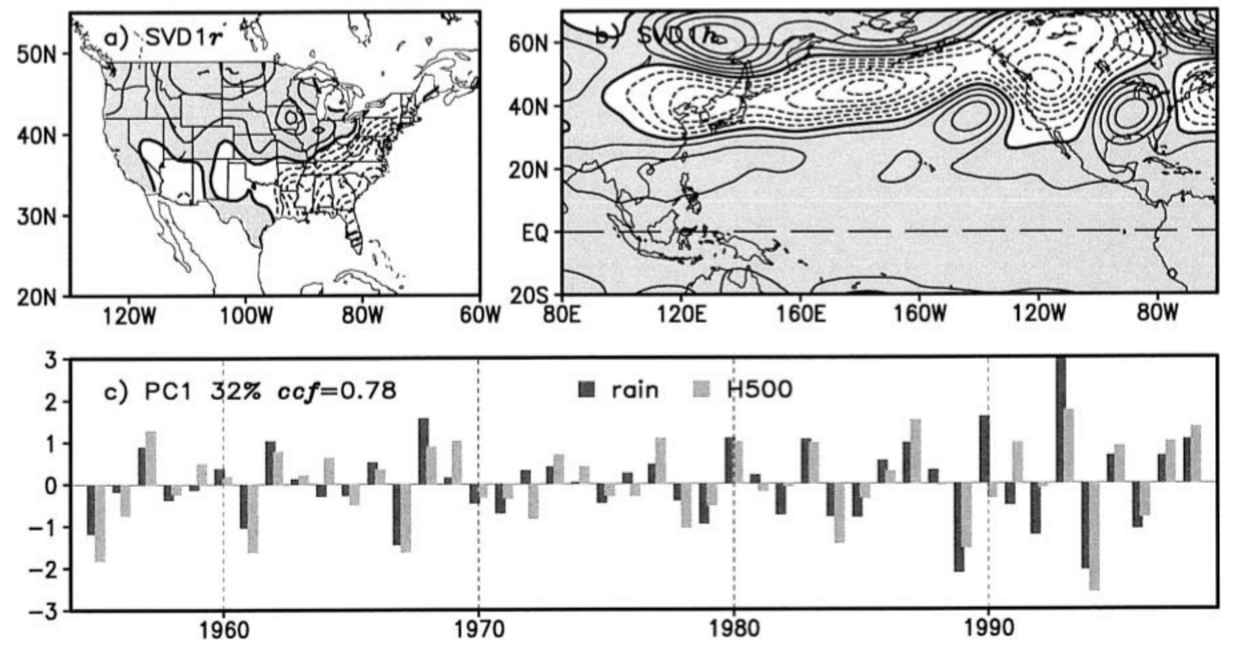


Fig. 2. Spatial patterns of SVD1 between JJA US rainfall and 500 hPa geopotential height for a) US rainfall, b) 500 hPa geopotential height, and c) corresponding principal components (non-dimensional). The contour interval is 10 mm in a), and 2 dm in b). Positive values are shaded.

From Lau and Weng 2002

Cluster Analysis

- Cluster analysis separates objects into groups in such a way that objects in the same group (called a cluster) are more similar (in some sense) to each other than to those in other groups (clusters).
- Application of cluster analysis in the climatological literature include grouping daily weather observations into synoptic types, defining weather regimes from upper-air flow patterns, grouping members of forecast ensembles, and defining climatic regions based on surface climate variables (Wilks 2011)
- In the example on the right, six clusters are defined based on the standardized precipitation and temperature. The number of clusters, however, needs to be prescribed by users for k-mean clustering.
    - The number can be somewhat subjective.

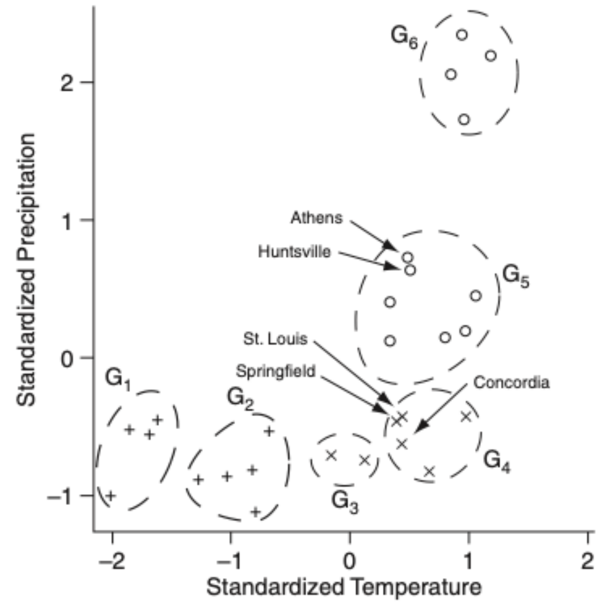


FIGURE 14.4 Scatterplot of the data in Table 13.1 expressed as standardized anomalies, with dashed lines showing the six groups defined in the cluster analysis tree diagram in Figure 14.3a. The five-group clustering would merge the central U.S. stations in Groups 3 and 4. The seven-group clustering would split the two central U.S. stations in Group 5 from six southeastern U.S. stations.

References

- NAS report: "*Assessment of Intraseasonal to Interannual Climate Prediction and Predictability* ", Section 2.3; Appendix A
- Lau, K-M., , and H. Weng, 2002: Recurrent teleconnection patterns linking summertime precipitation variability over East Asia and North America. *J. Meteor. Soc. Japan*, **80 ,** 1129–1147.
- Wilks, 2011: "Statistical Methods in the Atmospheric Sciences", Chapter 14