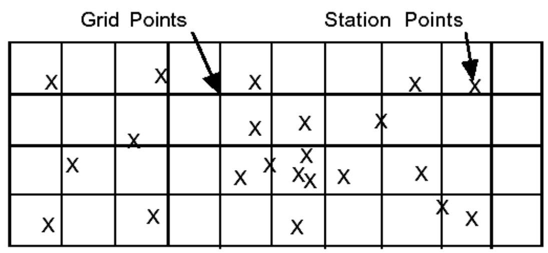
Part 6: Objective Analysis



(from Hartmann 2003)

- Optimum Interpolation
- Composite Analysis

ESS210B
Prof. Jin-Yi Yu

Purpose of Objective Analysis

- Observations are often available only at a few stations that are unevenly spaced in the domain of interests.
- In order to compute derivatives of the field variables, as would be required in diagnostic studies or in the initialization of a numerical model, or simply to perform a sensible averaging process, one often requires values of the variables at points on a regular grid.
- Assigning the best values at the grid points, given data at arbitrarily located stations and perhaps a first guess at regular grid points, is what has traditionally been called ***objective analysis***.

ESS210B
Prof. Jin-Yi Yu

When to Do the Gridding?

- The methods described are applicable to any problem where the data you are given do not fill the domain of interest fully, and/or where the data must be interpolated to a regular grid.
- The gridding can be in space, in time, or both.

ESS210B
Prof. Jin-Yi Yu

Polynomial Fitting Method

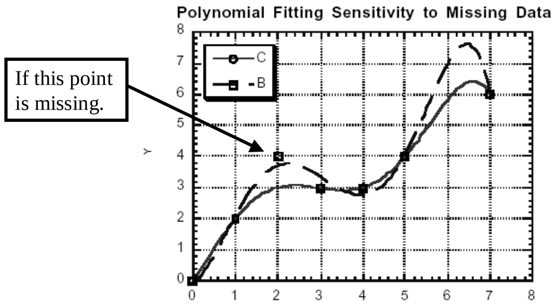
- We can use a polynomial function to fit the observational data and then use this function to generate data on regular grids.
- In meteorological applications, up to third order polynomial have been used. But usually, quadratic equation is sufficient for most purpose:

$$h(x, y) = a_0 + a_1 x + a_2 x^2 + a_3 y + a_4 y^2 + a_5 xy$$

- The problem now it to determine the values of the coefficients: a_1, a_2, …., and a_6.
- If we have 6 observations of h, then we can determine the 6 coefficients in the quadratic polynomial function.
- We can then use the polynomial function for gridding purpose.

ESS210B
Prof. Jin-Yi Yu

Problems With the Polynomial Method



(from Hartmann 2003)

- The instability of the polynomial fit is such that when one key data point is removed, the polynomial fit in that region may change radically.

Problems With the Polynomial Method

- Polynomial fits are unstable in the sense that the values the polynomials give at points between the stations vary greatly for small changes in the data at the station points, and especially so when data are missing.
- The problem gets worse as the order of the polynomial is increased. The method is nearly useless where the data are sparse.

Optimum Interpolation

- The optimum interpolation (OI) is a linear interpolation which requires its root-mean square error to be minimum.
- The discussion of this method focuses on the "deviations from a normal state":
- Let's say we have a variable ϕ and its normal state ϕ_{norm}. This normal state can be the climatological value of ϕ or a first guess of ϕ.
- Then we can define the deviation of ϕ from its normal state as ϕ':

$$\phi' = \phi - \phi_{\mathrm{norm}}; \quad \phi_{\mathrm{norm}} = \overline{\phi} \text{ or a first guess.}$$

Interpolation

- Now, we want to approximate the value of ϕ at a grid point (ϕ_g), in terms of a linear combination of the values of ϕ at neighboring station points (ϕ_i):

$$\phi_g' = \sum_{i=1}^{N} p_i \phi_i' ; \ \phi_g' = \text{grid value}; \ \phi_i' = \text{station values.}$$

- We want to determine the coefficients p_i by minimizing the mean squared error:

$$E = \overline{\left(\phi_g' - \sum_{i=1}^{N} p_i \phi_i'\right)^2}$$

Determine the Weightings

- We can normalized the error E:

$$\varepsilon \equiv \frac{E}{\phi_g'^2} = 1 - 2\sum_{i=1}^{N} p_i r_{gi} + \sum_{i=1}^{N}\sum_{j=1}^{N} p_i p_j r_{ij} \quad \text{where} \quad r_{gi} = \frac{\overline{\phi_g' \phi_i'}}{\phi_g'^2}; \quad r_{ij} = \frac{\overline{\phi_i' \phi_j'}}{\phi_g'^2}$$

- Now we can determine the weightings (***p**$_i$*) by asking:

$$\frac{\partial \varepsilon}{\partial p_i} = -2 r_{gi} + 2\sum_{j=1}^{N} p_j r_{ij} = 0 \qquad i = 1, 2, \ldots, N$$

- We then solve the *N* linear equations for the N *p*'s.
- It can be shown that the error obtained after fitting the coefficients is:

$$\varepsilon = 1 - \sum_{i=1}^{N} r_{gi} p_i$$

ESS210B
Prof. Jin-Yi Yu

What Do We Need to Get P_i?

- We need to know r_{ij} and r_{gi} in order to solve the N linear equation for the N Ps.

  ➔ But we don't data at grid points (can not calculated r_{gi} !).

- It is typical to assume that correlations between points depend only on the distance between them and not on location or direction.
- So we calculate the r_{ij} from station data and obtain r_{gi} from r_{ij} based on the distance between the grid points and the stations.

ESS210B
Prof. Jin-Yi Yu

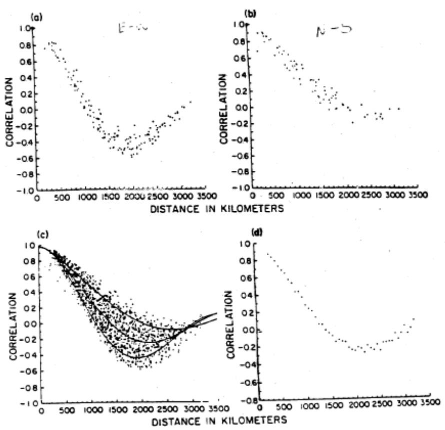


Fig. 5a . Example of anisotropy of geopotential correlation using 500 mb winter data versus between-station distance. a) Height-height correlations for East-West orientation, b) North-South orientation, c) Full array of correlations with top curve fitted to North-South variations and bottom curve fitted to East-West correlations, middle curve to whole array of correlations. D) average correlations values for 50 successive distance intervals of 62.5km. After H.J.Thiebeaux.

(from Hartmann 2003)

ESS210B
Prof. Jin-Yi Yu

Composite Analysis

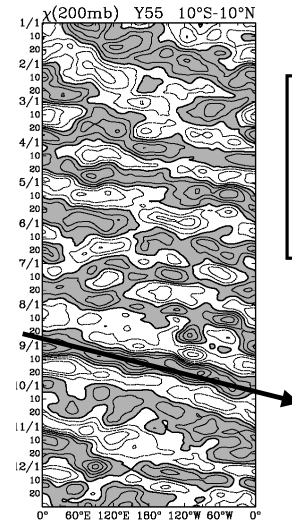
- Compositing analysis (also called superposed epoch analysis) is to sort time series into different categories (or phases) and to compare means in these categories.
- For example, how do Pacific SSTs evolve before, during, and after an El Nino event? In this case, there are three categories (before, during, and after) in the composite analysis.
- Compositing is useful when you have many observations of some event and you are looking for responses to that event that are combined with noise from a lot of other influences.
- The basic idea of the compositing analysis is that the averaging process will remove noise and keep the signals of interest.
- Often compositing will reveal periodic phenomena with fixed phase that cannot be extracted from spectral analysis if the signal is small compared to the noise.

ESS210B
Prof. Jin-Yi Yu

Steps in Compositing Analysis

- Select the basis for compositing and define the categories
  The categories might be related to the phase of some cyclic phenomenon or forcing, or to time or distance from some event. For example, we can use NINO3 index as the basis for compositing the ENSO cycle.
- Compute the means and statistics for each category
  We calculate the mean SST, wind stress, or heat flux for the onset, growing, and mature phases of the ENSO cycle.
- Organize and display the results
- Validate the results
  Validation of the results can be achieved in many ways. Statistical significance tests are only one of these.

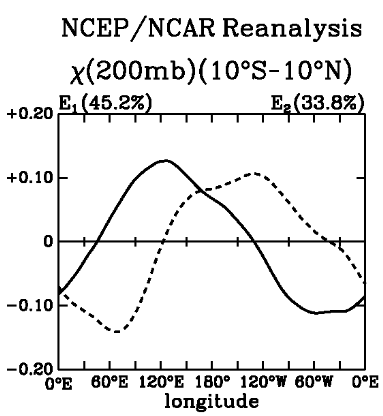
ESS210B
Prof. Jin-Yi Yu



An Example - the MJO Cycle

ESS210B
Prof. Jin-Yi Yu

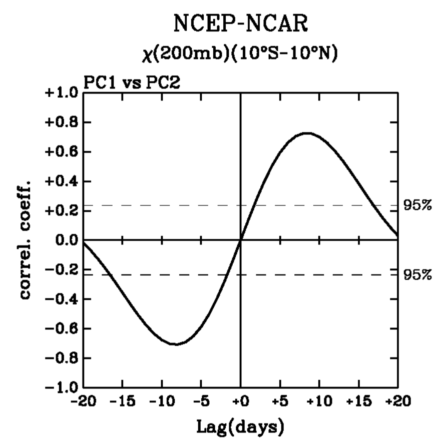
Step 1: Define Index



We use the principal components of the 1st and 2nd EOFs to define different phases of the MJO life cycle.

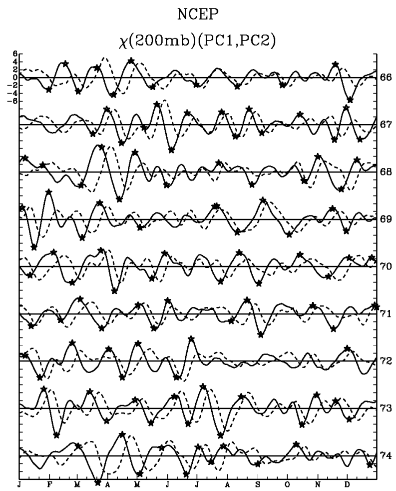
ESS210B
Prof. Jin-Yi Yu

Time-Lag Correlation Between PC1 and PC2



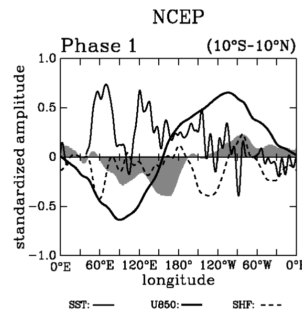
ESS210B
Prof. Jin-Yi Yu

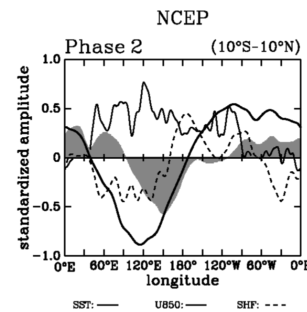
Time Series of the Indices



ESS210B
Prof. Jin-Yi Yu

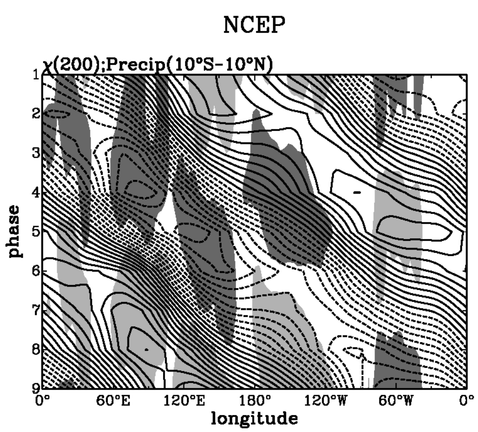
Step 2: Compute the Means





ESS210B
Prof. Jin-Yi Yu

Step 3: Display the MJO Life Cycle



ESS210B
Prof. Jin-Yi Yu