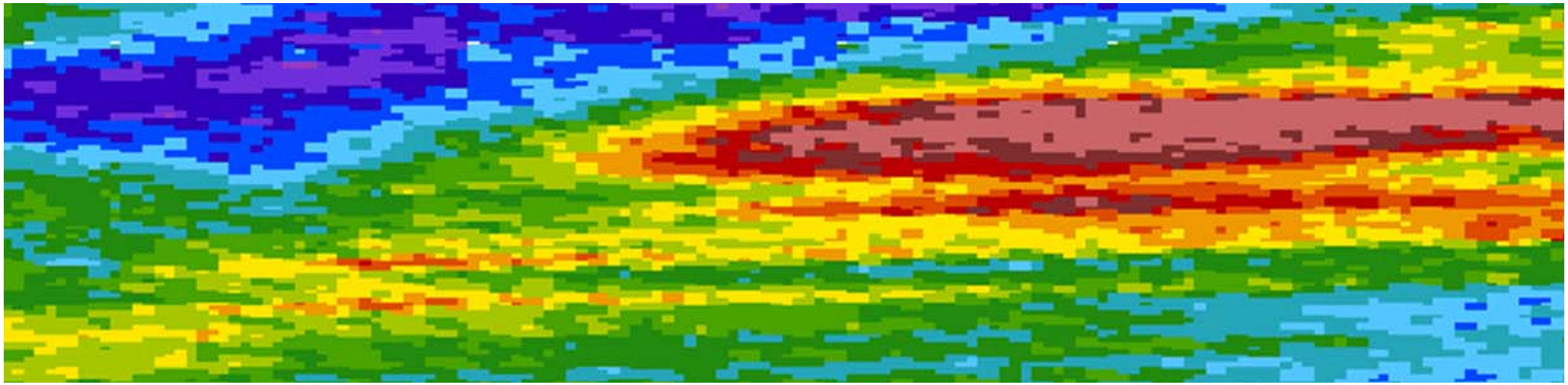


SIO 210: Data analysis, Fall 2019

1. Sampling and error
2. Basic statistical concepts
3. Time series analysis
4. Mapping
5. Filtering
6. Space-time data
7. Water mass analysis

Reading: DPO Chapter 6

Look only at the less mathematical parts
(skip sections on pdfs, least squares, EOFs and OMP except to note that the material is there)

6.1, 6.2, 6.3.1, 6.4,
6.5, 6.6.2, 6.7.1, 6.7.2

1. **Sampling** and error: definitions

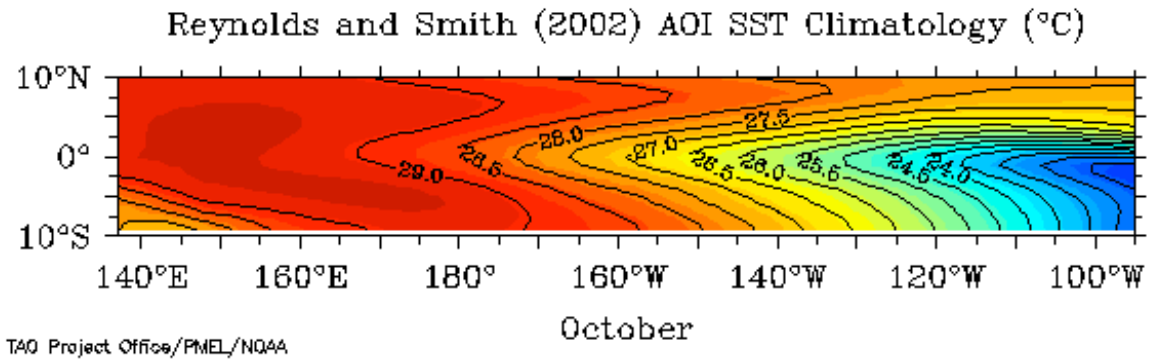
- Sampling (DPO Section 6.1)
  - Synoptic sampling
  - Climatology

  - Mean
  - Anomaly (difference between synoptic measurement and a mean)

1. Sampling: mean and anomaly

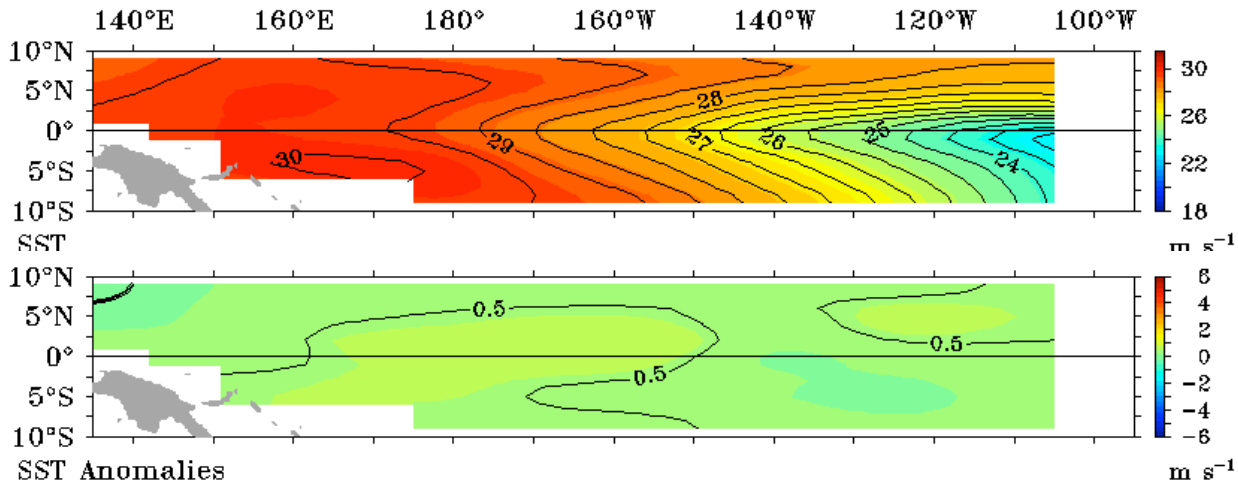
Climatology (mean): mean field based on large data base, usually covering many years

Climatology (Reynolds October Mean)



Observed (Oct. 2012)

Anomaly: Oct. 2012 Minus Reynolds Oct



http://www.pmel.noaa.gov/tao/

2. Basic statistical concepts: Mean, variance, standard deviation

(DPO Section 6.3.1)

Mean: $$\bar{x} = \frac{1}{N}\sum_{i=1}^{N} x_i$$

Anomaly: $x = x_{mean} + x_{anomaly}$; therefore $x_{anomaly} = x - x_{mean}$

Variance: $$\sigma^2 = \frac{1}{N-1}\sum_{i=1}^{N} (x_i - \bar{x})^2 = \frac{1}{N-1}\left[\sum_{i=1}^{N} (x_i)^2 - \frac{1}{N}\left(\sum_{i=1}^{N} x_i\right)^2\right]$$

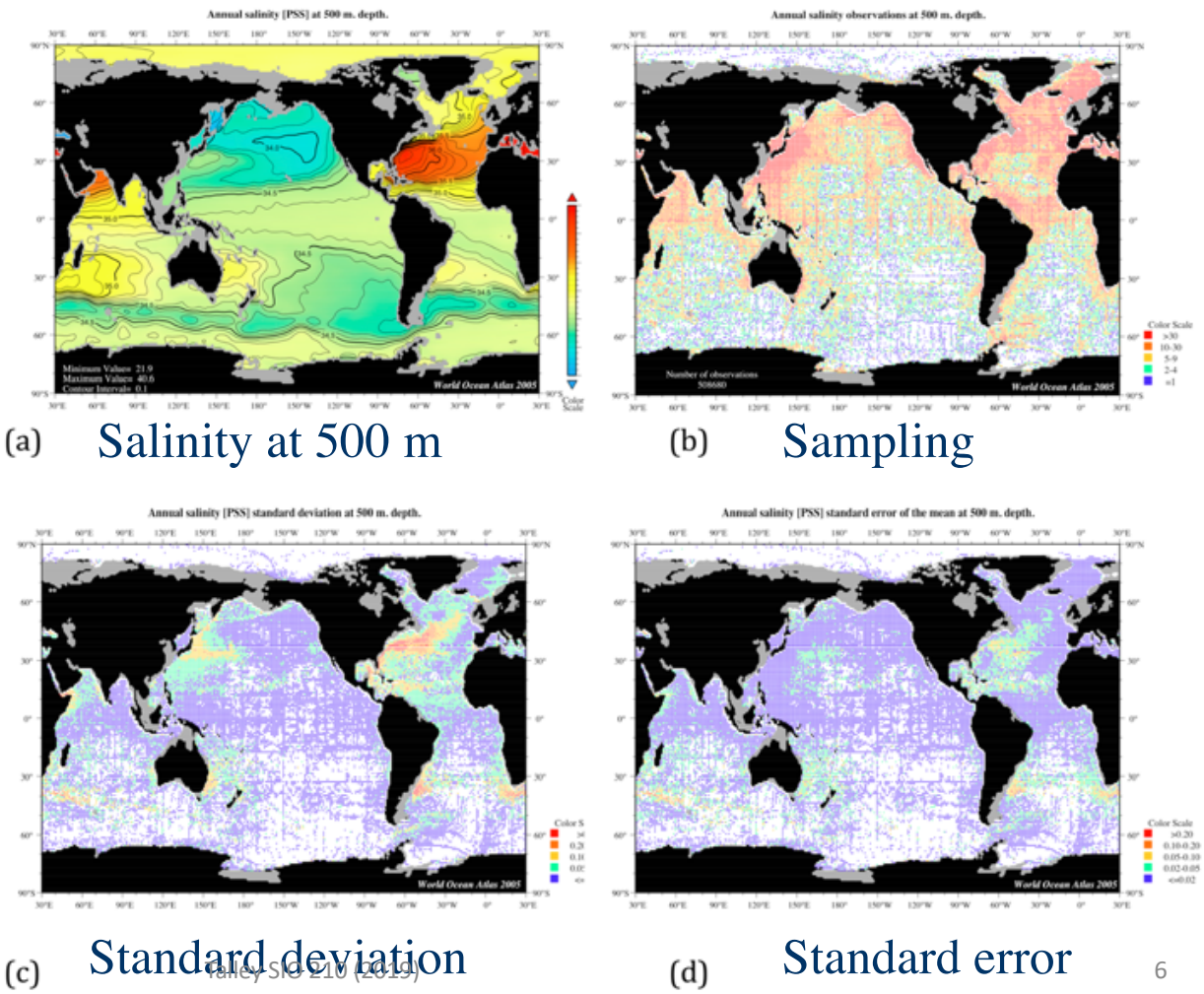
Standard deviation of the measurements: σ (characteristic of the phenomenon)

Standard error: $\sigma/\sqrt{N}$ (characteristic of the SAMPLING since it depends on N)

2. Basic statistical concepts: Standard deviation vs. standard error

Standard deviation is a measure of variability in the field that is measured.

Standard error is a measure of how well the field is sampled.



(a) Salinity at 500 m (b) Sampling

(c) Standard deviation (d) Standard error

DPO Fig. 6.13

3. Time series analysis (DPO 6.3.1, 6.3.3)

A time series is a data set collected as a function of time

Examples: current meter records, sea level records, temperature at the end of the SIO pier

Some common analysis methods:

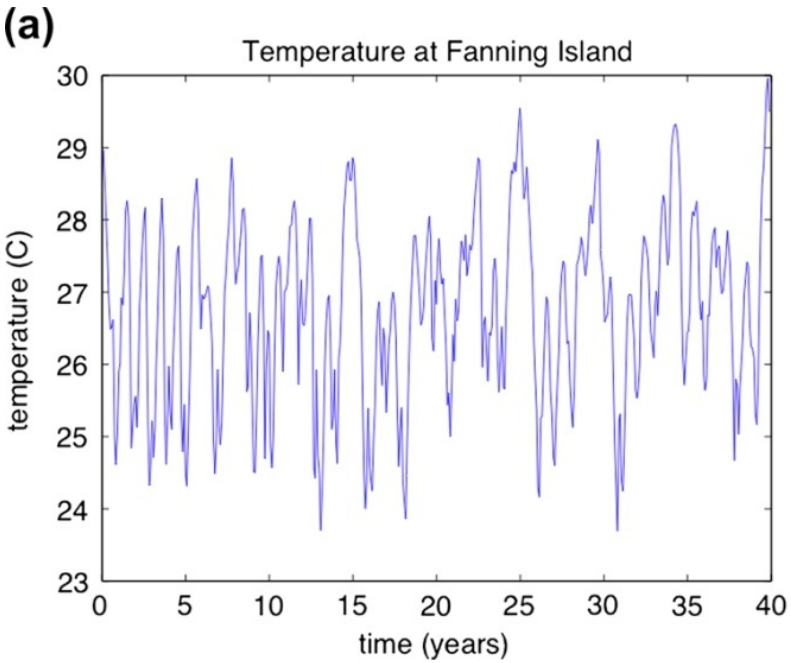
A. Display the data (simply plot)! (always useful, important)

B. Mean, variance, standard deviation, etc. (example not shown)

C. Covariance and correlation of different time series with each other

D. Spectral (Fourier) analysis

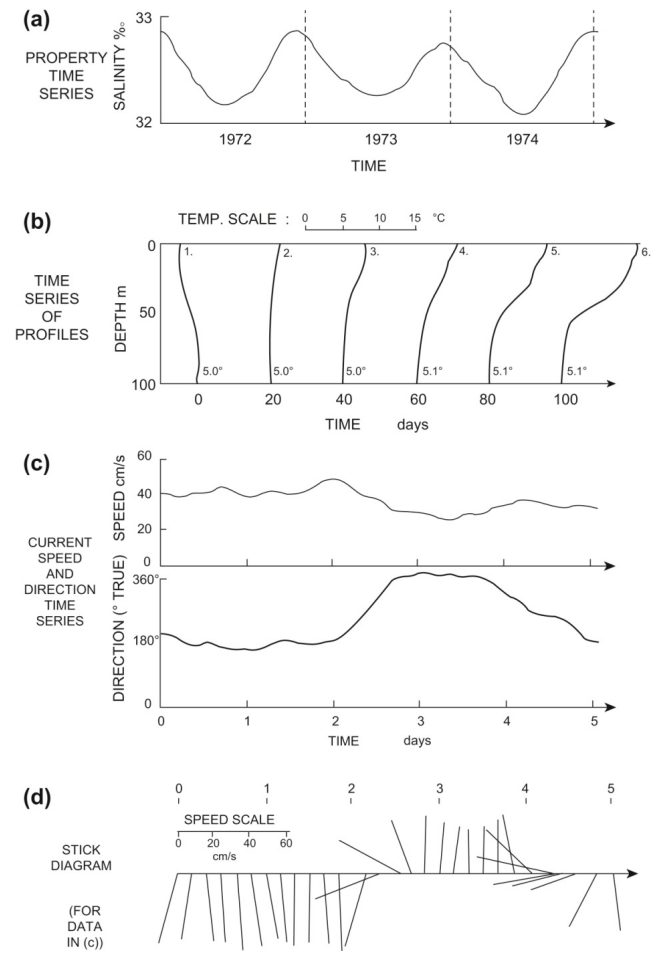
3.A. Display time series



(a) Time series of temperature at Fanning Island (Pacific Ocean) from the NCAR Community Ocean Model.

DPO FIGURE 6.2

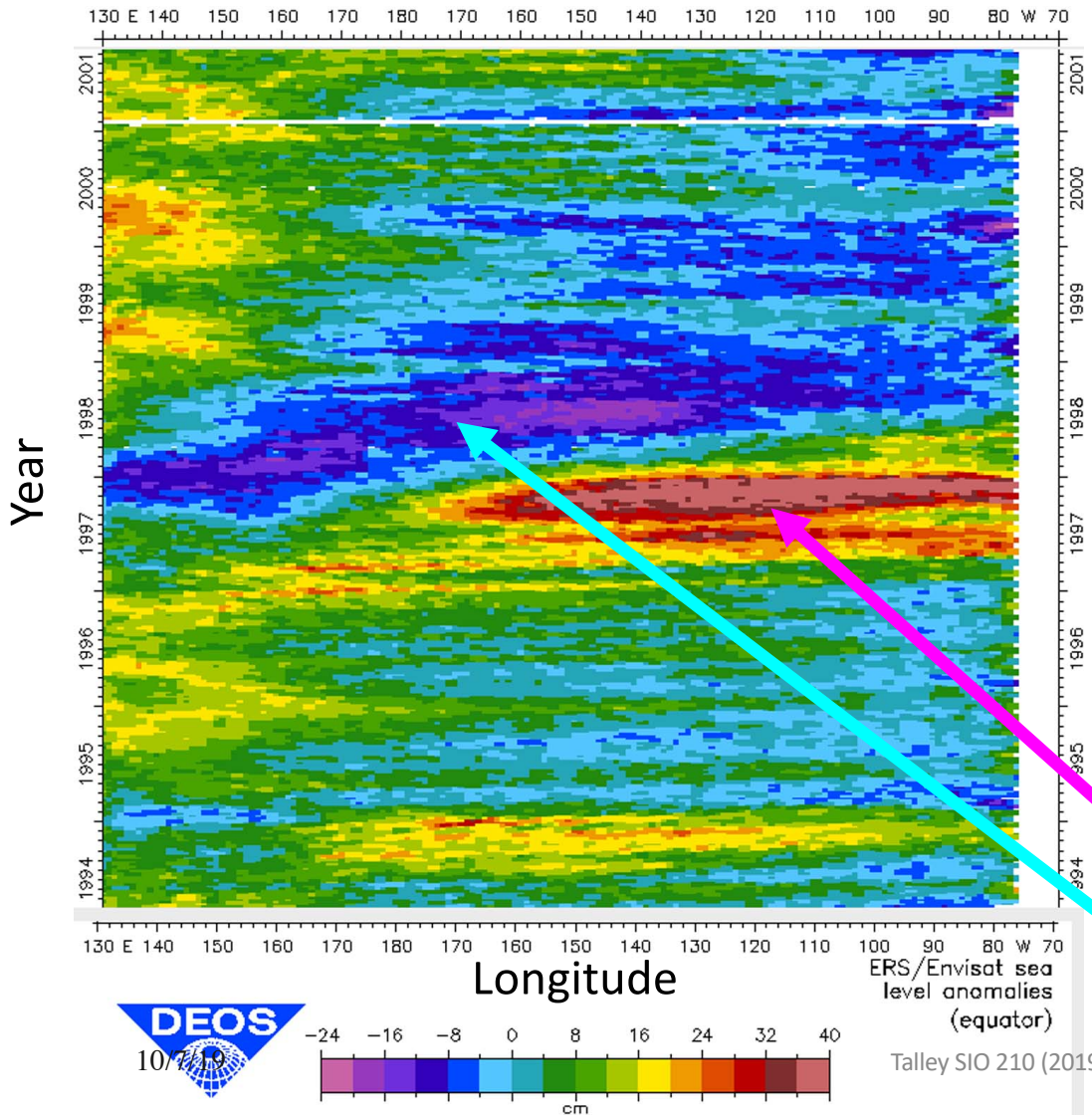
Examples of time series plots:
(a) property/time,
(b) time series of profiles,
(c) current speed and direction, and
(d) stick diagram for data of (c).



DPO FIGURE 6.6

3.A. Display time series

Hovmöller diagram (time on one axis, space on the other)



Equatorial Pacific sea level height anomaly from satellite

(High sea level anomaly - El Nino!)

(Low propagating west to east)

3.C. Time series analysis: covariance (DPO section 6.3.3)

Covariance and correlation: integral of the product of two time series, can be with a time lag. (Autocorrelation is the time series with itself with a time lag.)

- **Integral time scale**: time scale of the autocorrelation (definitions vary) – crude definition might be at what time lag does the autocorrelation drop to zero (which is the decorrelation timescale) (First calculate the autocorrelation for a large number of time lags.)

- **Degrees of freedom**: total length of record divided by the integral time scale – that gives how many realizations of the phenomenon you have. Good to have about 10.

3.C. Time series: covariance and correlation (DPO Section 6.3.3)

Covariance: how two variables are related statistically.

$$\boxed{\mathrm{cov}(x,y) = \frac{1}{N-1}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}$$

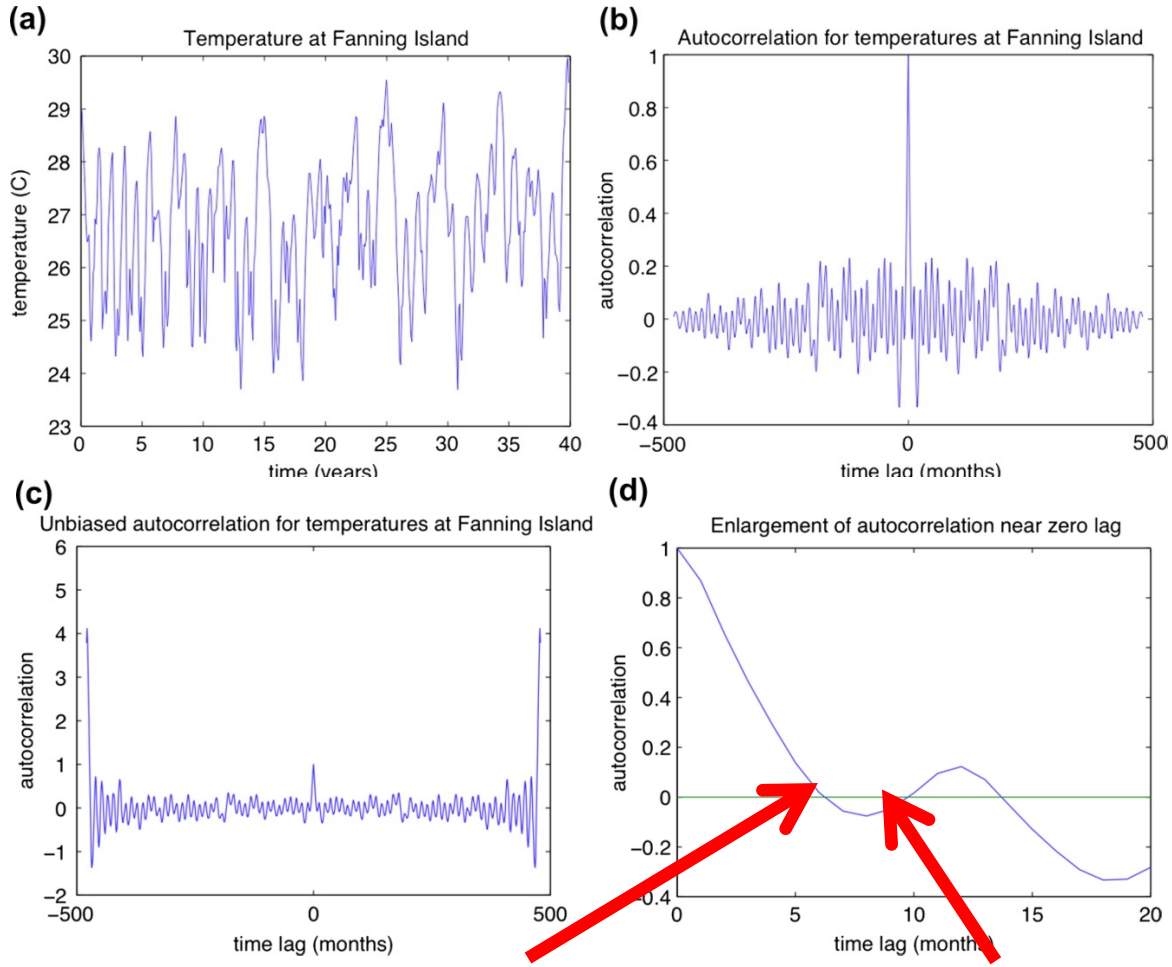
Correlation: covariance divided by the product of the (sample) standard deviations of the two variables.

$$\boxed{\rho_{x,y} = \frac{\mathrm{cov}(x,y)}{\sigma_x \sigma_y}}$$

Correlations have values from 0 to 1.

How to tell if a correlation is significant?

3.C. Time series: autocorrelation and timescales



(a) Time series of temperature at Fanning Island (Pacific Ocean) from the NCAR Community Ocean Model.

(b) Autocorrelation normalized to a maximum value of 1 (biased estimate with averages divided by N).

(c) and (d) Autocorrelation (unbiased estimate with averages divided by N−n). *Source: From Gille (2005).*

DPO FIGURE 6.2

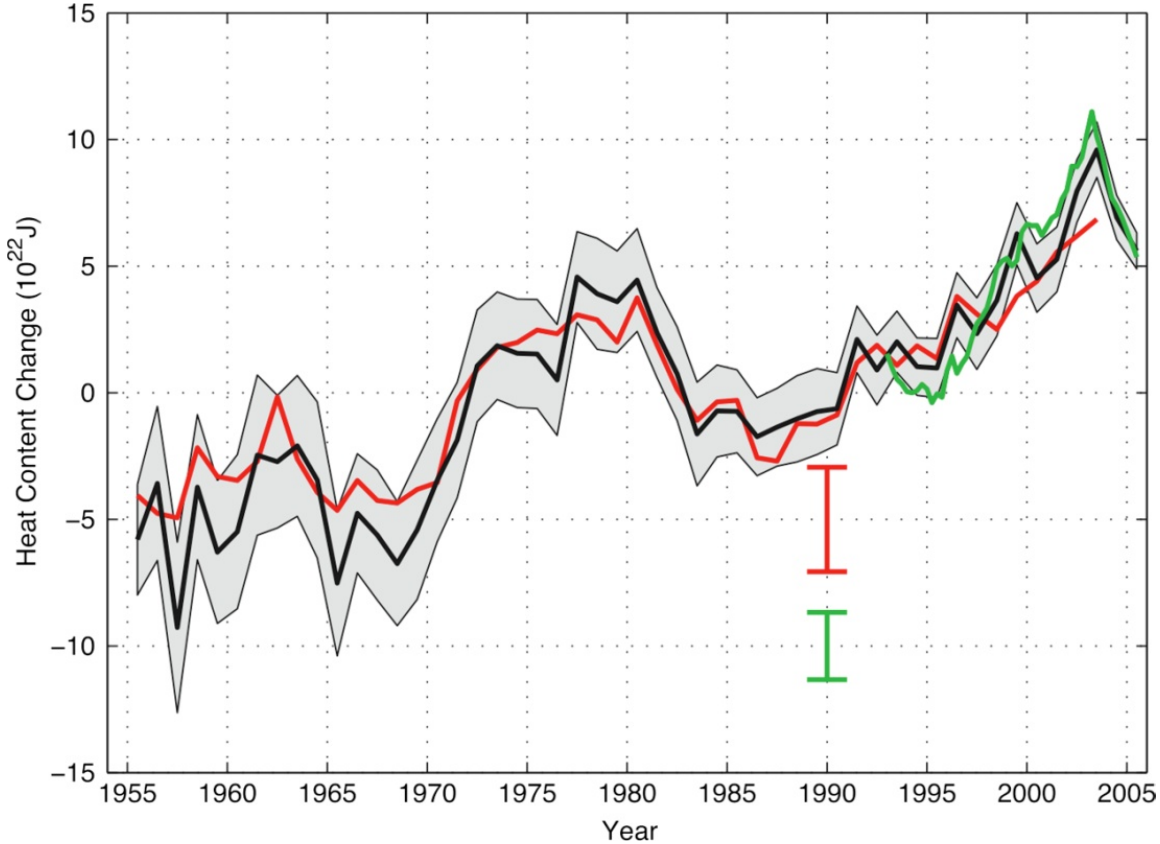
Decorrelation time scale: 1st 0 crossing. Integral time scale is a little longer (based on integral of the autocorrelation)

3.C. Time series analysis: confidence intervals (DPO Section 6.3.3)

- **Confidence intervals**: based on degrees of freedom and assumptions about statistical distribution

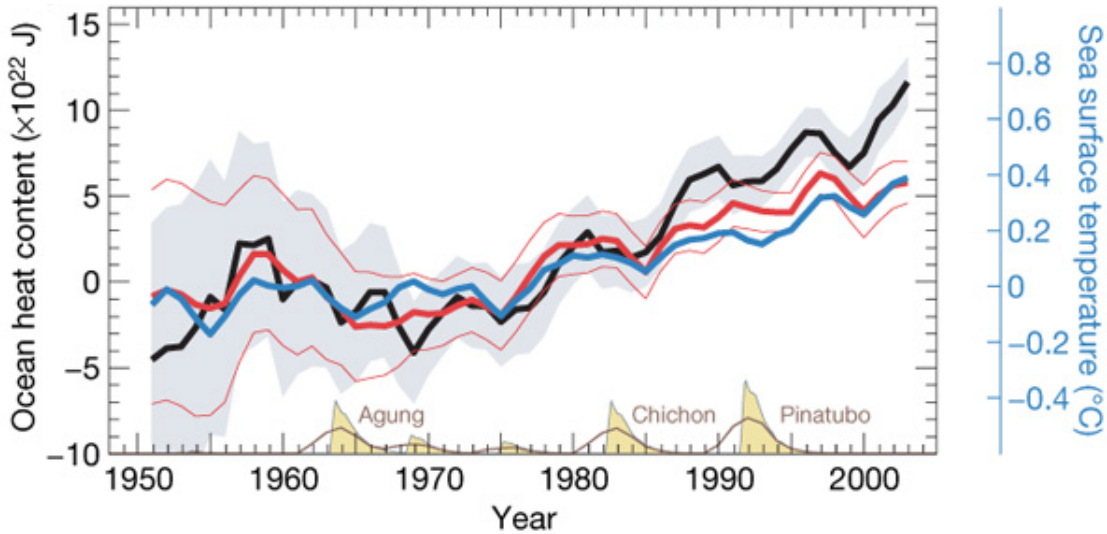
- **95% confidence interval**: 95% probability that a value is within the standard error of the mean (times a t-test quantity that you look up).

3.C. Time series analysis: confidence intervals
(DPO Section 6.3.3)



Example of time series with confidence intervals. Global ocean heat content (10^{22} J) for the 0 to 700m layer, based on Levitus et al. (2005a; black curve), Ishii et al. (2006; full record gray curve and larger error bar), and Willis et al. (2004; darker gray after 1993 and shorter error bar). Shading and error bars denote the 90% confidence interval. Compare with Figure S15.15 seen on the textbook Web site from Domingues et al. (2008) which uses improved observations:
NEXT SLIDE

3.C. Time series analysis: confidence intervals (DPO Section 6.3.3)



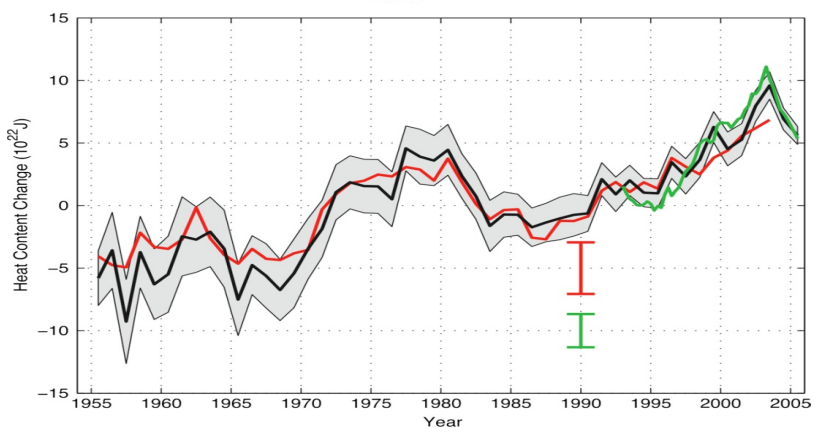


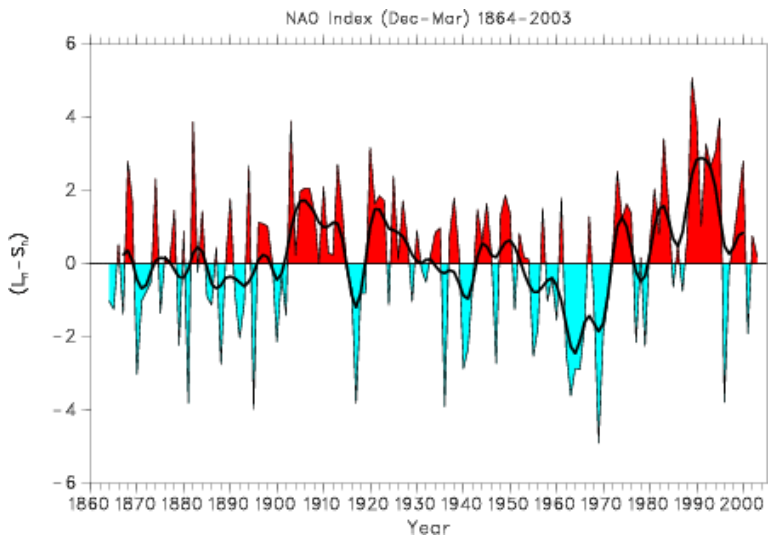
Figure S15.17
Global sea surface temperature

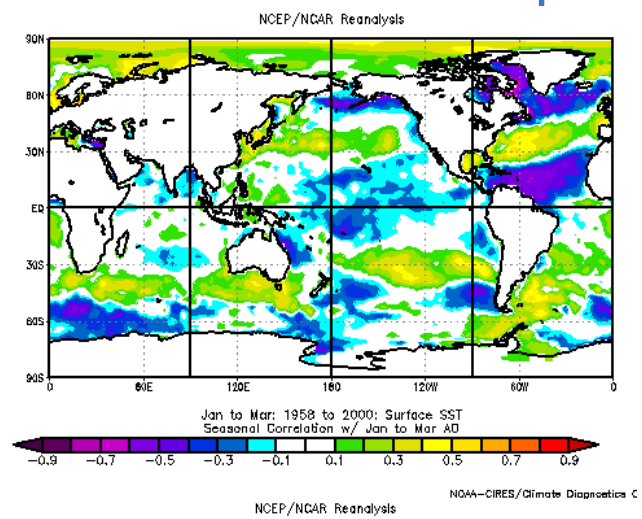
on the textbook Website from Domingues et al. (2008) which uses improved observations

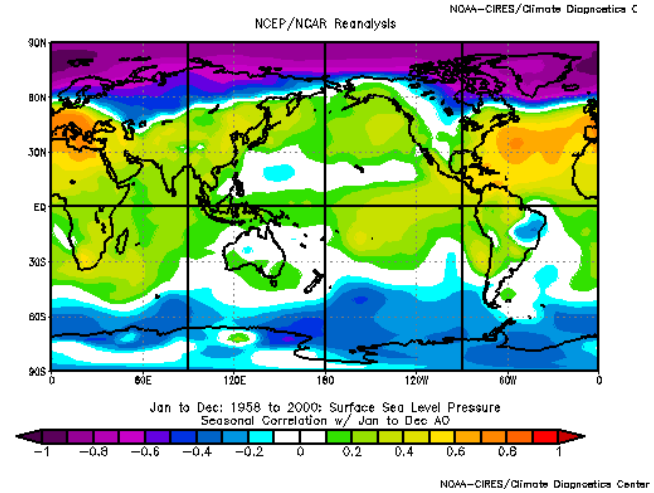
Bottom line: uncertainty estimates are only as good as the underlying data!

3.C. Time series analysis: correlation example

Example of correlation of different time series with each other







North Atlantic Oscillation index (time series)

Correlation with surface temperature and with surface pressure (done at each point, so each lat/lon gives a time series to correlate with NAO index)

Similar to DPO S15.2 and others S15 figures

3.D. Time series analysis: spectral analysis (DPO Section 6.5.3)

Spectral (Fourier) analysis of a time series is used to determine its frequency distributions.

Underlying concept: any time series can be decomposed into a continuous set of sinusoidal functions of varying frequency.

Spectrum: amplitude of each of the frequency components. Very useful for detecting things like tides, inertial motions, that have specific forcing frequencies.

3.D. Time series analysis: spectrum (DPO Section 6.5.3)

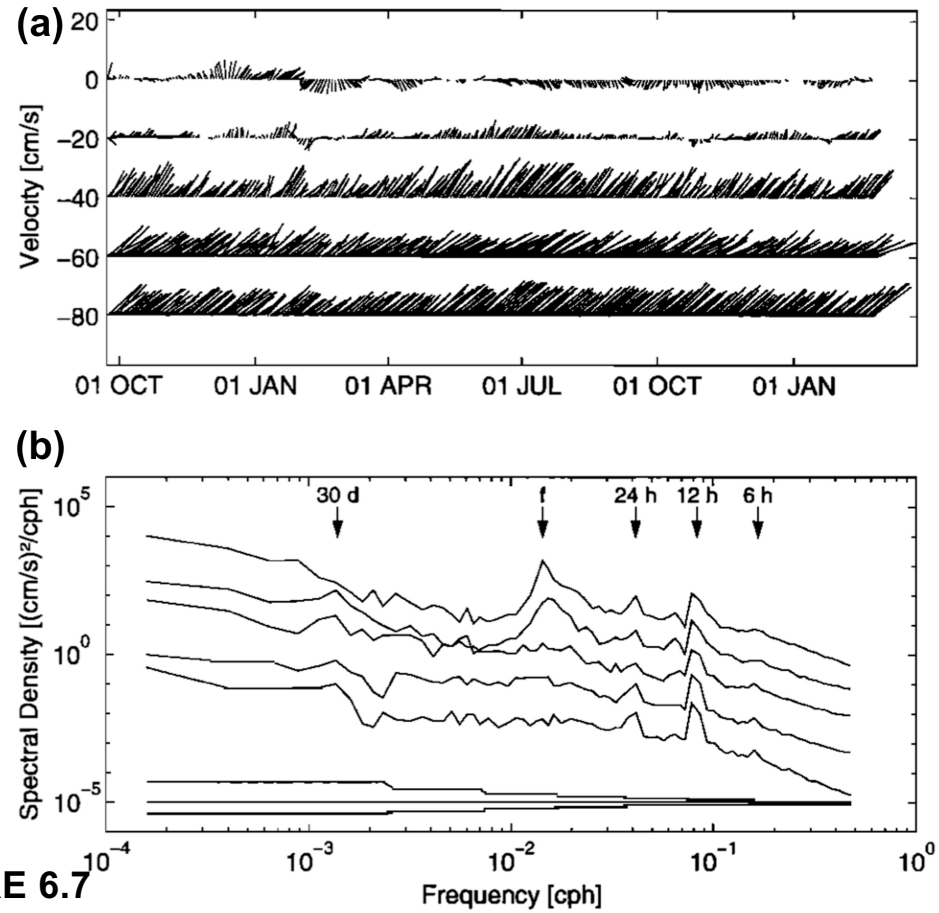


FIGURE 6.7

Example of time series, spectra, and spectral confidence intervals. (a) Velocity (cm/sec) stick plot, lowpassed at 100 hours, from 5 deep current meters at different depths on one mooring in the Deep Western Boundary Current in Samoan Passage (see Figure 10.16). The vertical direction is along the passage axis. (b) Spectra from the same current meters, offset by one decade. The 95% confidence intervals are shown at the bottom. *Source: From Rudnick (1997).*

3.D. Time series analysis: spectrum (DPO Section 6.5.3)

Spectra:

Limitations are imposed by

1. sampling interval. The highest frequency that can be resolved in the spectrum (the "Nyquist frequency"). Any energy at higher frequencies becomes "aliased" into lower frequencies, causing error in their amplitude estimates.

2. length of the total record. Generally should have about 10 realizations of a given frequency for the amplitude to be significant. Can prove difficult when trying to detect climate change for instance. This is related to "degrees of freedom".

Sampling:

N samples at time interval Δt

Length of time series is $T = N\Delta t$

- Nyquist frequency
  - The highest frequency that is resolved is
    - $f = 1/2\Delta t$
- Fundamental frequency
  - The lowest frequency that is resolved is
    - $f = 1/T = 1/N\Delta t$

3.D. Time series: Nyquist frequency and aliasing (DPO Section 6.5.3)

Aliasing: what happens to high frequencies that are above the Nyquist frequency cutoff?

They are in the record, and are sampled, but not resolvable because their frequency content is not resolved

They are aliased to a lower frequency.

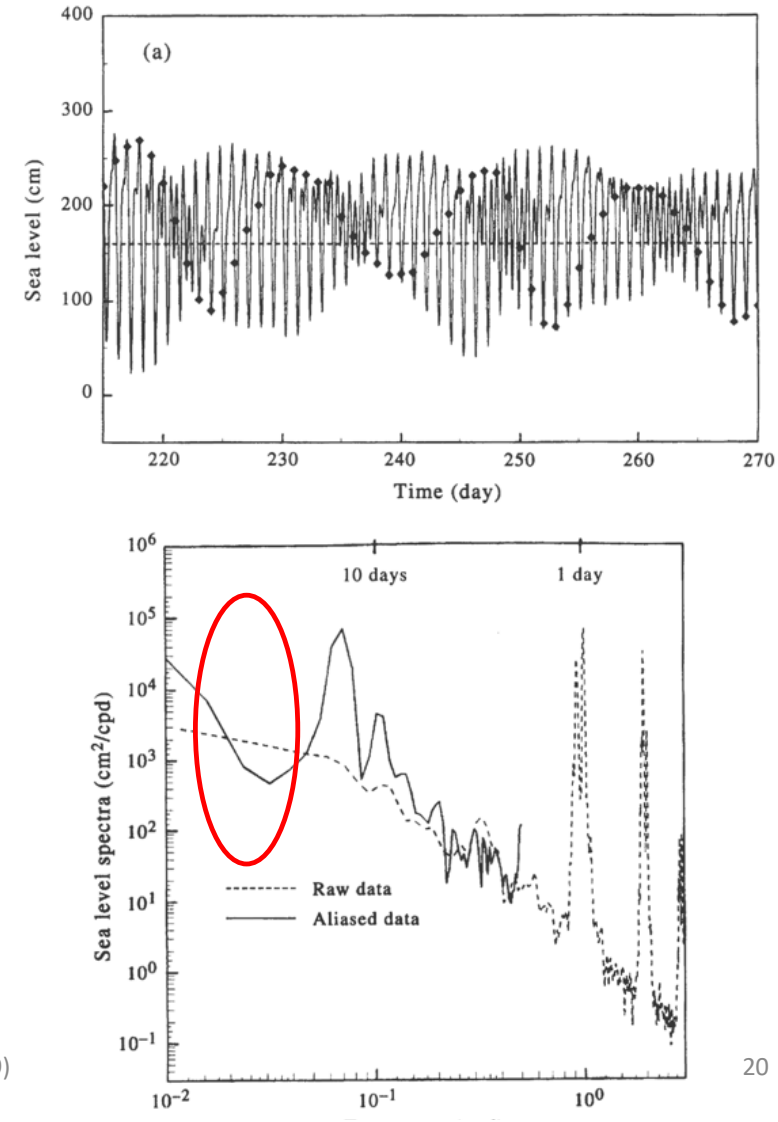


FIGURE 6.8

4. Mapping - Least squares methods (DPO Section 6.3.4)

How do we map data that are not regularly spaced (particularly in space)?

"Least squares methods" are often used for mapping etc.

A. Objective mapping

B. Data assimilation (merging data and computer models) (no info given here)

C. Inverse methods (finding "optimal" solution given sparse or incomplete data) (no info given here)

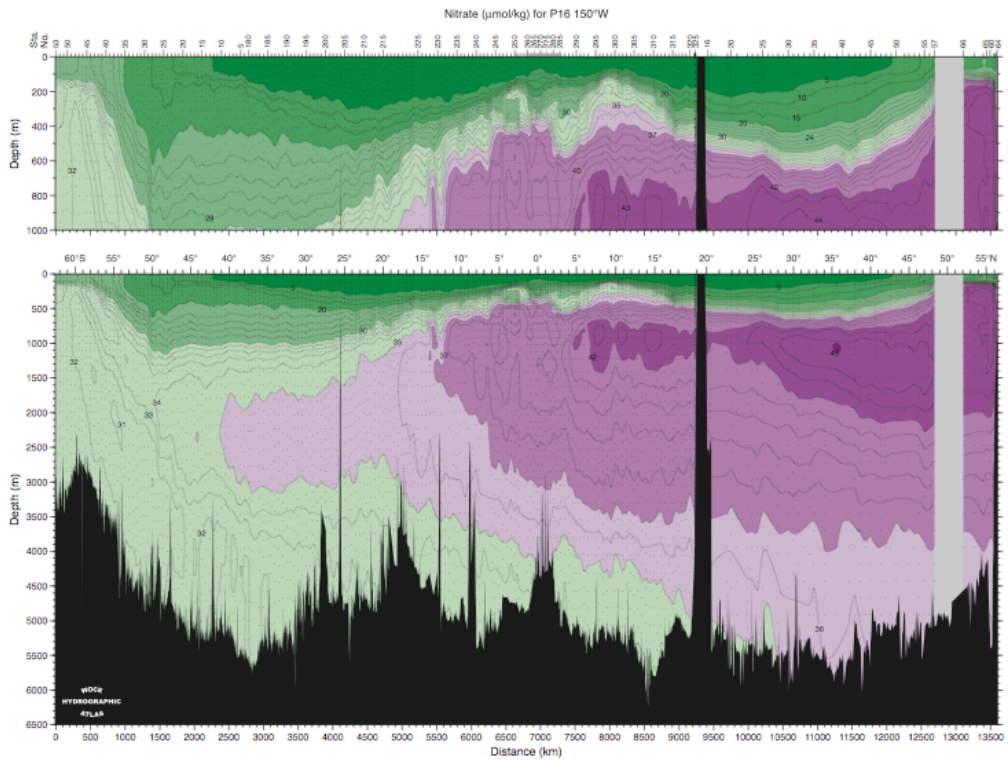
4.A. Mapping - Spatial sampling: Objective analysis (DPO Section 6.4)

Many oceanographic data are collected in several spatial dimensions.

Sample separations are very irregular.

Interpolate these in an optimal way to a grid for plotting and comparison with other data sets.

Objective analysis is a method for interpolating. It requires prior knowledge of the spatial correlation scales of the data to be mapped; since these would normally be calculated from the data themselves, they are often chosen in an ad hoc fashion.



Discrete bottle samples: objectively mapped to uniform grid in vertical (10 m) and horizontal (10 km)

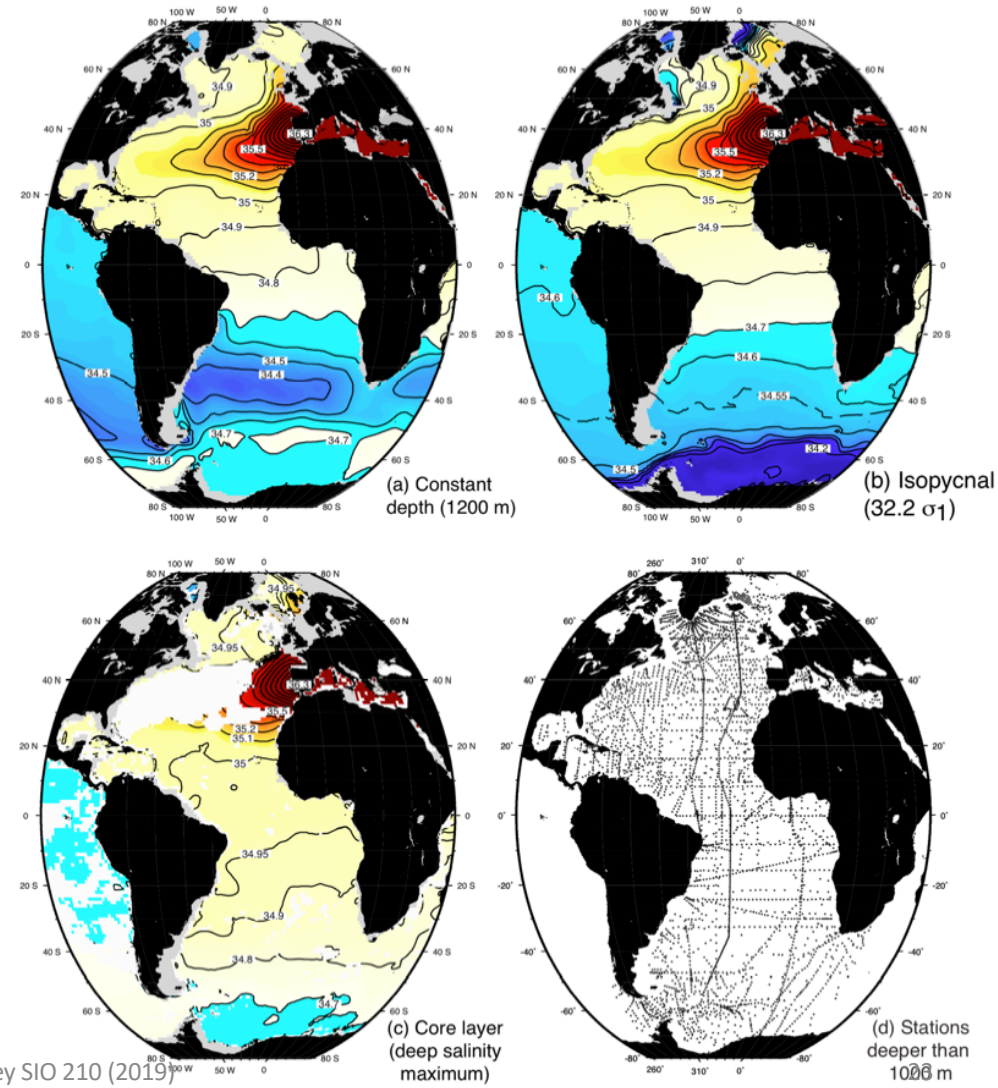
WOCE Pacific atlas http://woceatlas.ucsd.edu/index.html (Talley, 2007)

4.A. Oceanographic sampling: Mapping in the horizontal (DPO Section 6.4.2)

Original station data: objectively mapped to a regular grid (NOAA/NODC)

Different types of surfaces that are commonly used:

1. Constant depth
2. Isopycnal
3. "core layer" – here salinity maximum



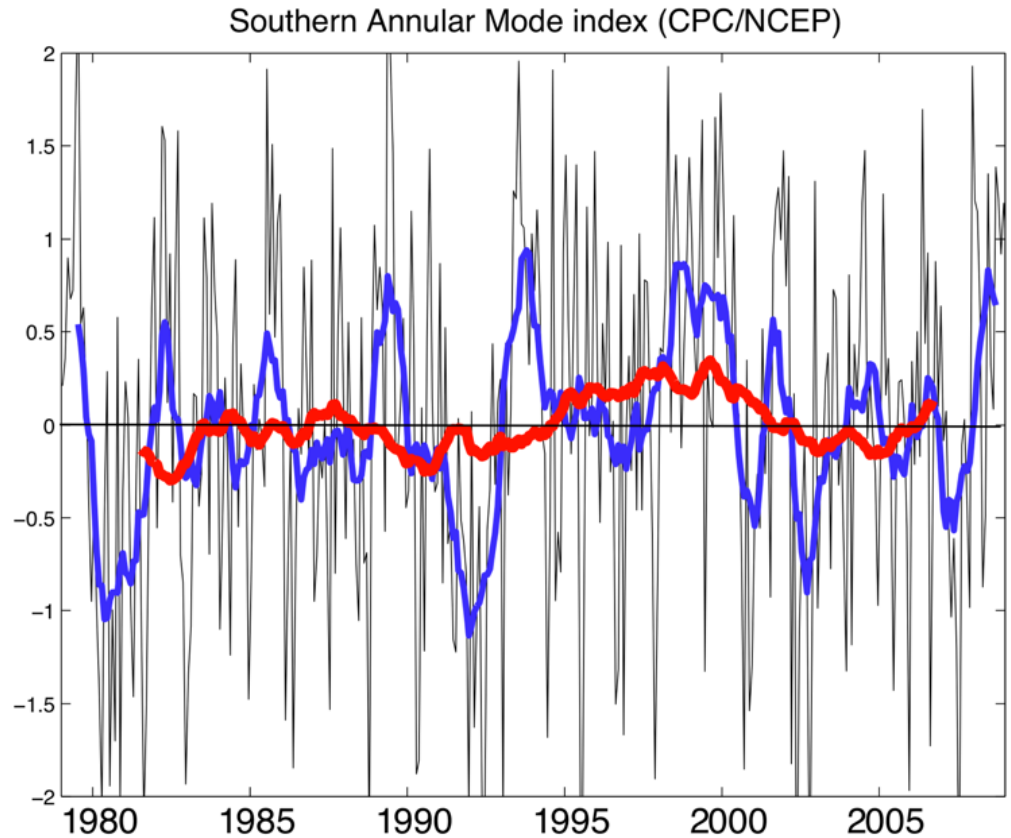
DPO Fig. 6.4

5. Filtering (DPO Section 6.5.4)

Time series or spatial sampling includes phenomena at frequencies that may not be of interest

(examples: want inertial mode but sampling includes seasonal, surface waves, etc.)

Use various filtering methods (running means with “windows” or more complicated methods of reconstructing time series) to isolate frequency of interest

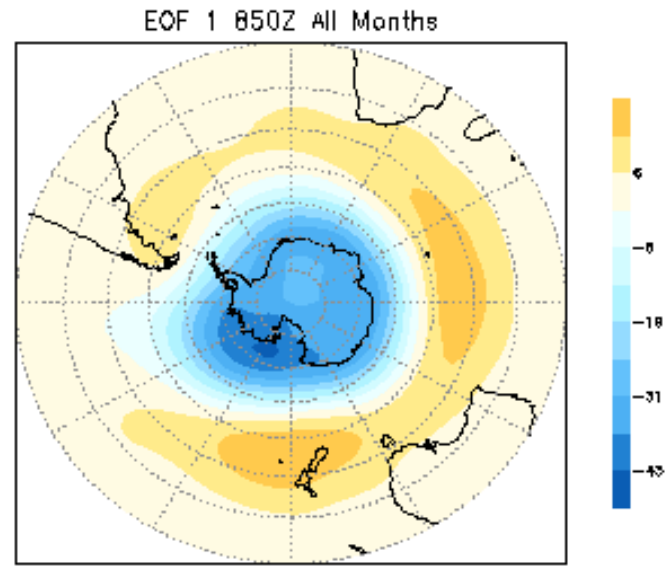


DPO Fig. 6.9

Example: Time series of a climate index with 1-year and 5-year running means as filters.

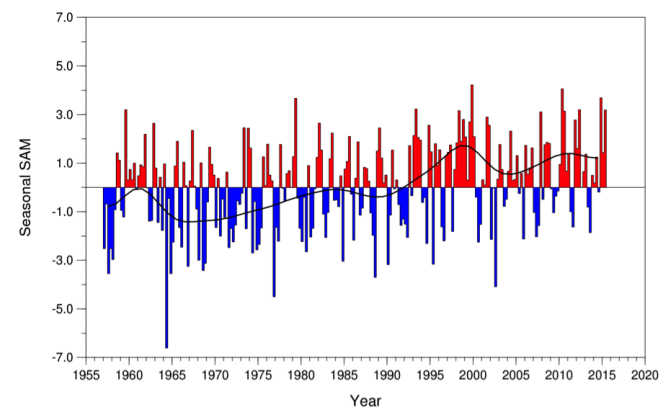
6. Space-time sampling: empirical orthogonal functions (DPO Section 6.6.1)

Southern Annular Mode (SAM)



Underlying physical processes usually not well characterized by sines and cosines, especially in the spatial domain. They may be better characterized by functions that look like the basin or circulation geometries.

Empirical orthogonal function analysis (EOFs): let the data decide what the basic (orthogonal) functions are that add to give the observed values.

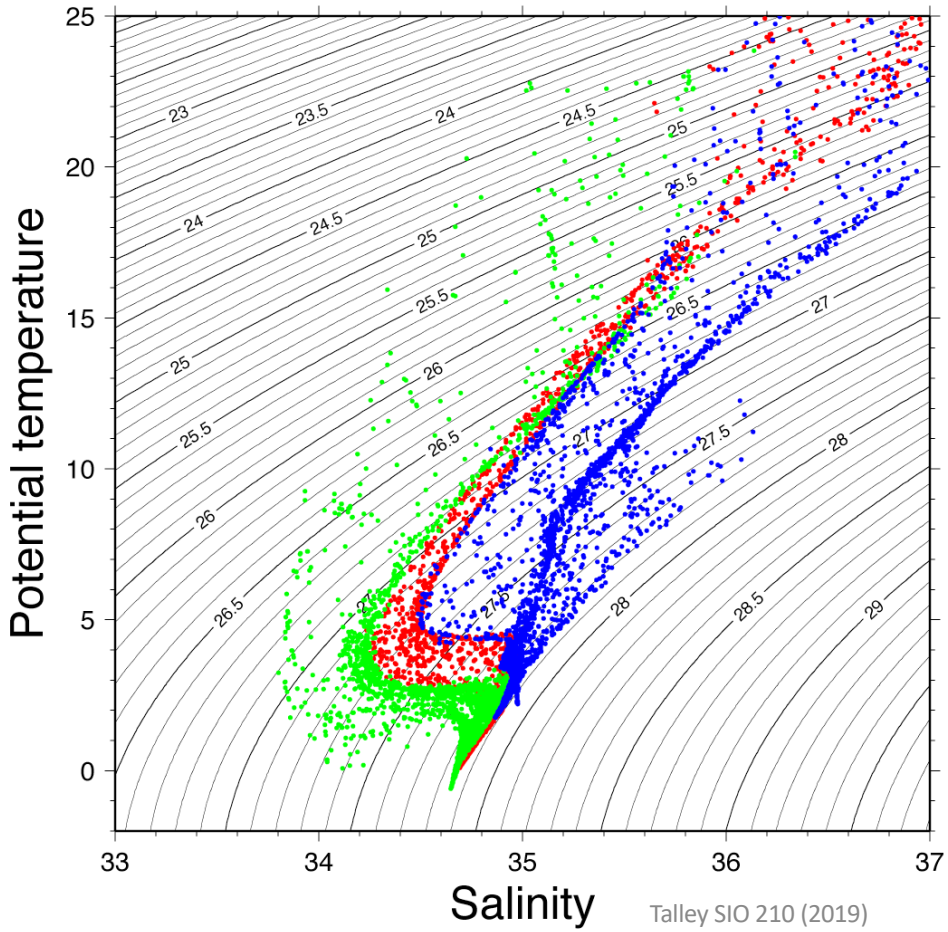


Basic EOF analysis: obtain spatial EOFs with a time series of amplitude for each EOF. (The time series itself could be Fourier-analyzed if desired.)

7. Water mass analysis (DPO Section 6.7.2)

- Property-property relations (e.g., theta-S)
- Volumetric property-property
- Optimum multiparameter analysis (OMP) (Section 2 only)

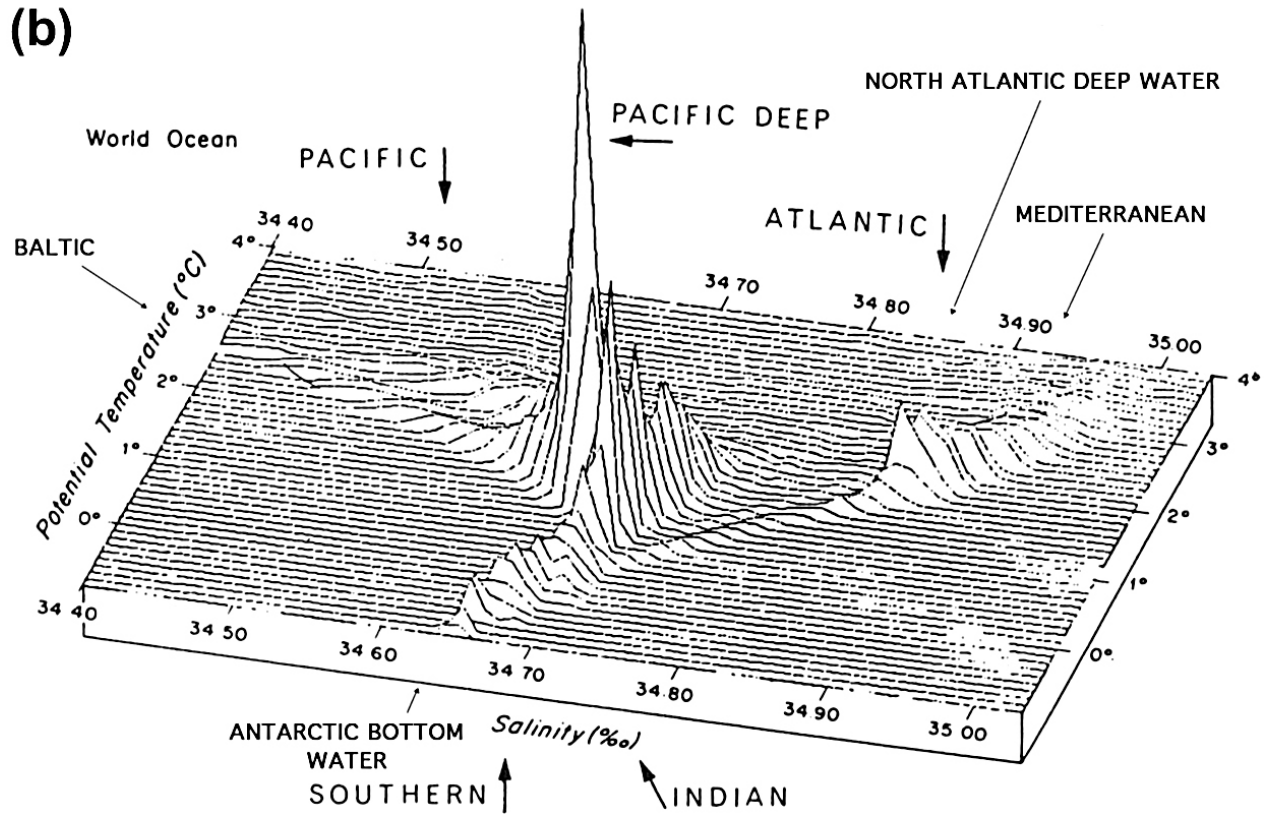
7. Water mass analysis: potential temperature-salinity diagrams (DPO Section 6.7.2)



Potential temperature-salinity diagram used in several lectures looking at Atlantic properties

DPO FIGURE 6.14

7. Water mass analysis: volumetric potential temperature-salinity (DPO Section 6.7.2)



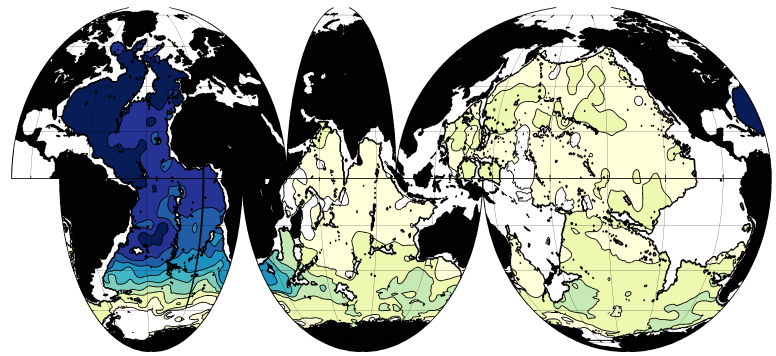
DPO 4.17b

7. Optimum multiparameter analysis (OMP) (DPO Section 6.7.3)

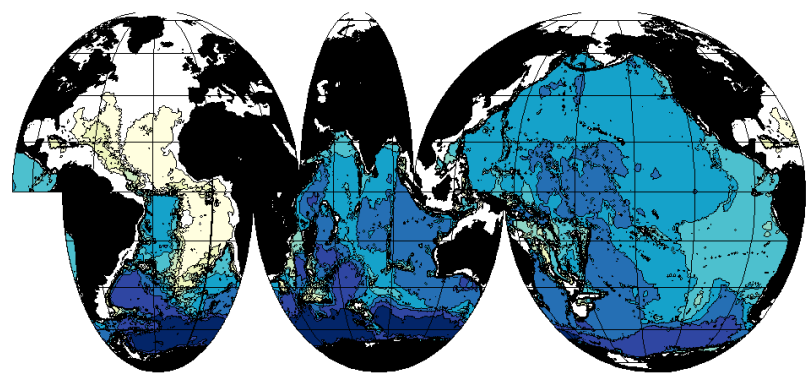
Use of water properties (T, S, oxygen, nutrients, other tracers): define source waters in terms of properties, then use a least squares analysis to assign every observation to a proportion of each source water

Example of NADW vs. AABW: fractions of each

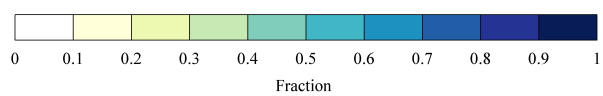
(Johnson et al., 2008)
DPO 14.15



(a) Fraction of NADW at γ^N=28.06 kg/m^3 (2500-3000 m)



(b) Fraction of AABW at ocean bottom



Summary - definitions

Sampling
- synoptic
- Climatology

Uncertainty
- Random error
- Systematic error (bias)

Statistical concepts
- Variance
- Standard deviation
- Standard error

Time series
- Hovmoller diagram
- Covariance and correlation
- Integral time scale
- Degrees of freedom
- Confidence intervals

Spectral analysis
- Spectrum
- Nyquist frequency
- Fundamental frequency
- Aliasing

Mapping
- Least squares methods
- Objective analysis

Filtering:
- Band pass filter

Empirical orthogonal functions

Water mass analysis
- Property-property plots
- Volumetric potential temperature-salinity
- Optimum multiparameter analysis