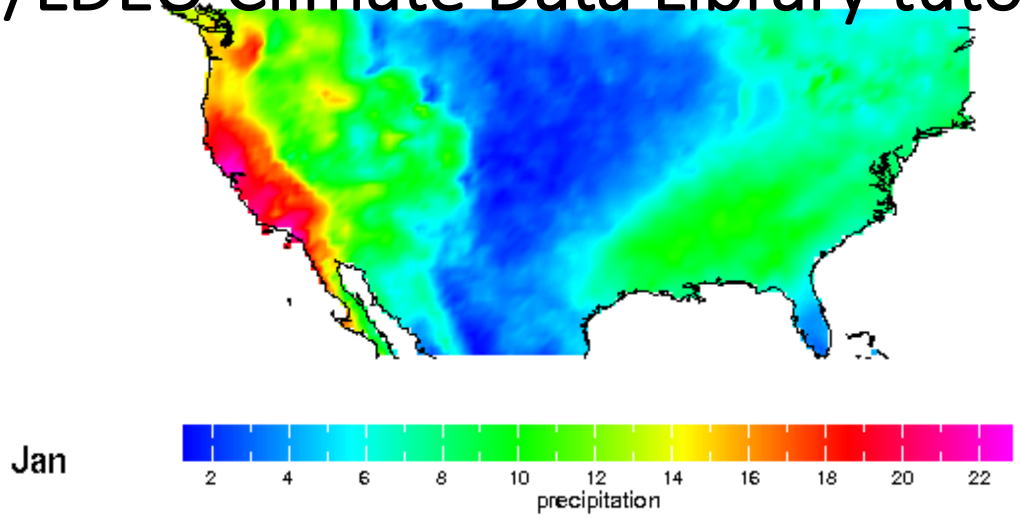
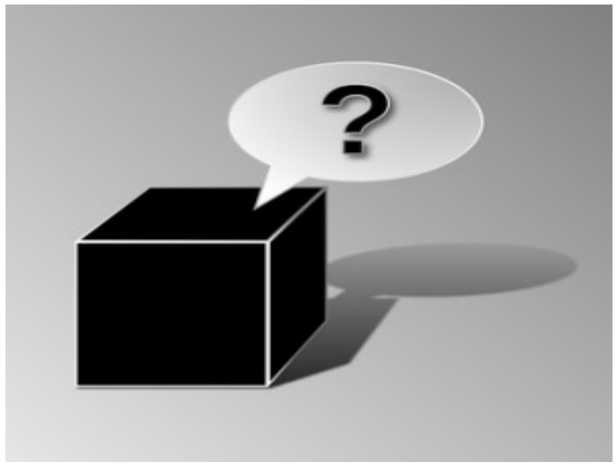
IRI/LDEO Climate Data Library tutorial



Announcements

- Fill out the survey if you haven't already!
- No class Thurs– work on HW1 (posted tonight!)
- No office hours this week

But first let's dive into the EOF black box!



Emperical Orthogonal Function analysis

- ***Emperical***: the eigenvectors are based on the observations and adapted to these observations; determines the modes of variability that best describe the data in terms of the data's covariance.
- ***Orthogonal:*** Each of the eigenvectors are orthogonal (perpendicular) to each other and thus independent of each other.

- similar to the Fourier analysis in that both decompose the time series into components that explain the variance of the time series.
- Recall: Fourier analysis breaks the time series up into a series of sine and cosine waves of different frequencies
- EOF analysis empirically finds a set of orthogonal functions which describe the covariance of the time series.
- Returns the spatial patterns of variability (EOFs), their time variation (PCs) and a measure of the "importance" of each pattern (eigenvalues, %)

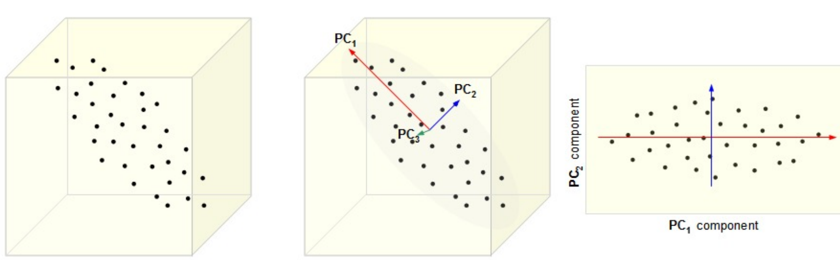
“Modes of variability”

- Important caveat: EOFs are *data modes* and not necessarily *physical modes*
- Physical meaning is a matter of subjective interpretation

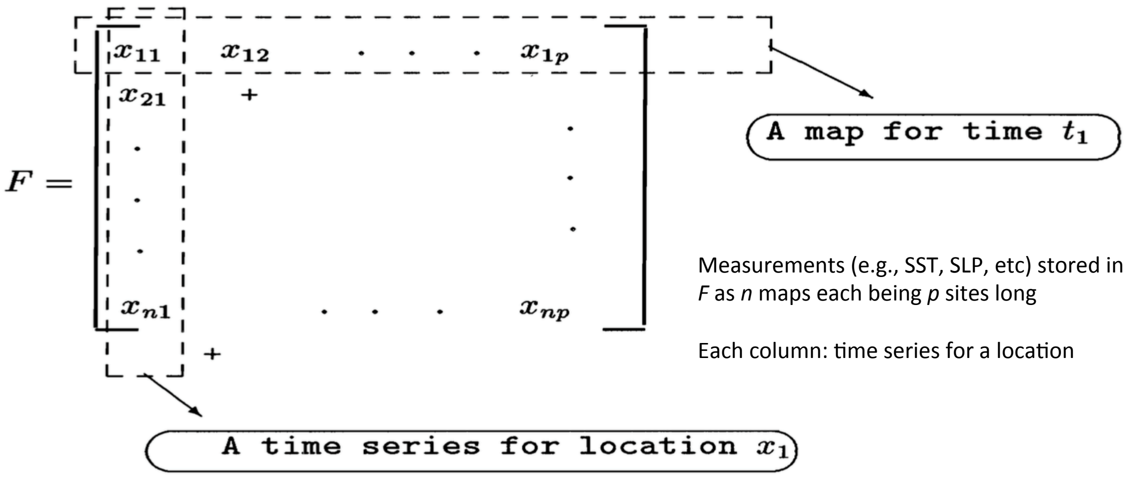
“I have learned the following rule to be useful when dealing with advanced methods. Such methods are often needed to find a signal in a vast noisy phase space, i.e. the needle in the haystack. But after having the needle in our hand, we should be able to identify the needle by simply looking at it. Whenever you are unable to do so there is a good chance that something is rotten in the analysis.” (Storch and Navarra)

3D view of our n x p matrix F

- Every observation makes one point in this space
- If the observations are totally random, the points would describe a blob in this space
- Any regularities in the date make points organize into clusters or along preferred directions
- EOFs tries to find optimal coordinate system through these points so each cluster of data points has a coordinate axis (a “mode” or EOF)



Behind the scenes…

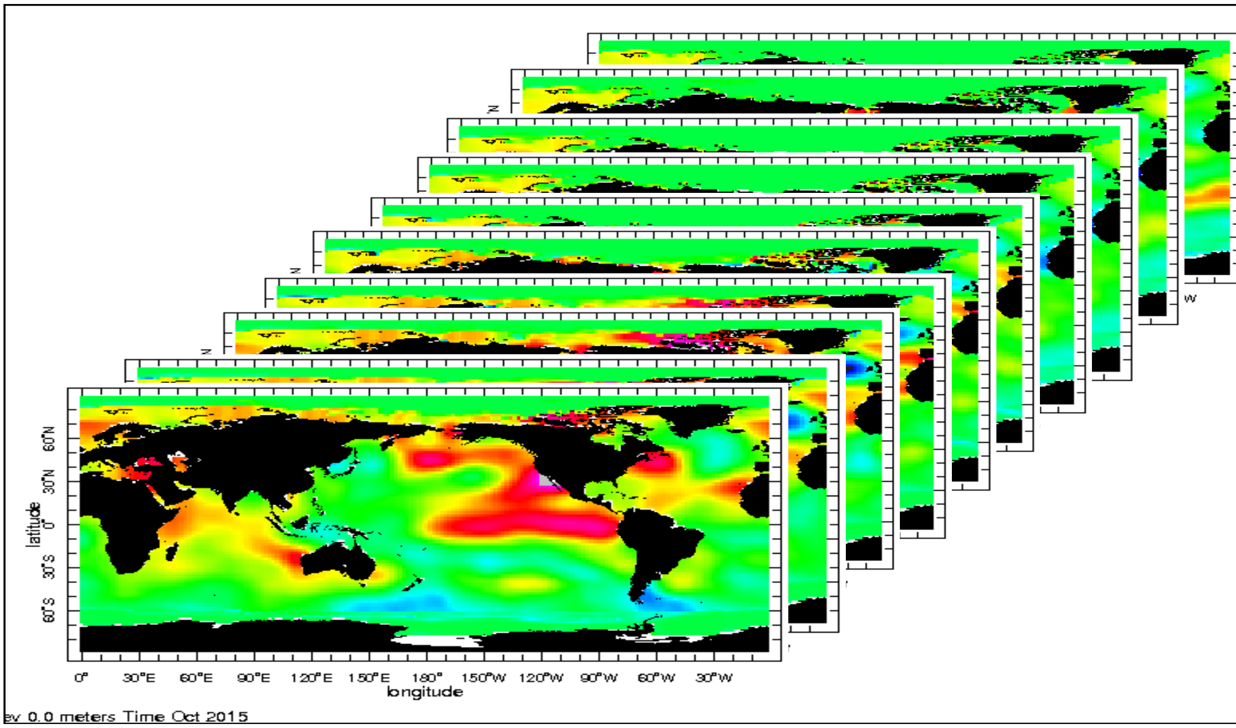


Measurements (e.g., SST, SLP, etc) stored in a matrix *F* as *n* maps each being *p* sites long

Each column: time series for a location

1. Form matrix *F* from the observations, and remove the time mean from each series

Björnsson and Venegas 1997

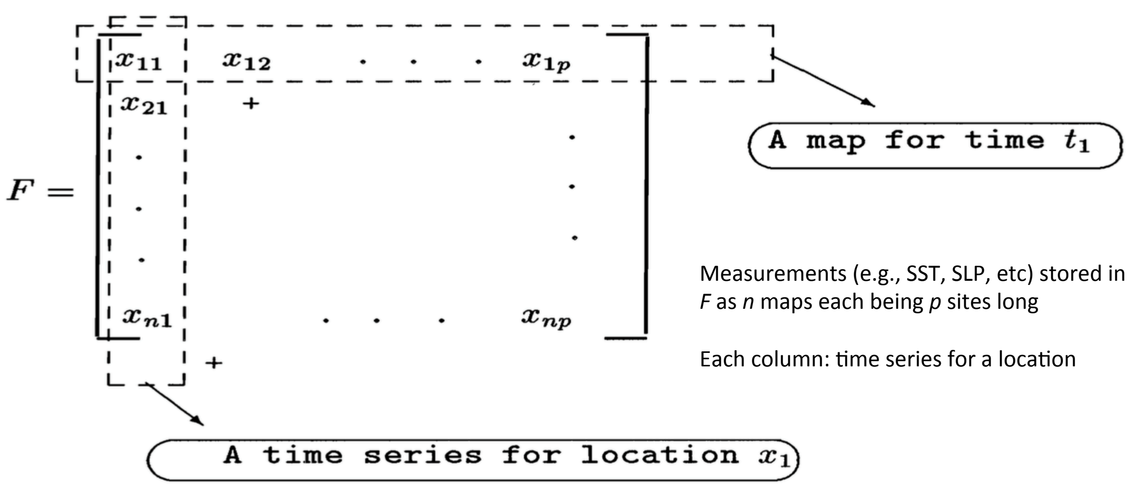


Behind the scenes (cont)…

2. Calculate the covariance matrix of *F*:

   $R = F^T F$

Behind the scenes…



Measurements (e.g., SST, SLP, etc) stored in a matrix *F* as *n* maps each being *p* sites long

Each column: time series for a location

1. Form matrix *F* from the observations, and remove the time mean from each series

Björnsson and Venegas 1997

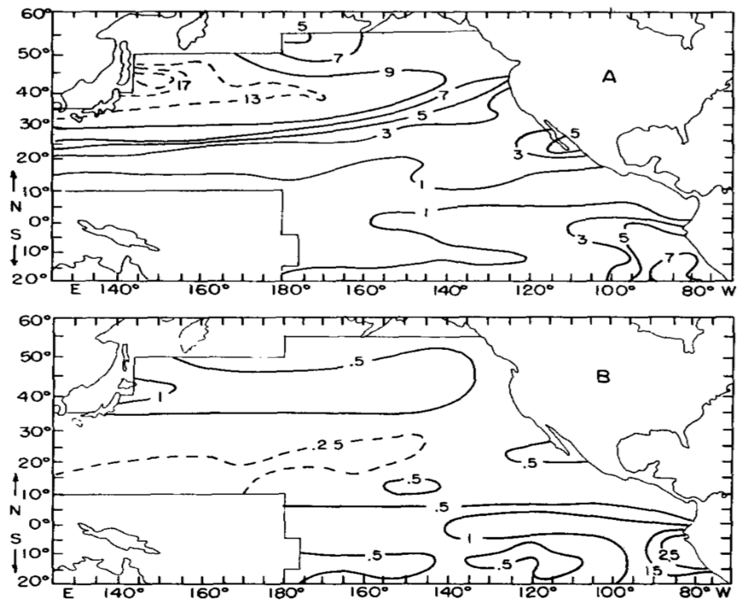
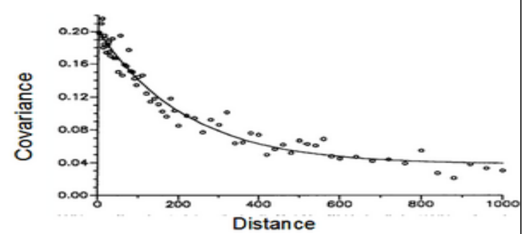


FIG. 2. Variances of deviations from long-term means for 1949–73: (a) deviations computed from mean annual average, (b) deviations computed from monthly means.

Covariance

R= [(x- xmean)^T * (y – ymean)]/ nt-1

For a *realization* (a finite set of observations, x & y) of length *nt* of two time series X and Y.
R is of size X by Y (nsites, nsites)
Diagonal: variance
Symmetrical matrix, as covariance $X_iY_i = Y_iX_i$



Weare et al. (1976)

Behind the scenes (cont)…

2. Calculate the covariance matrix of *F*:

   $R = F^TF$

3. Solve the eigenvalue problem:

   $RC = C\Lambda$

   where:

   - Λ is a diagonal matrix containing the eigenvalues λ_i of *R*. $\lambda_i / \Sigma\lambda_i$ gives the fraction of the total variance in *R* explained by each mode
   - c_i column vectors of C are the eigenvectors of R corresponding to the eigenvalues. Eigenvectors = EOFs = map of spatial pattern
   - Both are size p by p (nsites by nsites)
   - Each eigenvector has the property that $C^TC = CC^T = I$ (I is the identity matrix. I.e. EOFs are uncorrelated over space (orthogonal to each other)

4. To see how an EOF_j (C_j) evolve in time, we calculate the "PCs" as:

   $a_j = FC_j^T$ (the projection of F onto the j-th EOF)

5. Check our work! *F = C*a*

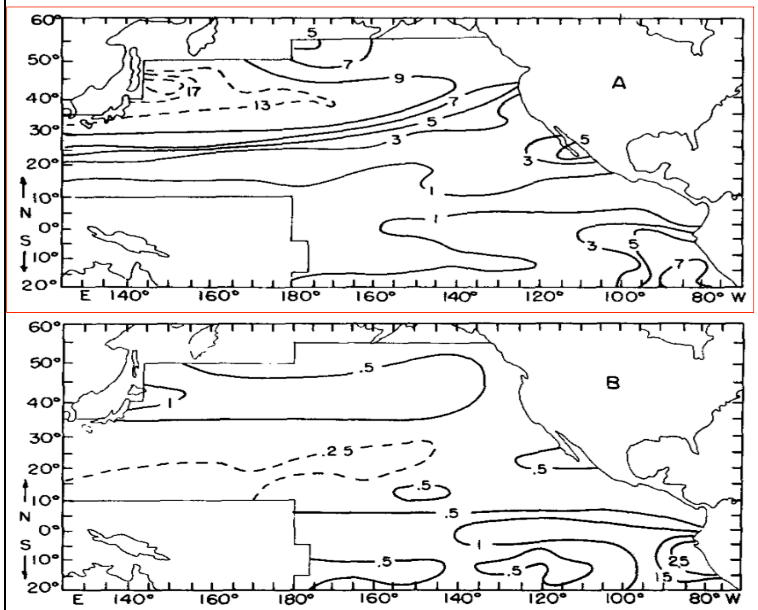
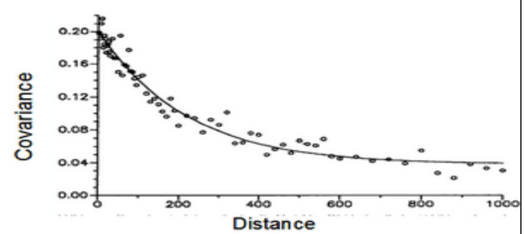


FIG. 2. Variances of deviations from long-term means for 1949–73: (a) deviations computed from mean annual average, (b) deviations computed from monthly means.

Covariance

R= [(x- xmean)^T * (y – ymean)]/ nt-1

For a *realization* (a finite set of observations, x & y) of length *nt* of
of two time series X and Y.
R is of size X (nsites) by Y (nsites)
Diagonal: variance
Symmetrical matrix, as covariance $X_iY_i = Y_iX_i$



Weare et al. (1976)

EOF1 of mean annual average SSTs

- We have a seasonal cycle!
- Explains 81.7% of the total variance

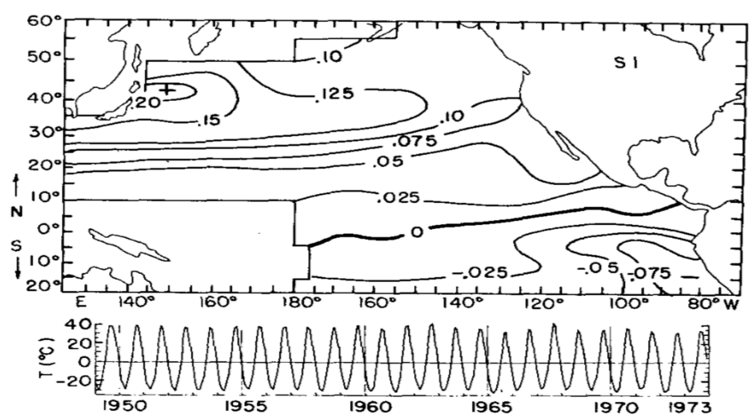


FIG. 3. Empirical orthogonal function S1, explaining 81.7% of total variance shown in Fig. 2a, together with its corresponding time coefficients. Function is based on departures from long-term annual averages. Tics in the time series identify January of the respective year starting with January 1949 and ending with December 1973.

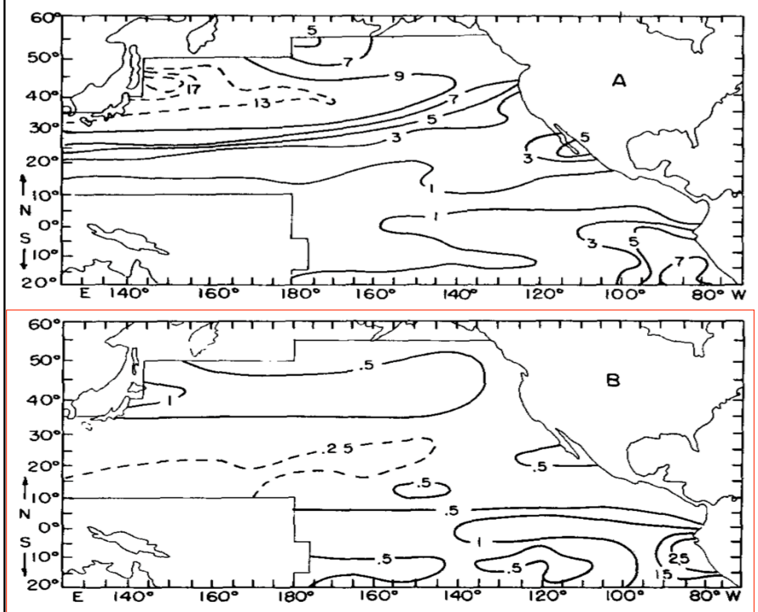
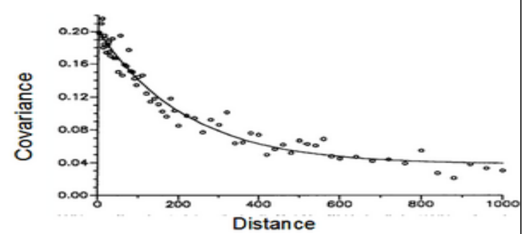


Fig. 2. Variances of deviations from long-term means for 1949–73: (a) deviations computed from mean annual average, (b) deviations computed from monthly means.

Covariance

$R = [(x - xmean)^T * (y - ymean)] / nt-1$

For a *realization* (a finite set of observations, x & y) of length *nt* of
of two time series X and Y.
R is of size X (nsites) by Y (nsites)
Diagonal: variance
Symmetrical matrix, as covariance $X_iY_i = Y_iX_i$



Weare et al. (1976)

EOF1 of SST anomalies

- ENSO!
- Explains 23.1% of variance

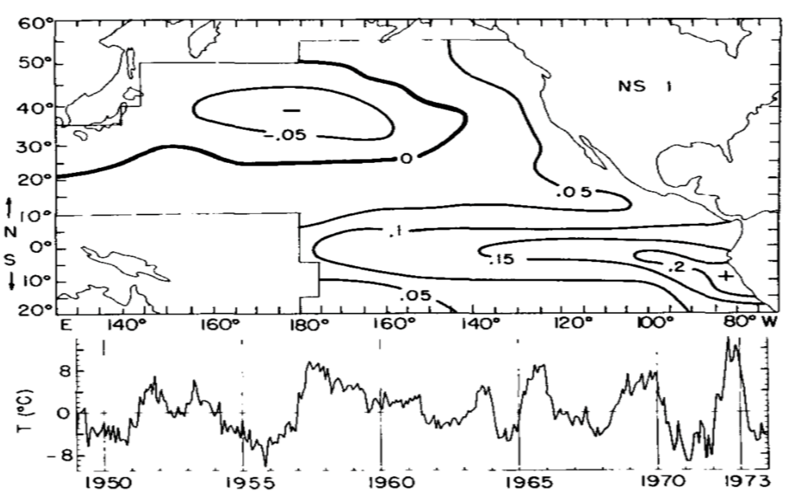
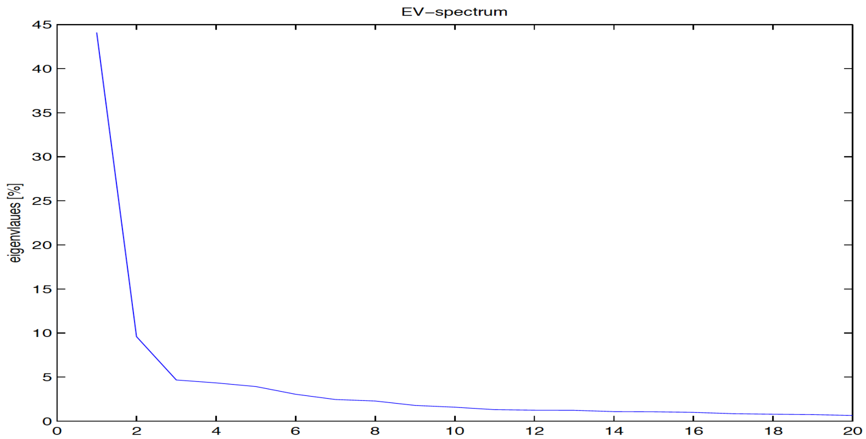
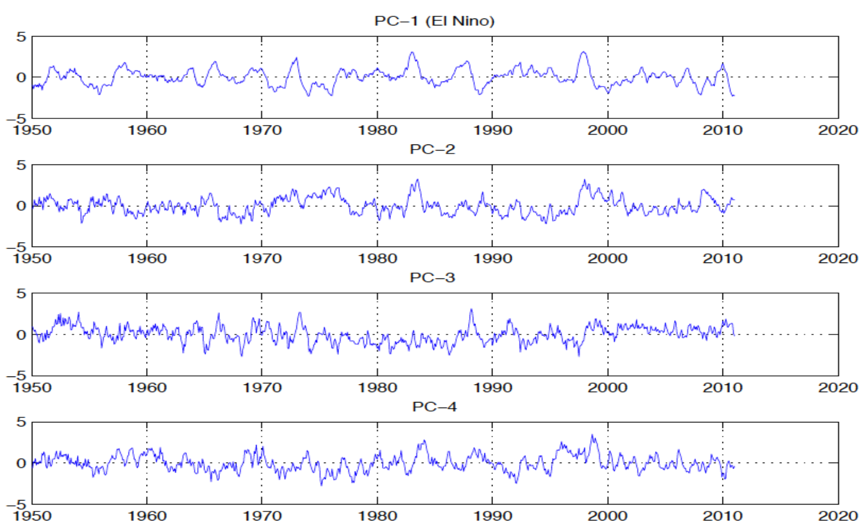


Fig. 7. As in Fig. 3 except for empirical orthogonal function NS1, explaining 23.1% of the total variance in the Fig. 2b.

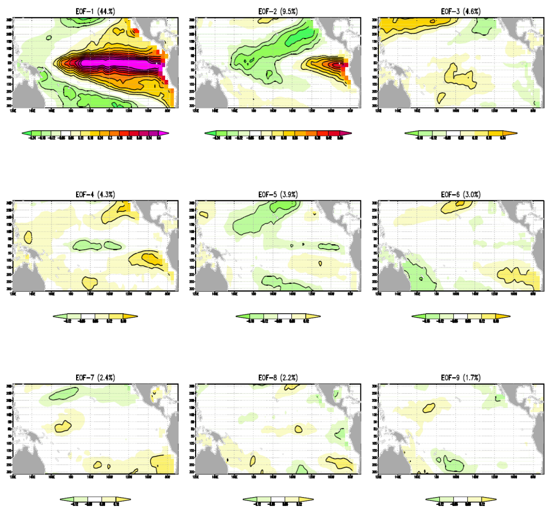
Example 1: Tropical Pacific SSTs
Eigenvalues



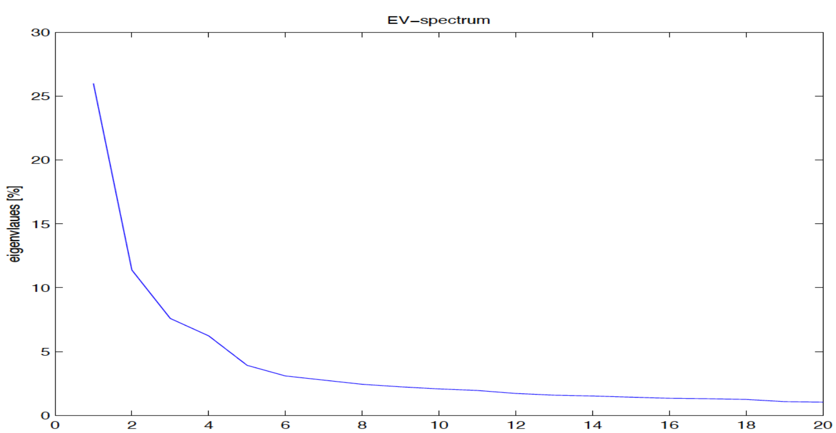
Example 1: Tropical Pacific SSTs
Principal Components



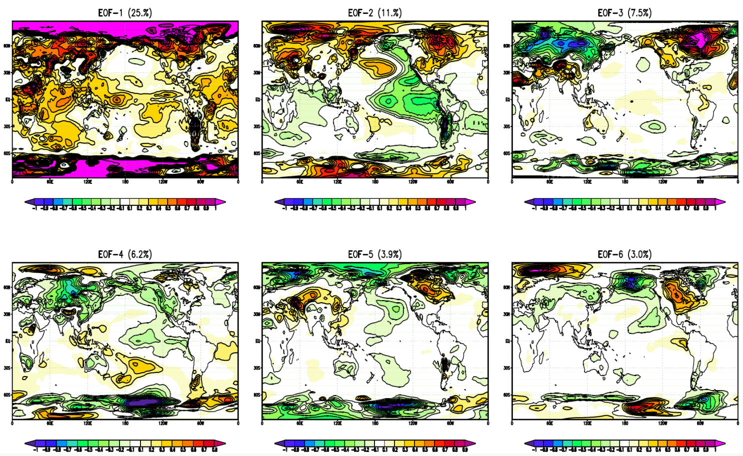
Example 1: Tropical Pacific SSTs EOF modes



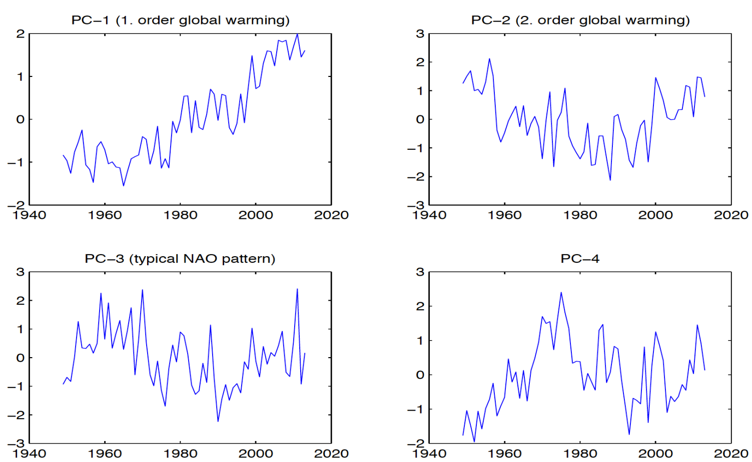
What if we look at global temperatures?



Top EOFs of global temperatures



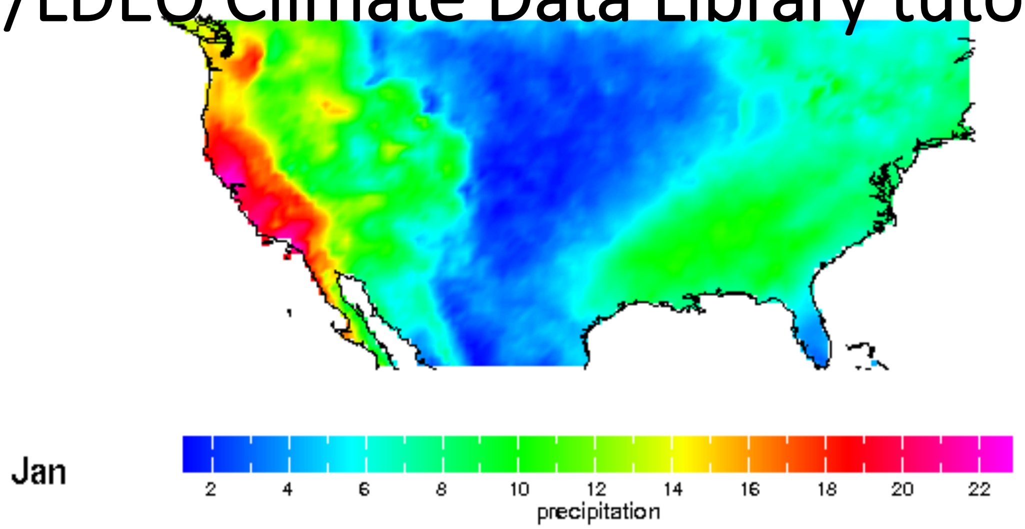
Top PCs of global temperatures



A note on rotated EOFs

- Physical climate processes are rarely truly independent → orthogonality constraint can limit the physical interpretation of (higher order) EOFs
- Rotated EOFs:
  1. Alleviate the orthogonality & uncorrelation constraints of EOFs & PCs, respectively
  2. Result in simple, often more physically interpretable, patterns

IRI/LDEO Climate Data Library tutorial

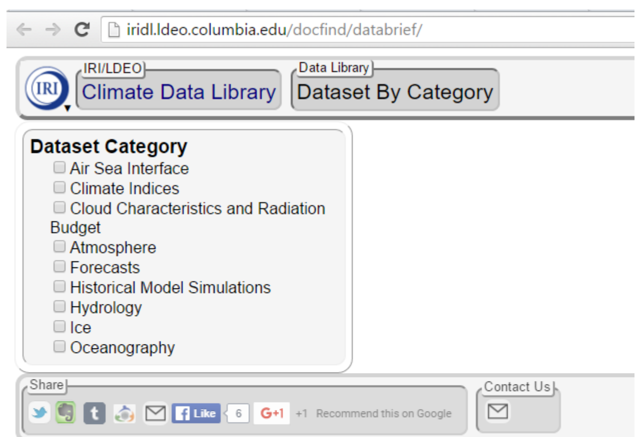


Finding data

1) a listing of some of the datasets by category
2) a complete list of datasets according to their source
3) a keyword search powered by Google

1. Data by category

- Sorted listing of a few of the datasets in the Data Library based on the type of data they contain
- A summary of each dataset, including a brief description, spatial and temporal resolutions, and spatial and temporal limits, is also offered.
- Browse the data categories.
- Note that the dataset names are also links to the datasets themselves.



2. Data by source

Data Library SOURCES | Language english

Description | Expert Mode

served from IRI/LDEO Climate Data Library

SOURCES

SOURCES

SOURCES: the IRI/LDEO collection of climate data.

Documents

overview an outline showing sub-datasets of this dataset

Datasets and Variables

ANEEL	Agência Nacional de Energia Elétrica.
ARCTIC	Oceanic station data for the Arctic Region.
Biosphere	Biosphere 2 Laboratory.
BoM	Australian Bureau of Meteorology.
BRF	Black Rock Forest.
CAC	Climatological, smoothed, and raw sea surface temperature data for the tropical Pacific Ocean from the Climate Analysis Center (now known as the Climate Prediction Center).
CARDONE	ECMWF five day average wind stress; Cardone Level 2B data.
Carton/Giese	CARTON-GIESE[**SODA**]
CAYAN	Heat flux and irradiance data analysis of COADS from Cayan of Scripps.

3. Data by search

IRI/LDEO Climate Data Library

Google™ Custom Search

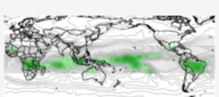
IRI/LDEO Climate Data Library

The IRI Data Library is a powerful and freely accessible online data repository and analysis tool that allows a user to view, analyze, and download hundreds of terabytes of climate-related data through a standard web browser.

It is a powerful tool that offers the following capabilities at no cost to the user:

- access any number of datasets;
- create analyses of data ranging from simple averaging to more advanced EOF analyses using the Ingrid Data Analysis Language;
- monitor present climate conditions with maps and analyses in the Maproom;
- create visual representations of data, including animations;
- download data in a variety of commonly-used formats, including GIS-compatible formats.

IRI Climate and Society Map Room



The climate and society maproom is a collection of maps and other figures that monitor climate and societal conditions at present and in the recent past. The maps and figures can be manipulated and are linked to the original data. Even if you are primarily interested in data rather than figures, this is a good place to see which datasets are particularly useful for monitoring current conditions.

Data by Source

Datasets organized by source, i.e. creator and/or provider.

Data By Category

Selected Datasets for particular topics

Dataset and Map Room Browser

Find datasets and maps

Navigating Through the IRI Data Library: A Tutorial

The goal of this tutorial is to introduce you to the structure of the Data Library and the many ways to navigate through it.

Statistical Techniques in the Data Library: A Tutorial

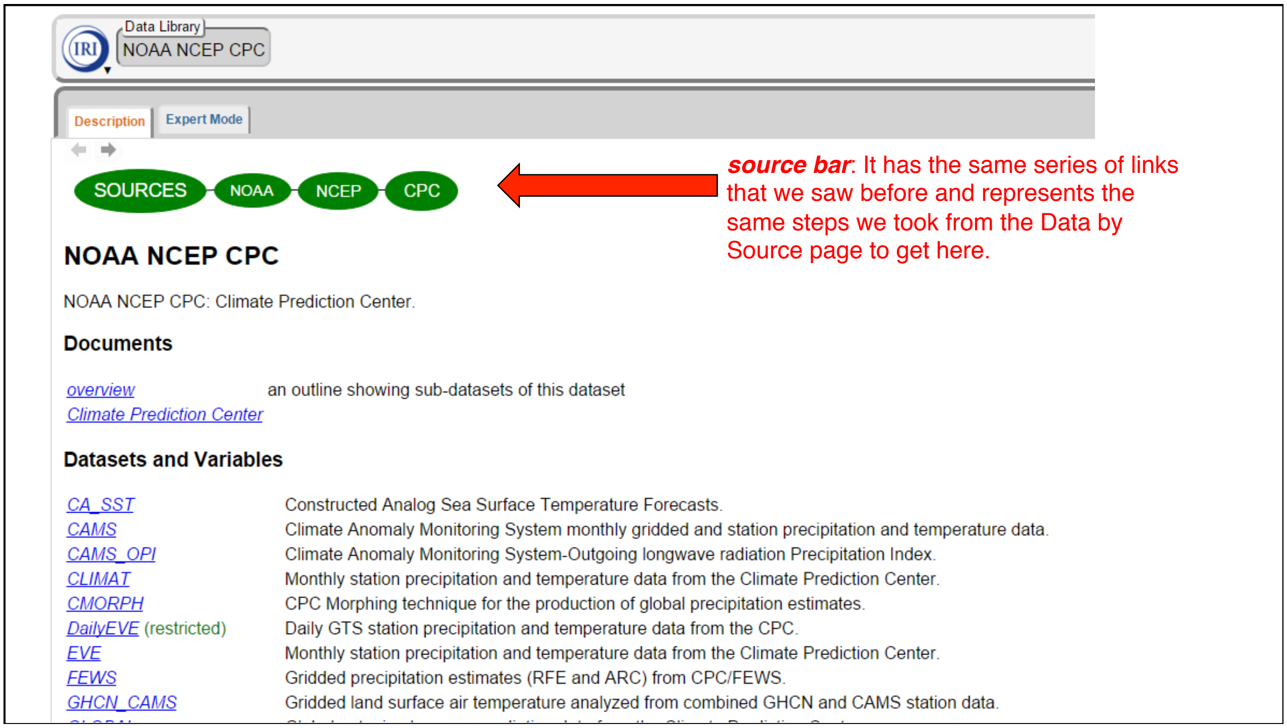
Statistical techniques are essential tools for analyzing large datasets; this statistics tutorial thus covers essential skills for many data library users.

Function Index

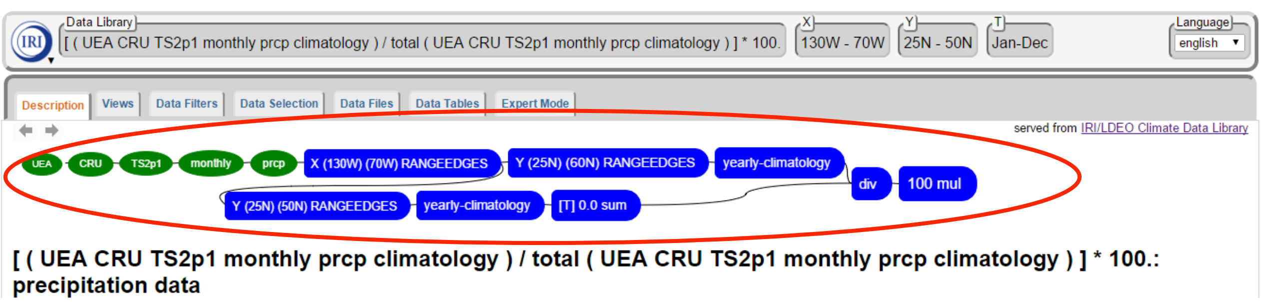
Index for functions that can be used to analyze data within the Data Library.

Help Resources

The Help Resources include basic and statistics tutorials, function documentation, and other resources to help you get the maximum utility out of the Data Library

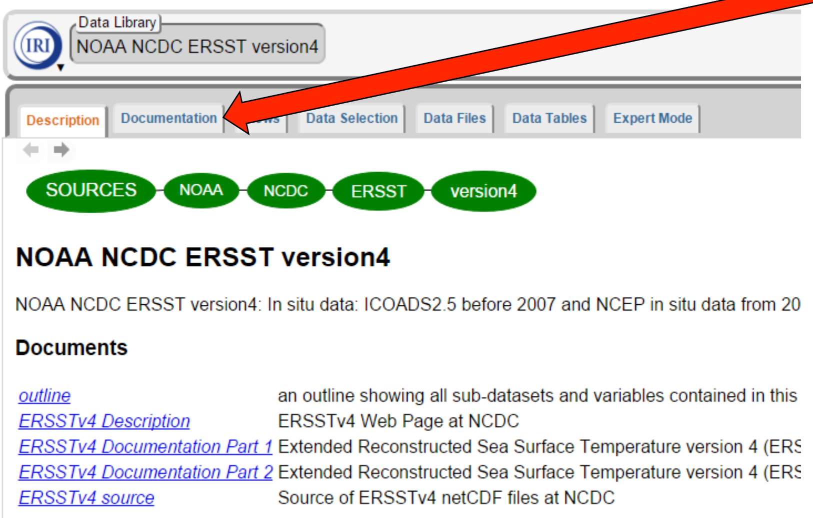


Dataset structure: source bar



- Use to track the location of the dataset in the Data library
- Click on any of the higher levels in the hierarchy to broaden your scope within the dataset
- Check the computations you have performed on the dataset using "expert mode" (Eg: percent of total annual precipitation falling in each month over North America)

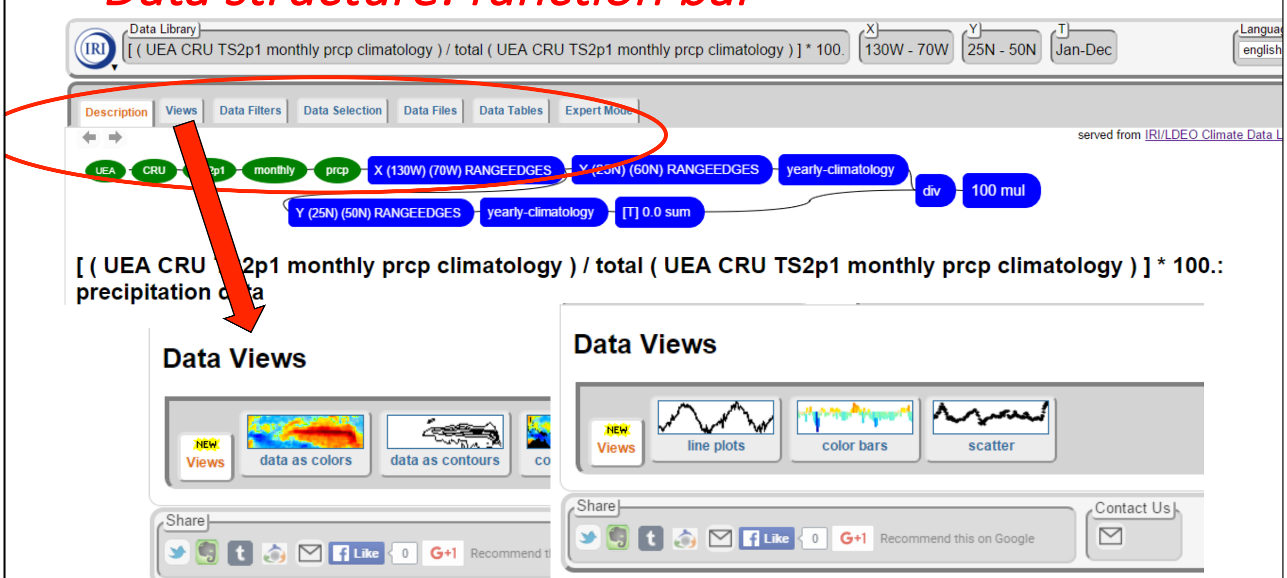
Data structure: function bar



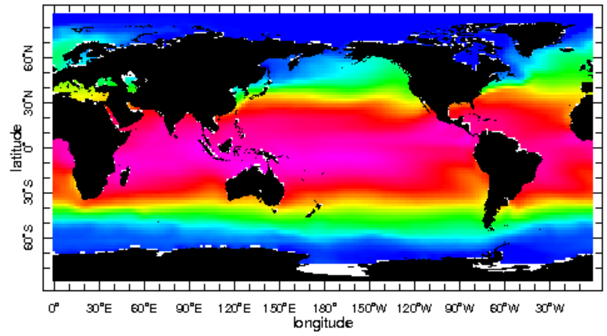
Documentation: description of datasets and key references for further information

Datasets and variables

Data structure: function bar

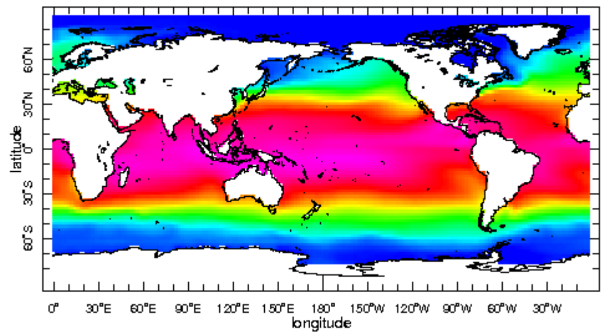


Views example 1: maps
colors with land



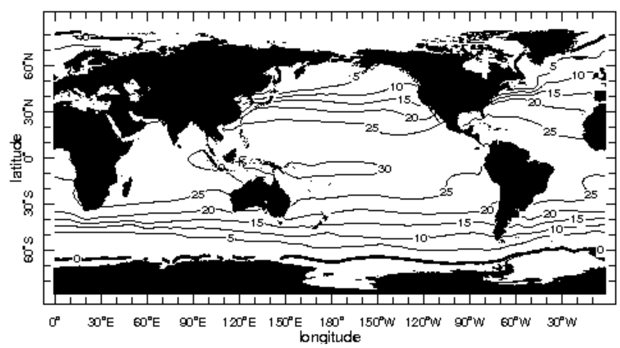
zlev 0.0 meters Time Jan 2016

colors with coasts

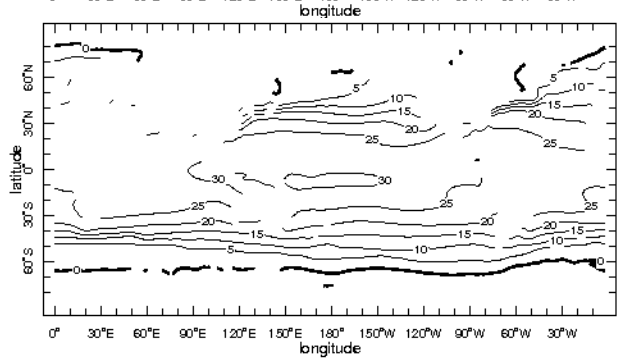


http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCDC/.ERSST
figviewer.html?map.url=X+Y+fig-+colors+land+-fig

Views example 1: maps
contours with land

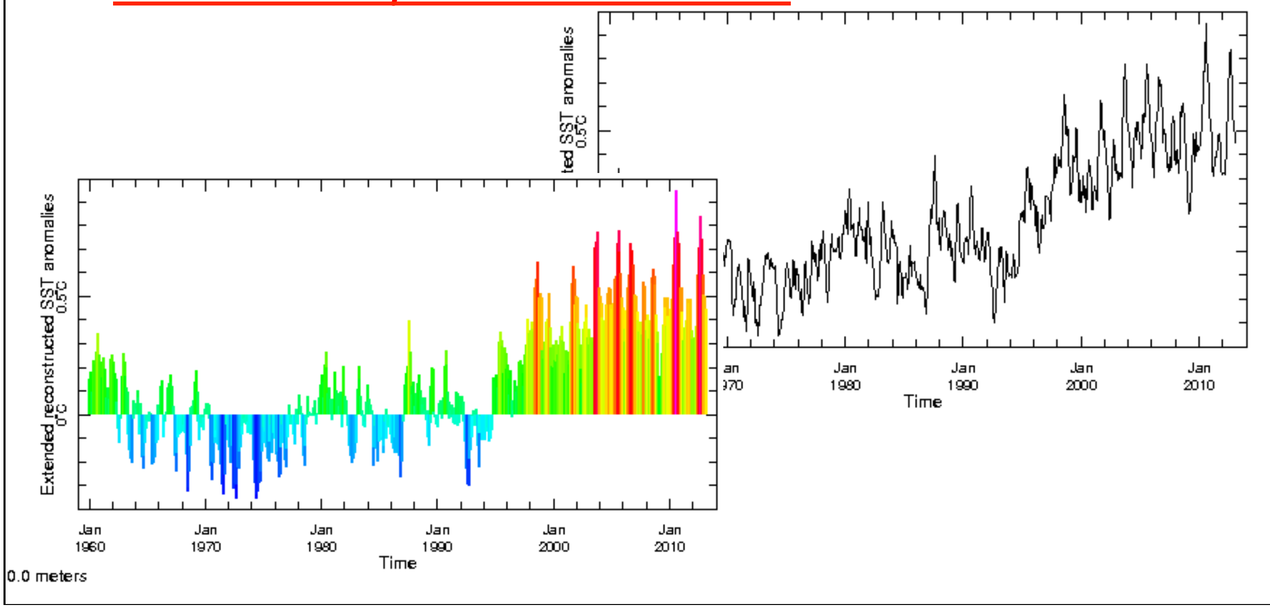


contours with coasts

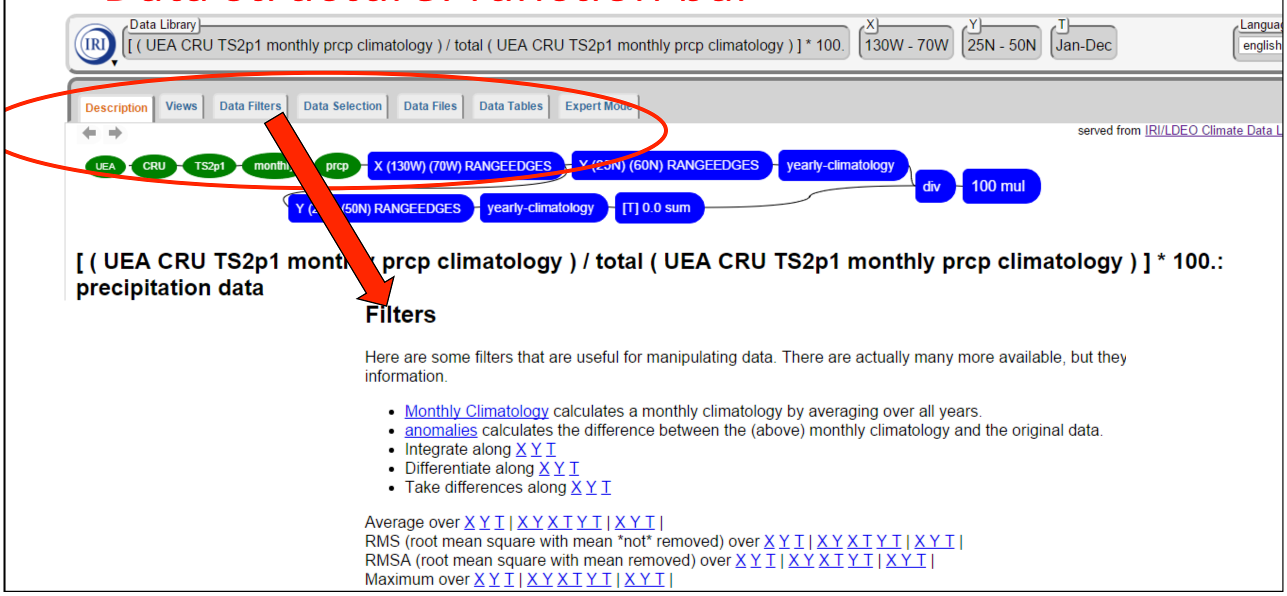


zlev 0.0 meters Time Jan 2016

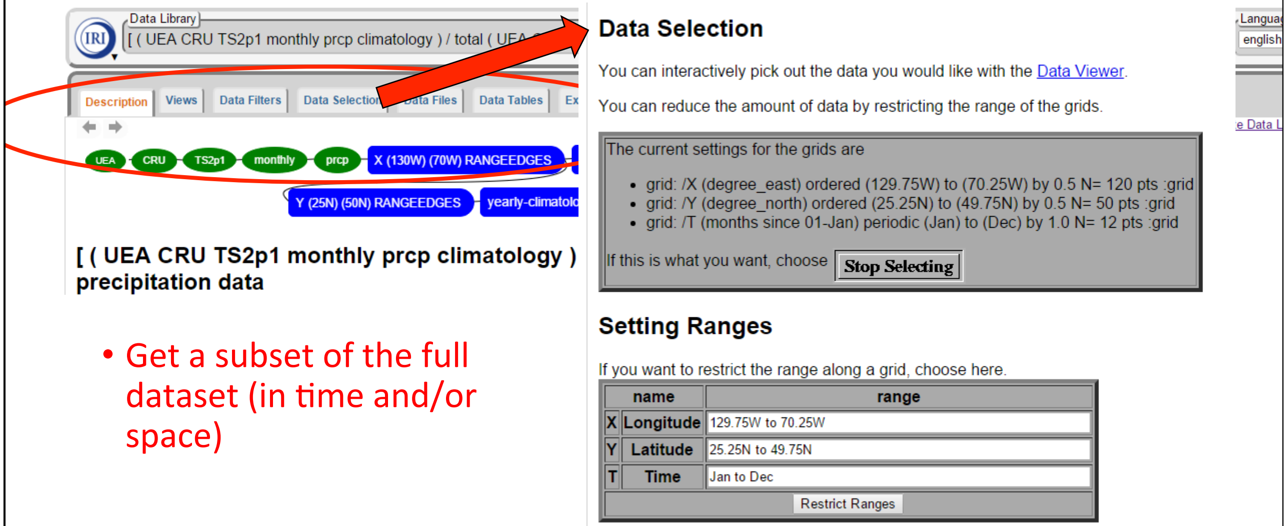
Views example 2: time series



Data structure: function bar



Data structure: function bar

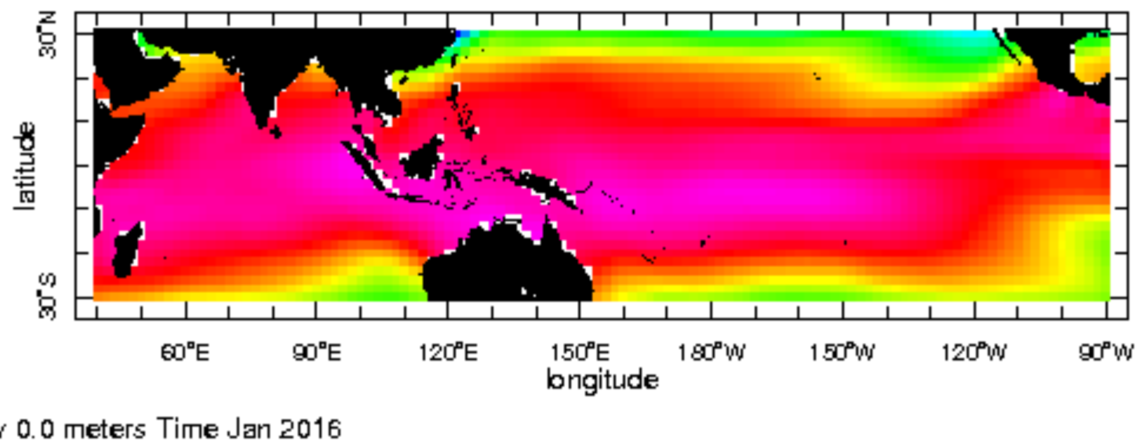


- Get a subset of the full dataset (in time and/or space)

Data selection hints

- longitude is best specified as west to east, two east values or two west values, otherwise you can end up with the wrong half of the world (e.g. 0.5E to 355.5E will work much better than 0.5E to 0.5W).
- order matters: reversing values will reverse the grid.
- when specifying time, some seasonal patterns work, i.e. Jan-Mar will select Jan-Mar of all years in the dataset, and Jan 1980-1990 will select Jan 1980, Jan 1981 through Jan 1990 (if those years exist in the dataset).
- a blank entry will get you back to the full grid.

Data selection example



Data selection example

Data Selection

You can interactively pick out the data you would like with the Data Viewer.

You can reduce the amount of data by restricting the range of the grids.

The current settings for the grids are

- grid: /X (degree_east) ordered (40E) to (90W) by 2.0 N= 116 pts :grid
- grid: /Y (degree_north) ordered (30S) to (30N) by 2.0 N= 31 pts :grid
- grid: /zlev (meters) ordered [(0.0)] :grid
- grid: /T (months since 1960-01-01) ordered (Jan 1854) to (Jan 2016) by 1.0 N= 1945 pts :grid

If this is what you want, choose **Stop Selecting**

Setting Ranges

If you want to restrict the range along a grid, choose here.

	name	range
X	longitude	40E to 90W
Y	latitude	30S to 30N

What happens if we enter the other way around??

Data Selection

You can interactively pick out the data you would like with the Data Viewer.

You can reduce the amount of data by restricting the range of the grids.

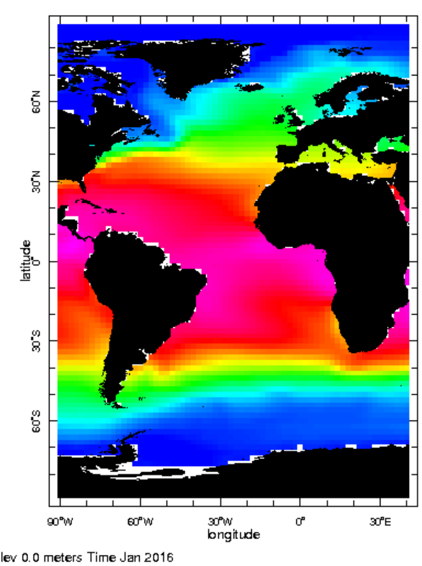
The current settings for the grids are

- grid: /X (degree_east) ordered (90W) to (40E) by 2.0 N= 66 pts :grid
- grid: /Y (degree_north) ordered (88S) to (88N) by 2.0 N= 89 pts :grid
- grid: /zlev (meters) ordered [(0.0)] :grid
- grid: /T (months since 1960-01-01) ordered (Jan 1854) to (Jan 2016) by 1.0 N= 1945 pts :grid

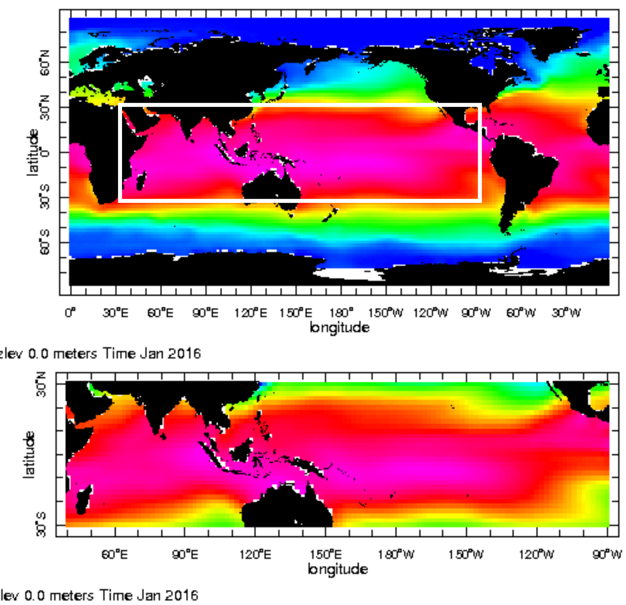
If this is what you want, choose **Stop Selecting**

What happens if we enter the other way around??

…..We get the Atlantic, not the Pacific!!

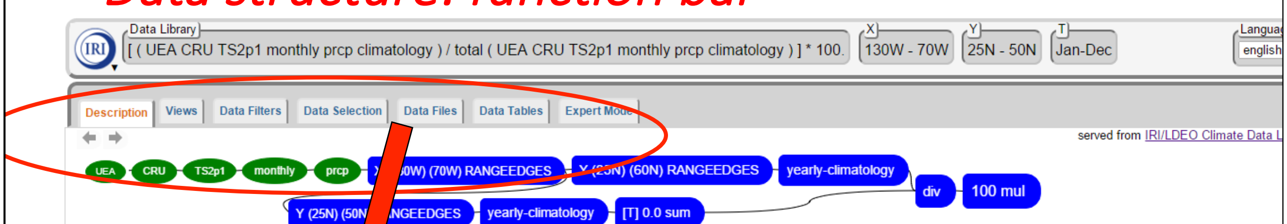


Data selection from view



http://iridl.ldeo.columbia.edu/SOURCES/.NOAA/.NCDC/.ERSST/.version4/.sst/
figviewer.html?map.url=X+Y+fig-+colors+land+-fig

Data structure: function bar

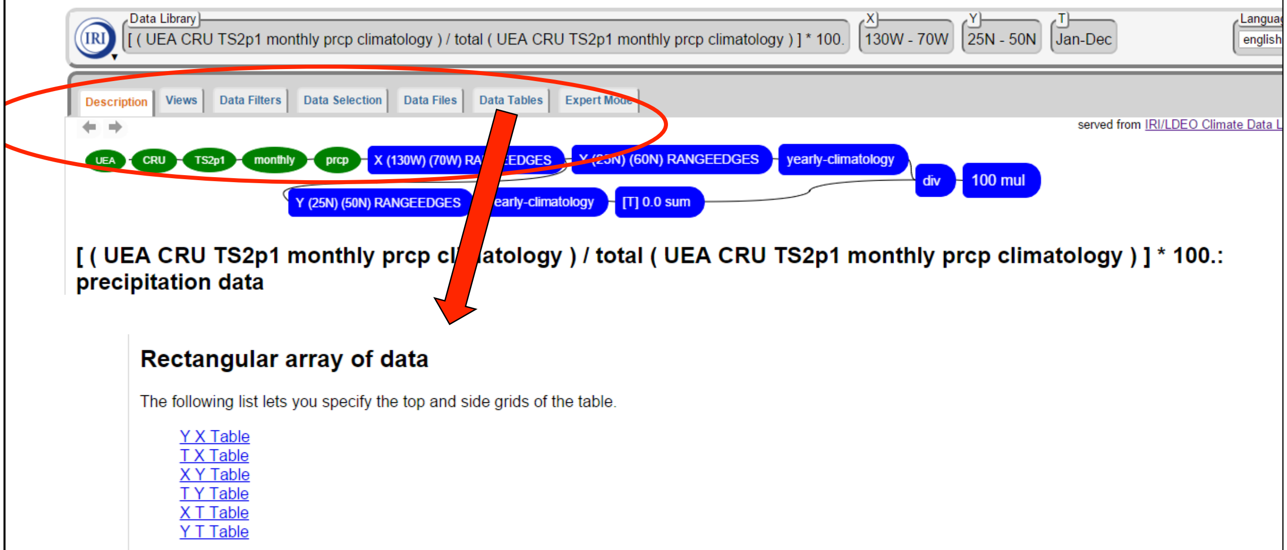


[(UEA CRU TS2p1 monthly prcp climatology) / total (UEA CRU TS2p1 monthly prcp climatology)] * 100.: precipitation data

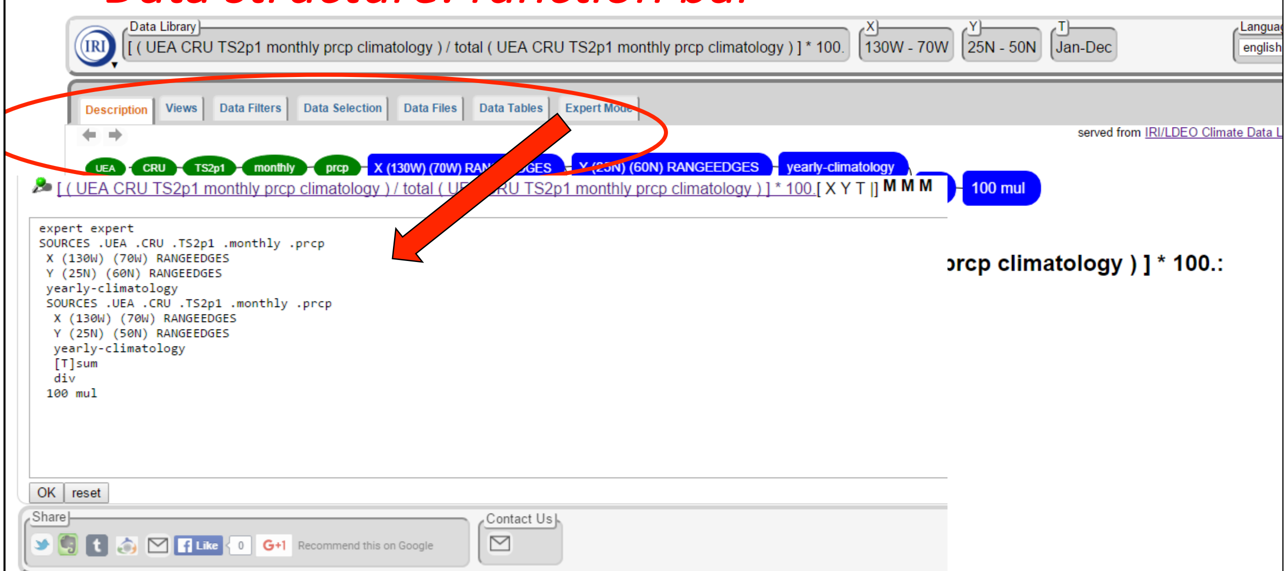
Download Data To Specific Software

ingrid	The Postscript-based software on which the Data Library is built.
CPT	Climate Predictability Tool More information
ferret	Interactive computer visualization and analysis software. More information
GrADS	Grid Analysis and Display System More information
matlab	Data analysis and visualization software. More information
NCL	NCAR Command Language More information
WinDisp	A public domain software package for the display and analysis of satellite images, maps and associated databases, with an emphasis on early warning for food security. More information

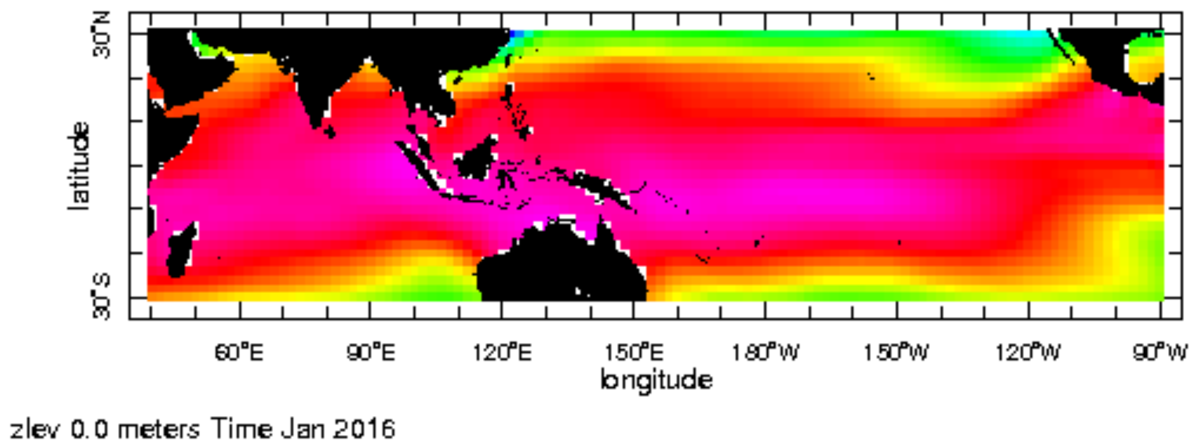
Data structure: function bar



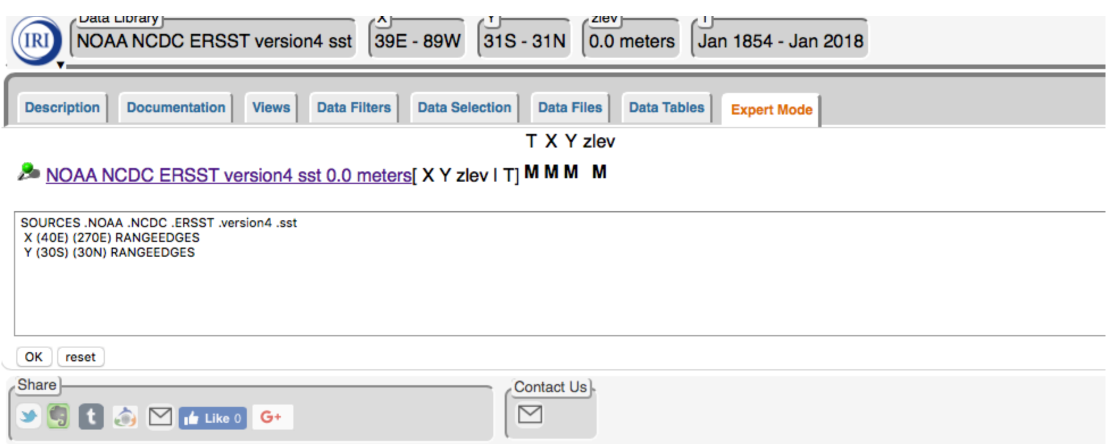
Data structure: function bar



Data selection in expert mode



Data selection in expert mode

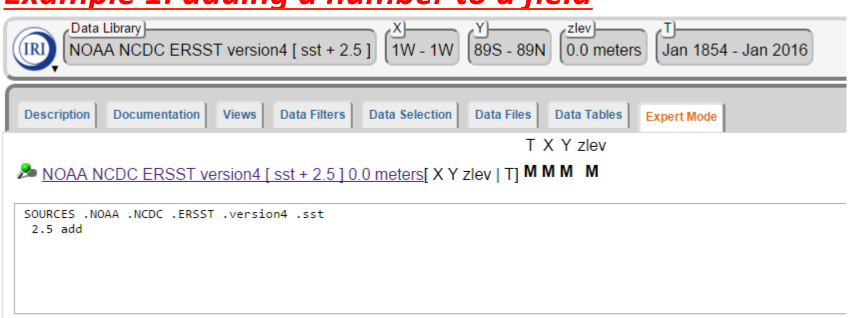


Manipulating data in expert mode

1. Arithmetic
2. Limits
3. Averages
4. Statistical
5. Other

Manipulating data: arithmetic

Example 1: adding a number to a field

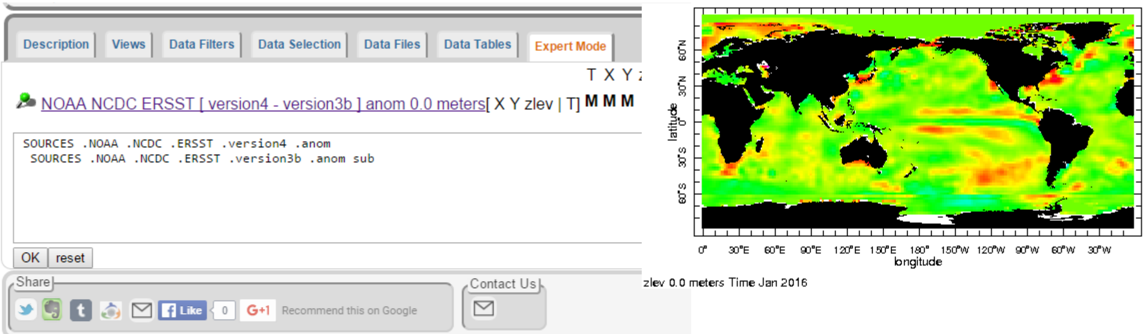


Note: units of that data variable are automatically used

NOAA NCDC ERSST version4 [sst + 2.5]: Extended reconstructed sea surface temperature data

Manipulating data: arithmetic

Example 2: subtracting one dataset from another



Subtracts bottom dataset from the top (ERSSTv4-ERSSTv3b)

Note: the spatial grids of these two variables must match!

Manipulating data: arithmetic

Basic arithmetic functions:

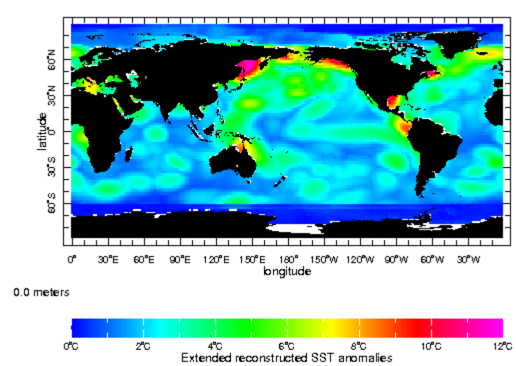
1. mul (multiply)
2. div (divide)
3. add (add)
4. sub (subtract)

Manipulating data: limits

- Set min or max limits (e.g. quality control)
- Select data that meets a certain criteria

Manipulating data: limits

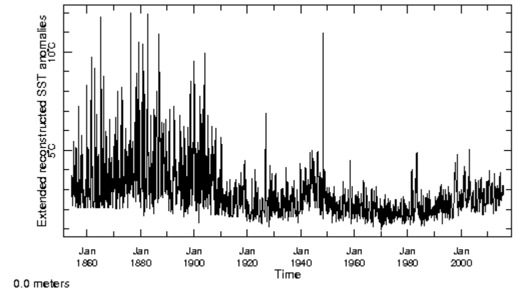
Example 1: find the largest SST for the entire time grid



[T] maxover or [T]minover

Manipulating data: limits

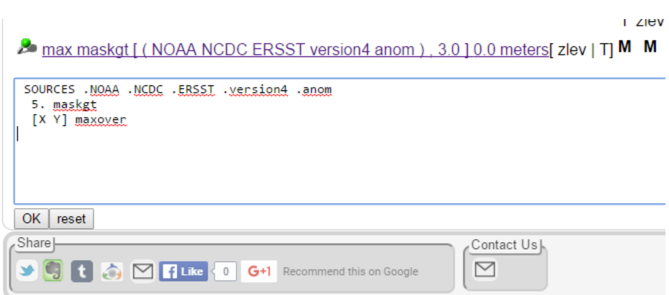
Example 2: find the largest SST for the entire spatial grid



[X Y] maxover or [X Y]minover

Manipulating data: limits

Example 2: creating a numerical mask

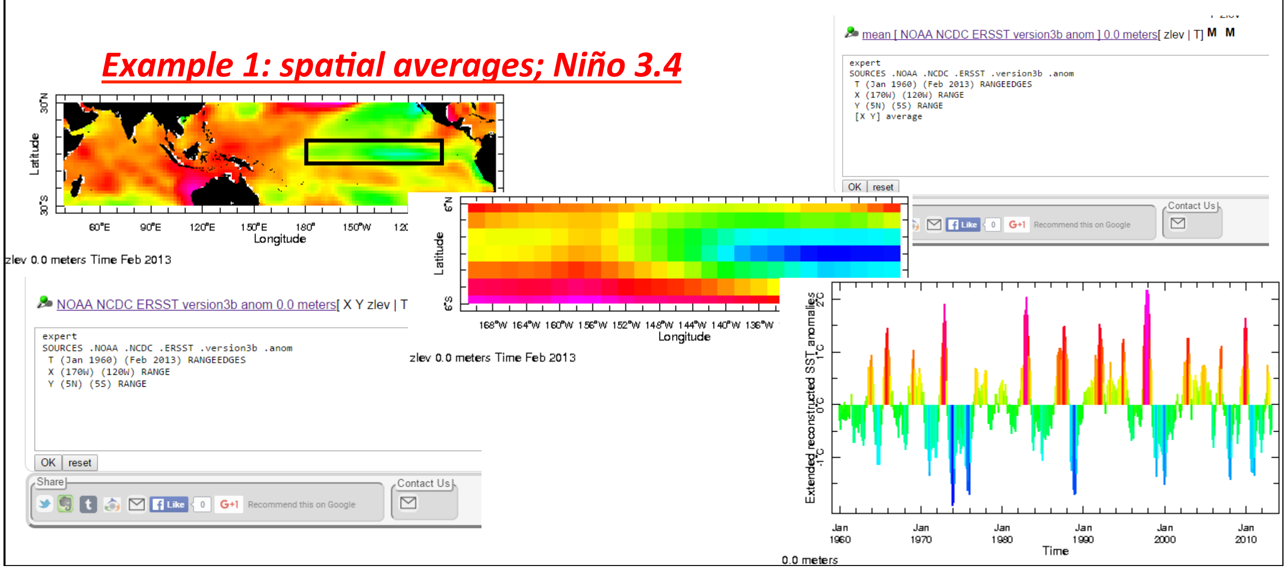


e.g., 5. maskgt or -1. masklt

Masks make data values that meet a particular threshold equal to NaN

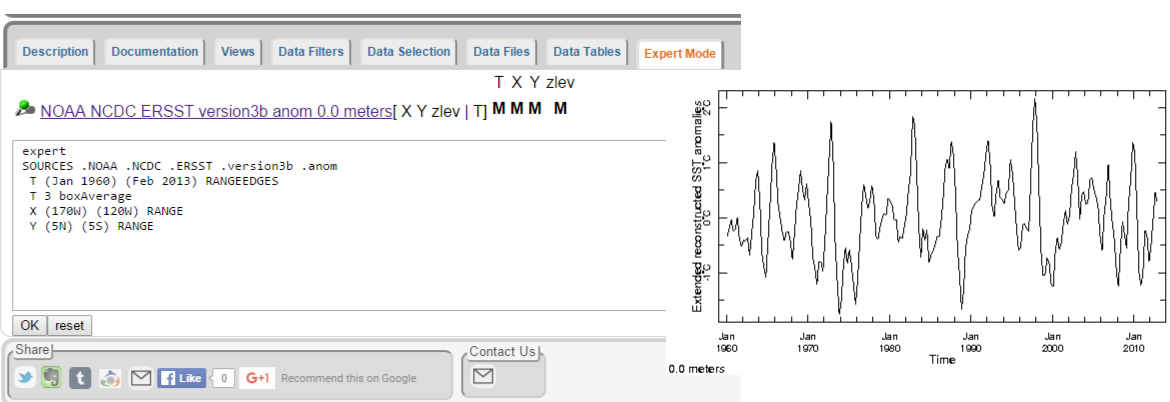
Manipulating data: averages

Example 1: spatial averages; Niño 3.4



Manipulating data: averages

Example 2: seasonal averages



Manipulating data: averages

Example 3: annual averages

Manipulating data in expert mode

Tips

- when specific time periods or locations are not selected, these operations are applied to the time and spatial grids in their entirety by default
- when multiple data variables are to be compared, as is common in the examples in this section, the time and spatial grids of those variables must be identical