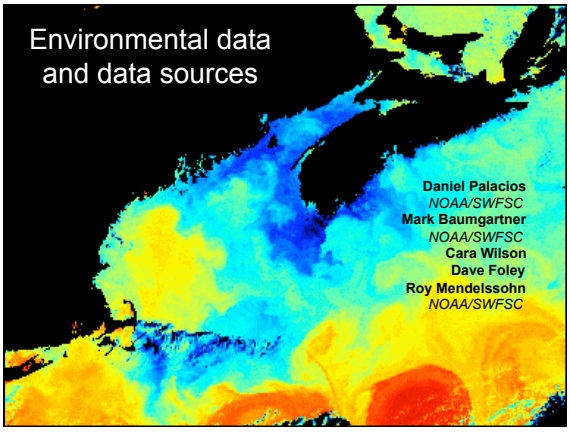
Environmental data and data sources



Daniel Palacios
NOAA/SWFSC
Mark Baumgartner
NOAA/SWFSC
Cara Wilson
Dave Foley
Roy Mendelssohn
NOAA/SWFSC

Outline

I. Data generalities
II. Environmental data
III. Data access

I. Data generalities

1. Data types
2. Data formats
3. Software packages for analysis

1. Data types

- Signed byte (8 bits)
- Unsigned char (8 bits)
- Signed short integer (16 bits)
- Signed long integer (32 bits)
- Single-precision floating point (32 bits)
- Double-precision floating point (64 bits)

A good data format must:

1. Be of easy and efficient storage and maintenance for long periods of time
2. Allow easy access in the long term

--National Center for Supercomputer Applications (NCSA)

Format characteristics:

- Is self descriptive (includes metadata)
- Is compact
- Supports sequential access
- Is adequate for a variety of storage technologies
- Is simple
- Access software is available and is easy to implement
- A rigorous format definition, API and I/O libraries are available
- The format is widely used
- Has long-term institutional support
- Is stable: changes little or nothing with time
- Supports the efficient operation of massive storage systems

--National Center for Supercomputer Applications (NCSA)

Metadata: representation of the data in the file, the structures contained in the file, the storage methods (compression, partition), information about the scientific content, time and date stamps, software versions, error change history

Metadata standards:
•Federal Geographic Data Committee Metadata Standard (www.fgdc.gov/index.html)
•National Center for Supercomputer Applications (www.ncsa.illinois.edu)

Kinds of things to include in your metadata:
•Project name
•Geographic area
•Species of interest
•Platform
•Who collected the data (e.g., name, organization)
•Kind of data (e.g., salinity)
•Type of instrument (what it is, model, serial number)
•What kind of processing has been done to the raw data

2. Data formats

- ascii
- Binary
- NetCDF (www.unidata.ucar.edu/software/netcdf)
- HDF5 (hdf.ncsa.uiuc.edu)

NetCDF: Is a stored list of multidimensional, numerical, named arrays called "variables." Their size is determined by zero or more named "dimensions." The variables may contain other named information called "attributes," and the file itself may have similar information called "global attributes." The variables are sub-referenced as "slabs," which consist of "corners" (starting indices), "count," and "strides" along the participating dimensions. A dimensionless variable can only store one value and is called a "scalar."

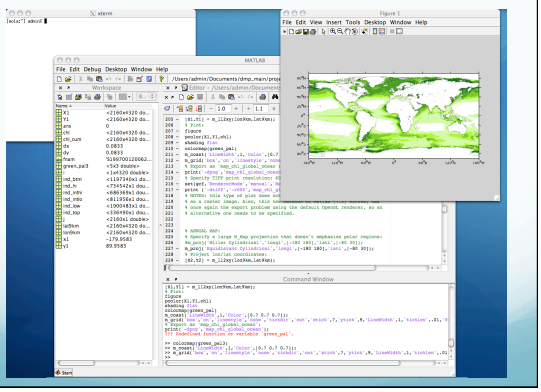
3. Software packages

- Low-level languages: compiled
  Fortran, C, C++
- High-level languages: interactive (environment)
  Perl, Matlab, IDL, Octave, R

A good package:

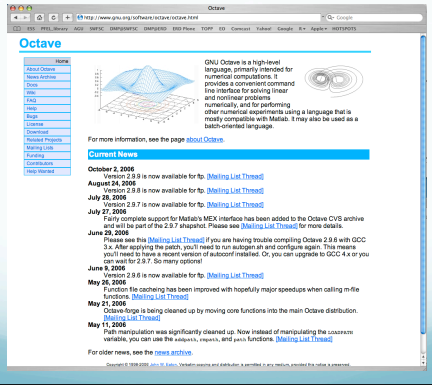
- Designed for numerical computing, solving linear algebra problems and non-linear equations, integration of ordinary functions, polynomial manipulation, integration of differential equations, matrix manipulation and *n*-dimensional arrays
- Expandable with functions written by the user and with modules written in C++, C, Fortran, java, and other languages
- Provides data visualization, signal processing, image processing, statistics
- Reads a wide variety of data formats
- Has ample documentation and an extensive user community

Matlab

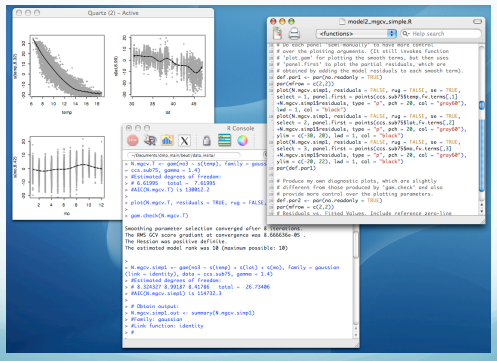


Matlab

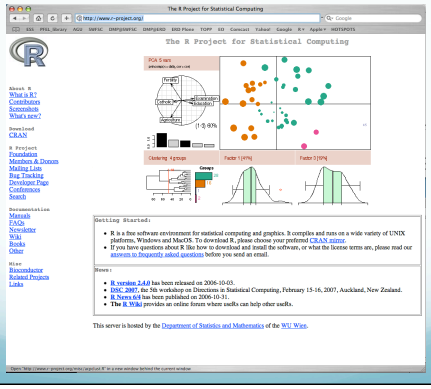
Octave



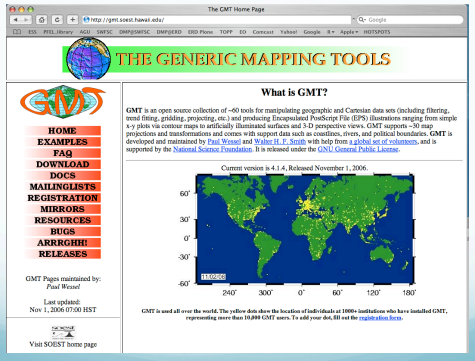
R



R



GMT

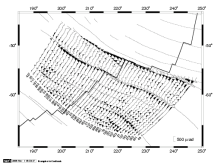


GMT

Examples of GMT output

Each of the 25 examples in the GMT Technical Reference and Cookbook may be viewed individually by clicking on the desired example. The images are reduced to speed up transmission; clicking on these images will open up a more detailed image. You will also have the option of viewing the responsible GMT script and download the complete PostScript illustration.

1. Contour map of global data set.
2. Color images of gridded data.
3. Spectral comparisons and x-y plots.
4. 3-D perspective, mesh-line surface
5. 3-D perspective, artificially illuminated grayshade
6. Two types of histograms.
7. A typical location map.
8. A 3-D bar graph.
9. Time/space-series plotted along tracks.
10. Superposition of 3-D bargraph and map.
11. The RGB color cube.
12. Delaunay triangulation, contouring, and imaging.
13. Plotting of vector fields.
14. Gridding, contouring, and trend surfaces.
15. Gridding with missing data.
16. More gridding options.
17. Blending images using clipping.
18. Volumes and spatial selections.
19. GMT color patterns
20. Extending GMT with custom symbols.
21. Demonstrating time-series annotations.
22. Map legend for global seismicity plot.
23. Spherical distances on the Earth.
24. Geospatial filtering of data.
25. Global distribution of antipodes.



Example 9. A variety of data sets can be described as a time-series collected along a track. Examples are magnetic anomalies and seismic traces. This illustration shows deflections of the vertical along several sub-parallel Geosat tracks over some south-Pacific fracture zones. The program pswiggle plots the anomalies, while psxy and pstext draw other features and annotates the plot.

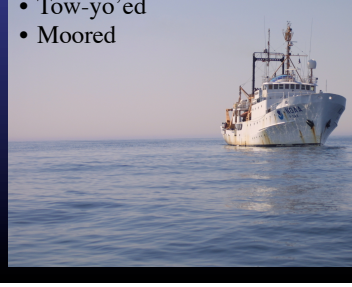
SCRIPT View GMT script POSTSCRIPT Download PostScript version. Back

II. Environmental data

- *In situ*
- Remotely sensed
- Climatologies
- Ocean Models
- Static

In-situ Data Sources

- Along-track
- Profiled
- Tow-yo'ed
- Moored



In-situ Data Sources

- Along-track
- Profiled
- Tow-yo'ed
- Moored

Primary Measurements
Water Depth
Temperature
Salinity
Currents
Fluorescence
Zooplankton
Acoustic Backscatter

In-situ Data Sources

- Along-track
- Profiled
- Tow-yo'ed
- Moored

Scientific Echosounder

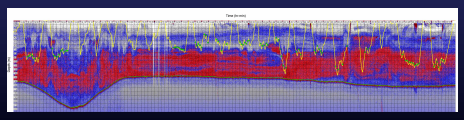
Thermosalinograph

Acoustic Doppler
Current Profiler

In-situ Data Sources

- Along-track
- Profiled
- Tow-yo'ed
- Moored

Scientific Echosounder



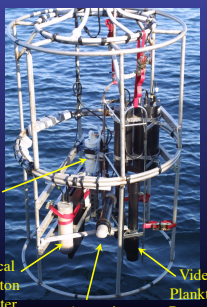
Following an animal
Goldbogen et al. (2008, JEB)

In-situ Data Sources

- Along-track
- Profiled
- Tow-yo'ed
- Moored

Fluorometer

Conductivity
Temperature
Depth Instrument
(CTD)



CTD

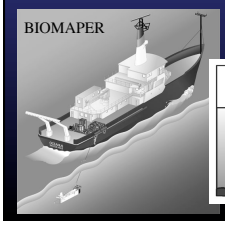
Optical
Plankton
Counter

Acoustic
Doppler Current
Profiler

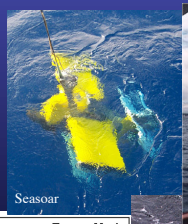
Video
Plankton
Recorder

In-situ Data Sources

- Along-track
- Profiled
- Tow-yo'ed
- Moored

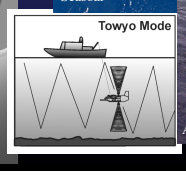


BIOMAPER

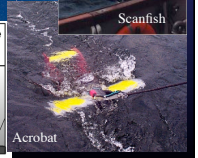


Seasoar

Scanfish



Towyo Mode

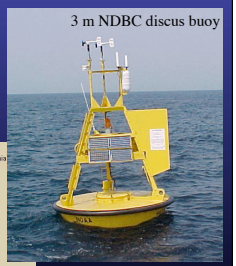


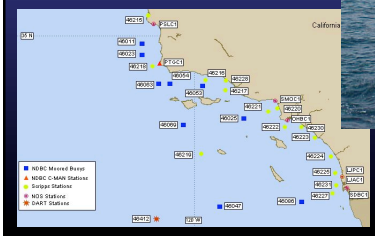
Acrobat

In-situ Data Sources

- Along-track
- Profiled
- Tow-yo'ed
- Moored

3 m NDBC discus buoy





In-situ Data Sources

Resources

- Instrumentation
  - → Oceanographers
  - → Instrument Manufacturers
- Data Archives
  - → National Oceanographic Data Center (www.nodc.noaa.gov)
  - → National Data Buoy Center (www.ndbc.noaa.gov)

II. Environmental data

- *In situ*
- Remotely sensed
- Climatologies
- Ocean Models
- Static

Remotely sensed Data Sources

- Passive
  - → Sea surface temperature (AVHRR, GOES)
  - → Frontal probability (AVHRR, GOES)
  - → Ocean color (CZCS, ADEOS, SeaWiFS, MODIS, VIIRS) & primary production
- Active
  - → Sea surface height (TOPEX/ Poseidon, JASON, ESR 1&2)
  - → Winds (QuikSCAT)
  - → Salinity (Aquarius)

Remotely sensed Data Sources

Resources

- Sea surface temperature
  - → JPL Data Archive (poet.jpl.nasa.gov; podaac-www.jpl.nasa.gov/sst)
  - → NOAA Coastwatch (coastwatch.noaa.gov)
- Ocean Color
  - → NASA Ocean Color Web (http://oceancolor.gsfc.nasa.gov/)
- Sea Surface Height
  - → JPL Data Archive (poet.jpl.nasa.gov)
  - → Aviso (www.aviso.oceanobs.com)

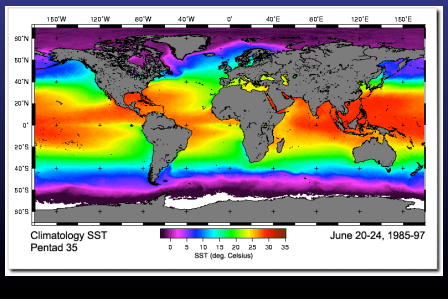
II. Environmental data

- *In situ*
- Remotely sensed
- Climatologies
- Ocean Models
- Static

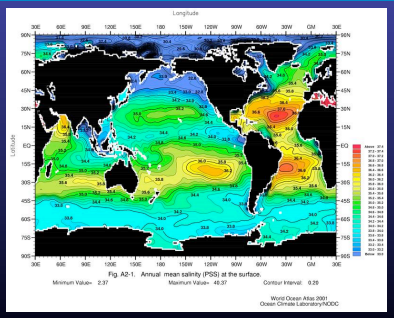
Climatological Data Sources

Climatology: Long-term average conditions

Anomaly: Deviation from the long-term average conditions



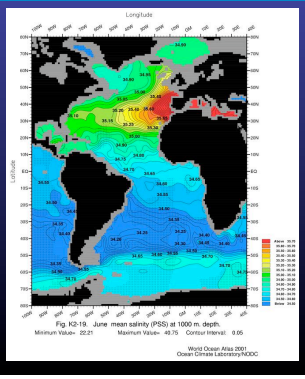
Climatological Data Sources



Surface Salinity Climatology
from World Ocean Atlas ("Levitus")

Climatological Data Sources

Salinity at 1000 m
Climatology from World
Ocean Atlas



Climatological Data Sources

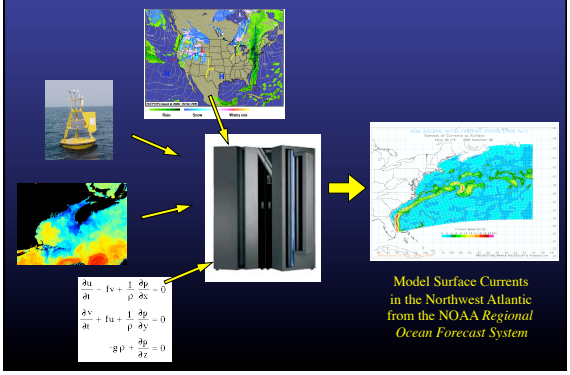
Resources

- Global Climatology
  - → World Ocean Database and World Ocean Atlas (www.nodc.noaa.gov/OC5/indprod.html)
- Regional or Local Climatology
  - → Find in the literature and request data from authors

II. Environmental data

- *In situ*
- Remotely sensed
- Climatologies
- Ocean Models
- Static

Ocean Model Data Sources



Model Surface Currents
in the Northwest Atlantic
from the NOAA *Regional*
Ocean Forecast System

Ocean Model Data Sources

Resources

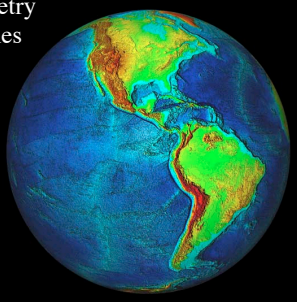
- Global Models
  - → U.S. Navy Models (www7320.nrlssc.navy.mil/global_nlom)
- Regional Models
  - → U.S. East Coast Regional Ocean Forecast Model (polar.ncep.noaa.gov/cofs)
  - → Regional Ocean Modeling System (ROMS)
- Local Models
  - → Find local modelers and build collaborations

II. Environmental data

- *In situ*
- Remotely sensed
- Climatologies
- Ocean Models
- Static

Static Data Sources

- Bathymetry
- Coastlines



Static Data Sources

Resources

- High-resolution bathymetry (ETOPO5, ETOPO2, SRTM30_PLUS)
  - → Scripps Institution of Oceanography (topex.ucsd.edu/WWW_html/mar_topo.html)
  - → U.S. National Geophysical Data Center (www.ngdc.noaa.gov/mgg/global/global.html)
- Coastlines
  - → U.S. National Geophysical Data Center (www.ngdc.noaa.gov/mgg/shorelines/shorelines.html)
- Rivers
  - → Digital Chart of the World (http://www.maproom.psu.edu/dcw/)

III. Data integration & delivery

At the Environmental Research Division
Of NOAA/NMFS/SWFSC

Roy Mendelssohn	NOAA/NMFS/SWFSC ERD
Lynn DeWitt	NOAA/NMFS/SWFSC ERD
Bob Simons	NOAA/NMFS/SWFSC ERD
Luke Spence	NOAA Corps
Dave Foley	JIMAR, University of Hawaii & NOAA/ERD
Cindy Bessey	JIMAR, University of Hawaii & NOAA/ERD (now at Florida International University)
Eoin Howlett	Applied Sciences Associates (ASA), Inc.
Chris Mueller	Applied Sciences Associates (ASA), Inc.
Kelly Knee	Applied Sciences Associates (ASA), Inc.

Background

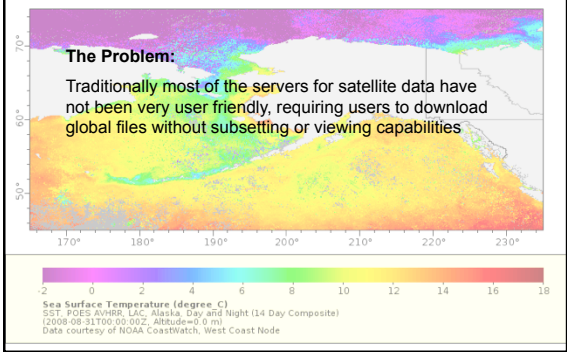
- The continuity, global coverage, and high temporal and spatial resolution of satellite data make them important assets for monitoring and characterizing marine ecosystems
- However oceanographic satellite data have been underutilized by both NOAA and marine resource managers in general
- The Environmental Research Division, ERD, (NOAA/SWFSC) has developed a variety of tools to make using satellite data an easier process

Why is satellite data underutilized?

- Satellite data can be difficult to access, manipulate and process, particularly for people who have never used it before
- Work required to get relevant parameters can be cumbersome, i.e.:
  - primary productivity from chlorophyll
  - front locations from SST fields
  - climatologies required to generate anomalies
  - rigorous 'data mining' needed to match up satellite data with telemetry records
  - "one-stop" shopping desired

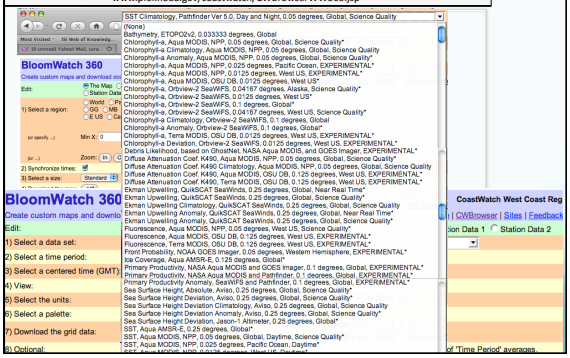
1. Data Servers
2. The EDC for ArcGIS (Environmental Data Connector)
3. Xtract-o-matic for Matlab and R

1. Data Servers

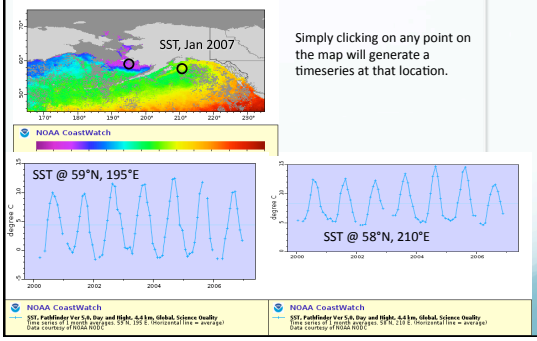


ERD Satellite Data Browser

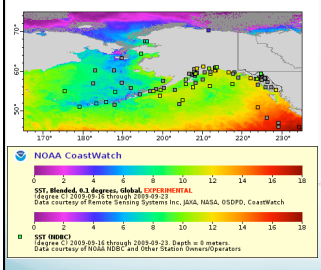
www.pfel.noaa.gov/coastwatch/CWbrowserWW360.jsp



Viewing Timeseries



Comparing Datasets

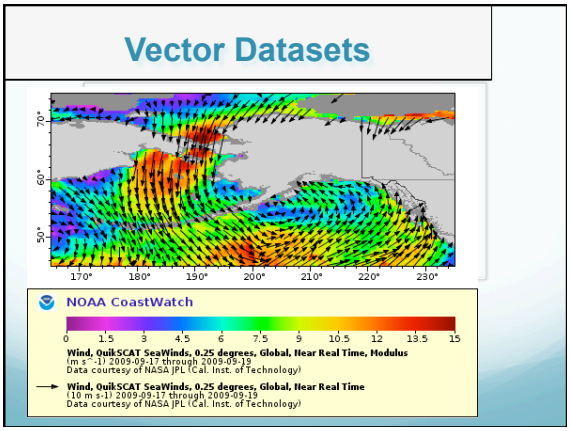


Blended SST from 9/16-9/23 2009

Overlay of in-situ SST data from NDBC buoys from the same time period

Ability created in response to needs of IOOS RAs

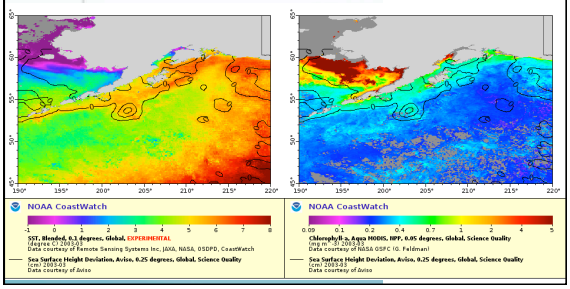
Vector Datasets



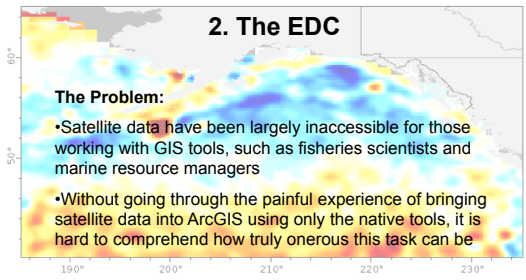
Overlaying Datasets

March 2003

SST Chlorophyll



2. The EDC



The Problem:

- Satellite data have been largely inaccessible for those working with GIS tools, such as fisheries scientists and marine resource managers
- Without going through the painful experience of bringing satellite data into ArcGIS using only the native tools, it is hard to comprehend how truly onerous this task can be

EDC™ Environmental Data Connector

an extension for connecting ArcGIS to Thredds/OPeNDAP

- Allows easy selection, and importation, of any gridded, non-projected, CF-compliant dataset served by Thredds/ OPeNDAP into ArcGIS 9.2 or 9.3
- Data can be subset temporally and spatially
- Developed by Applied Science Associates, Inc.
- Funded by NOAA's Satellite R&O project
- Free distribution

www.pfeg.noaa.gov/products/EDC

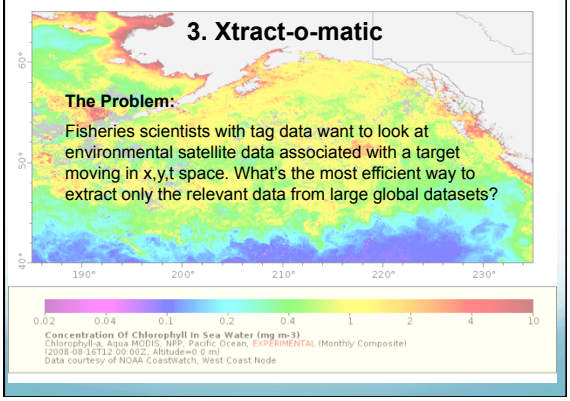
Standalone Module

EDC™ Environmental Data Connector

- Provides the front-end capabilities of the EDC independent of ArcGIS
- Provides a GUI to browse THREDDS catalogs or OPENDAP directories, to subset the selected data in space and time, and to download the data as a netcdf file
- The EDC Standalone Module will work on any computer with Java 1.5

www.pfeg.noaa.gov/products/EDC

3. Xtract-o-matic



The Problem:

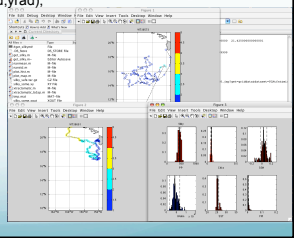
Fisheries scientists with tag data want to look at environmental satellite data associated with a target moving in x,y,t space. What's the most efficient way to extract only the relevant data from large global datasets?

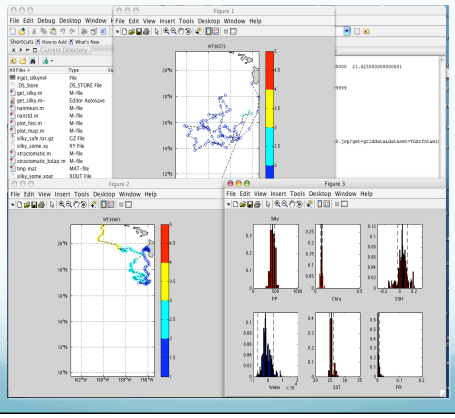
Xtract-o-matic

- Client-based tool to import 4-dimensional environmental data from the web directly into their own working environments, including Matlab, R and [coming soon] IDL
- Key features:
  - One option allows the selection of data in a user-selected region along arbitrary tracks. This allows the analyst to "swim" or "fly" with the animals through data fields while reducing strain on bandwidth and other technical infrastructure.
  - No fuss, no muss with file formats: the data is imported as a variable directly into the given workspace.
  - Access to a variety of data is built in, but it can potentially access data served by any interoperable web service (e.g., WCS and OPeNDAP) as recommended by the Integrated Ocean Observing System.
  - Based on simple URL queries, as opposed to OPeNDAP, which requires the installation of software specific to both platform and client software.
- http://coastwatch.pfel.noaa.gov/xtracto

Matlab Code

```
[chl] = xtractomatic(lon,lat,ftime,'62',xrad,yrad);
[sst] = xtractomatic(lon,lat,ftime,'18',xrad,yrad);
[pp] = xtractomatic(lon,lat,ftime,'41',xrad,yrad);
[wekm] = xtractomatic(lon,lat,ftime,'58',xrad,yrad);
[ssh] = xtractomatic(lon,lat,ftime,'48',xrad,yrad);
[tfnt] = xtractomatic(lon,lat,ftime,'65',xrad,yrad);
```





A demonstration

EDC™ Environmental Data Connector

an extension for connecting ArcGIS to Thredds/OPeNDAP

- A screen capture movie showing data being brought into ArcGIS
- Surface vector wind (SVW) data off of the California coast is selected and imported into ArcGIS
- Sea Surface Temperature data is then selected and overlaid on the SVW data
- The time slider is used to move through the different timesteps
- Datasets do not need to have identical time steps or bounding boxes