Ocean Data Quality Control [and Assimilation]

Select and assimilate “good” observations while rejecting “bad” observations when “good” and “bad” are system-dependent

JCSDA Summer Colloquium on Data Assimilation

Summer 2012

Kayo Ide

University of Maryland

Basic Ideas for Ocean Data Quality Control

- Quality control (QC) for ocean data from data assimilation perspective
  - Basic concepts are similar to atmospheric systems from theoretical (Bayesian) view
    - If observations are abundant, then buddy check (not only to reject but also to accept)
    - If observations are sparse, primary corroborative source of information may be forecast (in particular in the form of ensemble)
  - Details may require unique adjustment, development, and implementation
    - Ocean data (network) can be sparse/inhomogeneous and sporadic
      - » If erroneous values are assimilated, then they can cause immediate spurious overturning and/or be advected around for years afterwards (nothing else to correct)
    - Ocean data assimilation is relatively new and may have different emphases from atmospheric data assimilation.
    - Scales are different in both:
      - » Dynamics
      - » Observations

Background: Buddy-Check. Atmospheric Example

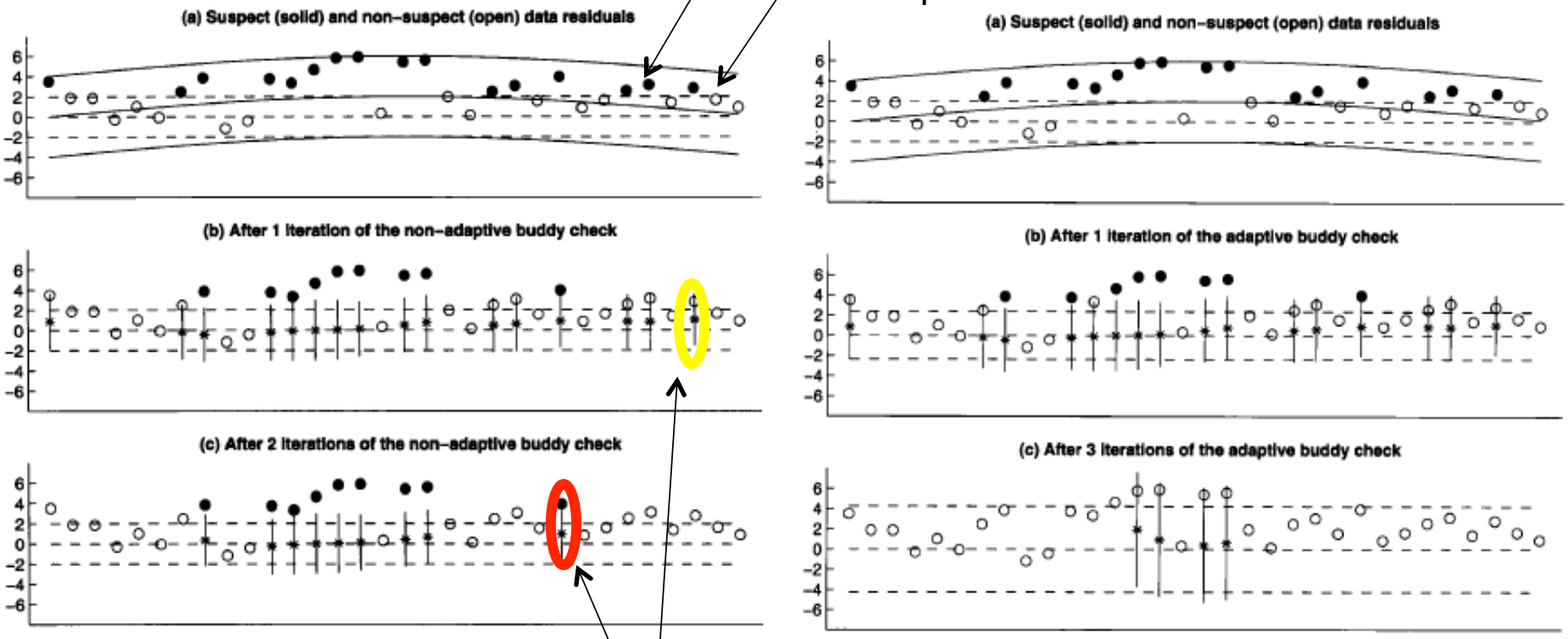
- Based on innovation $\mathbf{d}=\mathbf{y}^o-\mathbf{h}(\mathbf{x}^b)$ for dense observing network

Non-Adaptive

Suspects

Non-suspects

Adaptive

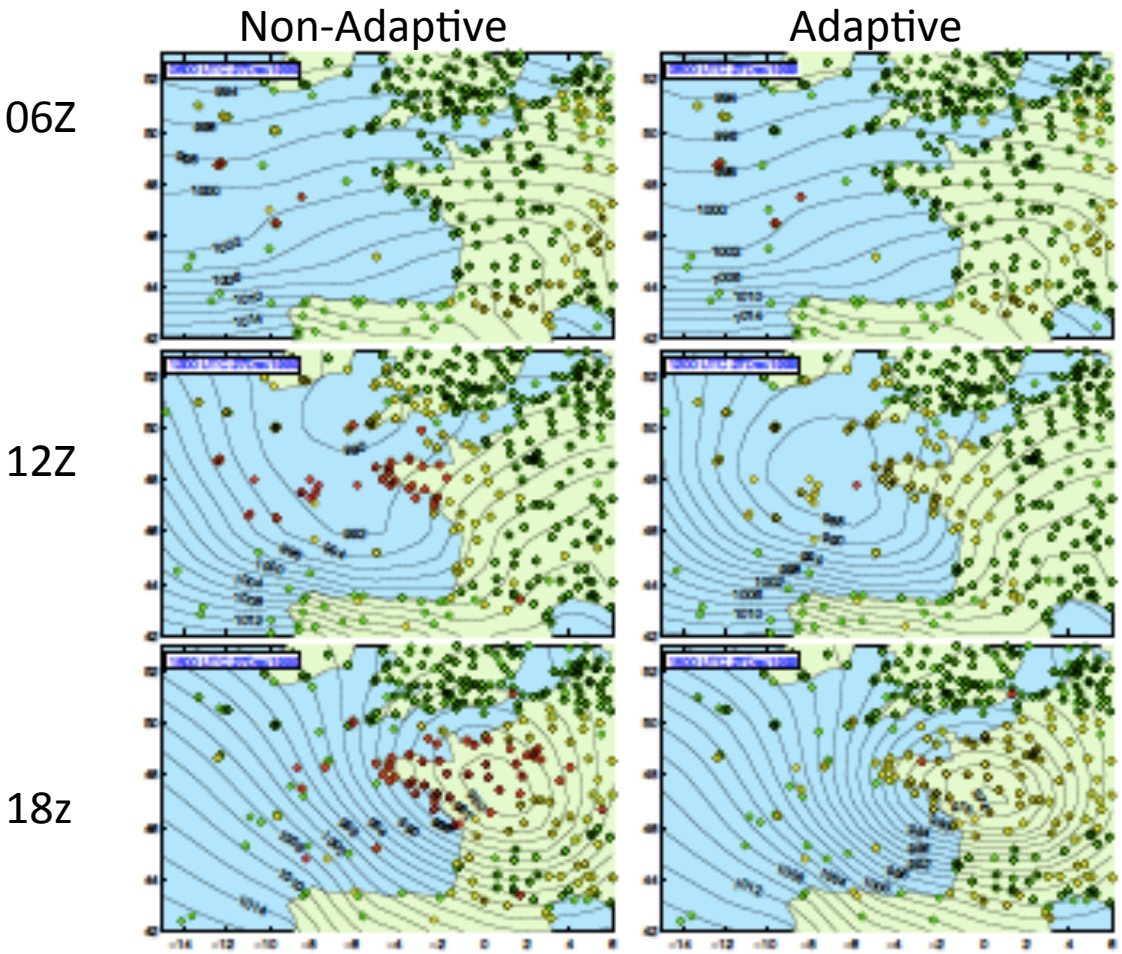


Given all non-suspects,
- conditional expectation
- conditional range with $\sqrt{3\sigma}$

Dee et al (2001)

Background: Buddy-Check

- Example: Intense storm over Europe December 27, 1999



Dee et al (2001)

Outline

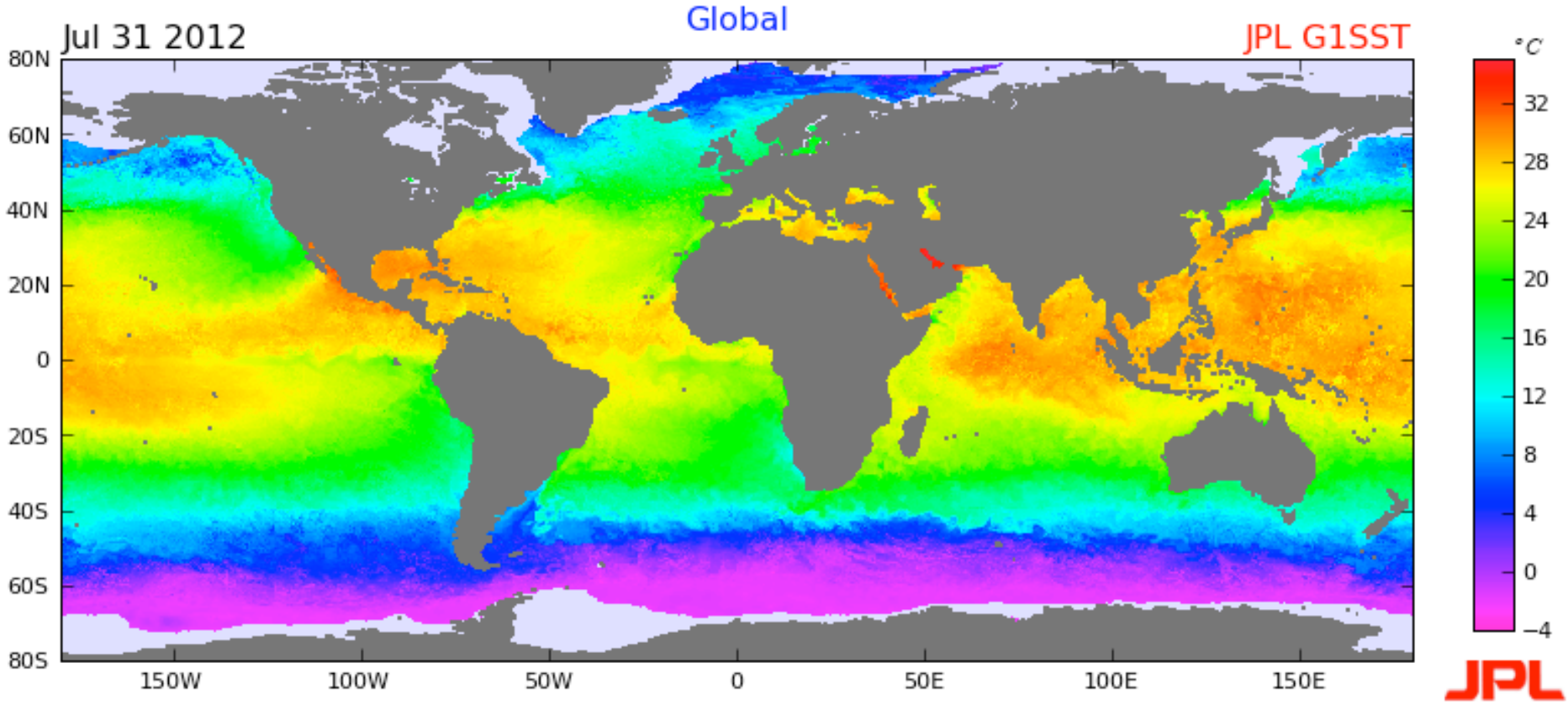
- Some examples of ocean data assimilation systems
  - Global Operational & Reanalysis
    - Navy Coupled Ocean Data Assimilation (NCODA) [real time]
    - ECMWF ORA & ORT-S4 [re-analysis and real-time]
    - NCEP Global Ocean Data Assimilation system (GODAS) [real time]
    - Simple Ocean Data Assimilation (SODA) [reanalysis]
  - Regional Operational/Real time
    - NCODA Relocatable system (RELO) [real time &Relocatable]
    - Southern California Coastal Ocean Observing System (SCCOOS) [near real-time]

- Types of ocean data used in the assimilation
  - Conventional platforms
    - In-situ [mainly T, S]
    - Remote sensing, satellite in particular, including Salinity
  - New types of platforms
    - HF (high frequency) radar [surface (u,v)]
    - Lagrangian data [trajectories]

- Quality control
  - Basic concepts
  - Examples
  - New perspective

Ocean Observation: Satellite

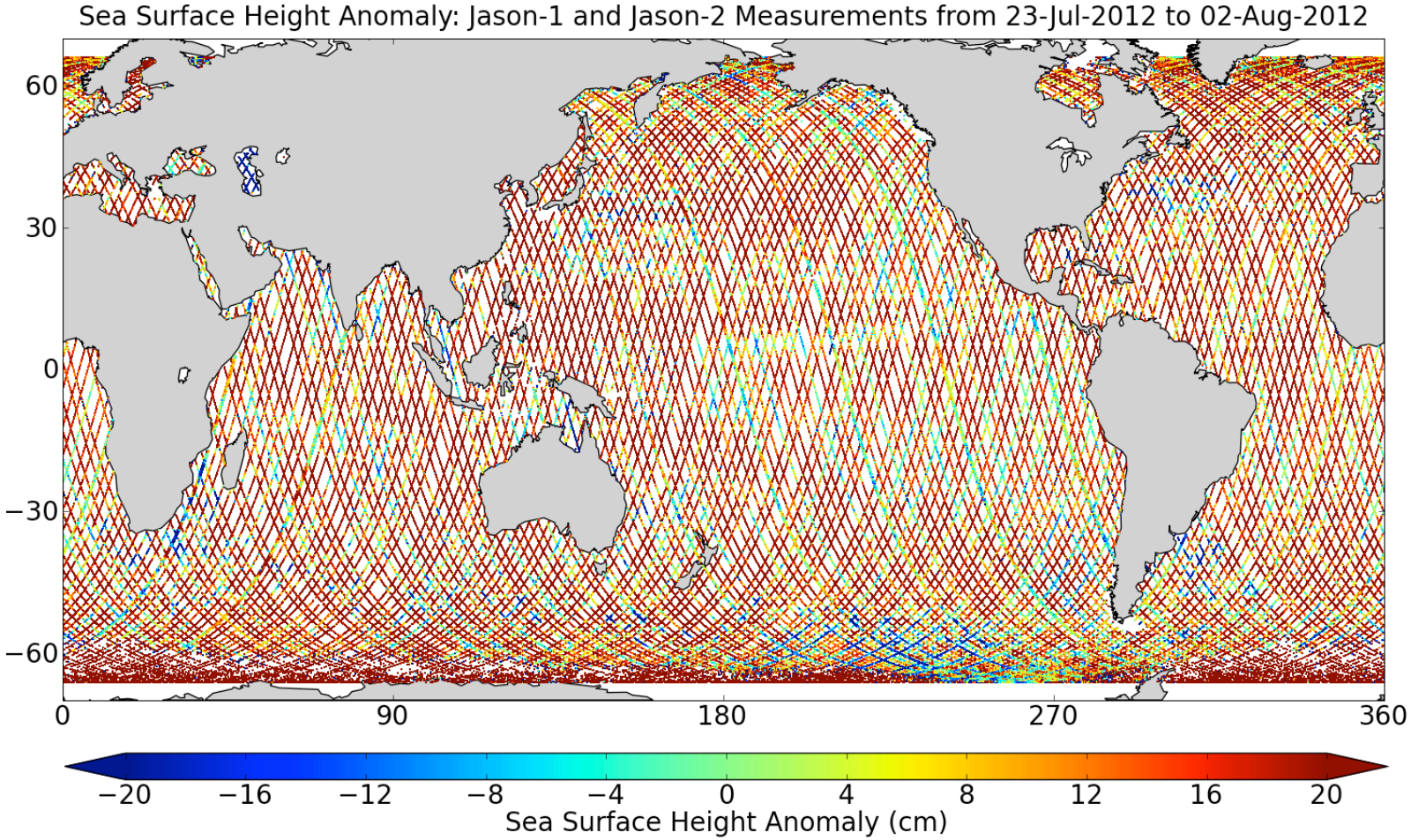
- Sea Surface Temperature



http://ourocean.jpl.nasa.gov/SST/#

Ocean Observation: Satellite

- Sea Surface Height

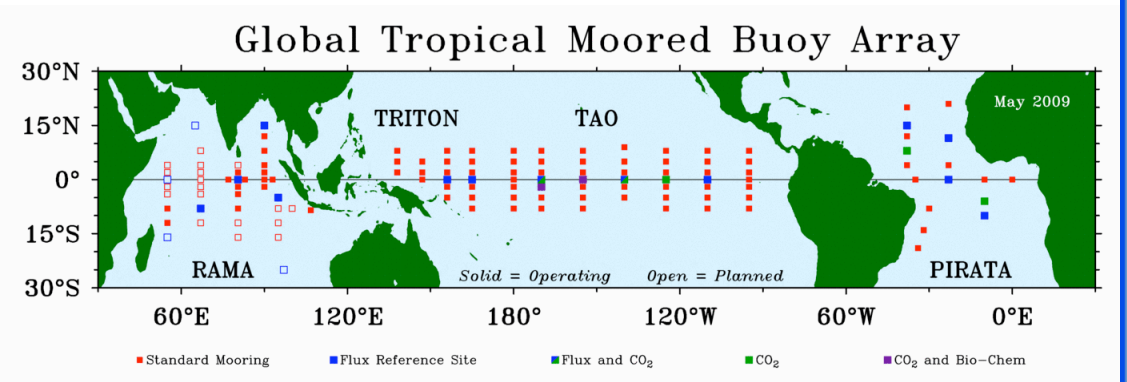


http://sealevel.jpl.nasa.gov/Science/datasources/ssha/

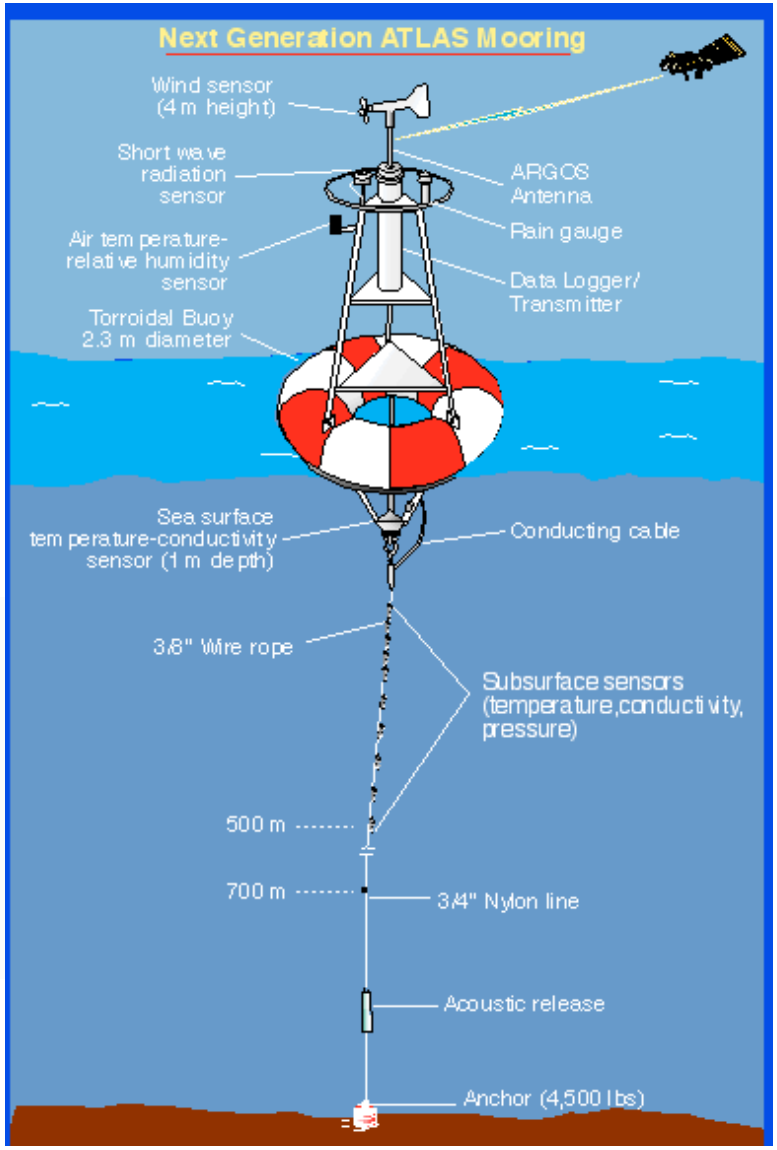
In-Situ Instrument. Profiling Station

- Mooring: CTD, Surface observations
  - TAO (Tropical Atmosphere Ocean):

    For improved detection, understanding and prediction of El Niño and La Niña.
  - TRITON
  - PRITA (small no of deep water buoys)

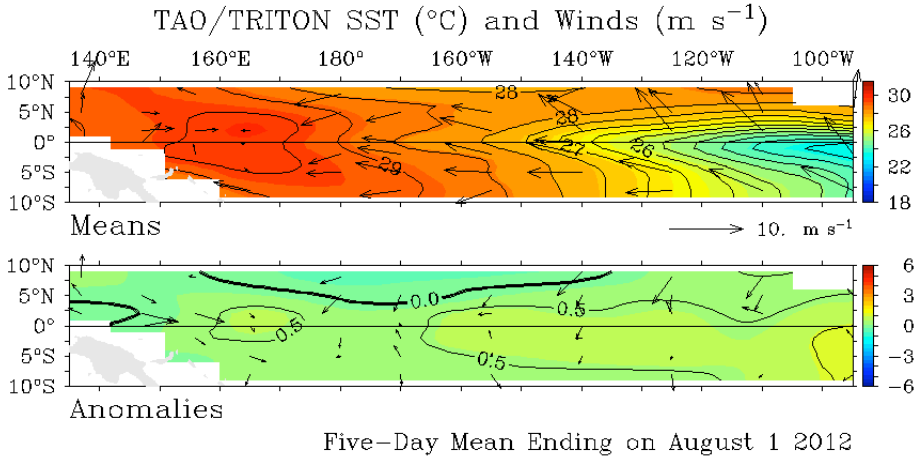


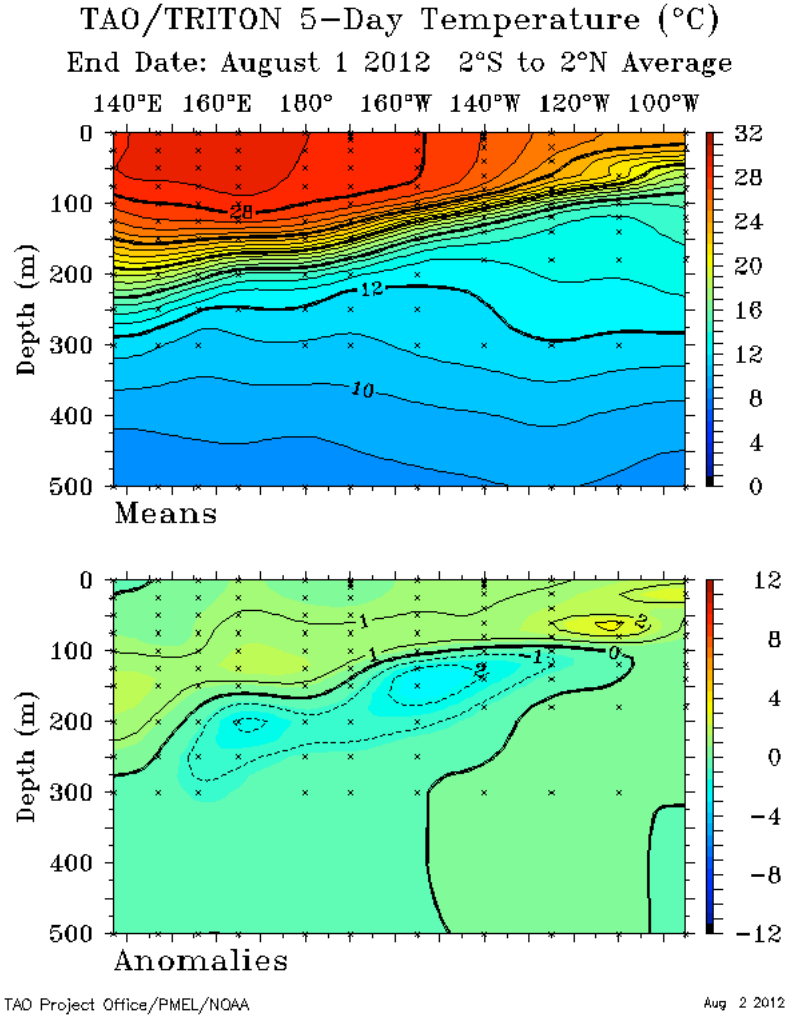
http://www.pmel.noaa.gov/tao/index.shtml



In-Situ Data

- Mooring Data





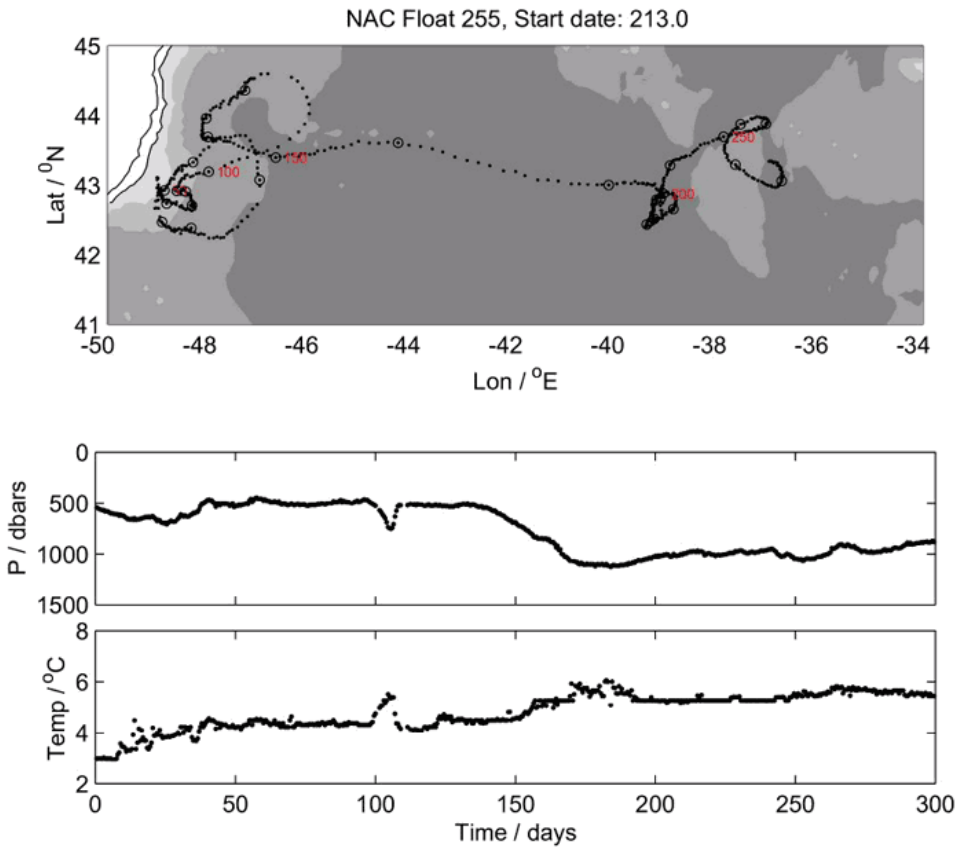
- Mooring data are temporally & vertically 'continuous'.

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

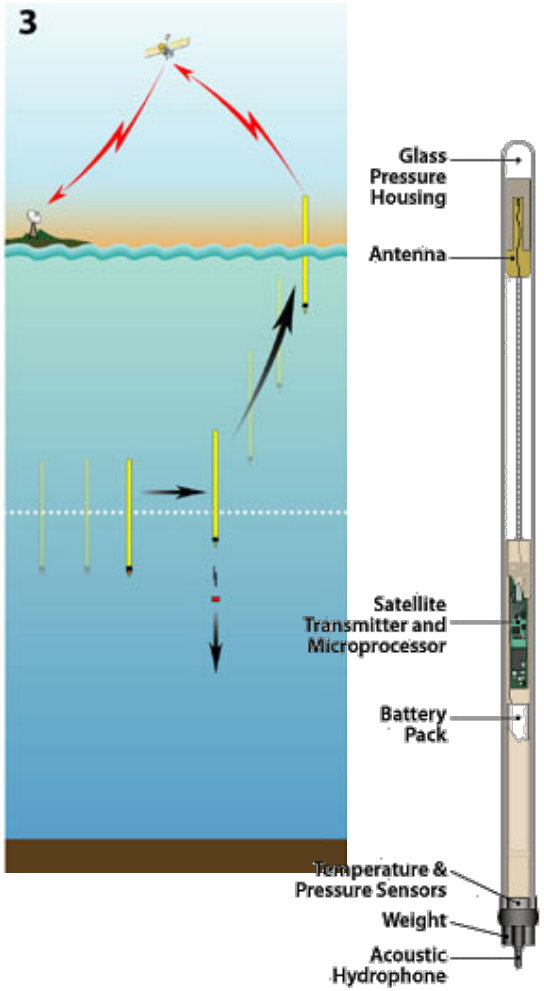
In-Situ Instruments(Movable Lagrangian) Platform. Floats

- Argo Floats. Observation on the isopyncnal surface
  - (*T,S*) by CTD
  - (u,v) derived from position

  along ($\mathbf{x}^{(2D)}(t_k)$, $p(\mathbf{x}^{(2D)}(t_k))$ in upper 2000m

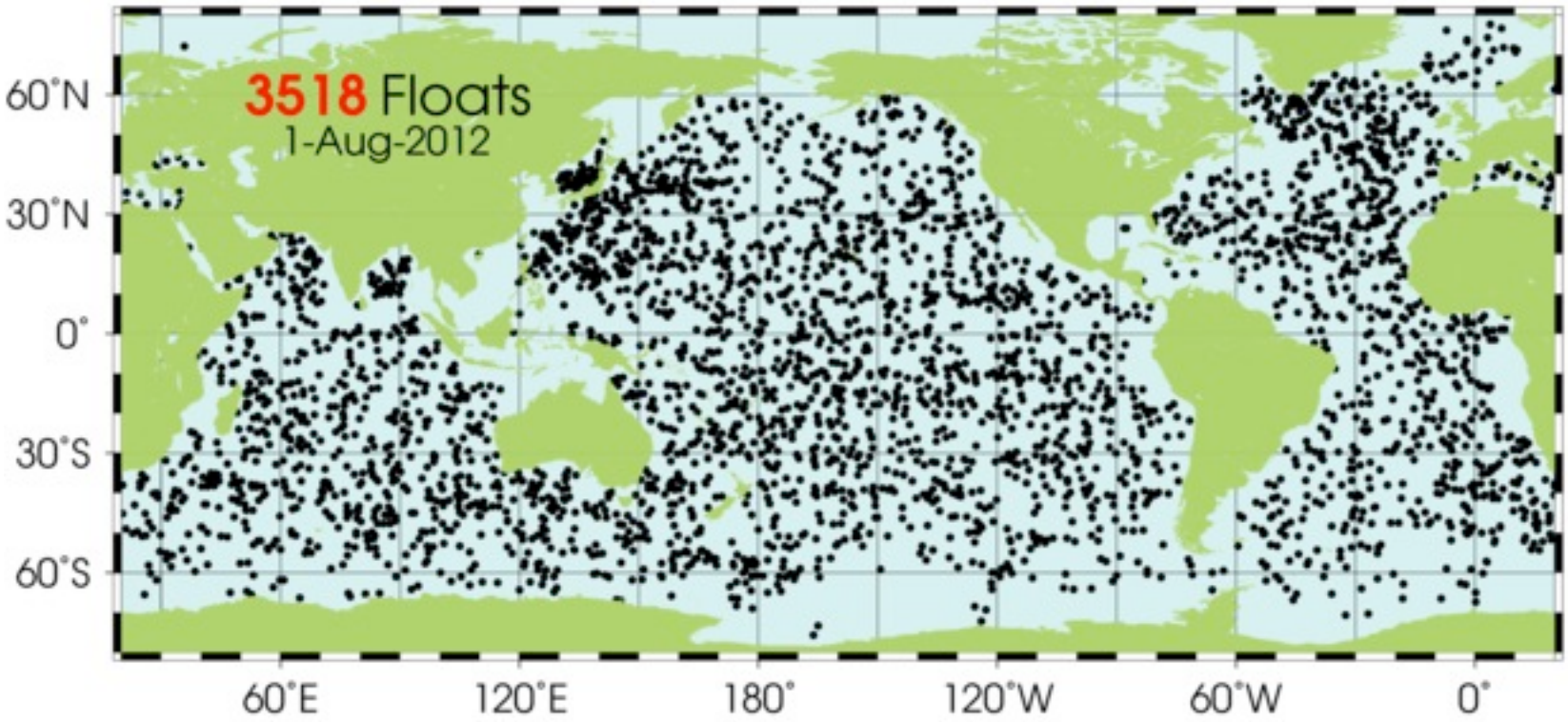


http://www.dosits.org/gallery/tech/ooct/rafos1.htm



http://www.whoi.edu/instruments/

Global Ocean Observing System by ARGO Floats



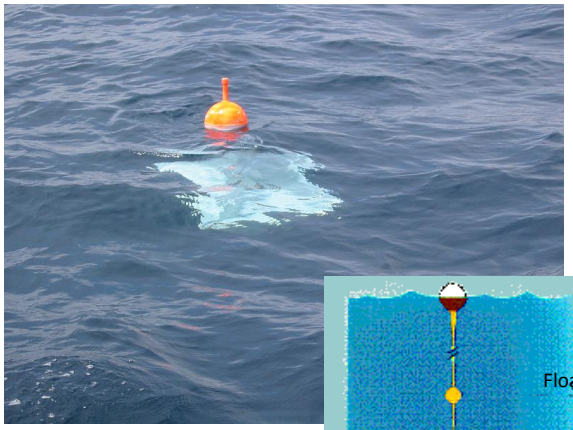
- Argo are used as the platform for continuous observation
  - Eulerian observations of *T, S,* and velocity
- By November 2007, Global observation network by drifters is 100% complete
  - ~3000 profiling at the 5^{o}x5^{o} resolution
  - 800 floats per year to maintain the level

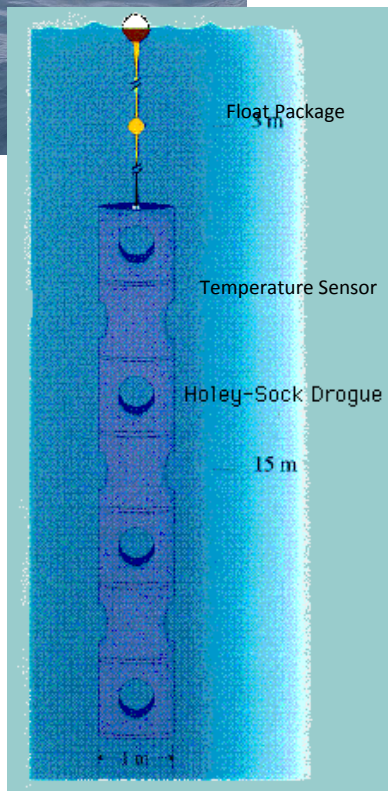
http://www-argo.ucsd.edu

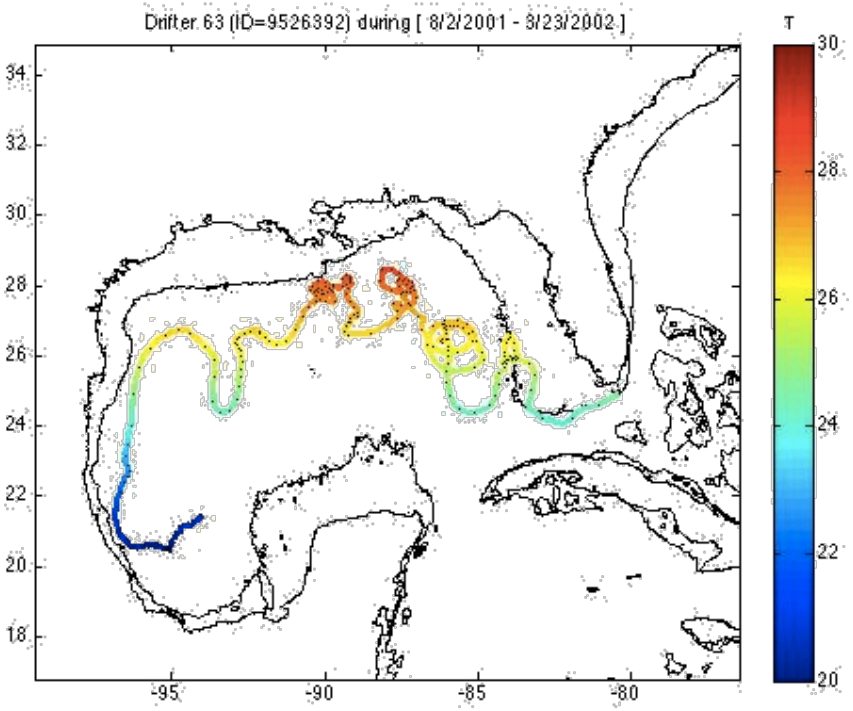
In-Situ Instruments(Movable Lagrangian) Platform. Drifters

- Observations at sea surface
  - T : Temperature

  along ($\mathbf{x}^{(2D)}$)(t_k)) at sea surface



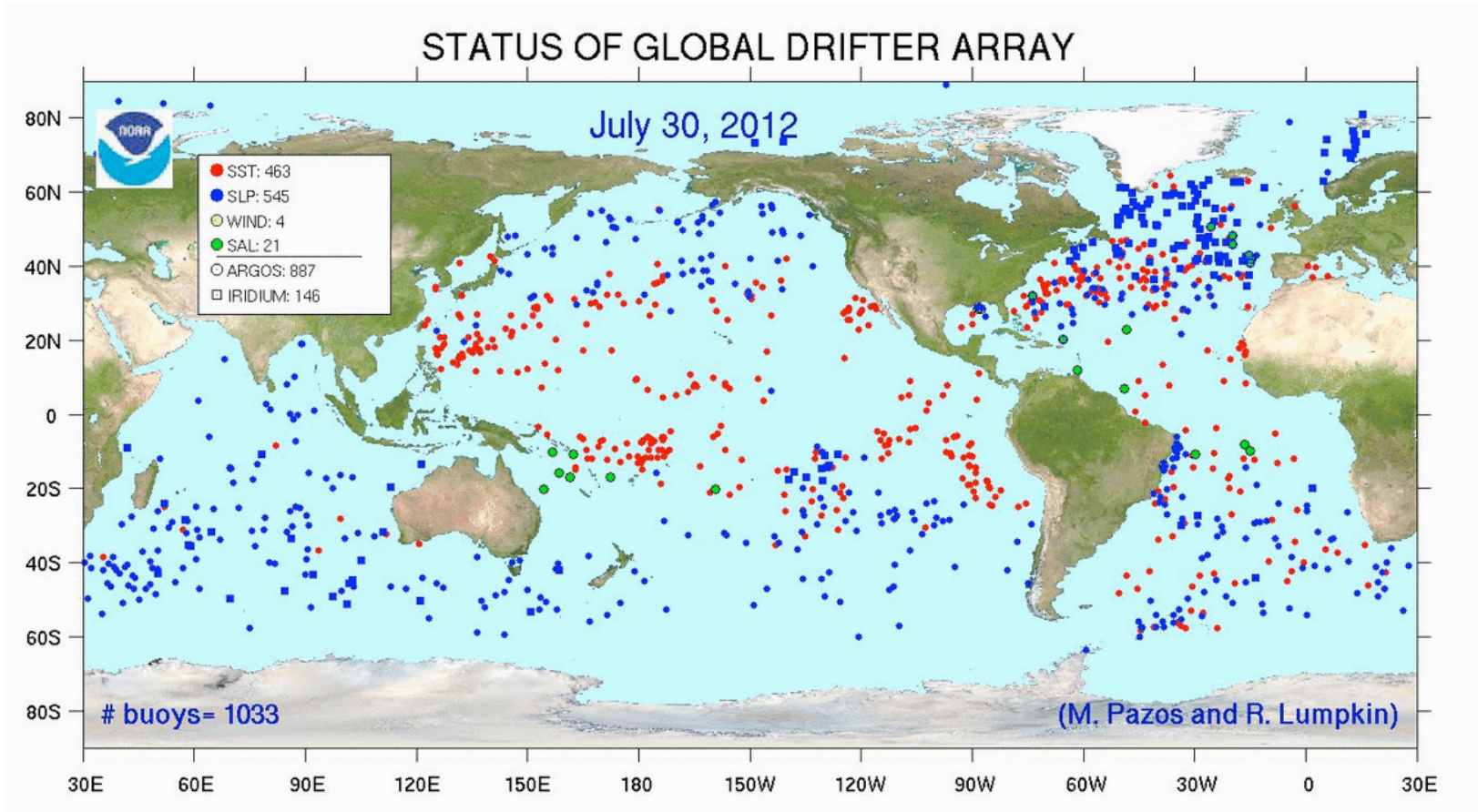




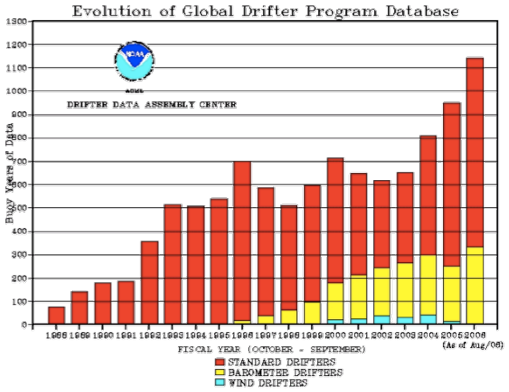
Data available from
http://www.aoml.noaa.gov/phod/dac/dacdata.html

http://www.drifters.doe.gov/design.html

Global Ocean Observing System by Drifters



- Global observation network by drifters
  - ~1200 drifters to cover at the 5°x5° resolution
- Drifters are used as the platform
  - Eulerian observations of *T* (SLP, Wind)



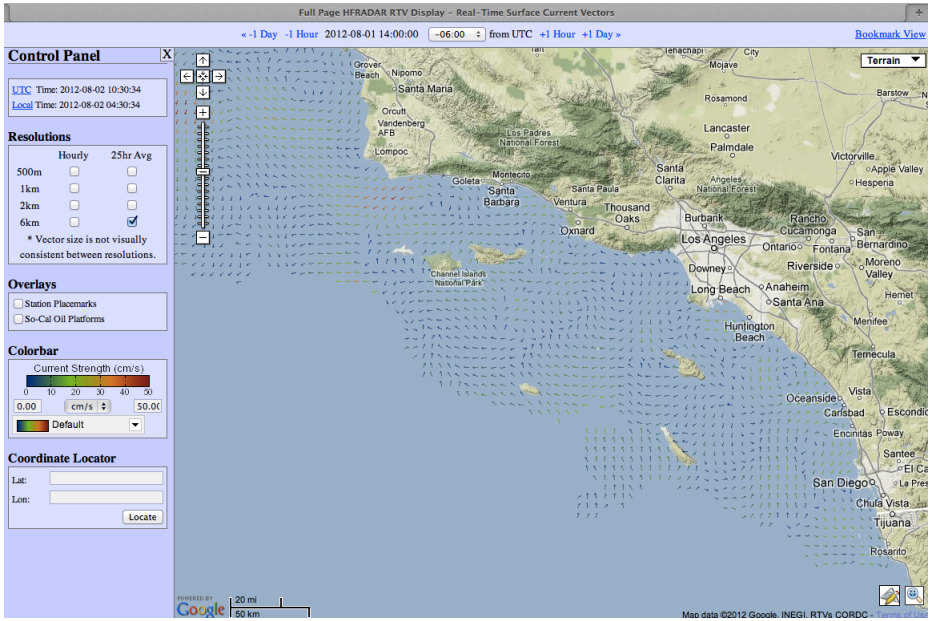
http://www.aoml.noaa.gov/phod/dac/gdp.html

Remote-Sensing by HF Radar Surface Current

- Actual measurement: Radial velocity
  - 2 observations makes 2D velocity

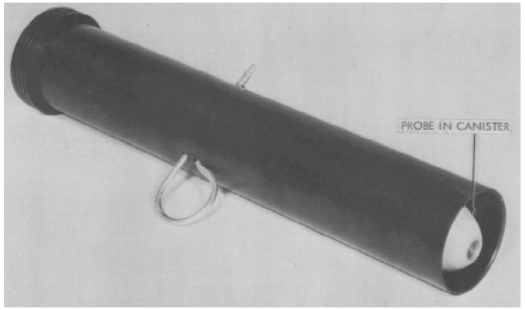
http://www.cocmp.org/

http://www.sccoos.org/data/hfrnet/



In-Situ Instruments. Vertical Profile

- BathyThermographs: Temperature recorders. σ^{o}_{T}~0.1-0.2°C
  - Mechanical (MBTs):
    - Lowered and then winched from the ship down to ~300m.
  - Expendable (XBTs):
    - Dropped from a ship;
    - Designed to fall at a constant rate..
    - Many goes to 460/760m, some goes to 1800m



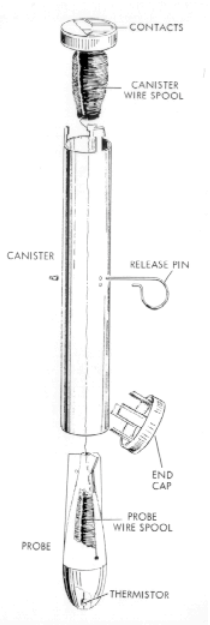
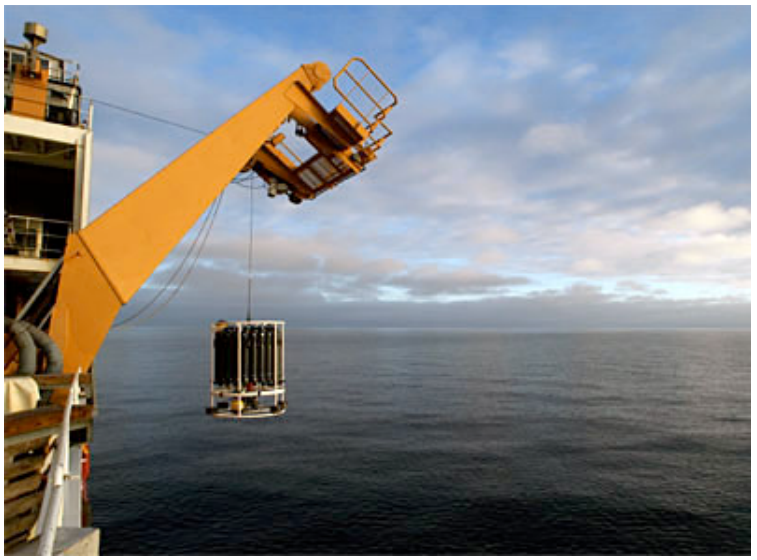
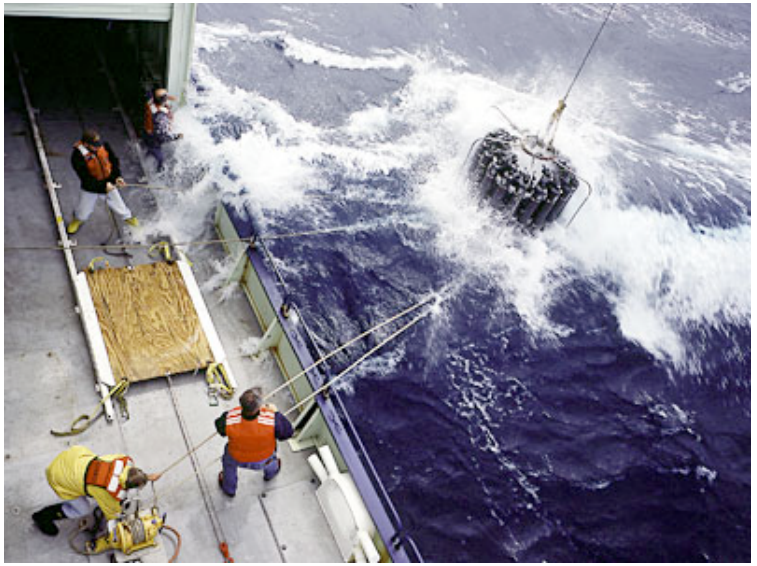


Fig. 1: XBT diagrams: Bathythermograph (probe) and exploded view.

In-Situ Instruments. Vertical Profile

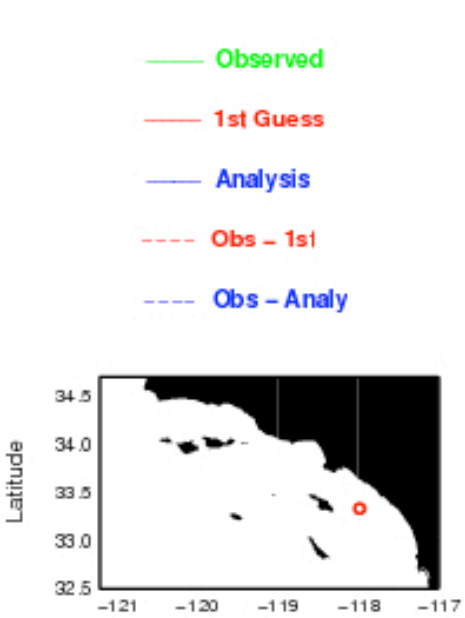
- Conductivity,Temperature, & Depth (CTD): High-quality T/S profile of 150 levels (σ^o_T, σ^o_S)~(0.002°C, 0.005psu)

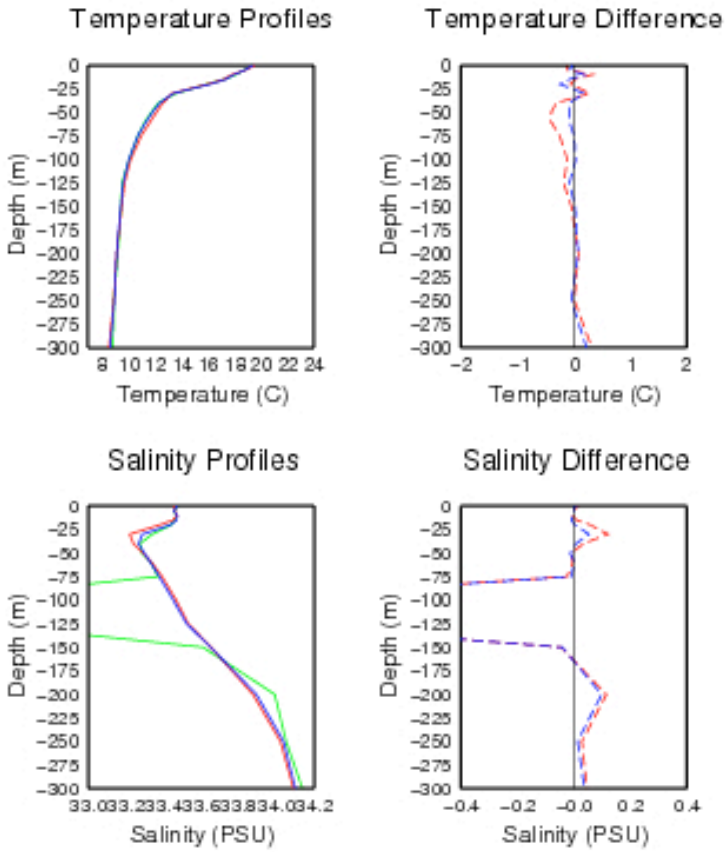




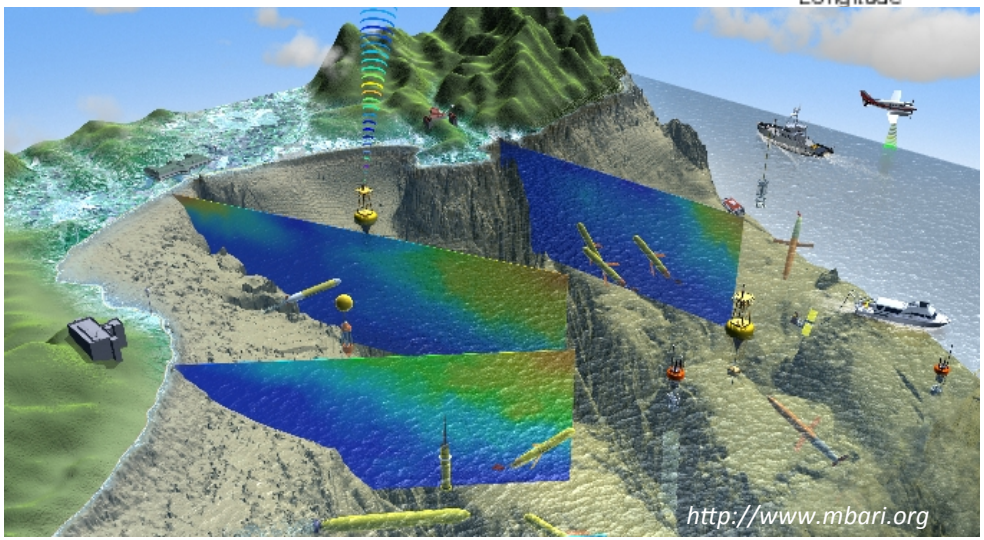
Ship deployed CTD
Woods Hole Oceanographic Institute

In Situ Instruments (Movable Platform): Glider





http://ourocean.jpl.nasa.gov/SCB



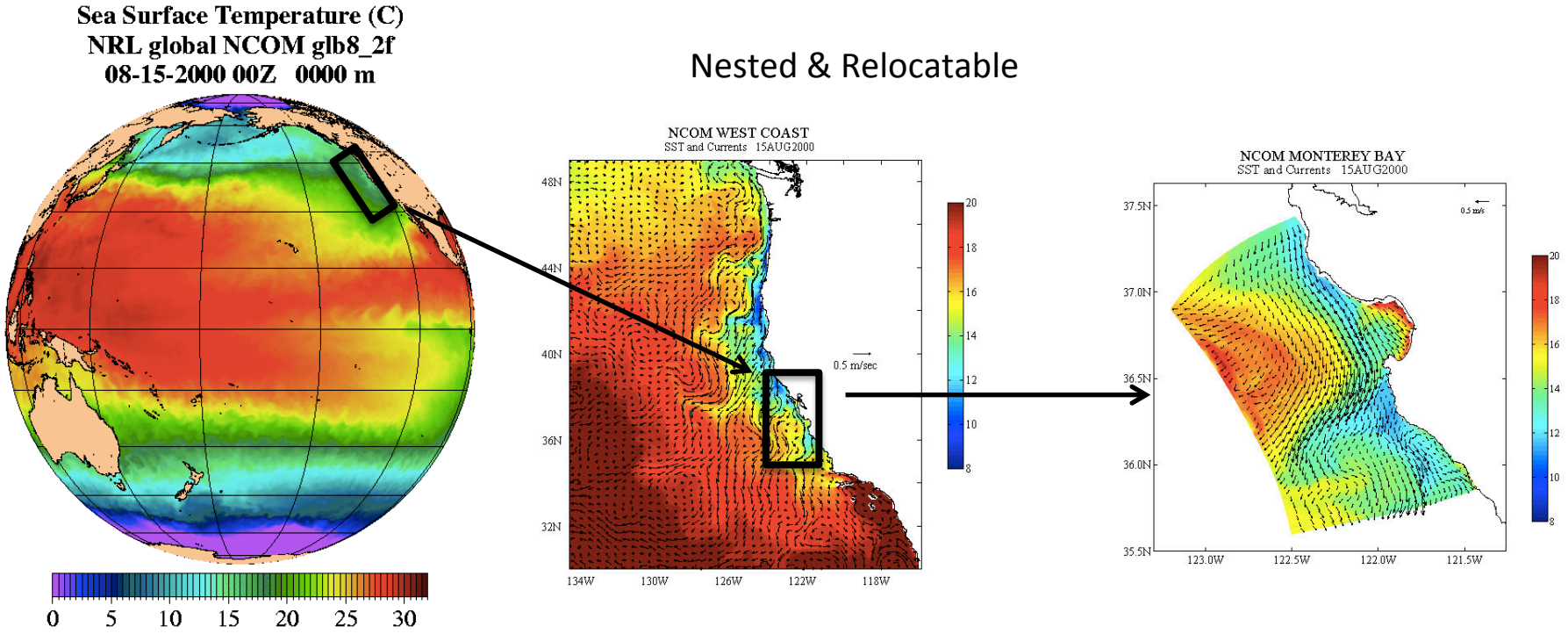
- Glider
  - 2D position, p (depth)
  - T, S

Basic Ideas for Ocean Data Quality Control

- Quality control (QC) for ocean data from data assimilation perspective
  - Basic concepts are similar to atmospheric systems from theoretical (Bayesian) view
    - If observations are abundant, then buddy check (not only to reject but also to accept)
    - If observations are sparse, primary corroborative source of information may be forecast (in particular in the form of ensemble)
  - Details may require unique adjustment, development, and implementation
    - Ocean data (network) can be sparse/inhomogeneous and sporadic
      - » If erroneous values are assimilated, then they can cause immediate spurious overturning and/or be advected around for years afterwards (nothing else to correct)
    - Ocean data assimilation is relatively new and may have different emphases from atmospheric data assimilation.
    - Scales are different in both:
      - » Dynamics
      - » Observations

Navy Coupled Ocean Data Assimilation

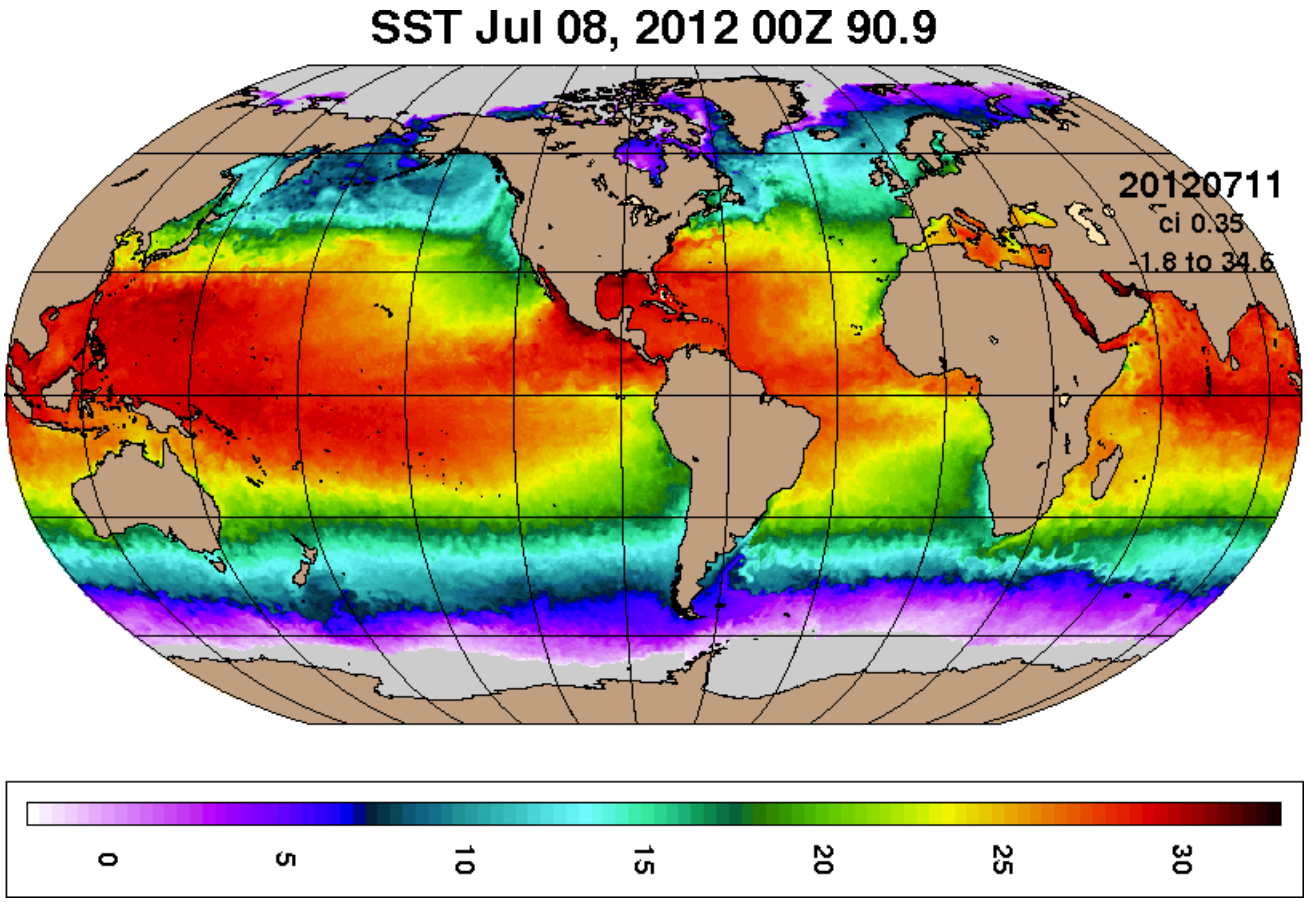
- Philosophy:
  - Approaches to ocean data assimilation vary widely both in terms of the sophistication of the method and the observations assimilated, and also in terms of specification of the forecast error covariances, model biases, observation errors, and quality-control procedures
  - Quality Control should be fully automated and performed in stages



Courtesy of H. Ngodock

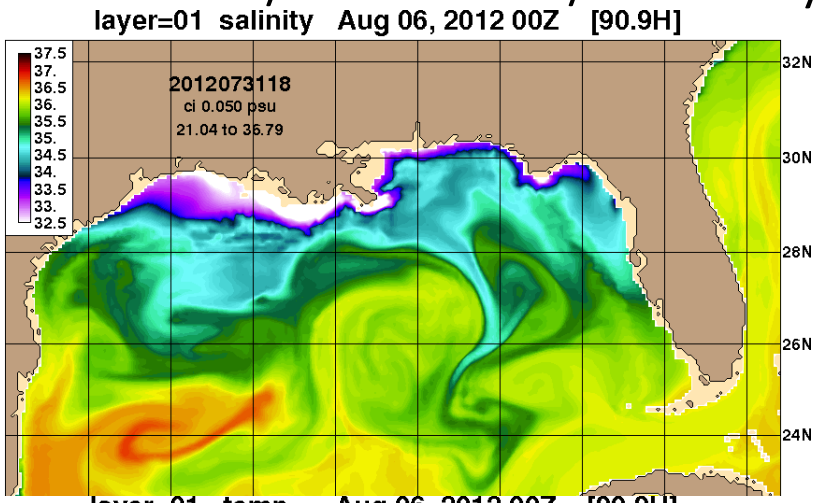
Navy Coupled Ocean Data Assimilation

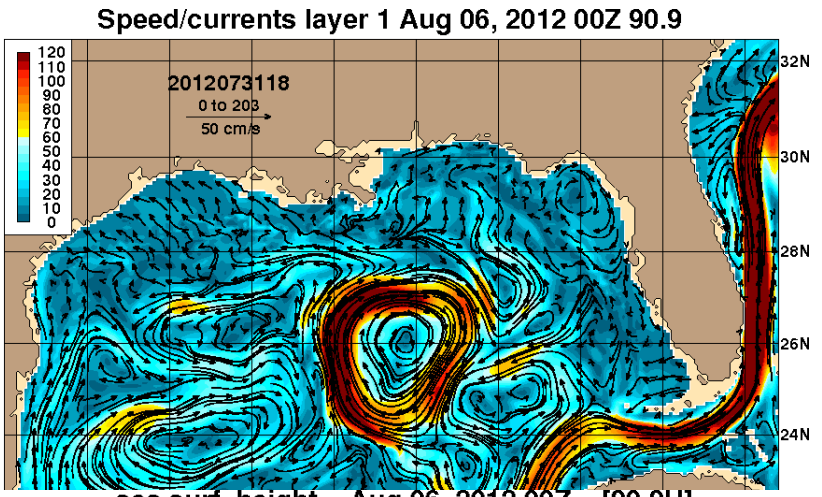
- Real-Time 1/12o Nowcast/forecast system: Global

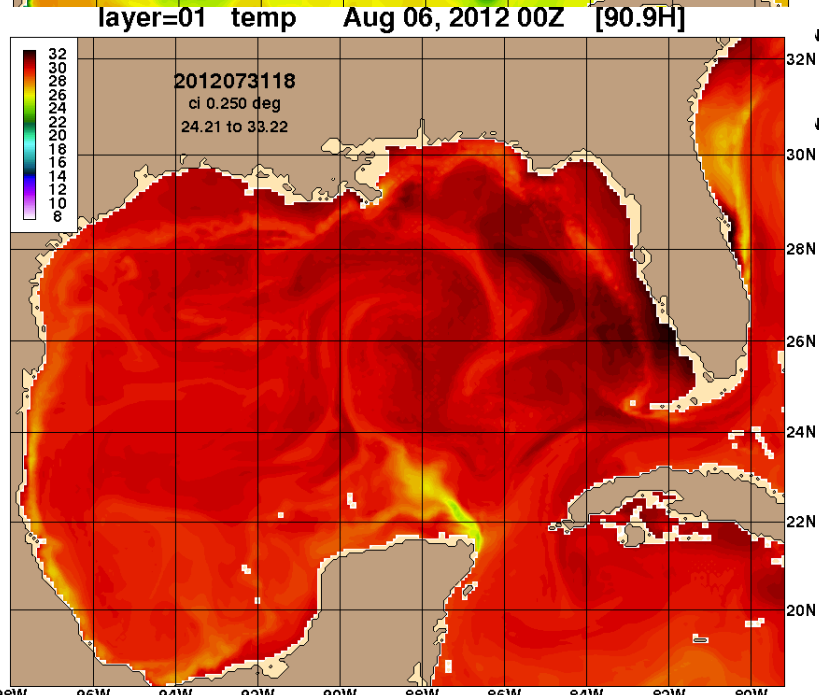


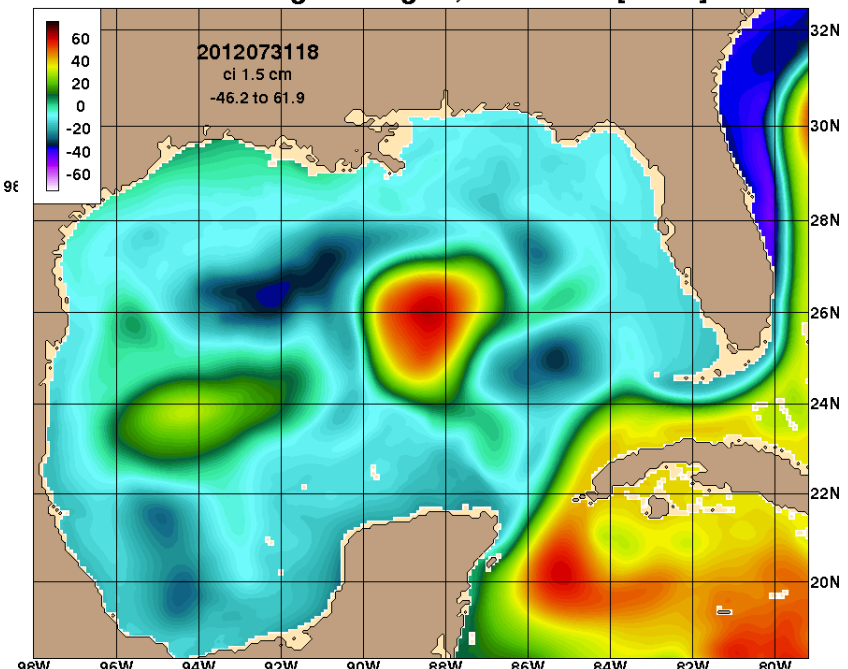
Navy Coupled Ocean Data Assimilation

- Real-Time 1/12o Nowcast/forecast system









Navy Coupled Ocean Data Assimilation

- Quality Control in 3 Stages according to Ingleby and Huddleston (2007)
  - Stage I. Sensibility Check
    - Land/sea boundary checks
    - Location (speed) test for drifters and ship (aircraft) observations
    - Exact and near-dup duplications

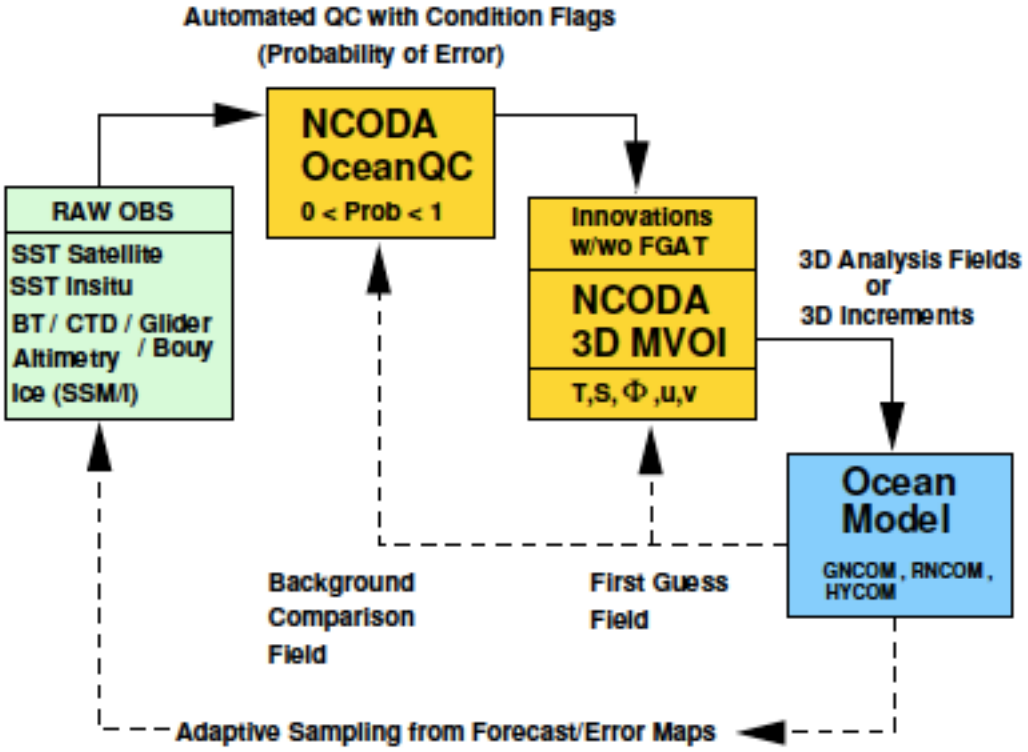
    →*Observations failing sensibility checks are removed*
  - Stage II. Gross error checks
    - Instrumentation error, vertical gradient, static stability checks
    - Cross validation checks (e.g., SST vs sea ice)
    - Background field checks against climate, previous analysis/forecast
    - Bouy (float) sensor drift
    - Satellite SST large scale aersol/dust bias detection

    →*Observations failing sensibiliyt checks are flagged*

    →*All gross error checks performed before decision to accept/reject*
  - Stage III. Consistency checks
    - Innovation error checks performed within analysis

Navy Coupled Ocean Data Assimilation

- NCODA data flow chart



Lunde et al (2009)

ECMWF ORA-S4 (Ocean Re-Analysis System 4)

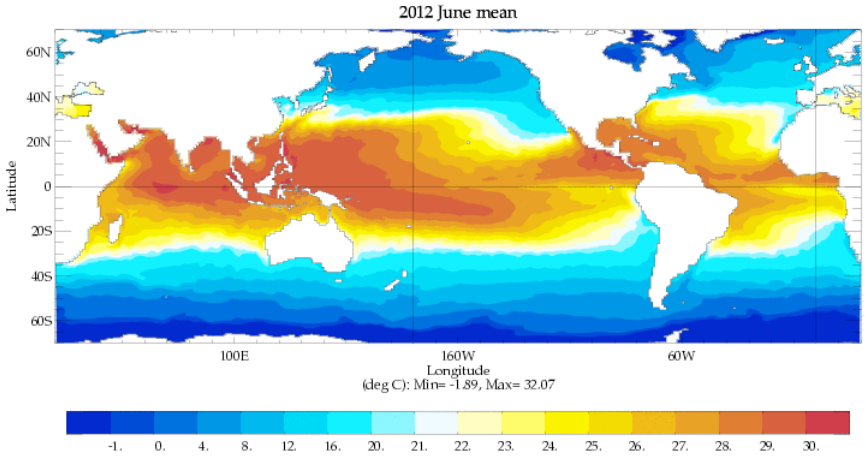
- Global ocean near real-time analysis & reanalysis since 2012
  - Daily starting from September 1957 & continuously maintained up to 10 days behind real time at 1^{o}x1^{o}x42 level resolution
  - Main purposes: to provide
    - Initial conditions for seasonal/monthly forecasts
    - Historical representation of ocean for climate studies
    - Uncertainty estimates by ensemble (5 total)
  - Featuring
    - Online bias-correction algorithm
    - Assimilation of salinity data
    - Assimilation of altimeter-derived sea level anomalies
  - Designed to reduce spurious climate variability due to observing system change while taking advantages of the new observations
- Previous systems
  - System 1 (ORA-S1) starting 1997
  - System 2 (ORA-S2) introduced in 2000
  - Syetsm 3 (ORA-S3) introduced in 2006

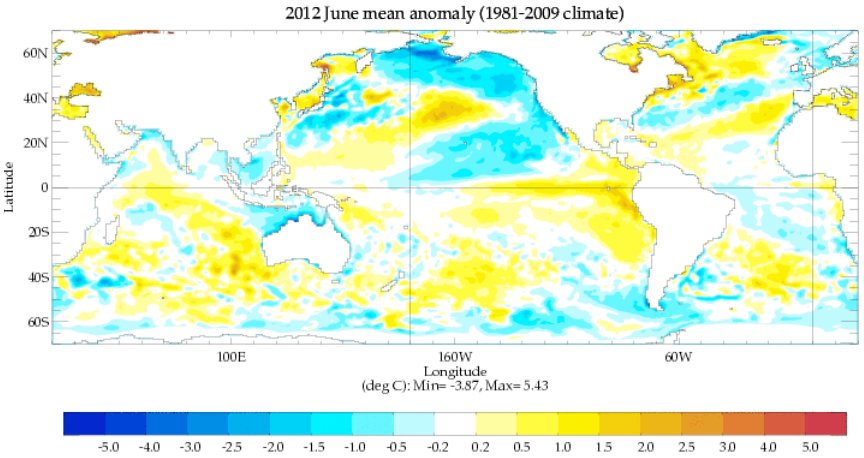
http://www.ecmwf.int/products/forecasts/d/charts/ocean/

Balmaeseda, Vidard, Anderson (2008)

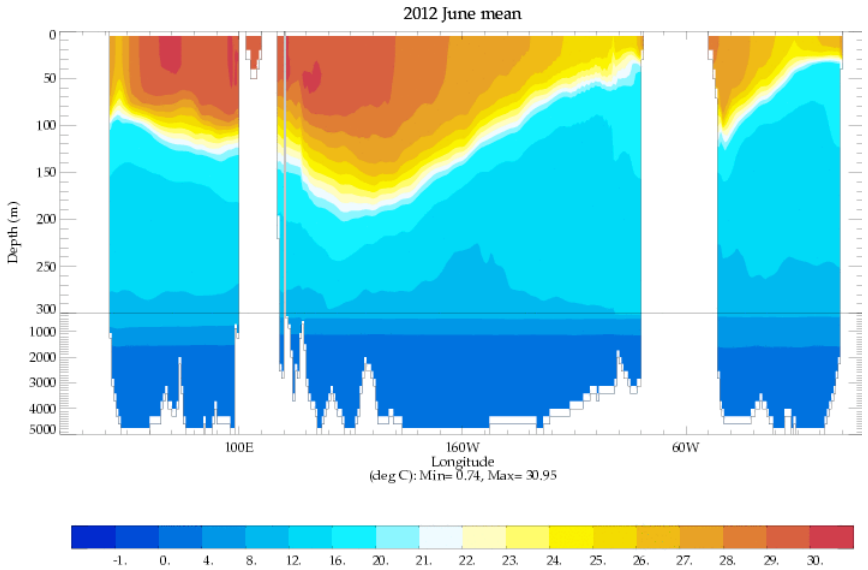
ORA-S4 : 3D evolution of ocean state

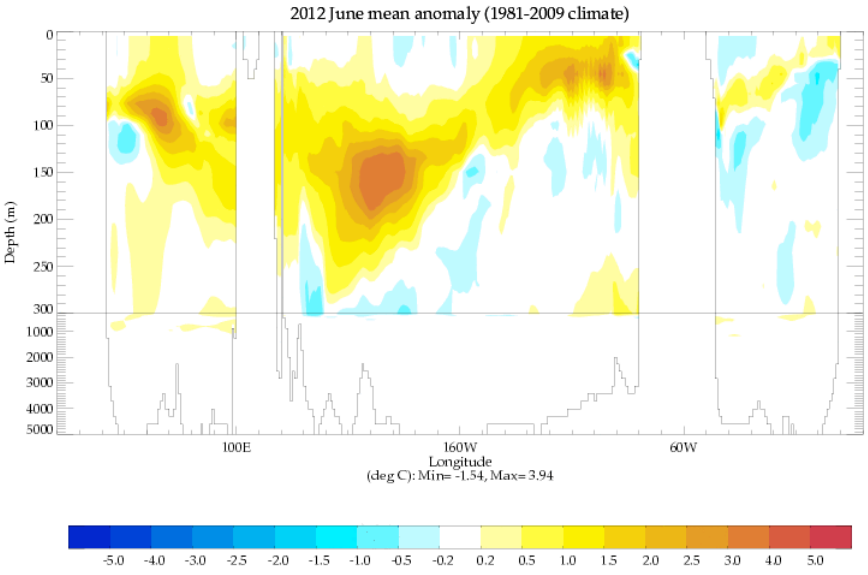
- Sea Surface Temperature





- Cross-Section at the Equator

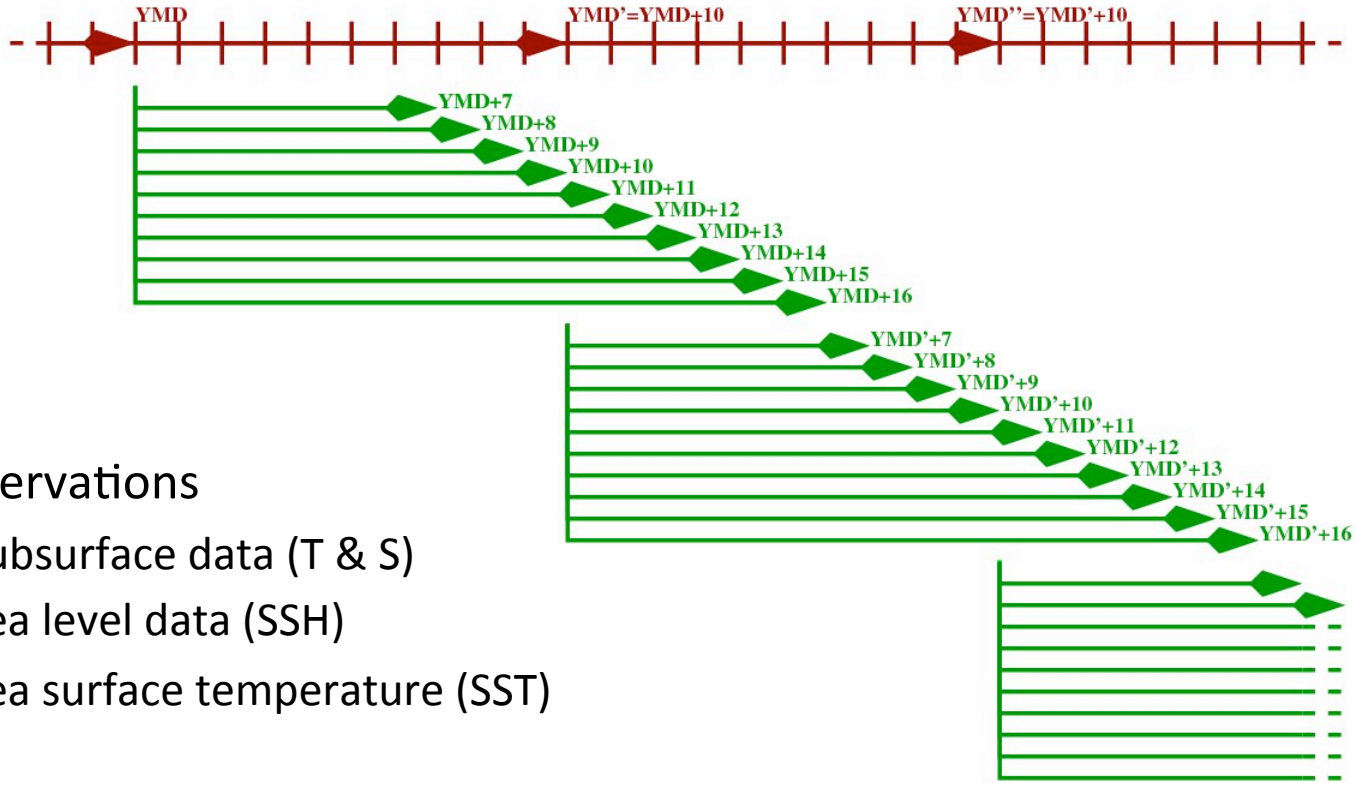




http://www.ecmwf.int/products/forecasts/d/charts/oras4/reanalysis

ECMWF ORAS4 & ORTS4 (Ocean Re-Analysis & Real Time)

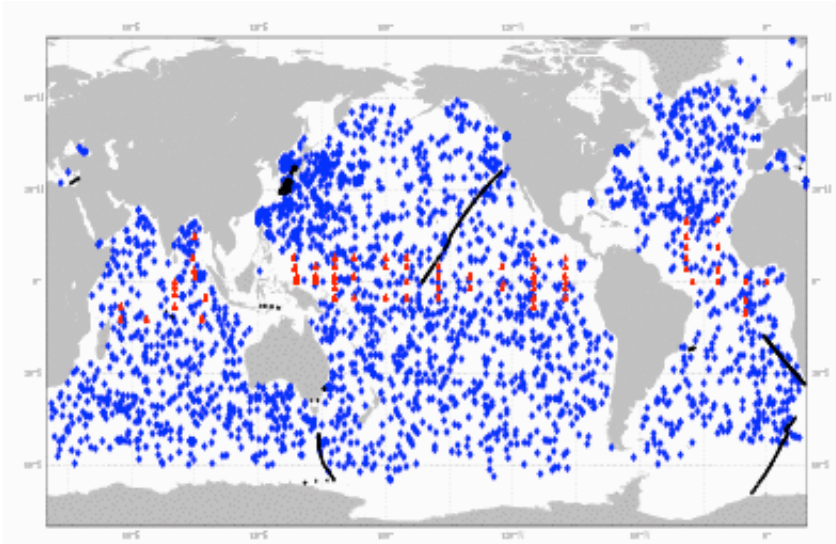
- Operational schedule for the production of ORAS4 & ORTS4
  - Reanalsysis: every 10 days, with a delay of 6 days to wait for the arrival of ocean observations (in particular, retrieval of the sea-level products)
  - Real-time: to create real-time initial condition on a daily basis, ORTS4 brings the latest ORAS4 state up to real time every day, using available observations in a variable assimilation windows

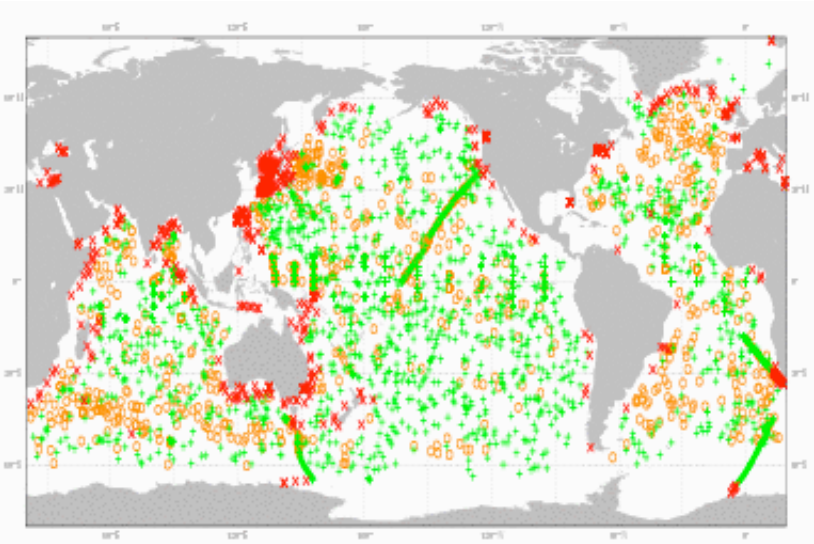


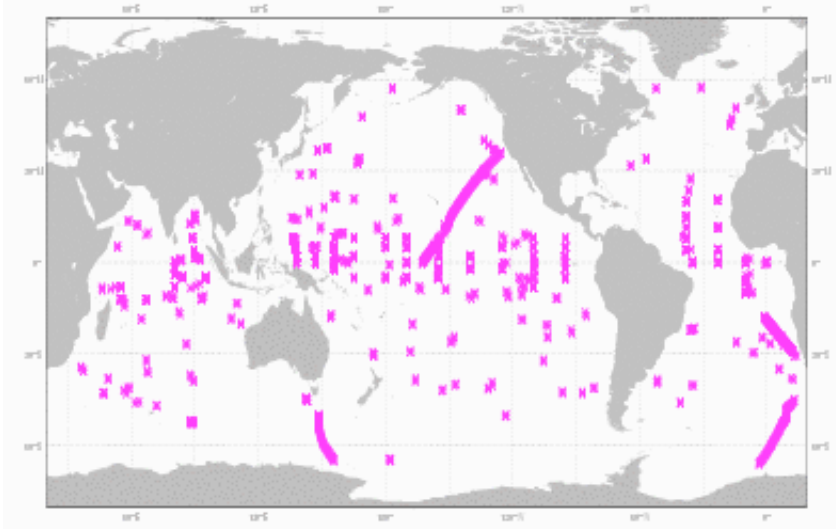
- Observations
  - Subsurface data (T & S)
  - Sea level data (SSH)
  - Sea surface temperature (SST)

http://www.ecmwf.int/products/forecasts/ocean/oras4_documentation/Streams.html

ECMWF ORAS4 & ORTS4: Data Coverage (Subsurface T)







XBT probes: 500 profiles

Argo floats: 2719 profiles

Moorings: 891 profiles

Partially Accepted: 947 profiles

Fully Accepted: 2398 profiles

Fully Rejected: 765 profiles

SuperObs: 1566 profiles

(at least one per profile)

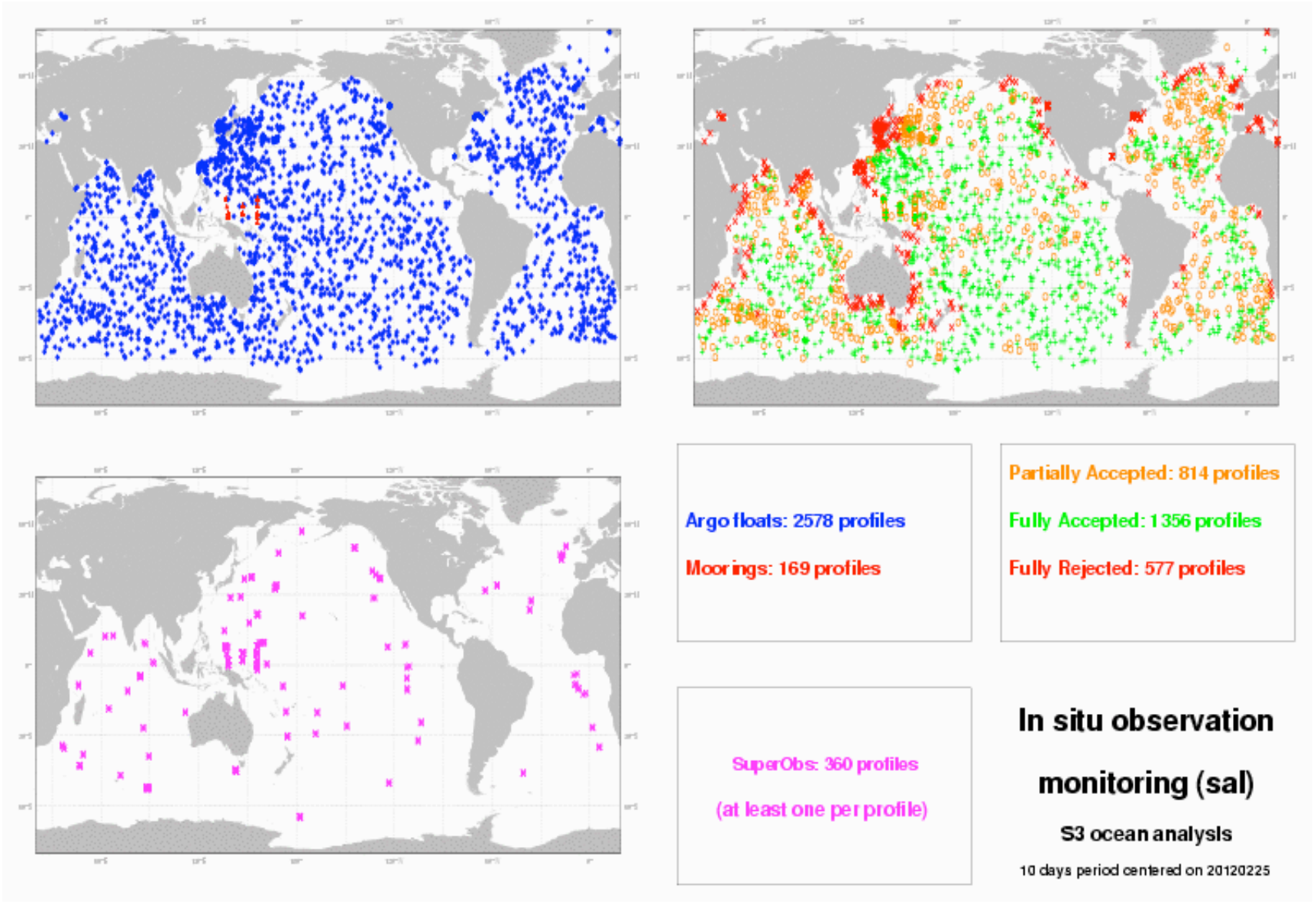
In situ observation monitoring (temp)

S3 ocean analysis

10 days period centered on 20120225

http://www.ecmwf.int/products/forecasts/d/charts/ocean/real_time/obsmap!20120225!Temperature!/

ECMWF ORAS4 & ORTS4: Data Coverage (Subsurface T)



http://www.ecmwf.int/products/forecasts/d/charts/ocean/real_time/obsmap!20120225!Salinity/

Ocean Observation and Quality Control for ORA & ORT-S4

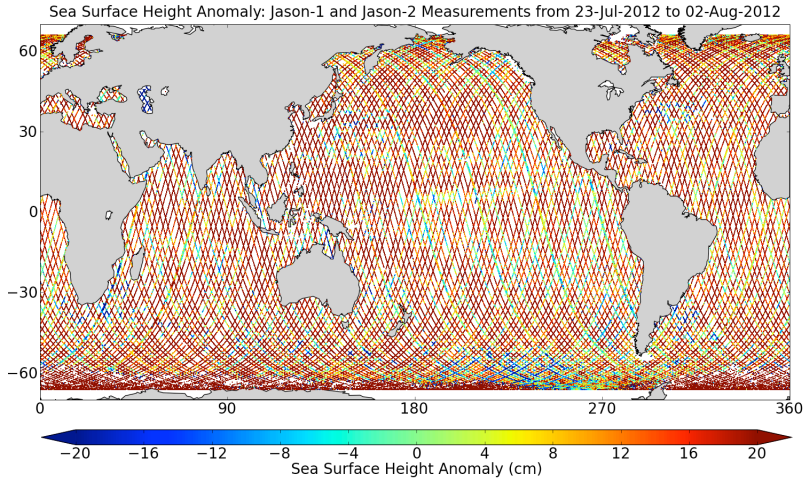
- Automatic Quality Control for subsurface data
  1. *Daily averaging.* If some site reports more frequently than once per day, daily averages are are created.
  2. *Blacklist of coastal observations.* Data in the vicinity of the coast are rejected, as a way of accounting for representativeness error.
  3. *Background check.* A level-by-level check between the distance between model values and observations in relation to the error statistics.
  4. *Buddy check.* A consistency test between observations is performed.
  5. *Super-obbing*. Profiles which are close in space and time are superobbed
  6. *Completness of profiles*. A profile is considered incomplete, and therefore rejected, if the sparsity of the remaining observations in the vertical is judged insufficient to resolve the vertical temperature gradients. (Ex: An observation profile will be rejected if the temperature difference between consecutive levels is larger than 5 deg C or if it contains a vertical temperature gradient larger than 0.1 deg C/m).

Extension of Smith N., J Blomley, and G Meyers, Prog in Oceanography,1990.

Satellite Data Quality Control at ECMWF

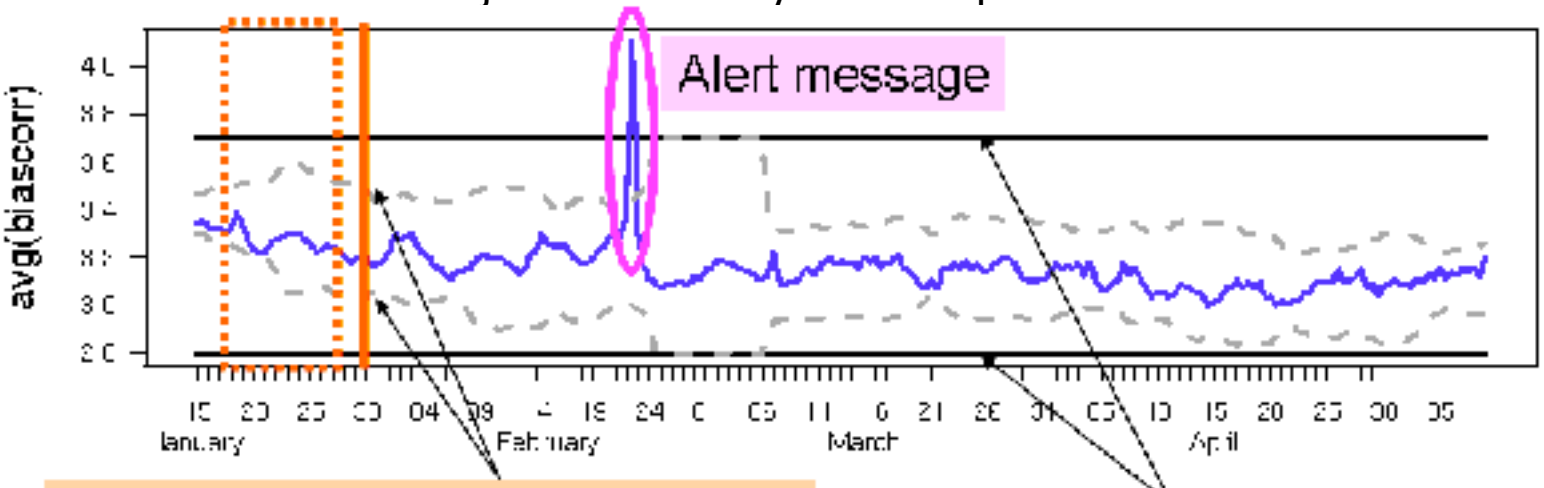
- Automatic checking – Experimental
  - Recently implemented at ECMWF
  - It triggers alarm messages if an anomaly is detected in the quality or the availability of the satellite data assimilated by the model.
  - Selected statistical parameters
    - number of observations
    - bias correction
    - mean bias-corrected background and analysis departures

    are checked against an expected range.
  - Currently, the automatic checking is limited to data passing through the minimization process.



Satellite Data Quality Control at ECMWF

- Procedure
  - An appropriate alert message (including a time series plot) is generated if statistics are outside the specified ranges.
  - A severity level (slight, considerable, severe) is assigned to each message depending on how far statistics are from the expected values.
  - Two kinds of ranges are used by the automatic checking:
    - Soft limits are updated automatically using statistics from the last twenty days
    - Hard limits are adjusted manually when required.



http://www.ecmwf.int/products/forecasts/satellite_check//do/get/satcheck/2908

Satellite Data Quali

- Operational Checking

AIRS

AQUA AIRS 2097 clear radiances

(6 times in last 10 days for

airs_784_11_2097_210.png

Slightly: avg(biascorr

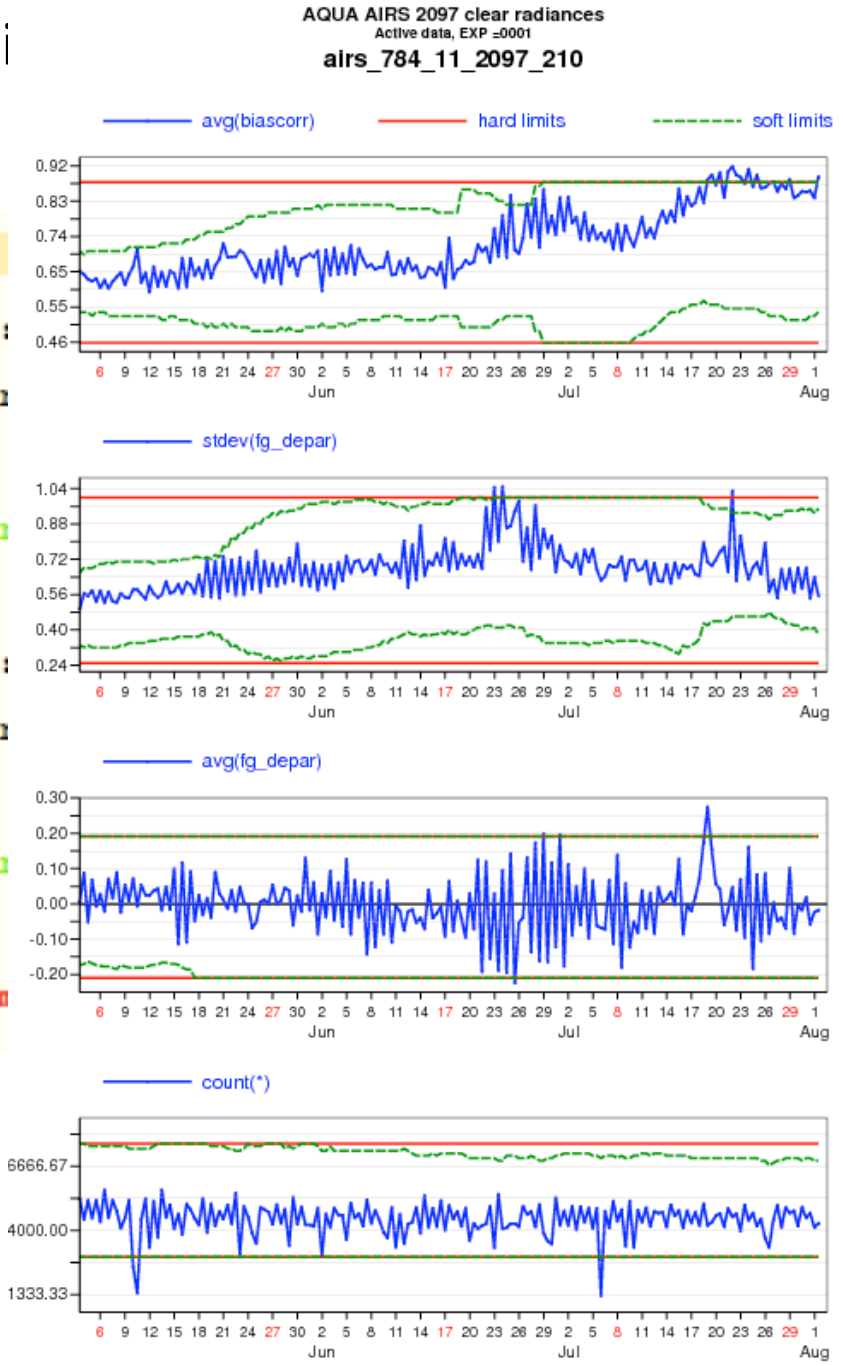
AQUA AIRS 2098 clear radiances

(7 times in last 10 days for

airs_784_11_2098_210.png

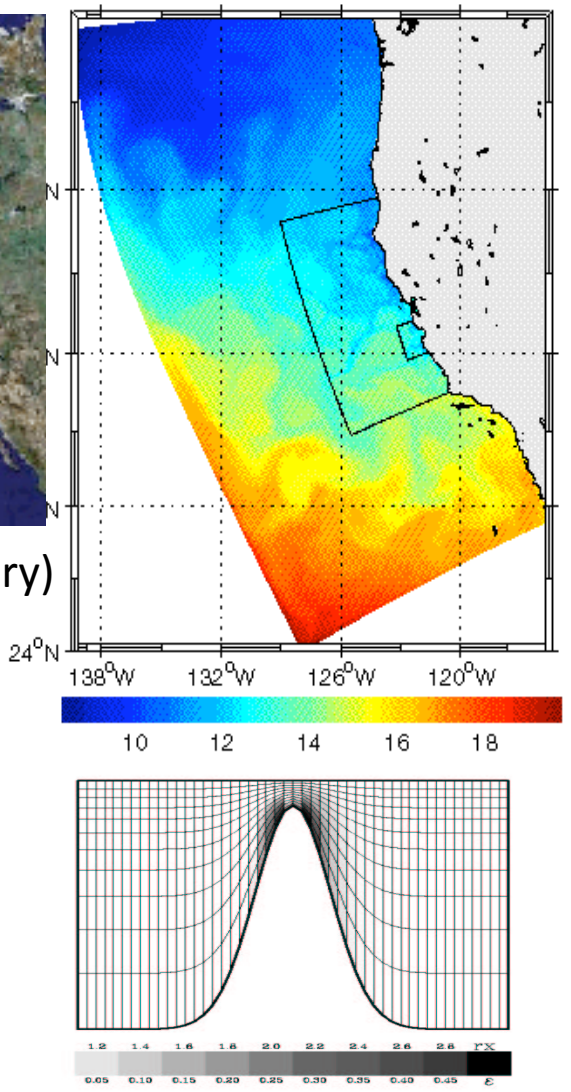
Slightly: avg(biascorr

AQUA AIRS 1119 clear radiances



ROMS (Regional Ocean Modeling System) 3D-Var

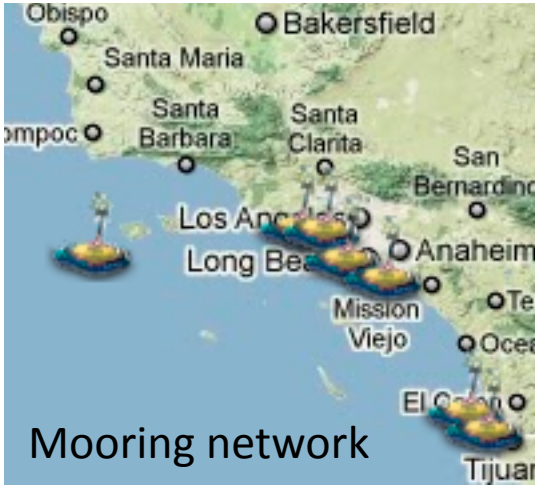
- Regional Ocean Modeling System (ROMS)
  - One-way nested configuration
    - Pacific basin for largest domain
    - Nested coastal configuration
      - » 15km-5km-1.5km-0.5km
      - » Vertical levels >24
  - Relocatable, in the future
  - Forcing by COAMPS (Coupled Ocean/Atmosphere Mesoscale Prediction System by Naval Research Laboratory)



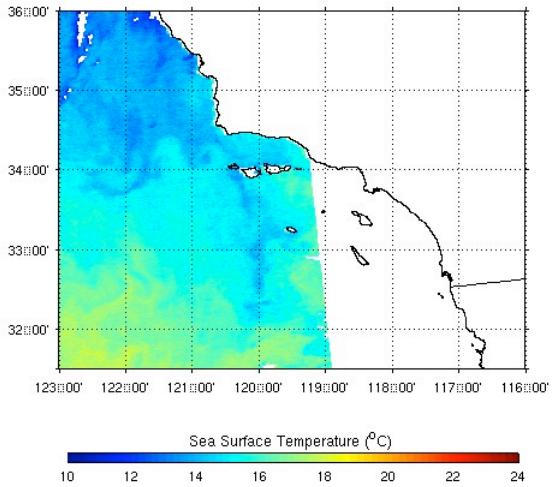
http://ourocean.jpl.nasa.gov/

Costal Ocean Observation Network

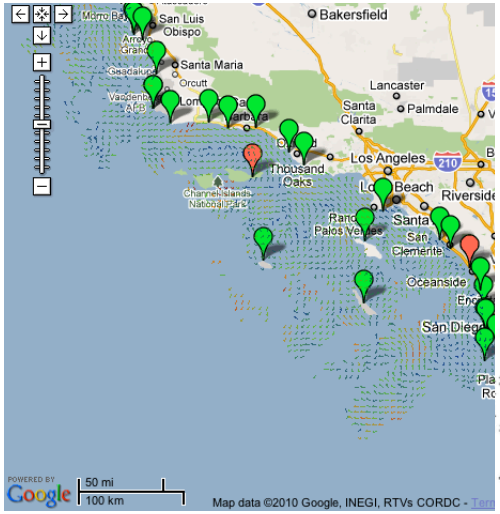
- Spatial distribution of observing network is highly inhomogeneous
  - Satellite images (SST) can be as high-resolution as the model state in horizontal
  - HR radar (surface velocity) can be highly concentrated and high-resolution
  - Others can be extremely sparse

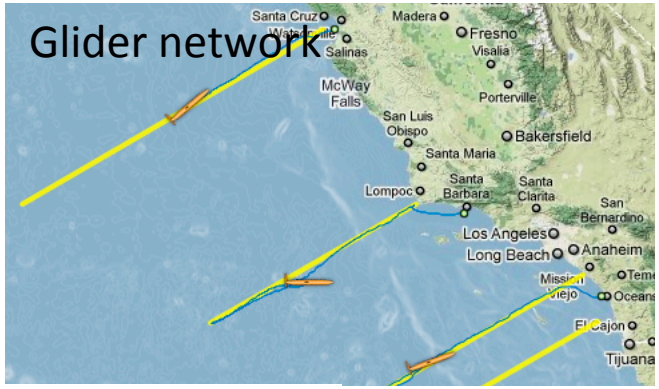


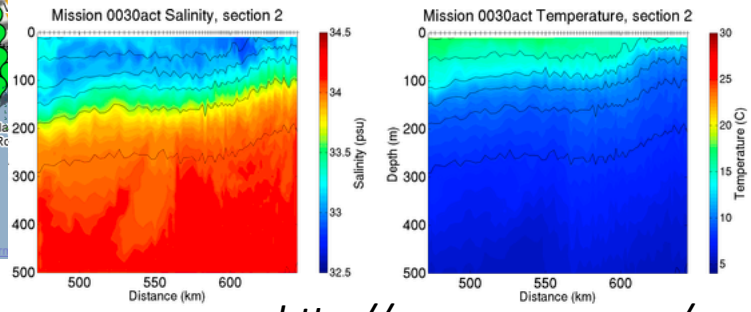
SST on a lucky day



HR radar network



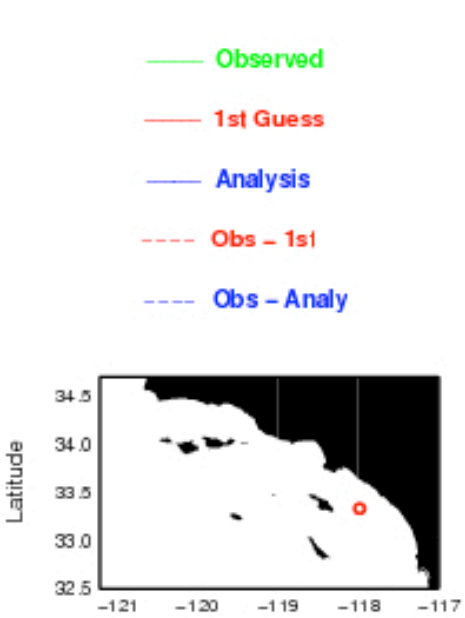


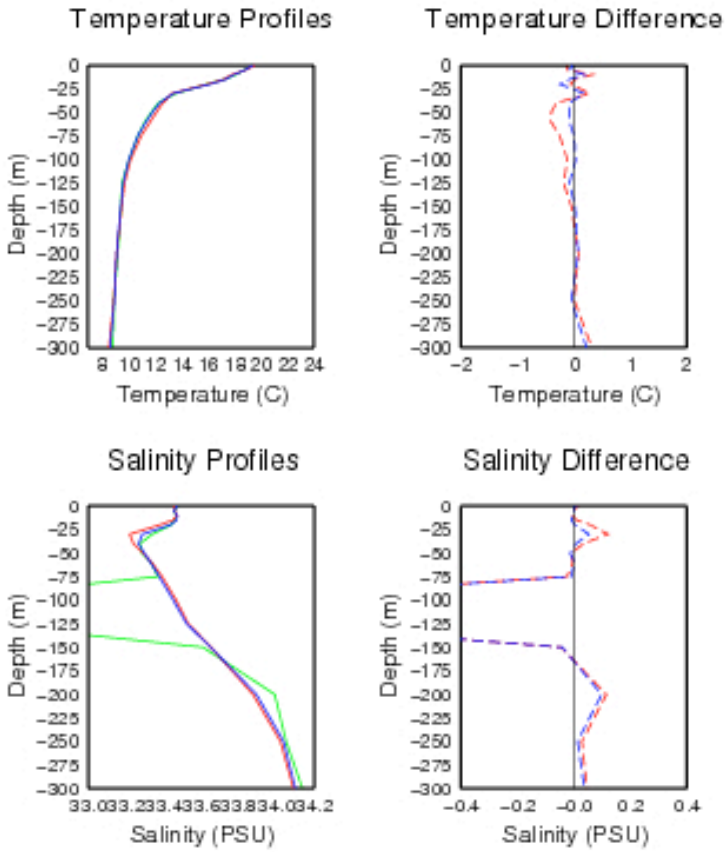


http://ouroceans.jpl.nasa.gov

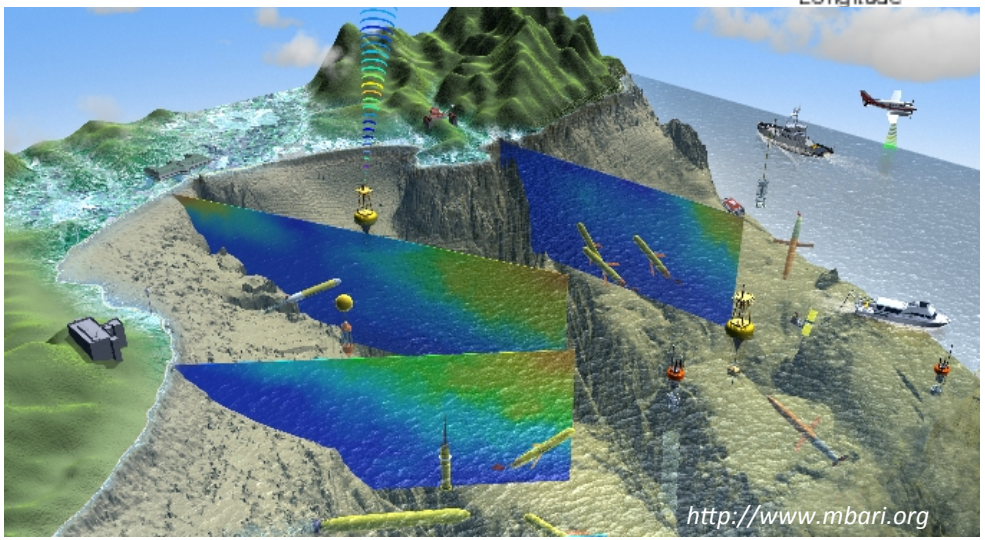
http://www.sccoos.org/

In Situ Instruments (Movable Platform): Glider





http://ourocean.jpl.nasa.gov/SCB



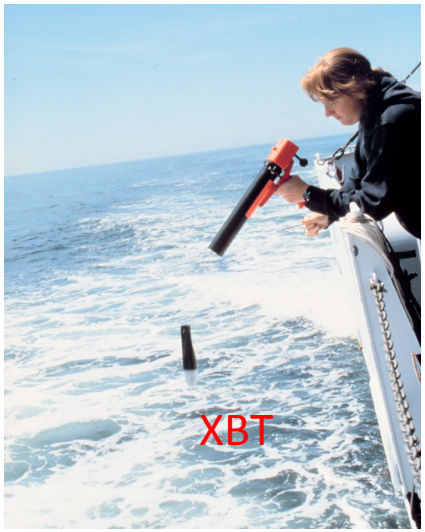
- Glider
  - 2D position, p (depth)
  - T, S

Quality Control of In-Situ T/S

- Developed by Ingleby & Huddleston (2007) at UKMO
- Principles for building the QC system
  - The system has to be automated to cope with the data volume involved
  - Original, reported values should be used as long as possible (flagged, rather than rejected)
  - Any decisions taken by the system should be traceable
  - The system is designed to support data assimilation
  - Tools to monitor system performance and individual cases are available
  - The generic checks and processing use code shared the UKMO atmospheric QC
  - The generic checks have a clear theoretical basis in probability theory
- QC Main Steps
  - Data specific check
  - Background and buddy check

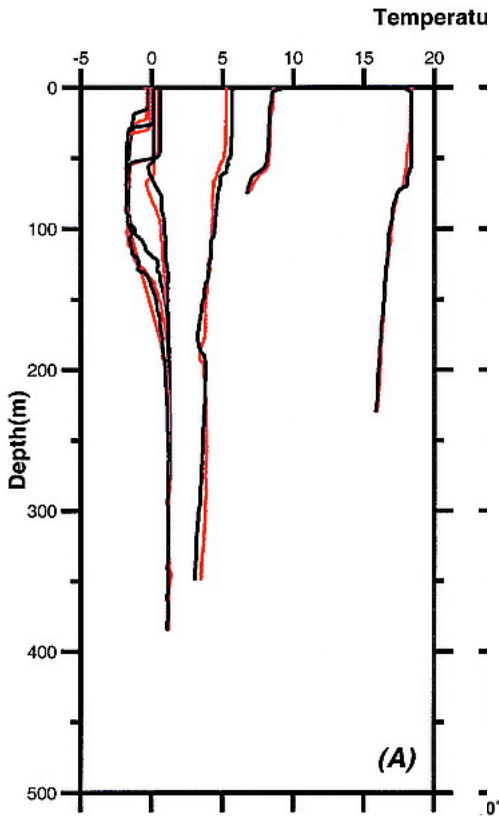
Pre-QC for XBT

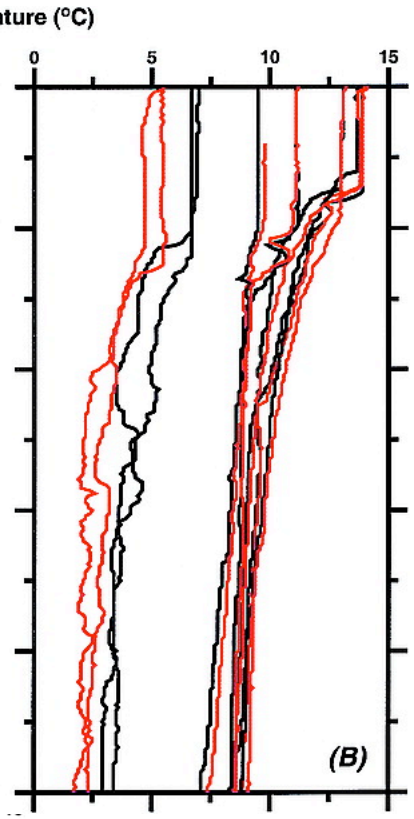
- Prior rejects for XBT below 1000m due to inaccuracy of the instruments.
- XBT depth correction: XBT depth is computed based on the time of the release.
  - All XBTs are designed to fall at the known rate, according to the manufacturer's design (formula, or equation for depth vs time from the release).
  - Many won't; revised equation (linear correction) is suggested by Hanawa et al (1995).
  - Depth for the profiles taken before Hanawa et al (1995) were corrected.
  - Additional difficulties:
    - Older data may not have record for type of XBT used
    - In the cold sea, viscosity is higher & drop rate changes.



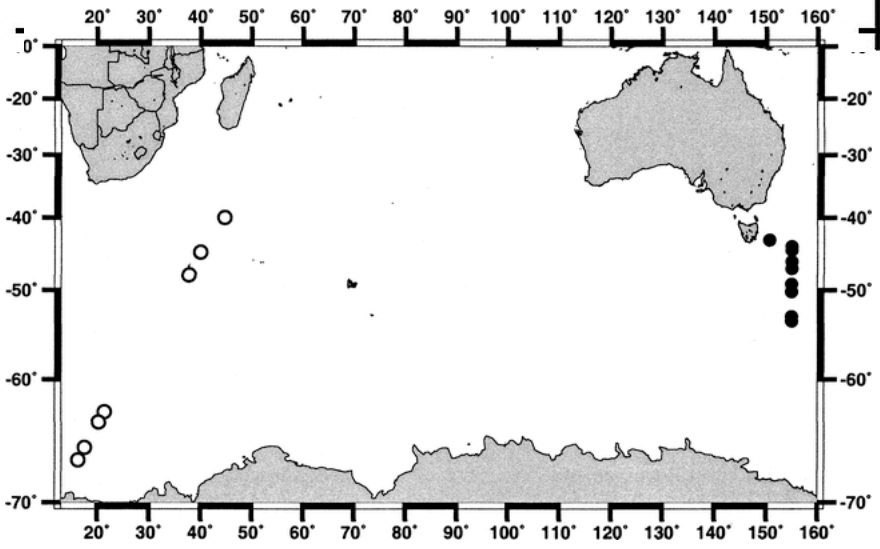
validated against

In Situ Data: XBT vs CTD





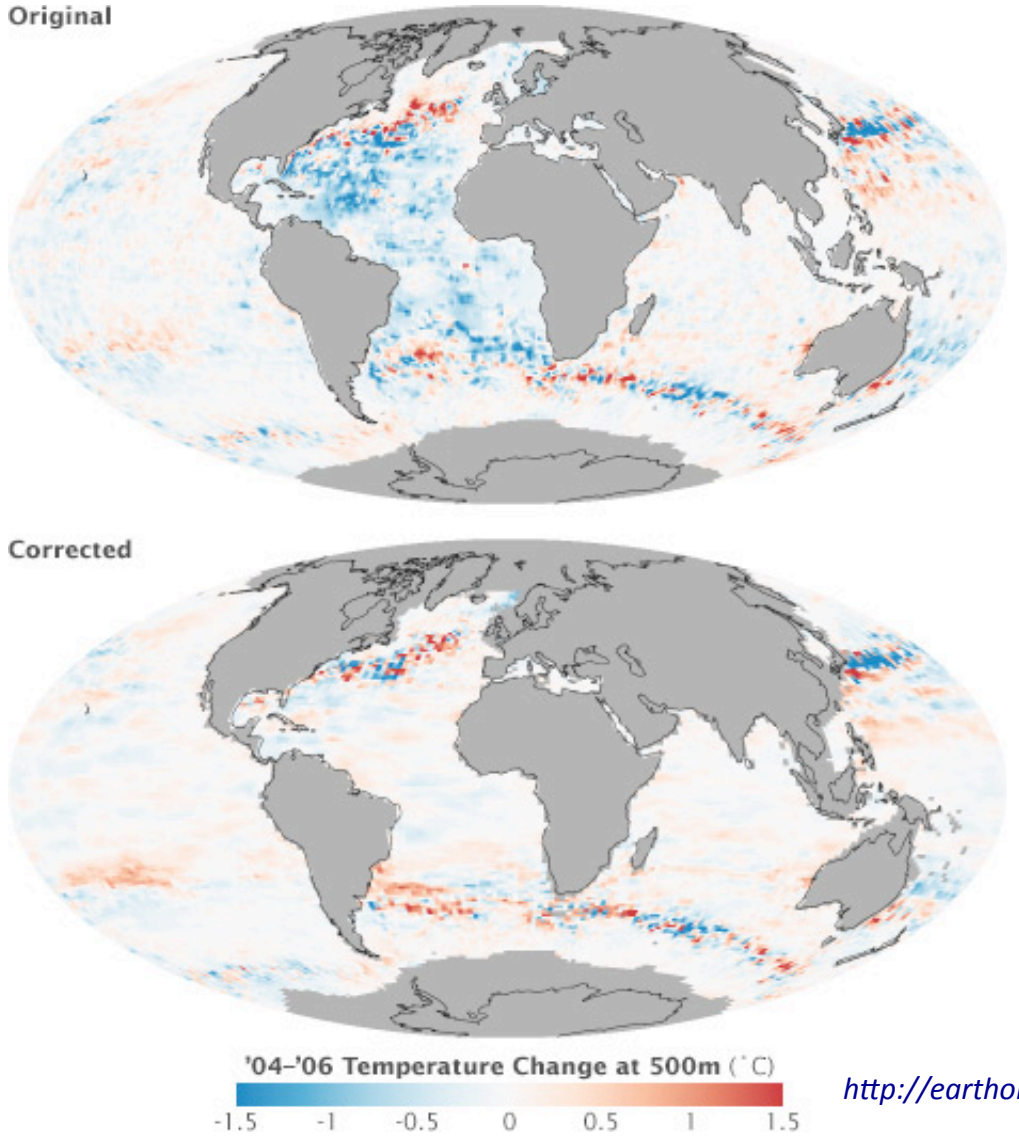
Composite profiles of
- XBT (red line) and
- CTD (black line)



Thadathil et al (2002)

Importance of QC / Bias Correction

- Ocean temperature change from 2004 to 2006



Correction in temperature due to bad data from the Argo floats and XBTs

http://earthobservatory.nasa.gov/Features/OceanCooling/page3.php

QC for Data Specific Check

- Movable instruments
  - Vertical check for the value of data: against constant, spike or step.
  - Horizontal position of the data along the track for each identifier: kinks or jumps.

- Superobs for mooring in time: TRITON are quasi-hourly but formed into daily averages.

- Stability check for T/S based on the density for vertical profiles.

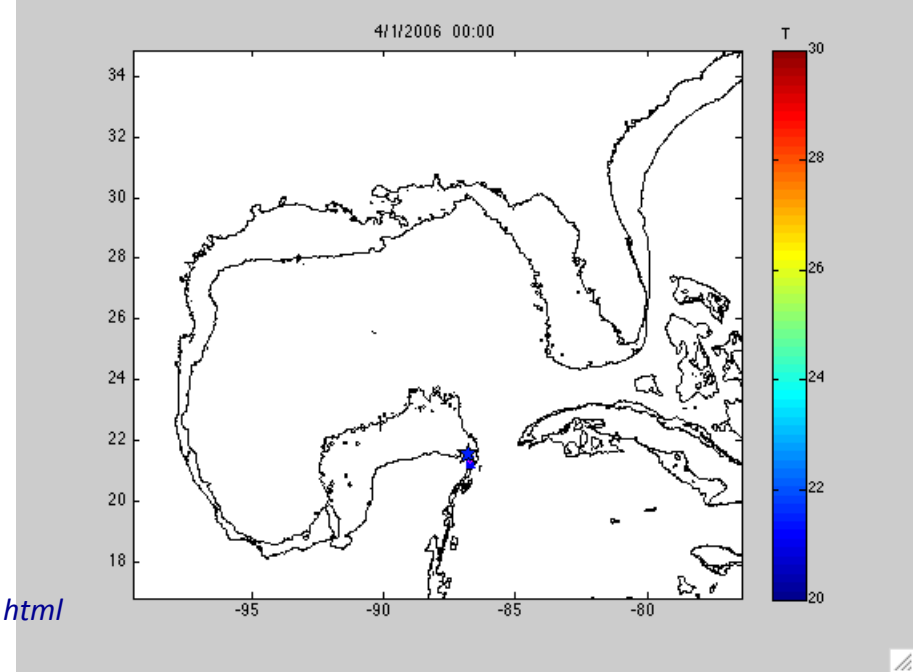
- Duplicate check and thinning in space.

Background and Observation Buddy Checks are Similar to Atmos QC

New Types of Observations

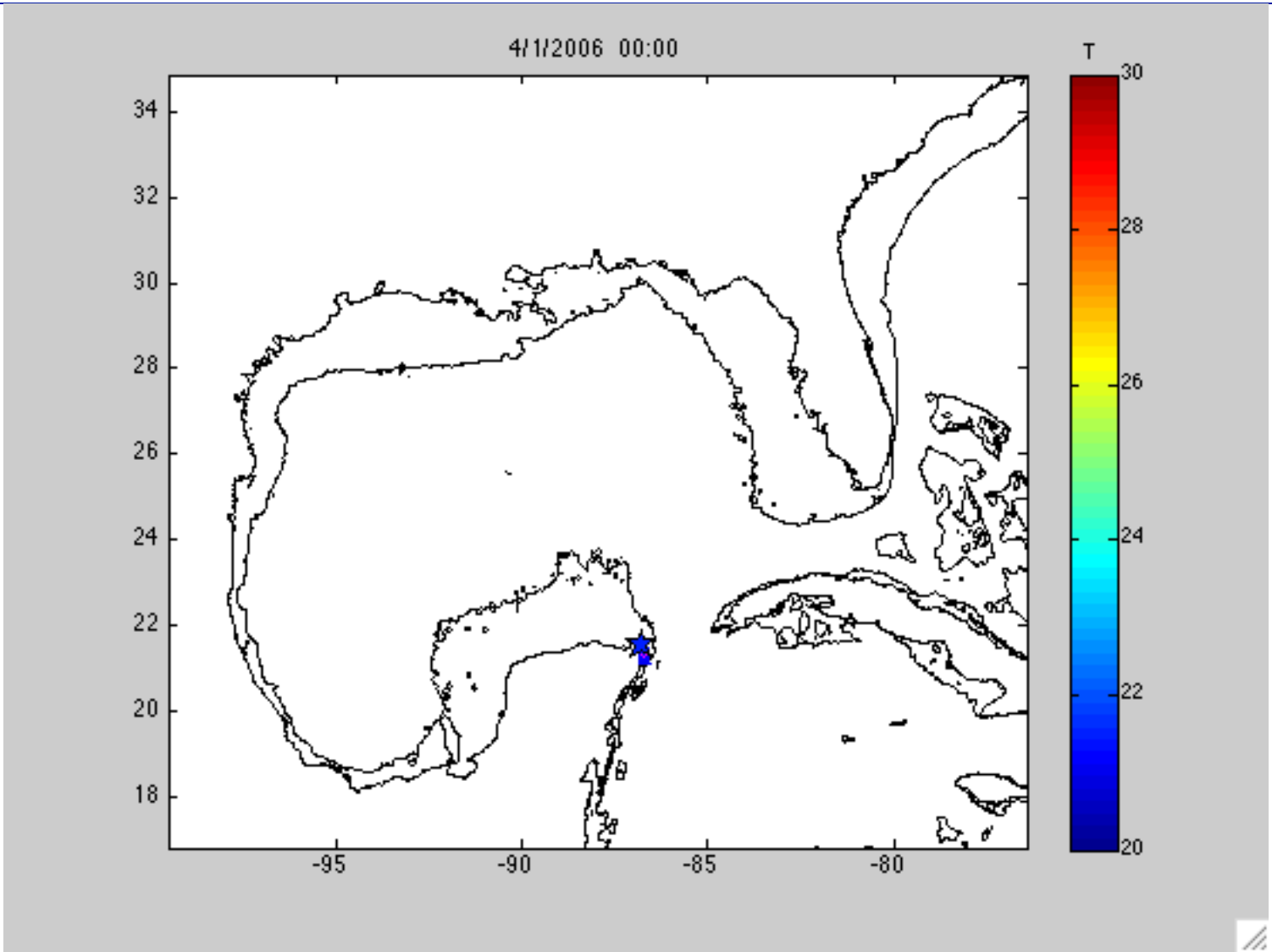
- Ocean observations are (with respected to atmospheric observations)
  - Sporadic
  - Inhomogeneous
  - Limited to upper ocean
- New types of observations that are promising & challenging
  - HF radar observations
  - Lagrangian observations (Position observations along the trajectory)
    - Drifters
    - Floats
    - Gliders (Lagrangian-like)

    with Targeting in mind



Data available from http://www.aoml.noaa.gov/phod/dac/dacdata.html

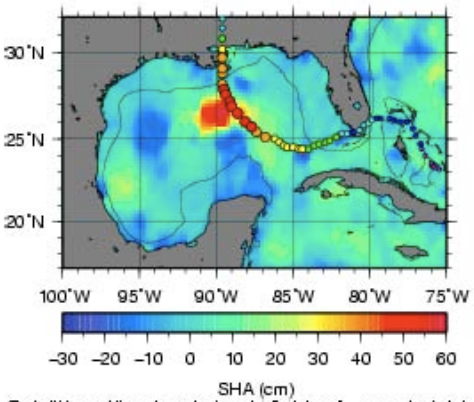
Lagrangian Data



Data available from http://www.aoml.noaa.gov/phod/dac/dacdata.html

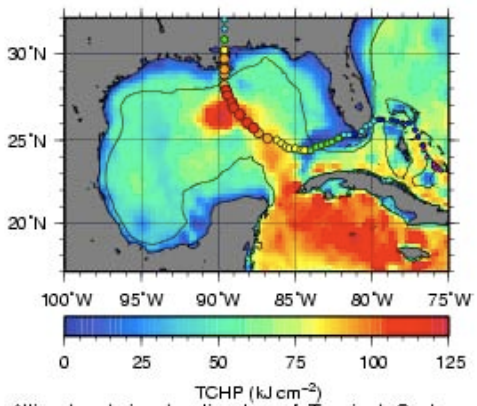
Additional Motivation: Hurricane Prediction

Aug 28



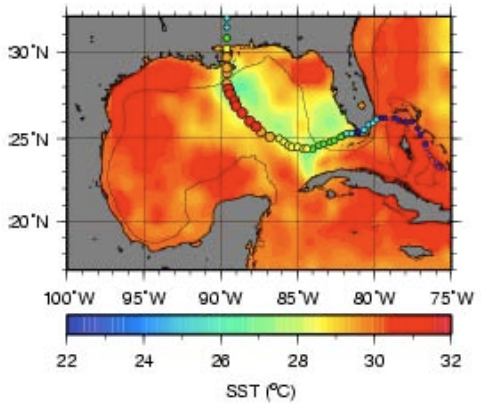
Satellite altimetry-derived field of sea height anomaly (SHA) on August 28, 2005, in the Gulf of Mexico. The large values (red) of SHA in the center of the Gulf are indicative of the presence of a warm anticyclonic ring. The circles of different colors indicate the track and intensity of Hurricane Katrina. The isobath of 200m is superimposed.

Aug 28



Altimeter-derived estimates of Tropical Cyclone Heat Potential (TCHP) for August 28, 2005. The Loop Current and a large warm anticyclonic ring have the largest amount of heat stored in the region. The circles of different colors indicate the track and intensity of Hurricane Katrina. The isobath of 200m is superimposed.

Aug 31

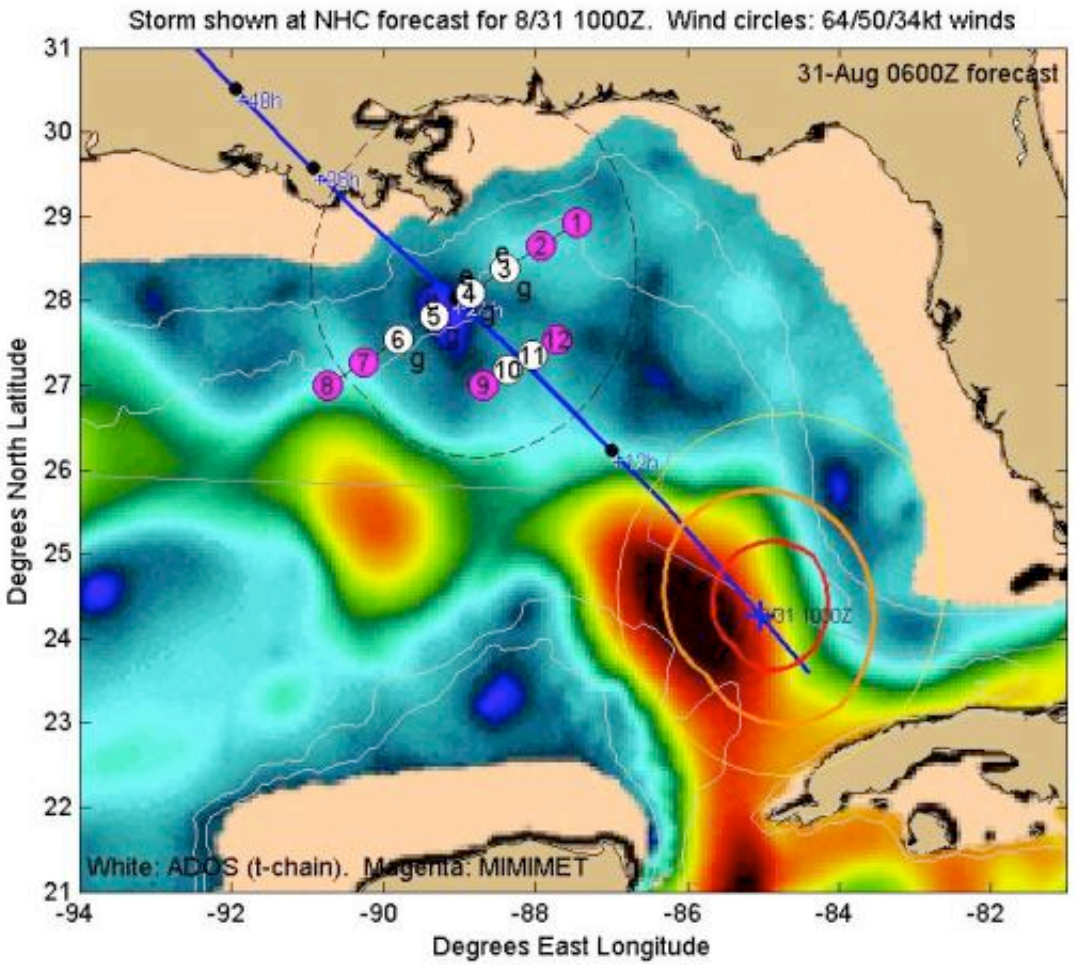


NOAA GOM surface dynamics report for Katrina

http://www.aoml.noaa.gov/phod/altimetry/katrina1.pdf

Adaptive Sampling by Drifters for Better Hurricane Prediction. 1

- Hurricane Gustav
  - 12 drifters deployed in the forecast path of Gustav on August 31, 2008
  - All survived and transmitted data



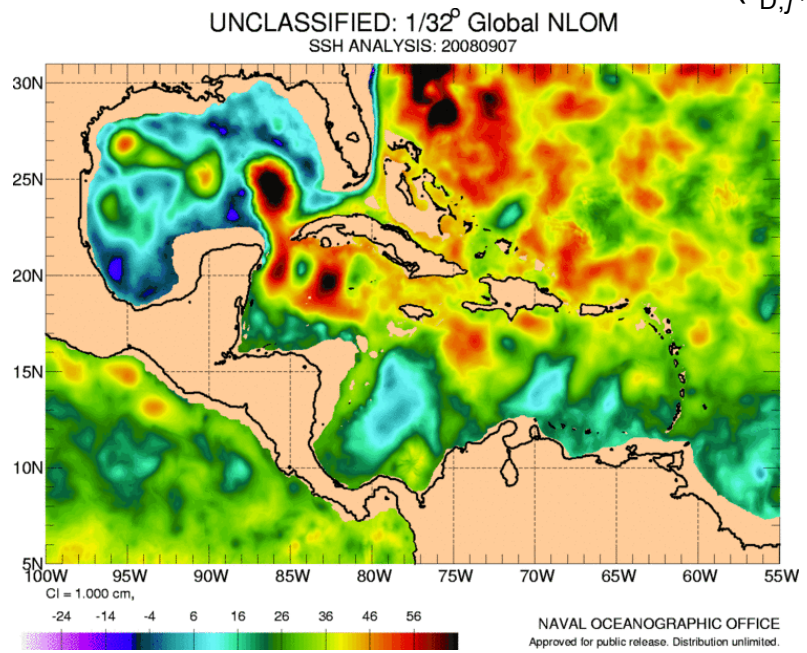
http://www.aoml.noaa.gov/phod/dac/gdp_information.html

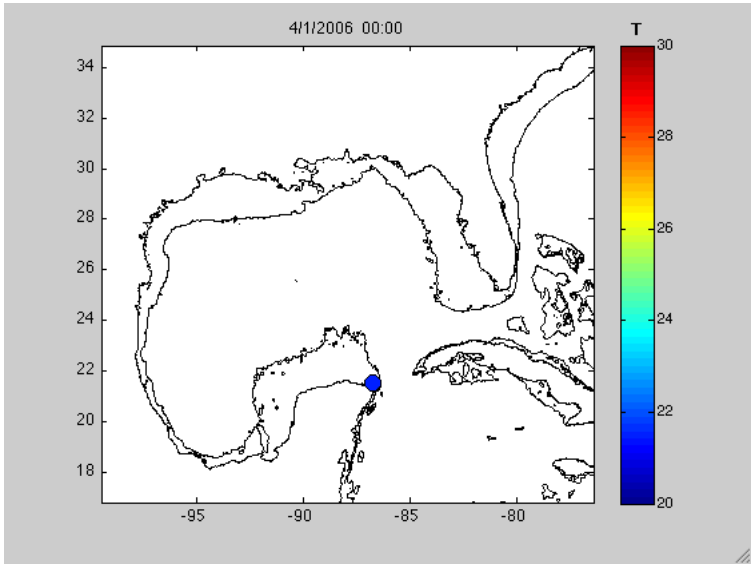
Ocean Flow and Lagrangian Dynamics

- Data assimilation attempts to estimate and forecast the evolution of the system (ocean)
- Ocean flow drives Lagrangian dynamics (drifter motion)
- How do we do LaDA? Is it effective for estimation and prediction of the flow evolution?

http://www7320.nrlssc.navy.mil/global_nlom32/navo/IAS_nlomw12912mooper.gif

$$\mathbf{u}(\mathbf{r}_{\mathrm{D},j}, t) = \frac{d}{dt}\mathbf{r}_{\mathrm{D},j}$$

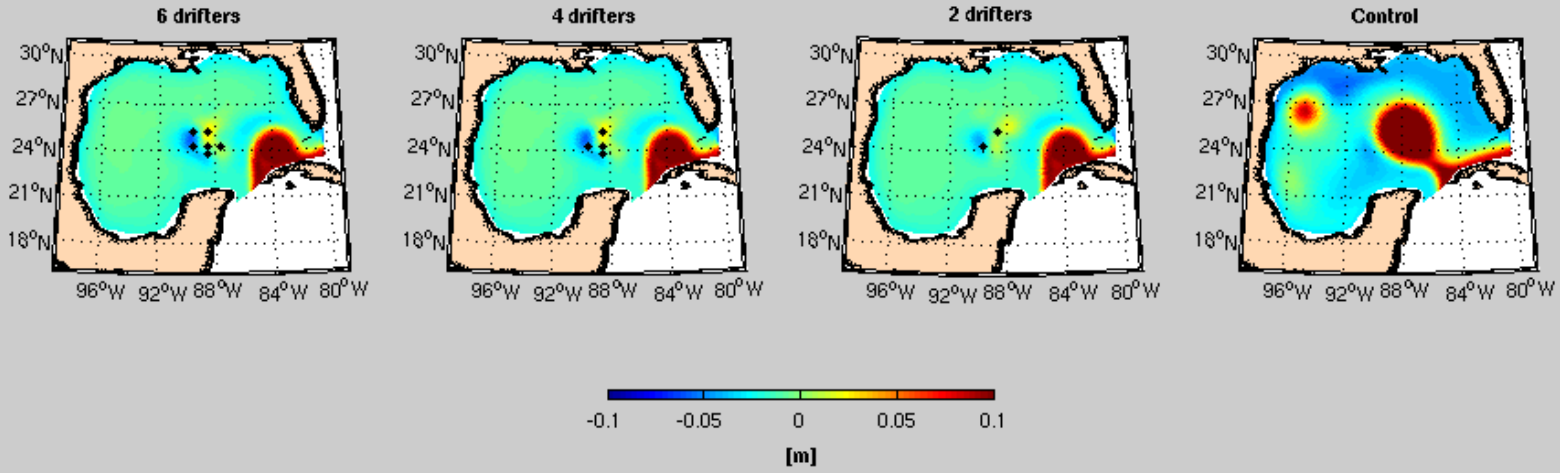




$$\int_{t_{k-1}}^{t_k} \mathbf{u}(\mathbf{r}_{\mathrm{D},j}, \tau)\, d\tau = \mathbf{r}_{\mathrm{D},j}(t_k) - \mathbf{r}_{\mathrm{D},j}(t_{k-1})$$

Data: http://www.aoml.noaa.gov/phod/dac/dacdata.html
Special Thanks to Ed Ryan

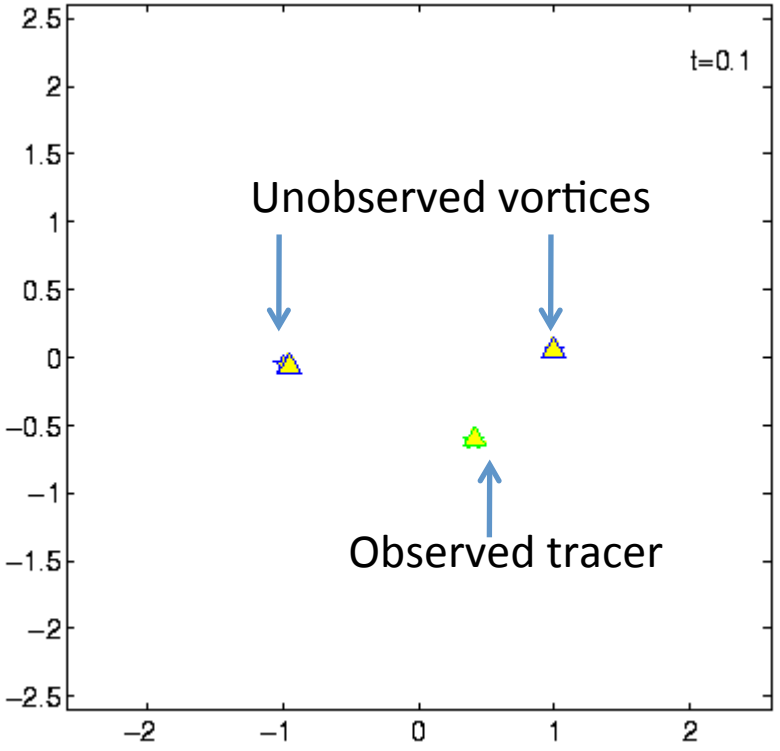
Proof of Concept



Challenge: Handling of Nonlinearity in Observation Operator

- Data assimilation grew up with quasi-linearity in mind
  - Nonlinearity in dynamics: Ensemble Kalman filter, Particle filter
  - Nonlinearity in observation operator: Handling of Chaos
    - Quality control requires extra steps

- Simple demonstration



Quality Control of Data Assimilation Products

- Background: Global Ocean Data Assimilation Experiments (GODAE)
  - Initiated to lead the way in establishing global operational oceanography in 1997 as the acknowledgment of the need for better ocean observations and ocean forecasts and with the scientific and technical opportunity that readily available satellite data had delivered.

- Main Objectives

1. To demonstrate operation ocean system in action
2. To access the quality of hindcast ocean products
3. To perform the intercomparison among the different operational centers
4. To promote validation and intercomparison as part of future ensemble multi model forecast applications.

Quality Control of Data Assimilation Products

- Quality control in GODAE framework
  - Various ocean data quality-control procedures are used by GODAE systems to ensure that erroneous data are not assimilated. Some systems use externally processed observations, while other systems have developed their own automatic quality-control procedures.
  - For systems executed in reanalysis mode, the observational data have often undergone more extensive delayed-mode, scientific quality-control procedures that are not available in near-real time.
- New perspective: quality control of data assimilation product
  - Intercomparison-metric

Summary

- Ocean data assimilation is relatively new, and may have different
  - Principal goals
  - Scales
  - Observations
- Conceptually, ocean data QC and atmospheric data QC are similar, we have to handle fpr
  - Individual
  - Data sets
  - Data types
- Observations are extremely important in ocean data assimilation
  - Number is limited
  - Inhomogeneous in time and space
- New types of observations are available
  - New types of QC: for example, QC for chaotic data (e.g., Lagrangian data)
  - Observing system design can be as important as the observed data themselves
- New prospect: Quality control of data assimilation products (overall)