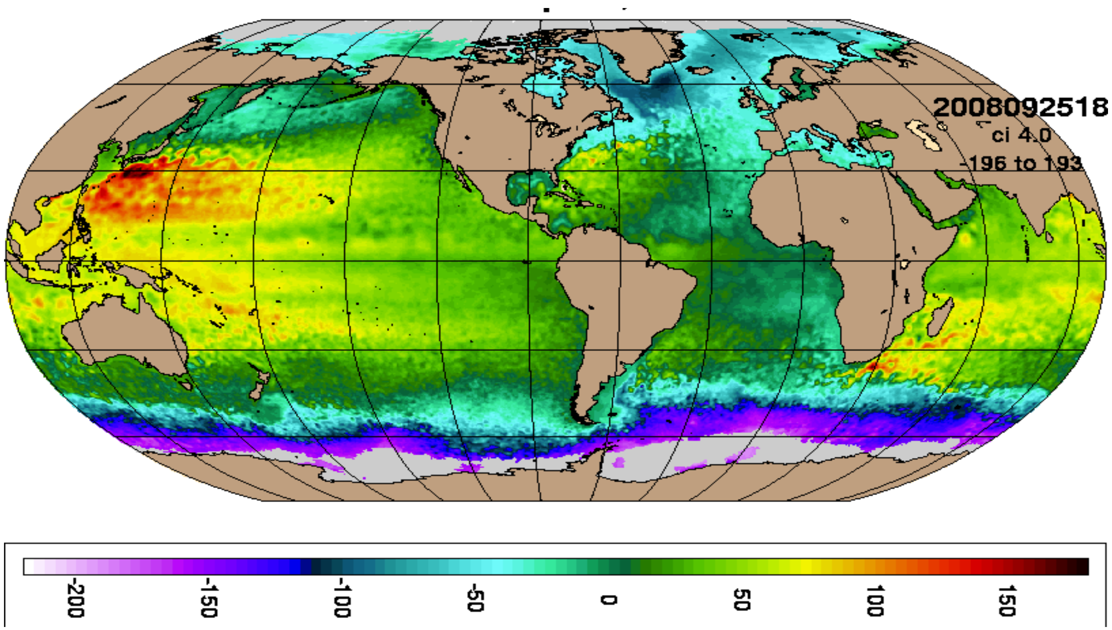
What can Data Assimilation Tools tell us about the Ocean Observing System

James Cummings
IMSG at NWS/NCEP/EMC



10th OOMD Community Workshop
NOAA Ocean Observing and Monitoring Division
9-11 May 2017 NOAA Science Center

Operational Oceanography: GODAE definition

The activity of *systematic* and *routine* measurements of the oceans and their interpretation through data assimilation and numerical modeling.

Data Assimilation:

- process of combining new observations with background information provided by a short-range ocean model forecast
- forecast background carries forward information from earlier assimilated observations
- differences between observations and the forecast provide the new information – *the innovations*

Why are the Innovations Not Zero?

Ocean models have errors:

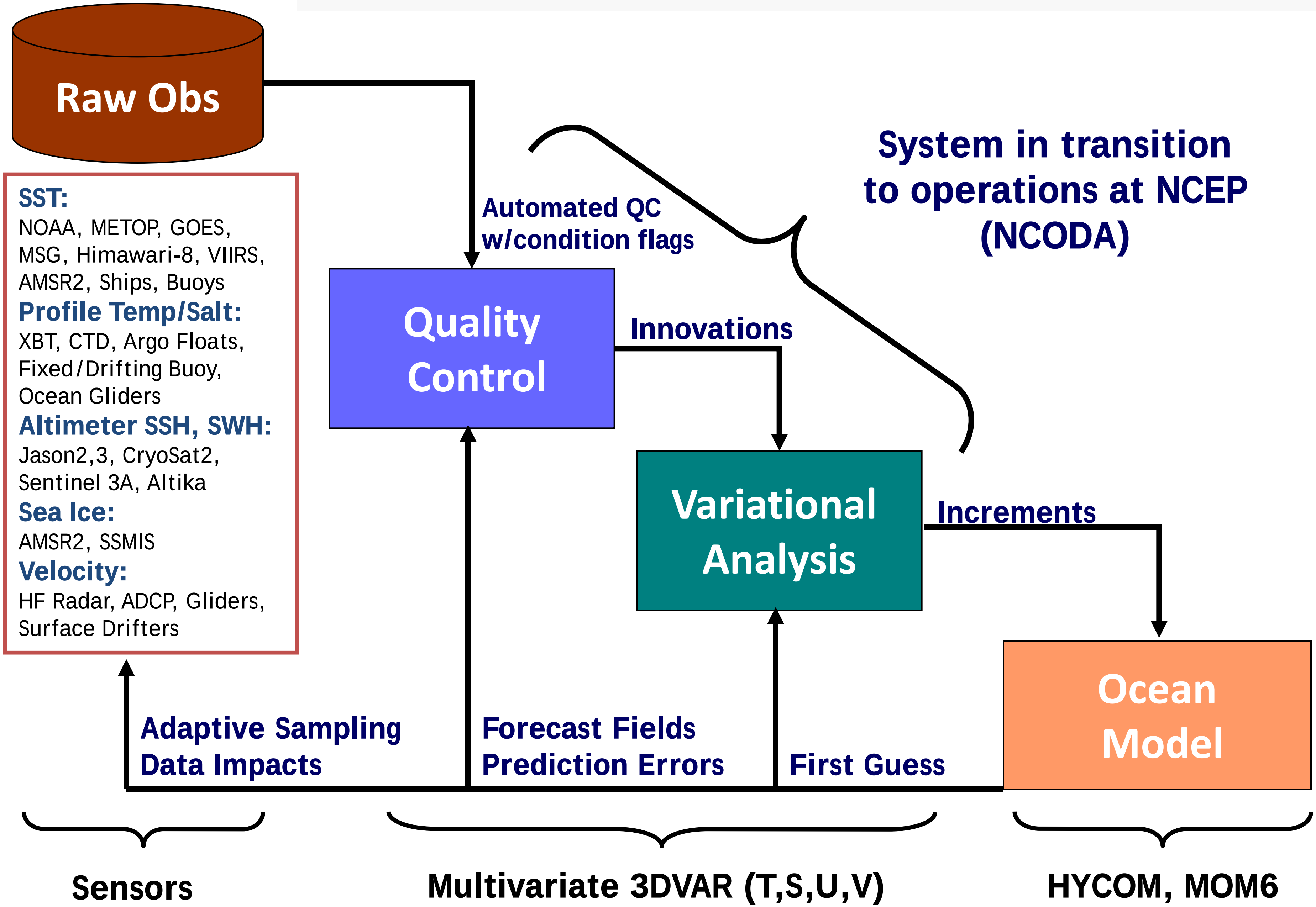
- due to initial conditions
- due to model parameterization, discretization
- due to atmospheric forcing

Ocean model errors accumulate with time:

- without data constraints ocean model forecasts become increasingly inaccurate

Observations are key!

Ocean Data Assimilation: Data Flow



Ocean Data Quality Control: Real Time

Fundamental component of any ocean forecast system:

- accepting erroneous data can cause incorrect analysis
- rejecting extreme (but good) data can miss important events

QC performed in stages:

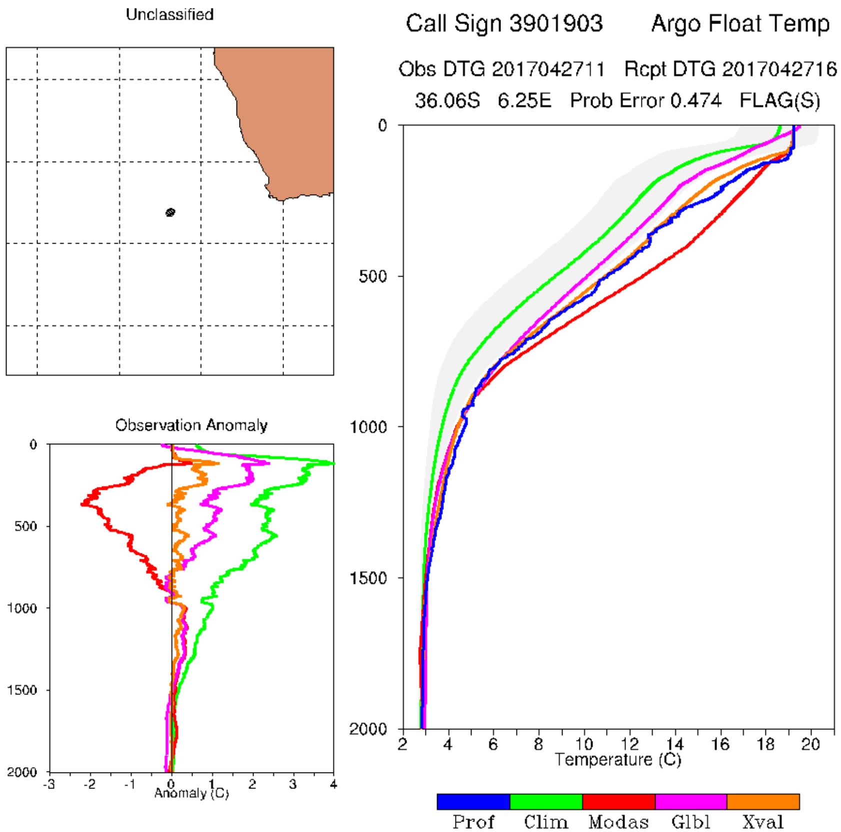
1. **Sensibility checks**: land/sea boundary, location (speed) test, exact duplicate, future observation time, valid value range checks
2. **Error checks**: profile instrumentation, vertical gradients, static stability; cross variable and cross validation; background fields (climate, analysis, forecast)
3. **Consistency checks**: calculated during variational minimization using assimilation machinery

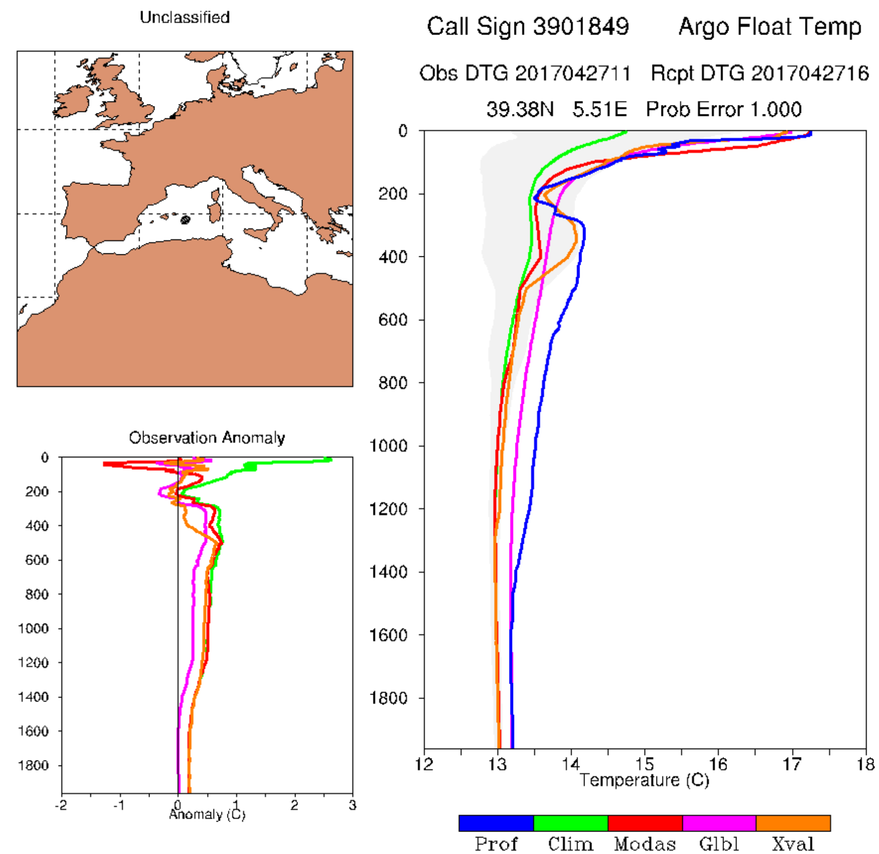
QC outcomes should be easily interpretable, associated with the data values, and shared with the data providers.

QC Error Checks: Cross Validation

OI analysis at observation location using nearby observations

first guess is climate: **cross validation** profile is corrected climatology





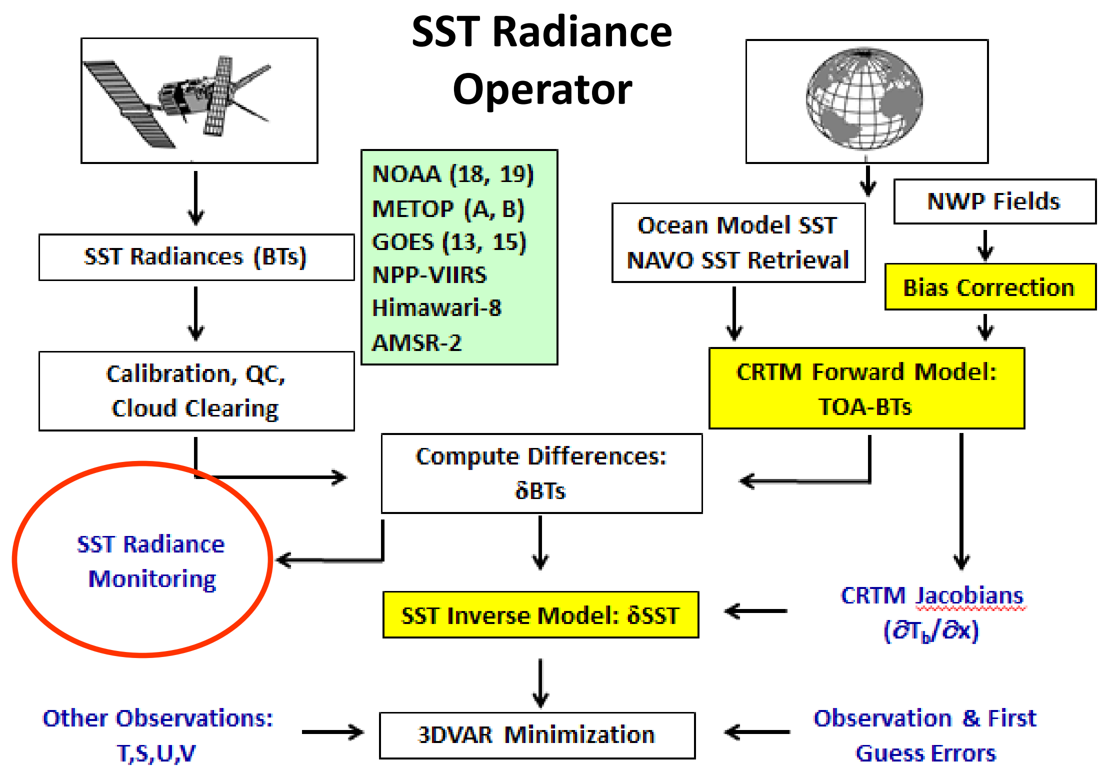
2-4°C warm climate anomaly at depth consistent with other observations

No observations available to verify anomalously warm water at depths > 400 m

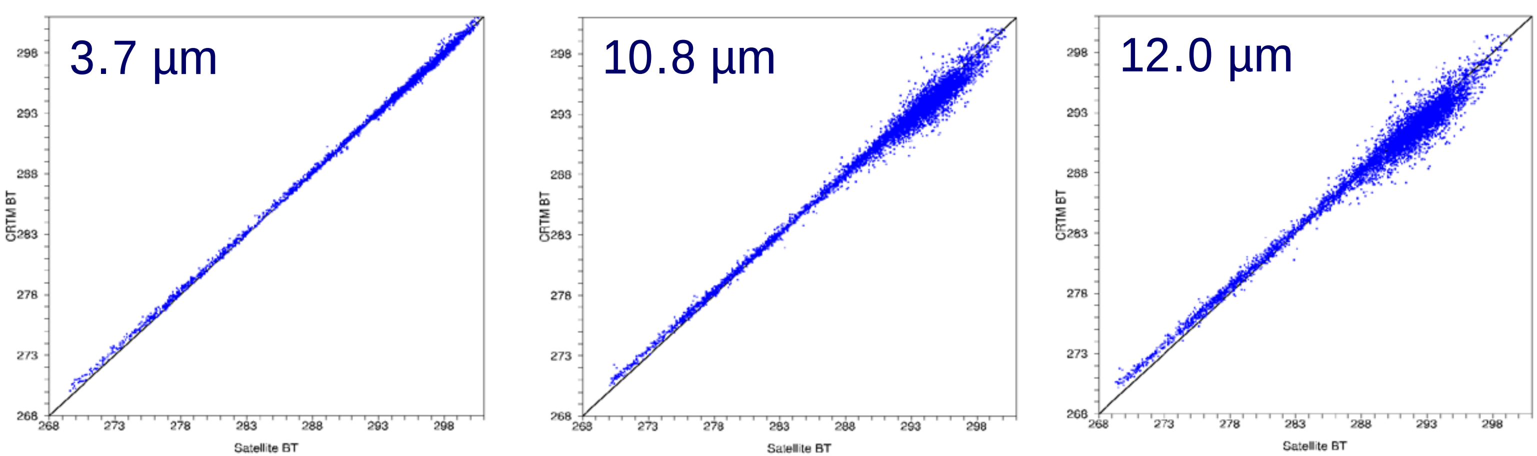
QC Error Checks: Cross Variable

SST Radiance Assimilation:

- simulate satellite SST radiances using CRTM and NWP fields (T_a, Q_a)
- monitor differences between observed and simulated radiances
- provide checks on satellite calibration drift and SST retrieval quality



METOP-A
17 Nov 2016



Analysis QC: Consistency Check

Purpose: check if a suspect observation is likely or unlikely with respect to all other observations in the analysis

define normalized innovation **d*** as:

$$A = HBH^T + R$$

$$d^* = A^{-1/2}[y - H(x_b)]$$

where **A** is the full multivariate forecast error covariance between all observations and [$\mathbf{y-H(x_b)}$] is the innovation vector

- after a few iterations of the descent algorithm check **d*** against original suspect observation innovation: $\mathbf{d\hat{}=diag(A^{-1/2})[y-H(x_b)]}$
- if **d*** < **d^** then the analysis solution is drawing to the suspect observation: it is *consistent* with the other observations
- if **d*** > **d^** then the analysis solution is moving away from the suspect observation: it is *inconsistent* and rejected

Method applied within the variational DA system

Ocean Data Assimilation: Observation Impacts

Ocean observations do not have equal value:

- in terms of reducing ocean model forecast error
- how can we ***quantify*** the impact of each observation?

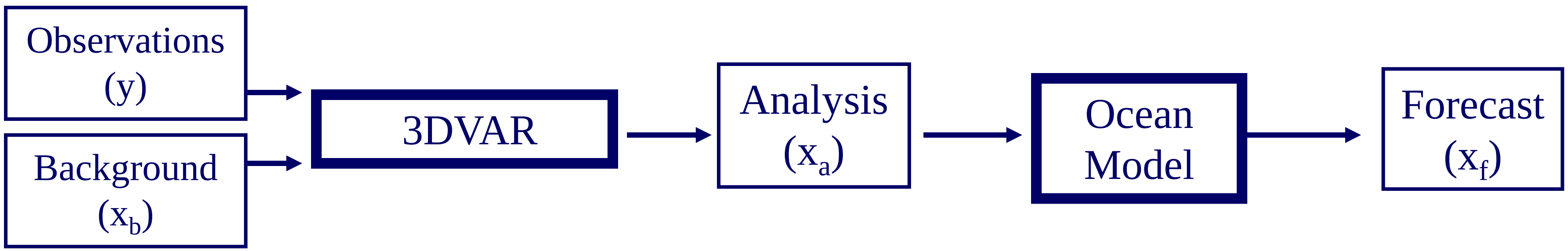
Ocean observing systems are in continuous evolution:

- how can we provide ***routine*** assessments of data impact?
- system must be computationally efficient and run in NRT

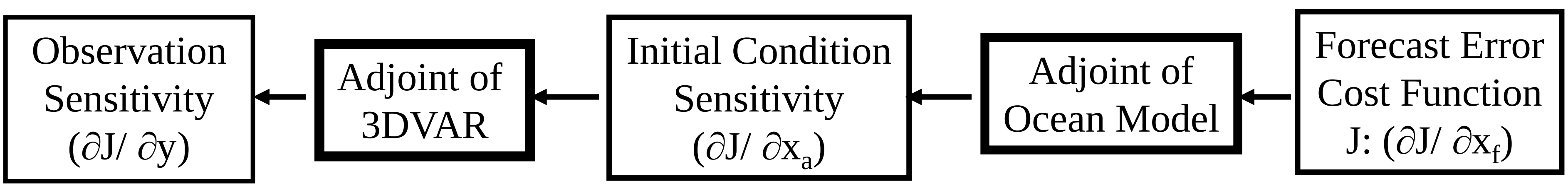
Impacts of observation subsets must be easily quantifiable:

- instrument type (with traceability to individual platforms)
- measurement variable (temperature, salinity, velocity)
- geographic region
- vertical depth level

Analysis – Forecast System



Data Impact System



$$\delta e_f^g = \langle (y - Hx_b), \frac{\partial J}{\partial y} \rangle \longleftarrow$$

What is the impact of observation (y) on measure of forecast error (J) ?

Observation Impact Equation
Langland and Baker (2004)

Observation Impact Equation: Interpretation

For any observation assimilated, if …

$\delta e_f^g < 0.0$ the observation is **BENEFICIAL** - forecast errors decreased from the assimilation

$\delta e_f^g > 0.0$ the observation is **NON-BENEFICIAL** - forecast errors increased from the assimilation

Non-beneficial impacts:

- not expected, assimilation should decrease forecast error

- if it occurs look for problems in data QC, instrument calibration, model error or lack of predictability

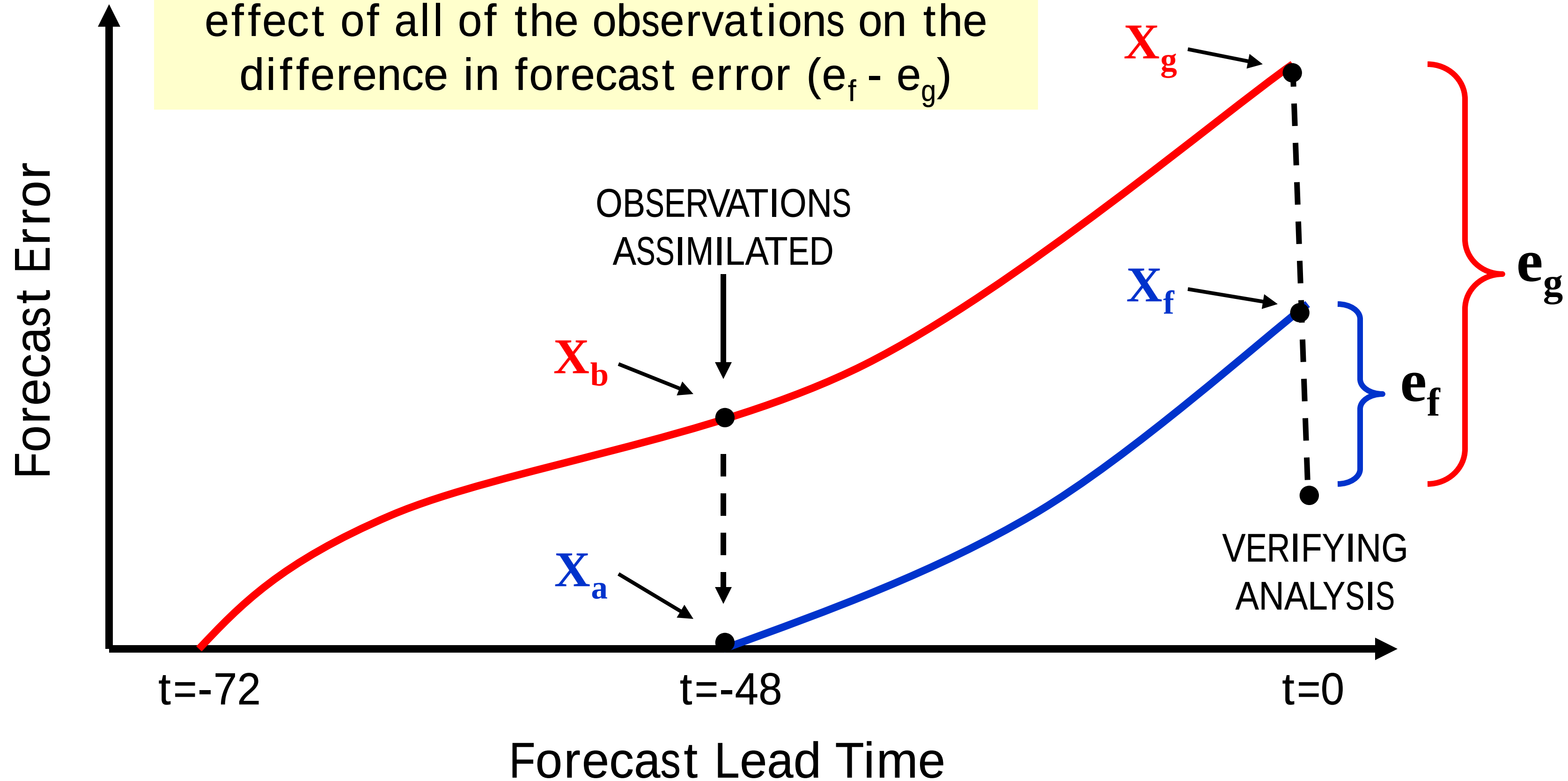
❖ Best Outcome:

- many observations that produce equal or similar impacts, not few, isolated observations that produce large impacts

Observation Impact: Concept

Observations move the forecast from the background trajectory (X_b) to the trajectory starting from the new analysis (X_a)

"Observation impact" is the combined effect of all of the observations on the difference in forecast error (e_f - e_g)



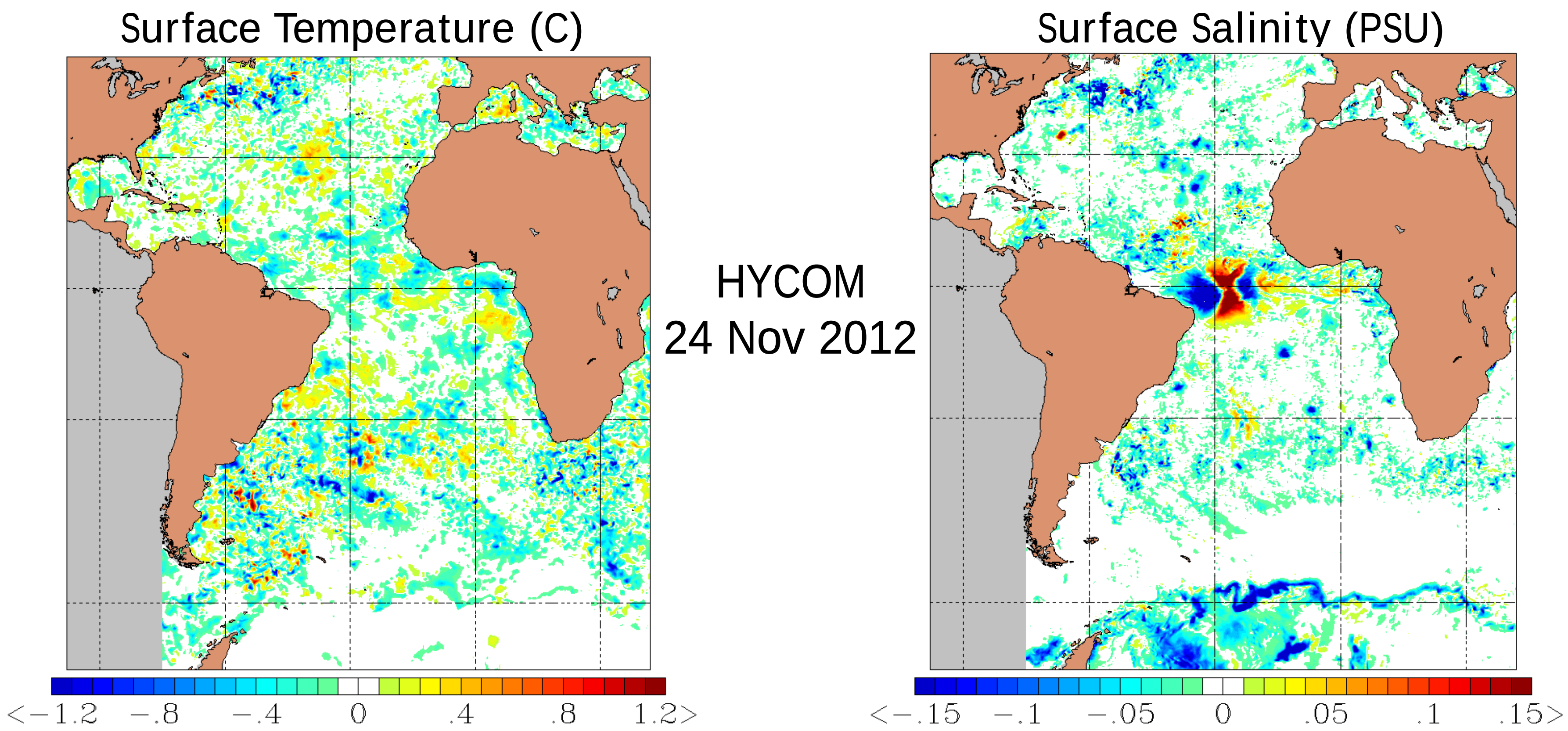
Note: X_a and X_b trajectories are the same if no observations are assimilated: $e_f - e_g = 0$

Forecast Error Gradients: Instantaneous Fields

$$\frac{\partial J}{\partial x_f} = e_f - e_g$$

$$e_f = (x_{48} - x_0)(x_{48} - x_0)$$
$$e_g = (x_{72} - x_0)(x_{72} - x_0)$$

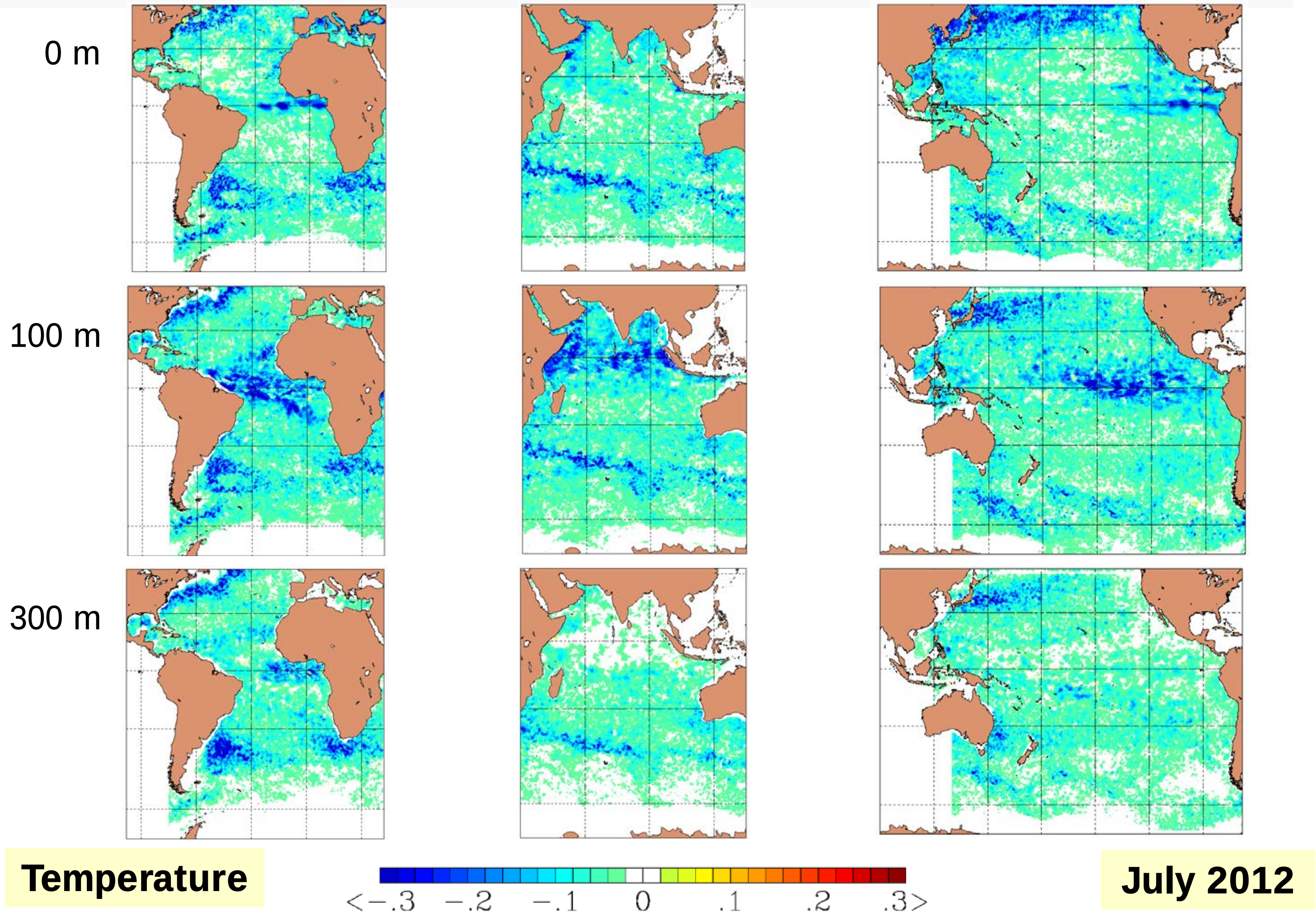
Difference between two forecasts valid at same time relative to a verifying analysis



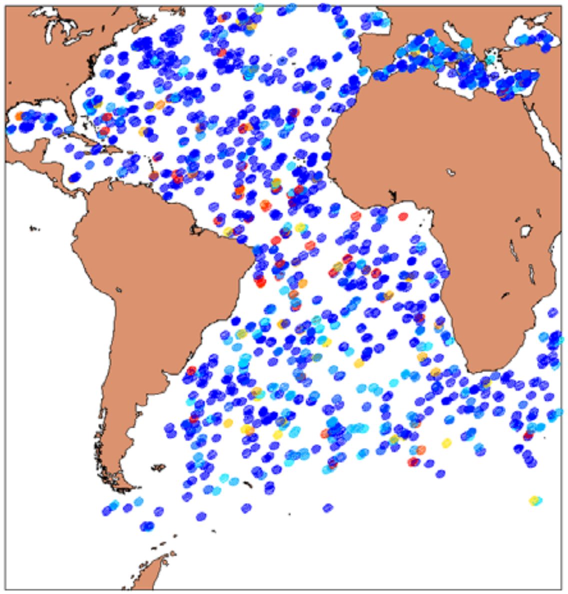
- negative values indicate forecast ***error reduction***
- positive values indicate forecast ***error growth***

$$\delta e_{48}^{72}$$

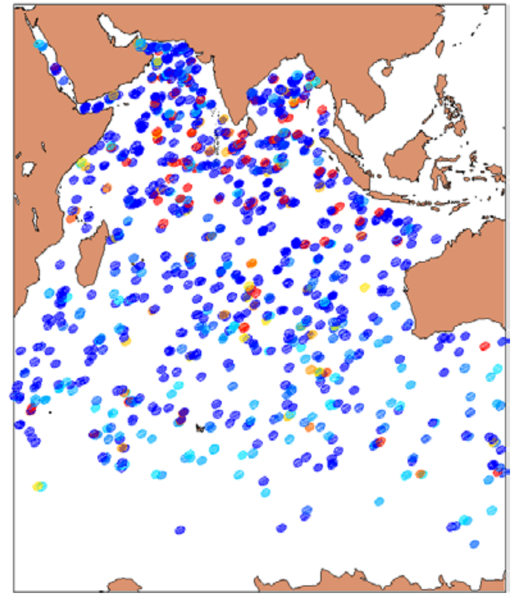
Forecast Error Gradients: Time Averaged Fields



Observation Impact: Argo Temperature

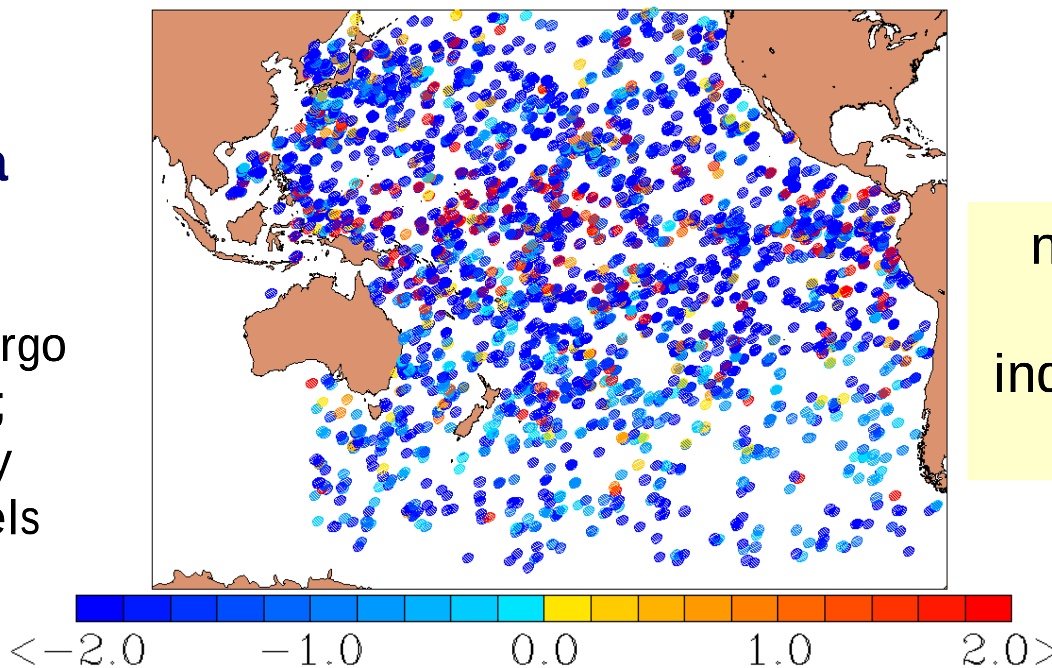


Global HYCOM
August 2016



Profile Data Impacts (C)

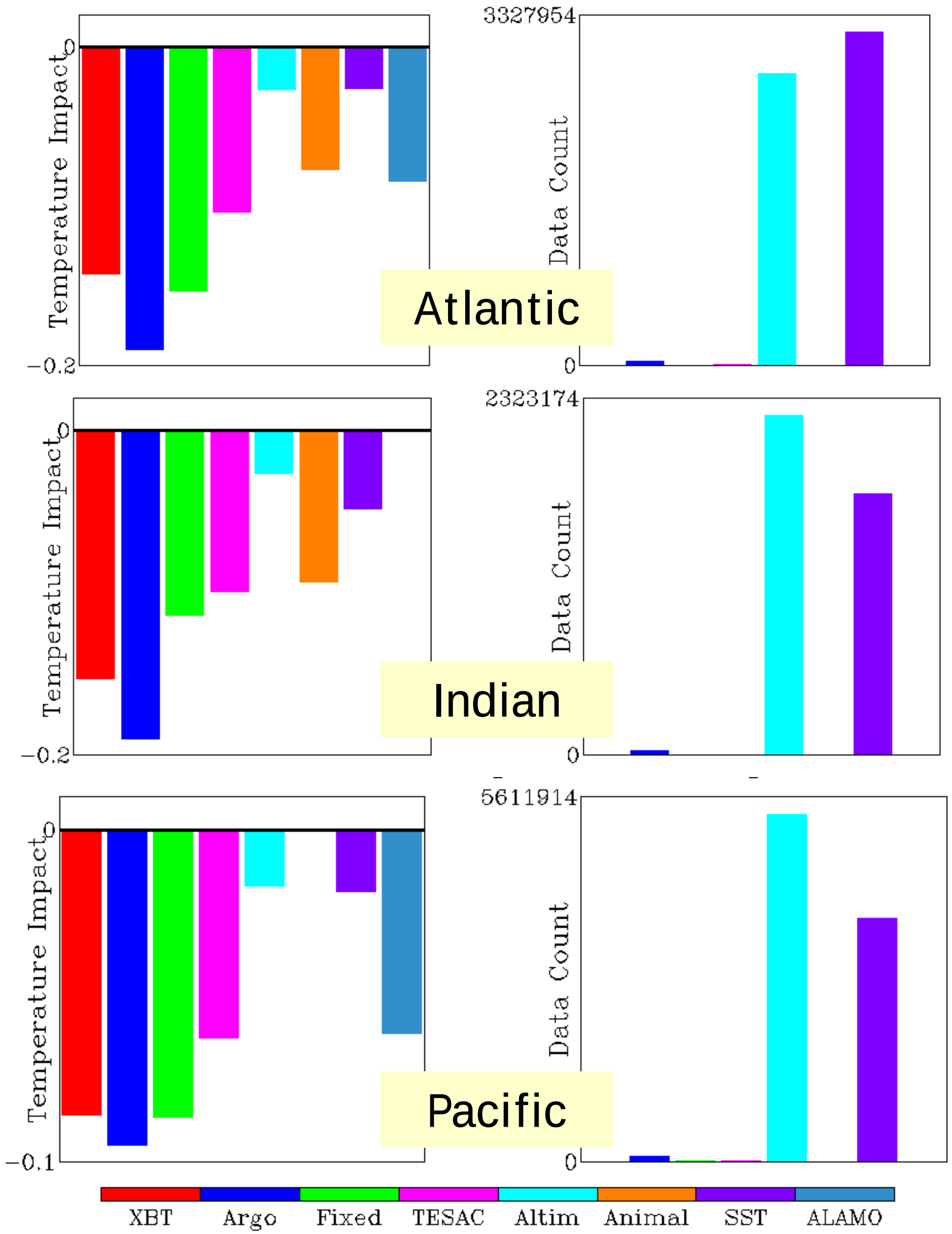
averaged over Argo profile levels; normalized by number of levels



$$\delta e_{48}^{72}$$

negative values (cool colors) indicate beneficial impacts

Observation Impact: Observing System Rankings



Global HYCOM - August 2016

- averaged over observing systems within ocean basins
- all observing systems assimilated have beneficial impacts
- Argo most important observing system in all ocean basins
- TESAC includes ocean gliders
- animal borne sensors (elephant seals) are important
- most numerous observing systems (SST and SSH) have low impacts on a per-ob basis

$$\delta e_{48}^{72}$$

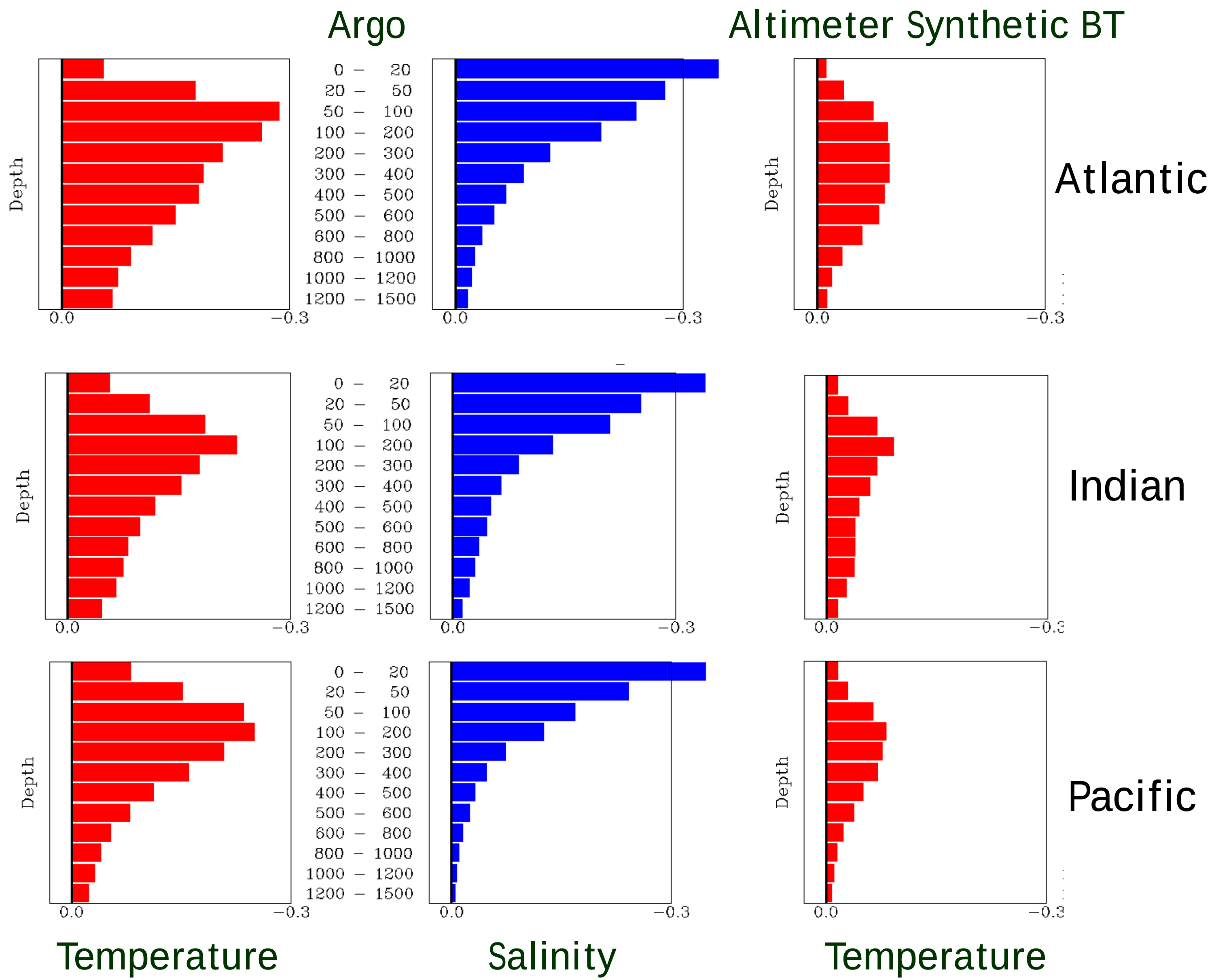
Observation Impact: Vertical Level

Global HYCOM

August 2016

averaged within observing system, variable, depth intervals, and ocean basins

$$\delta e_{48}^{72}$$



Applications: Multiple Uses of Impact System

Application	Innovations	Cost Function	Purpose
Observation Impact	Real observations (all known)	Forecast error (known)	Evaluate impacts of observations on forecast error
Observing System Design	Real and simulated observations (some known, some unknown)	Forecast error or model variable (known)	Develop more optimal configurations of observing systems
Targeted Observing	Simulated observations (all unknown)	Proxy for forecast error (unknown)	Impact of adding new observations at some future time

Observing System Design and Targeted Observing are variants of core "Observation Impact" system

Ocean Data Assimilation Tools: Conclusions

Ocean Data Quality Control:

- fully automated, real-time
- feedback to data providers on their data quality?

Adjoint-Based Ocean Data Impacts:

- provides routine assessment of data impacts:
  - useful when new observations are deployed
- but method has limitations – data impacts depend on:
  - assimilation system and forecast model
  - forecast error cost function metric
- need other estimates of data impacts for comparison:
  - OSE (data denial) studies
  - different forecast systems assimilating the same data

END

Questions?