Ocean Data Assimilation and Assimilation of Altimetry data

Imranali M. Momin, Scientist
National Centre for Medium Range Weather Forecasting
Ministry of Earth Sciences
A-50, Sector-62 Noida-201309, UP, India
E-mail: imranali@ncmrwf.gov.in

Joint IMD-WMO Group Fellowship Training Course on
"Numerical Weather Prediction"
Managed by
IMD, Pune, 04th October - 10th November 2021

Weather and Climate Prediction

- Numerical Weather Prediction (NWP) is initial/boundary value problem.

- Given initial/boundary conditions

1. Present state of the atmosphere/ocean (Initial Condition)
2. Appropriate surface and lateral boundary conditions

- NWP model simulates or forecasts the evolution of the atmosphere/ocean state.

More and more accurate initial states lead to better quality of model forecasts.

Primitive Equations

$$\underbrace{\frac{du}{dt}}_{\text{TOT}} = \underbrace{\frac{\partial u}{\partial t}}_{\text{LOC}} + \underbrace{u\frac{\partial u}{\partial x} + v\frac{\partial u}{\partial y}}_{\text{ADV}} + \omega\underbrace{\frac{\partial u}{\partial p}}_{\text{VER}} = -\underbrace{\frac{\partial \phi}{\partial x}}_{\text{PGF}} + \underbrace{f\,v}_{\text{COR}}$$

Advection

Pressure Gradient Force

Momentum Equations; (Conservation of momentum)

$$\underbrace{\frac{dv}{dt}}_{\text{TOT}} = \underbrace{\frac{\partial v}{\partial t}}_{\text{LOC}} + \underbrace{u\frac{\partial v}{\partial x} + v\frac{\partial v}{\partial y}}_{\text{ADV}} + \omega\underbrace{\frac{\partial v}{\partial p}}_{\text{VER}} = -\underbrace{\frac{\partial \phi}{\partial y}}_{\text{PGF}} - \underbrace{f\,u}_{\text{COR}}$$

Coriolis force

Vertical Velocity

$$\frac{\partial T}{\partial t} = -\nabla \cdot (T\,\mathbf{U}) + D^T + F^T$$

Divergence of T/S

Surface Forcings

Tracer Equations (Conservation of Heat and Salt)

$$\frac{\partial S}{\partial t} = -\nabla \cdot (S\,\mathbf{U}) + D^S + F^S$$

Parameterization of small scale physics

Continuity equation for incompressible flow (constant density)

$$\nabla \cdot \mathbf{u} = 0$$

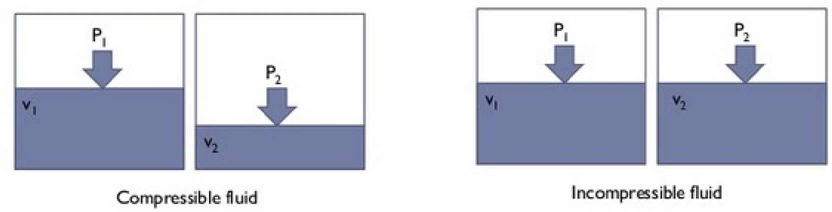
where **u** is the velocity vector

$$\nabla = \left(\frac{\partial}{\partial x}, \frac{\partial}{\partial y}, \frac{\partial}{\partial z}\right)$$

u, v, w are velocities in *x, y,* and z directions

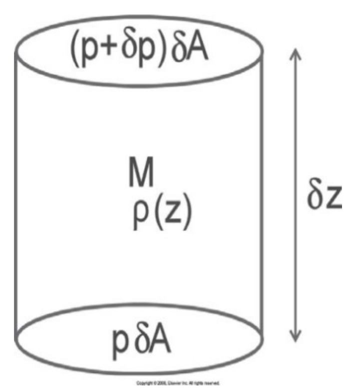
$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} + \frac{\partial w}{\partial z} = 0.$$

Compressible and Incompressible flows



Hydrostatic Approximation

Vertical force balance: Pressure gradient is proportional to density.



$$\text{PGF}_{(\text{vertical})} = \text{Gravity}$$

$$-\frac{1}{\rho}\frac{dp}{dz} = g$$

$$\frac{dp}{dz} = -\rho g$$

Time scales of Weather and Climate Prediction

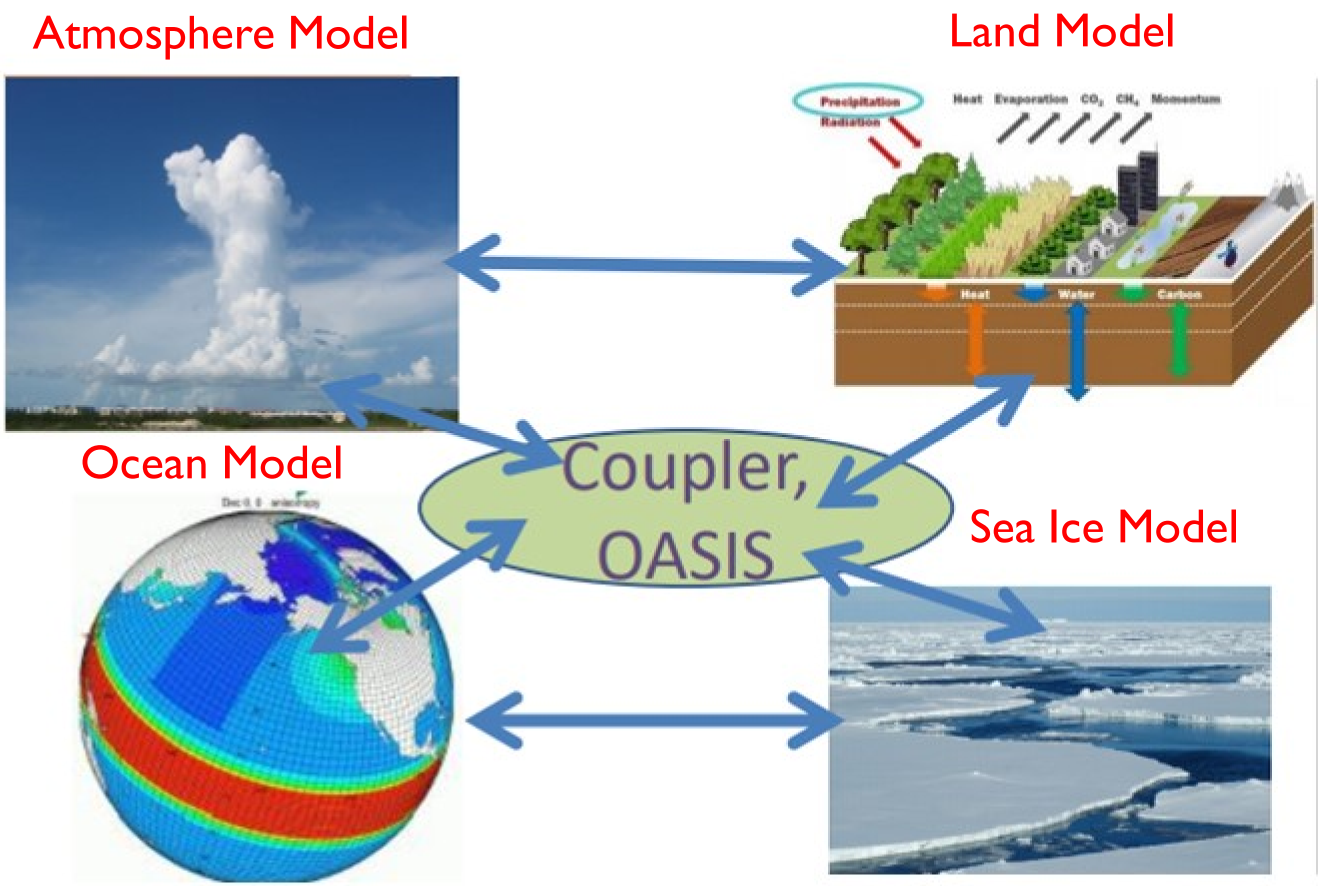
Time horizon

Hour | Day | Week | Month | Year | Decade | Century

Nowcasting | Weather Forecasting | Extended Range Forecast | Seasonal Foreacst | Decadal Forecast | Climate Projection

Range	Model	Interested
Nowcasting & Short range weather forecast (up to 3 days); with uncertainty	Regional Model; Regional Ensemble Model	Quantity+uncertainty (probability) (Fog, Therderstorm, heavy rainfall, flood forecast etc.)
Medium Range Weather Forecast (day-3 to day-10); with uncertainty	Global Model & Global Ensemble Model	Quantity+uncertainty (probability) (Cyclone track, intensity, low-depression, heat-cold waves, western disturbance etc.)
Extended Range up to month, Seasonal forecast up to 6 months; Decadal forecast	Global Coupled Model & Global Ensemble Model	Tendency with uncertainty (active-break cycle, MJO, seasonal monsoon rainfall forecast, global warming, sea level rise etc)

Global Coupled Model



Accuracy of Ocean Initial Condition

- Ocean model driven by surface fluxes(momentum, heat and fresh water from atmospheric analysis/ reanalysis)

    1) Which Ocean models (regional, global)

    2) Error characteristics of ocean models(resolution, boundary fluxes, and parameterization)

- Ocean Observations

    1) Which data?(SST, Subsurface temperature and salinity, sea level)

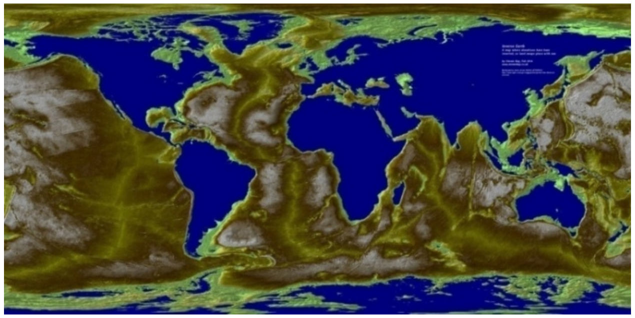
    2) Which instruments ? (TAO, XBTs, and ARGO)

    3) Which frequency, error statistics, balance relationship…. ?

- Assimilation Method

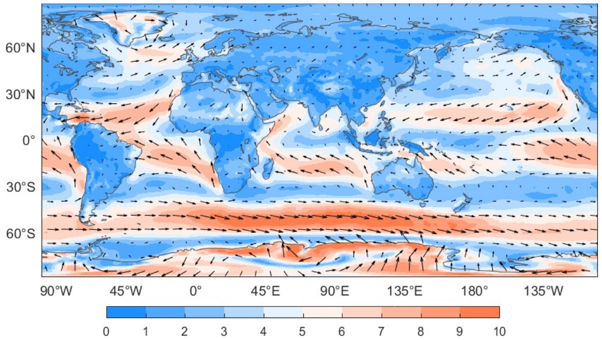
    1) Which assimilation method ? (OI, 3DVar, 4DVar, or EnKF)

Surface Boundary Condition



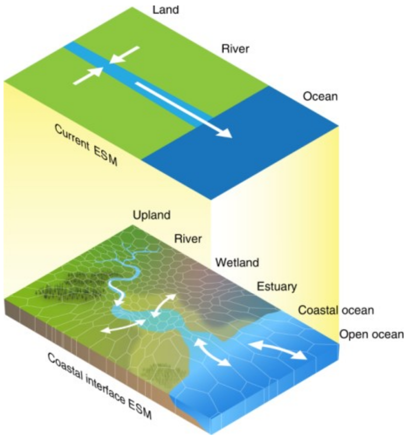
Solid Earth -Ocean Interface

No heat and salt flux exchange across the solid boundaries.



Atmosphere - Ocean interface

Atmosphere and Ocean exchanges the boundary fluxes such as horizontal momentum(wind stress), fresh water and heat



Land-Ocean interface

Freshwater continental land to the ocean through river runoff

Sea Ice - Ocean interface

Sea Ice and Ocean exchange heat, and fresh water fluxes.

Ocean Data Assimilation

Data Assimilation System

- Data assimilation is the mathematical process to combine the observations and the atmosphere/ocean models to extract the most important information.

- We never know the true state of the atmosphere/Ocean.

- We know that, the model forecast is not perfect due to the resolution, numerical truncation error & various physical parameterizations.

- Observations have limitation due to the spatio-temporal distribution and having instrument errors.

Requirement for Ocean/Atmospheric Data Assimilation

- **Background information**

1. Complete coverage
2. Filter out noise
3. Background (Climatology or Dynamical Model Forecast)

- **Observational Data**

1. Noisy
2. Scattered, non-uniformly distributed (space, & time)
3. Insufficient to determine the Initial Condition(or Analysis)
4. Different types of Observation (Radiance, Reflectivity, etc)

Observational Error	Background Error
1. Representation error	1. Due to Initial Condition (unknown true state)
2. Instrument error or measurement error (random error or systematic error)	2. Imperfect representation of circulation (model spatial/temporal resolution, physical parameterization, boundary conditions, computational truncation)

Input/Output of Data assimilation

- The basic input of any assimilation system is the innovation which is the difference between the observations and the model prediction of the observed variables **(Observation-Model forecast).**

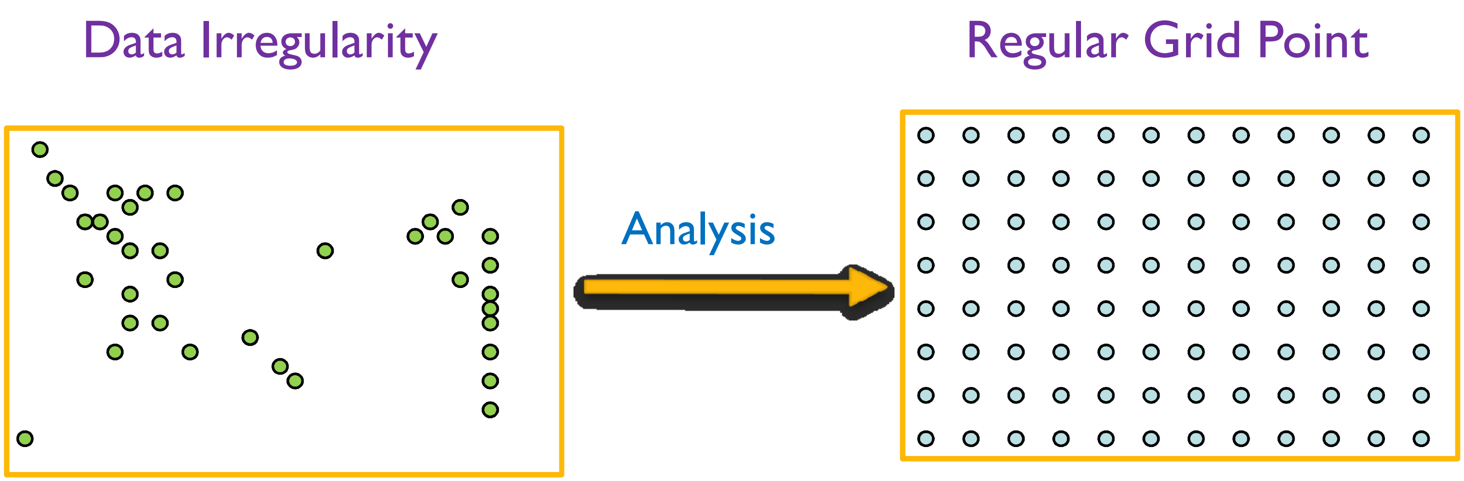
- **Innovations** measure the model errors at the updated cycle interval.

- The basic output of any data assimilation is the residual which is the difference between the analyzed field and the observations after the data assimilation **(Analysis-Observation).**

- **Residuals** measure the fit of the analysis to the observations.

Why we need Data Assimilation

- Range of observations(Ships, Satellite, Radiosonde, Radar, Gliders, GPS radio occultation etc.)
- Range of techniques(Direct measurement, radiance, reflectivity, refractivity; Scattering)
- Different Errors
- Data gaps
- Quantities not measured



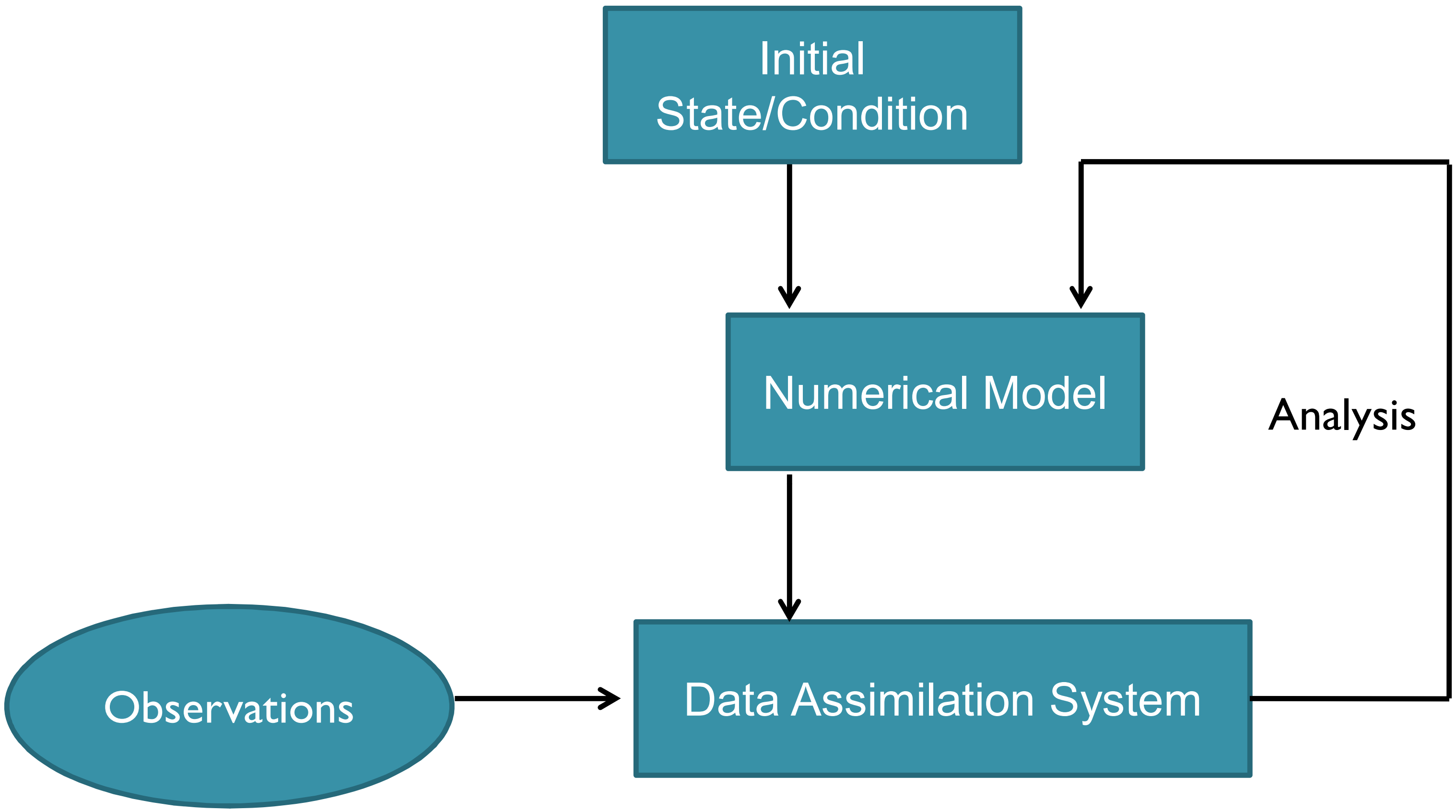
Used of Data Assimilation

Weather and Ocean forecasting	Satellite retrievals
Seasonal/Climate forecasting	Surface flux estimation
Land surface processes	Global climate datasets
Model parameters estimation (IMD-NCMRWF merged rainfall; INCOIS objective analysis of T/S etc.)	Planning satellite measurements

Benefit of Data Assimilation

- Evaluate error in model and observations
- Filling data gaps
- Designing observational network
- Quality control
- Estimating unobserved quantities

Data Assimilation System



Assimilation Cycle: 06 hour for Atmosphere
24 hour for Ocean

Assimilation Methods

➢ Least Square Method

- **Measuring two sets of temperature** from two different thermometers.

$$T_1 = T_t + \varepsilon_1 \qquad T_2 = T_t + \varepsilon_2$$

The analysis is estimated from a linear combination of the observations

$$T_a = a_1 T_1 + a_2 T_2$$

1. MEAN

$$\overline{T_a} = a_1(\overline{T}_t + \overline{\varepsilon_1}) + a_2(\overline{T}_t + \overline{\varepsilon_2})$$

we assume that analysis errors are unbiased.

$$\overline{T_a} = \overline{T}_t$$

$$\overline{T_t} = a_1(\overline{T}_t + 0) + a_2(\overline{T}_t + 0)$$

This leads to $a_1 + a_2 = 1$

- We assume that the **error in measurement is unbiased**.

$$\overline{\varepsilon_1} = 0; \overline{\varepsilon_2} = 0$$

$$\overline{T_1} = \overline{T_t} + \overline{\varepsilon_1} = \overline{T_t};$$

$$\overline{T_2} = \overline{T_t} + \overline{\varepsilon_2} = \overline{T_t};$$

Assimilation Methods

2. VARIANCE:

T_a will be best estimates of T_t, if t are chosen to minimize the mean square error of T_a.

$$\sigma_a^2 = \overline{((a_1 T_1 + a_2 T_2) - (a_1 \overline{T_1} + a_2 \overline{T_2}))^2}$$

$$\sigma_a^2 = \overline{(a_1 (T_1 - T_t) + a_2 (T_2 - T_t))^2}$$

$$\sigma_a^2 = \overline{(a_1 \varepsilon_1 + a_2 \varepsilon_2)^2} = a_1^2 \overline{\varepsilon_1^2} + a_2^2 \overline{\varepsilon_2^2} + 2 a_1^2 a_2^2 \overline{\varepsilon_1^2}\,\overline{\varepsilon_2^2}$$

$$\sigma_a^2 = a_1^2 \overline{\varepsilon_1^2} + a_2^2 \overline{\varepsilon_2^2}$$

$$\sigma_a^2 = a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 = a_1^2 \sigma_1^2 + (1 - a_1)^2 \sigma_2^2$$

To minimize above equation w.r.t. a_1, we require

$$\partial \sigma_a^2 / \partial a_1 = 0$$

subject to the constraint $a_1 + a_2 = 1$.

$$a_1 = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}; a_2 = \frac{\sigma_1^2}{\sigma_1^2 + \sigma_2^2}$$

$$\sigma_a^2 = \frac{\sigma_1^2 \sigma_2^2}{\sigma_1^2 + \sigma_2^2}$$

we know their variances

$$\overline{\varepsilon_1^2} = \sigma_1^2; \overline{\varepsilon_2^2} = \sigma_2^2$$

The errors of two thermometers are uncorrelated

$$\overline{\varepsilon_1}\,\overline{\varepsilon_2} = 0$$

Assimilation Methods

➢ Optimum Interpolation (OI)

- The analysis is defined as linear combination of two observation

$$T_a = a_1 T_1 + a_2 T_2$$

subject to the constraint $a_1 + a_2 = 1$.

$$a_2 = (1 - a_1)$$

$$or$$

$$a_1 = (1 - a_2)$$

$$T_a = a_1 T_1 + (1 - a_1) T_2 = T_2 + a_1 (T_1 - T_2)$$

$$or$$

$$T_a = (1 - a_2) T_1 + a_2 T_2 = T_1 + a_2 (T_2 - T_1)$$

$$T_b = T_1 ; T_o = T_2 ; a_2 = W$$

$$T_a = T_b + W (T_o - T_b)$$

T_b – Background; T_o - Observation;

Optimal weight $W = \dfrac{\sigma_b^2}{\sigma_b^2 + \sigma_o^2}$;

$$\sigma_1^2 = \sigma_b^2 \qquad \sigma_2^2 = \sigma_o^2$$

➢ Kalman Filter: It is a sequential form of OI

Analysis = Model forecast + **Kalman Gain** (observation – Model forecast at Observation)

$$x^a(t_i) = x^f(t_i) + K(y_i - H[x^f(t_i)]) \qquad where,$$

$$T_b = x^f(t_i); T_o = y_i; T_b = H[x^f(t_i)])$$

Assimilation Methods

➢ **3D variational method (3DVar)**

- Background & Observation are fixed in time.
- Background and Observation are assumed as normal distribution with observational error (R) and background error (B)
- The cost function is proportional to the square of the distance between analysis and both the background and the observations.

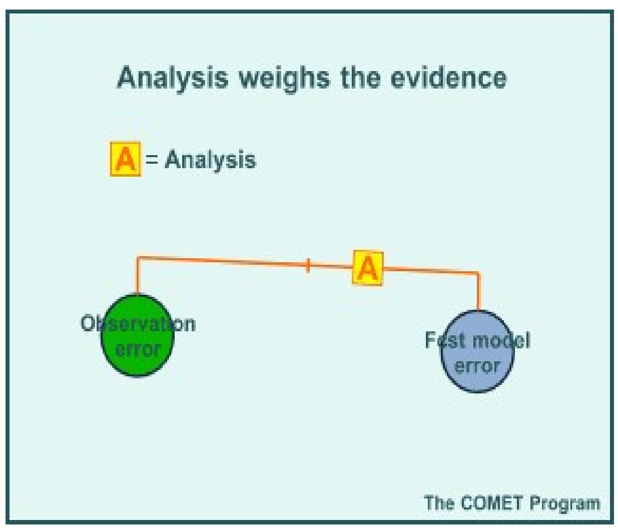
Cost function
$$J(x) = \frac{1}{2}(x - x_b)^T B^{-1}(x - x_b) + \frac{1}{2}(y - H[x])^T R^{-1}(y - H[x])$$

Distance of x to the background xb   Distance of x to observations

Where Background(forecast field) - xb; Analysis:-x; Analysis increment - x-xb; Background error covariance - B; Observation error covariance - R

3D Variational method (3DVar)

- The cost function J measures:
- The distance of a field x to the background (xb; first term).
- The distance of a field x to the observations(y;second term)
- The distances are scaled by the observation error covariance R and by the background error covariance B respectively.
- The minimum of the cost function is obtained for x = xa, which is defined as the analysis or initial condition.



NEMO variational Data Assimilation

Model	**Nucleus European Modelling of Ocean v3.4; Los-Alamos Sea Ice (CICE) v4.1**
Horizontal Resolution	0.25x0.25 (~25 km)
Vertical Levels	75 levels
Time for dynamics	20 min.
Data Assimilation	3D-Var FGAT
Assimilation Window	24 hour
Model forcings	NCMRWF Unified Model (~12 km)
Observations Assimilated	**SST** from satellite and in-site, **SLA** from satellite, **T/S profiles** from GTS; **Sea Ice Concentration** from satellite,

(Waters et al., 2014)

Variational Data Assimilation

Observation pre-processing

↓

Background trajectory: observation Operator Step

24 hour model run which calculates the model trajectory and the mismatch between the model and observations at their correct times, the innovations.

↓

Inner Loop: 3D-VAR

Minimises and incremental formulation of the cost function which gives the increments.

↓

Incremental Analysis Update

24 hour model run where the analysis is generated by gradually applying the increments over several model time steps

Run Next Day

- NEMOVAR is a multi-variate incremental 3D-VAR, first-guess-at-appropriate-time (FGAT) data assimilation scheme (¼ degree resolution and 75 vertical levels.

(Waters et al., 2014)

State Vectors

The state vector in NEMOVAR is:

•Temperature (T)

3D-Field

• Salinity (S)

• Sea Surface height (η)

2D-Field

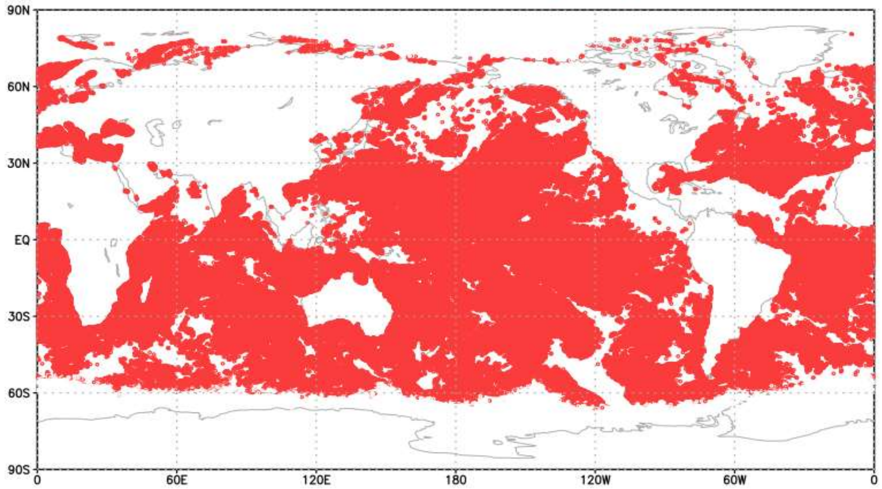
• Sea Ice

[Sea ice is currently treated as a totally unbalanced variable (as univariate)]

Velocity data (U,V) is not assimilated in NEMOVAR

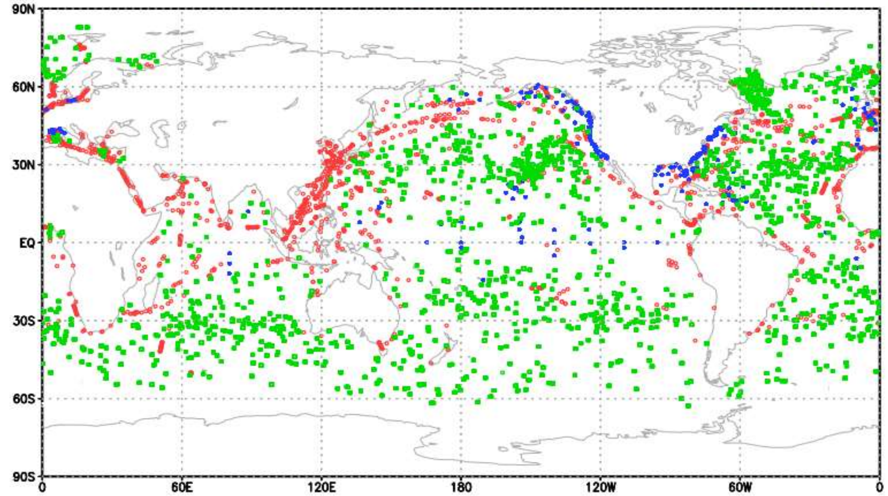
Data Coverage:Satellite SST(METOP); 20200627

Number of Observations --(574964)



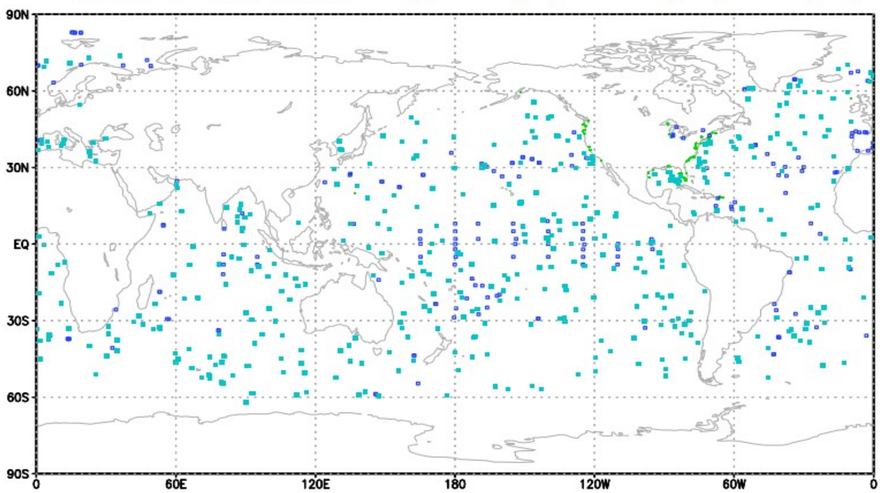
Data Coverage:Surface Observations; 20200627

Number of Observations: SHIP(2241) MOORED(21390) DRIFTER(34145)



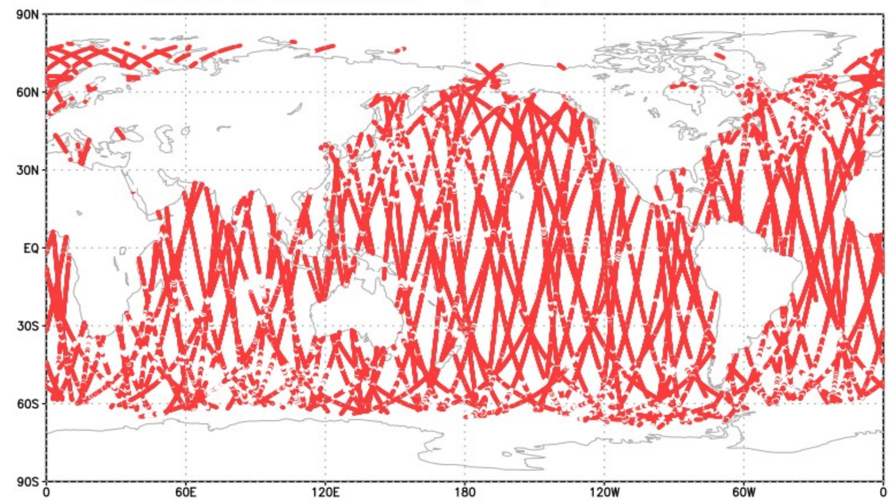
Data Coverage:Profile Observations; 20200627

Number of Observations: XBT(0) TESAC(69) MOORED(194) ARGO(451)

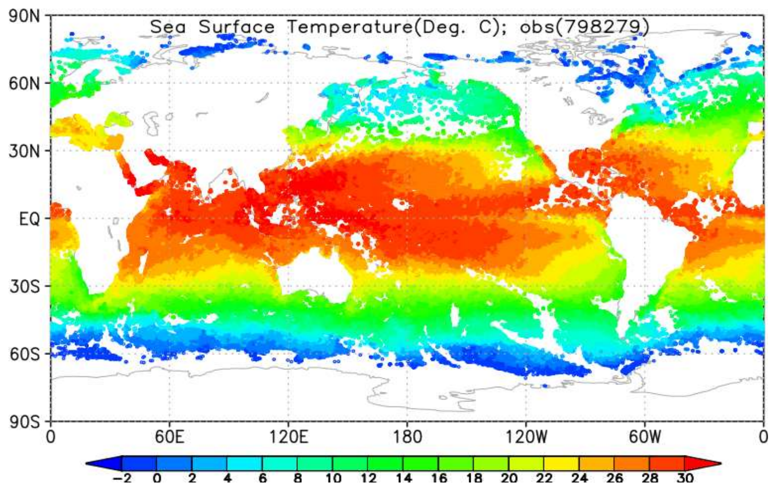


Data Coverage:SLA Observations; 20200627

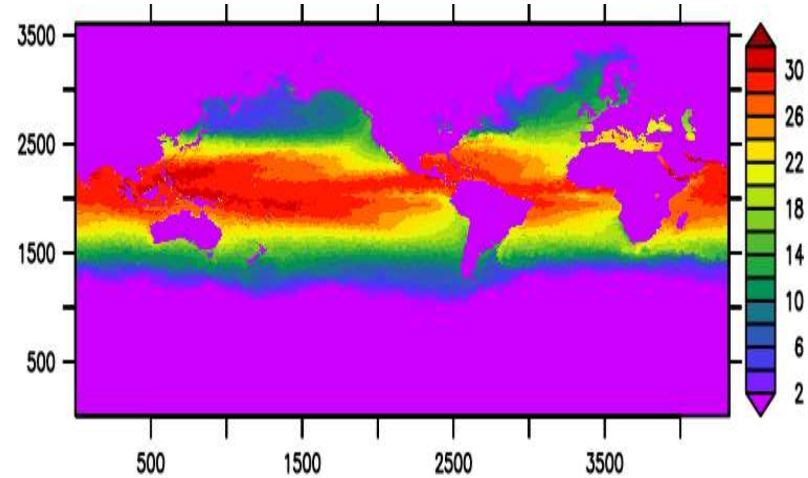
Number of Observations: SLA(90199)



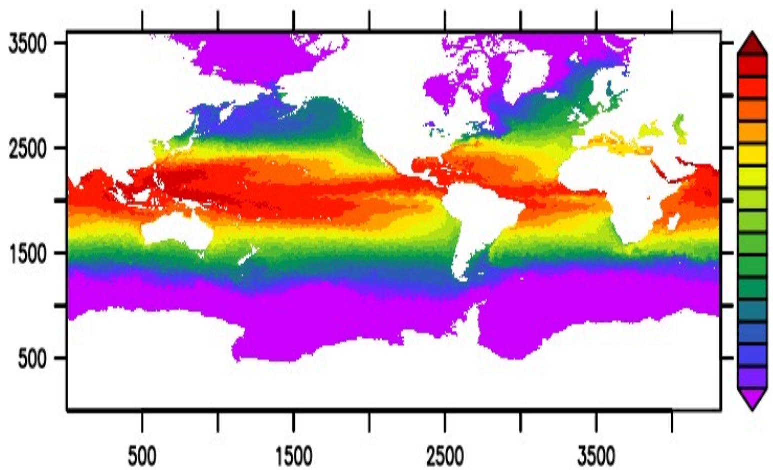
Assimilation of Sea Surface Temperature



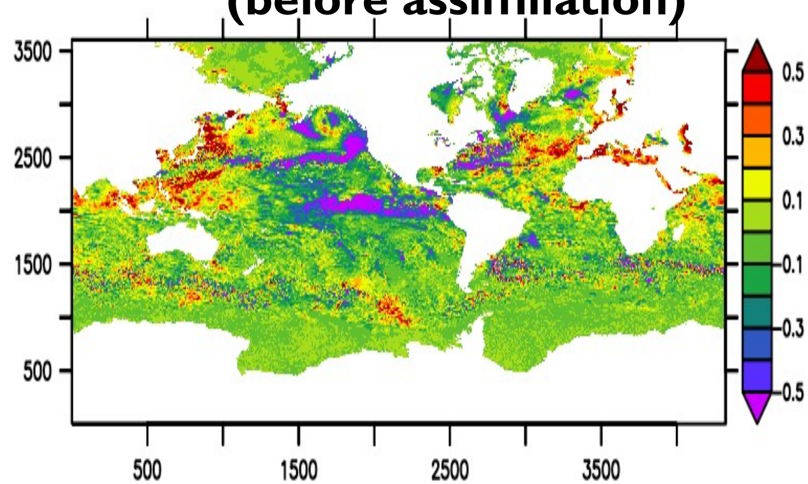
Observation



Background or Model forecast (before assimilation)



Analysis (after assimilation)



Increments (Analysis-Background)

Analysis – Background = Innovation (errors in model forecast)

Analysis – Observations = Residual (how observation fit to the background field)

Balance Operator for Ocean State Variables

Linear Balance Relationships

$$\delta T = \delta T$$

$$\delta S = K_{ST}\,\delta T + \delta S_u$$

$$\delta \eta = K_{\eta\rho}\,\delta\rho + \delta\eta_u$$

$$\delta u = K_{p\rho}\,\delta p + \delta u_u$$

$$\delta v = K_{p\rho}\,\delta p + \delta v_u$$

T-S balance

Hydrostatic balance

Geostrophic balance

Where

$$\delta\rho = \mathrm{K}_{\rho T}\,\delta T + \mathrm{K}_{\rho S}\,\delta S$$

$$\delta p = \mathrm{K}_{p\rho}\,\delta\rho + \mathrm{K}_{\rho\eta}\,\delta\eta$$

Altimetry Data Assimilation

Principles of Radar Altimeter

- Satellite Altimeter is active microwave sensor.

- Satellite Altimeter is nadir-viewing radar which emits pulses and records the travel time, magnitude and shape of the return signal (Reflection from the Earth's Surface).

- Basic measurement is distance between the satellite and the mean sea surface, surface roughness and sea surface variability.

- Altitude is defined as $h = c * t / 2$; c – speed of light, t – travel time of radar pulse

Frequencies Used for Radar Altimeters

- Ku band (13.6 GHz): It is most commonly used microwave frequency for radar altimeter (Topex/Poseidon, Jason, Envisat, and ERS etc). The availability of bandwidth, sensitivity to atmosphere and low perturbation by ionosphere are advantages of Ku band.

- Ka band (35 GHz): Ka band radar (SARAL/Altika) is better estimated Ice, rain, coastal zones, land masses and wave heights due to larger bandwidth and high power. However, the measurement is not possible due to high attenuation by the water or water vapour.

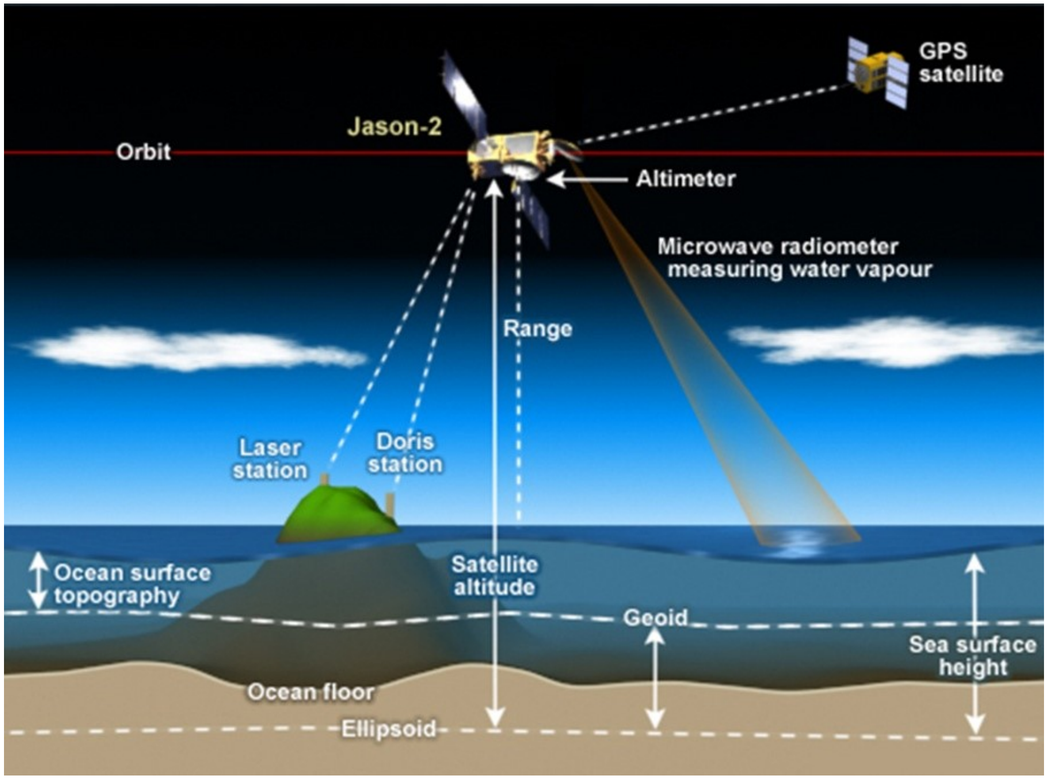
- C band (5.3 GHz): It is more sensitive to ionosphere and less sensitive to the atmosphere compare to Ku band. It main function is to correct ionosphere delay in combination with Ku band.

- S band (3.2 GHz): S band is also used in combination to Ku band to correct ionosphere delay (same as C band).

Applications of Radar Altimeter

- .Ocean circulation and Sea level variability
- Ocean surface wind speed
- Ocean wave parameters
- Mixed layer depth
- Land and Sea Ice
- Assimilation of sea level into Ocean Model
- Coastal application

Principles of Radar Altimeter



Courtesy: https://sealevel.jpl.nasa.gov/

SSH = Altitude – Range

Altitude is distance between the satellite and reference ellipsoid.

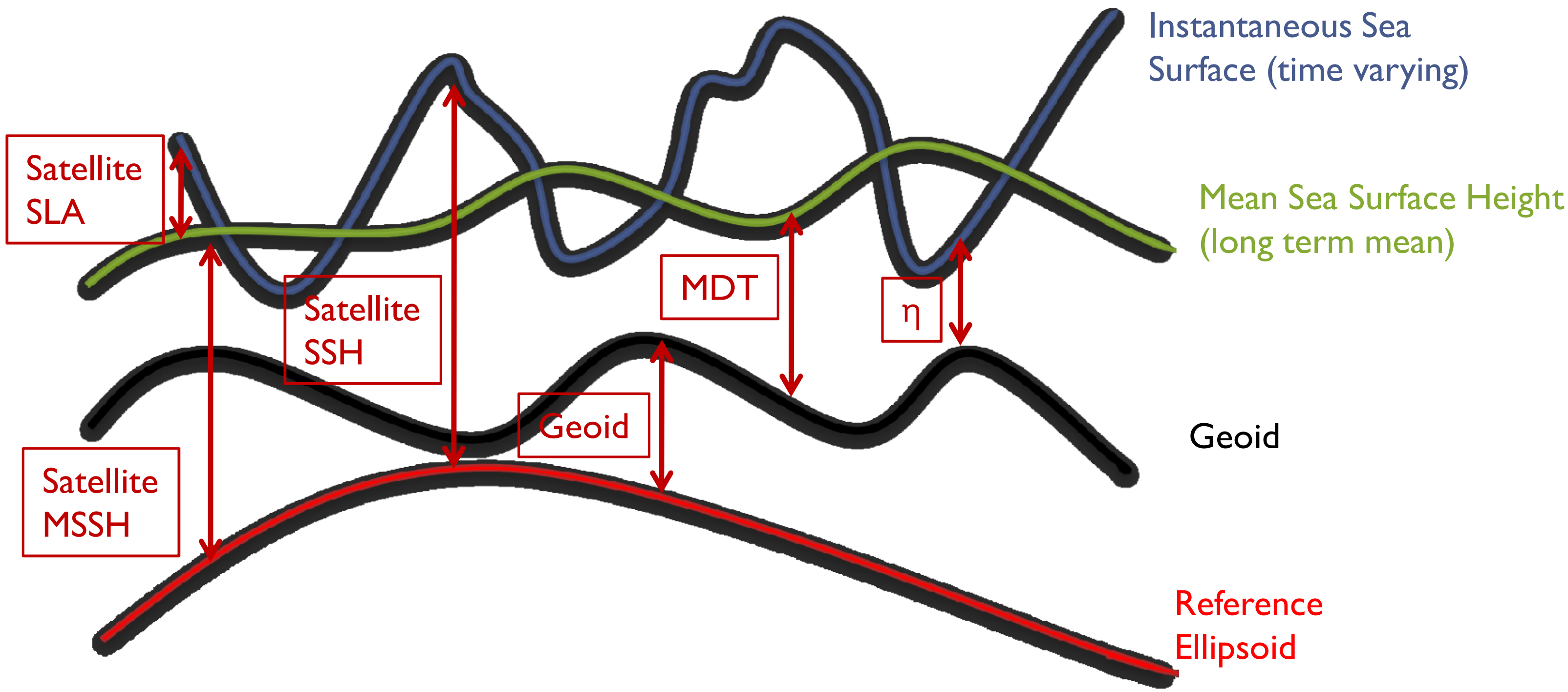
Range is the distance between the satellite and mean sea surface.

SSH is the distance between sea surface with respect to reference ellipsoid.

Reference Ellipsoid is the shape of the earth which is perfectly sphere (no gravity effect). Satellite estimates the SSH with respect to Reference ellipsoid.

Geoid is the shape of the earth under the influence of the gravity of Earth. Model estimates the SSH (η) with respect to Geoid.

Assimilation of Sea Level Anomaly



Satellite SLA = SSH – MSSH

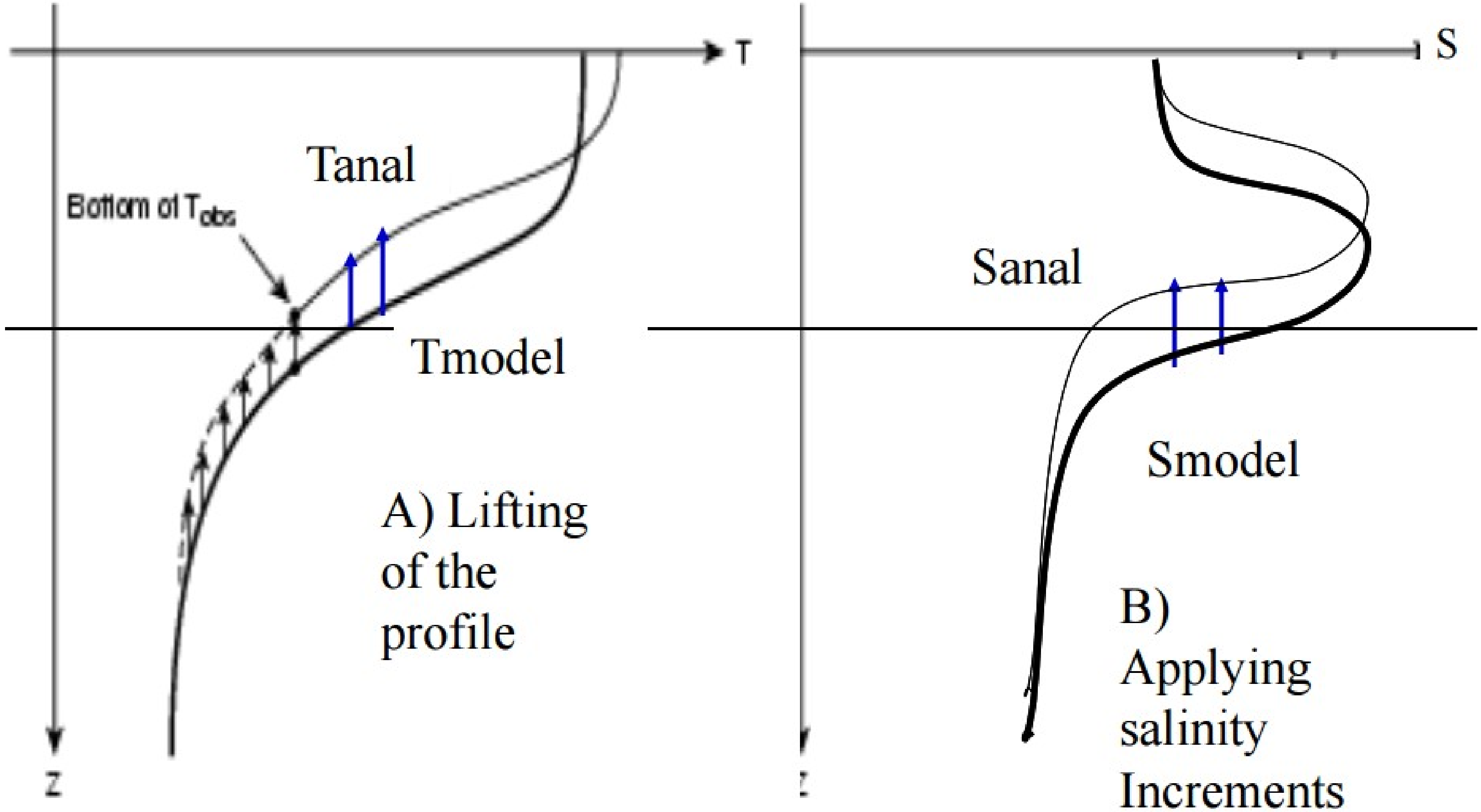
Where the MSSH is a long term mean sea surface height from the satellite altimeter.

Satellite estimates SSH with respect to Geoid (η) = **MDT** + **SLA**

Where, **MDT = MSSH - Geoid**

Geoid is estimated with help of Gravitational satellite mission (GRACE, & CHAMP).

T/S/SSH balance: Effective vertical displacement



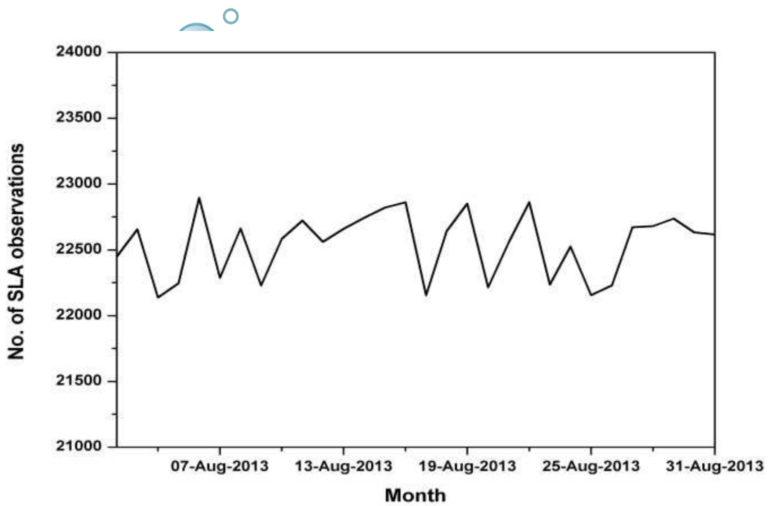
*Hydrostatic Pressure P = - rho * g * h*

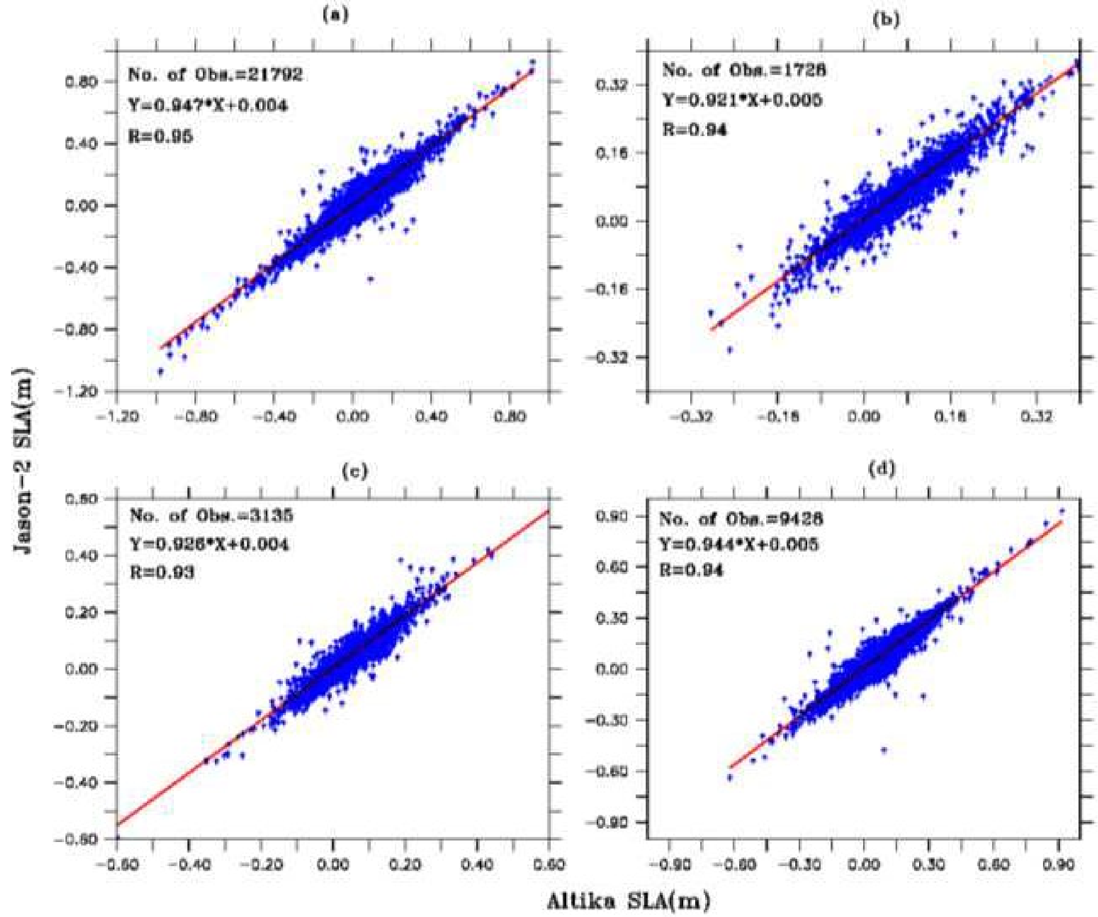
where h – surface height;   rho – density of the ocean

The sea surface height assimilation affects the density of the ocean through the variation in temperature and salinity.

Impact of Sea Level Anomaly (SLA) observations

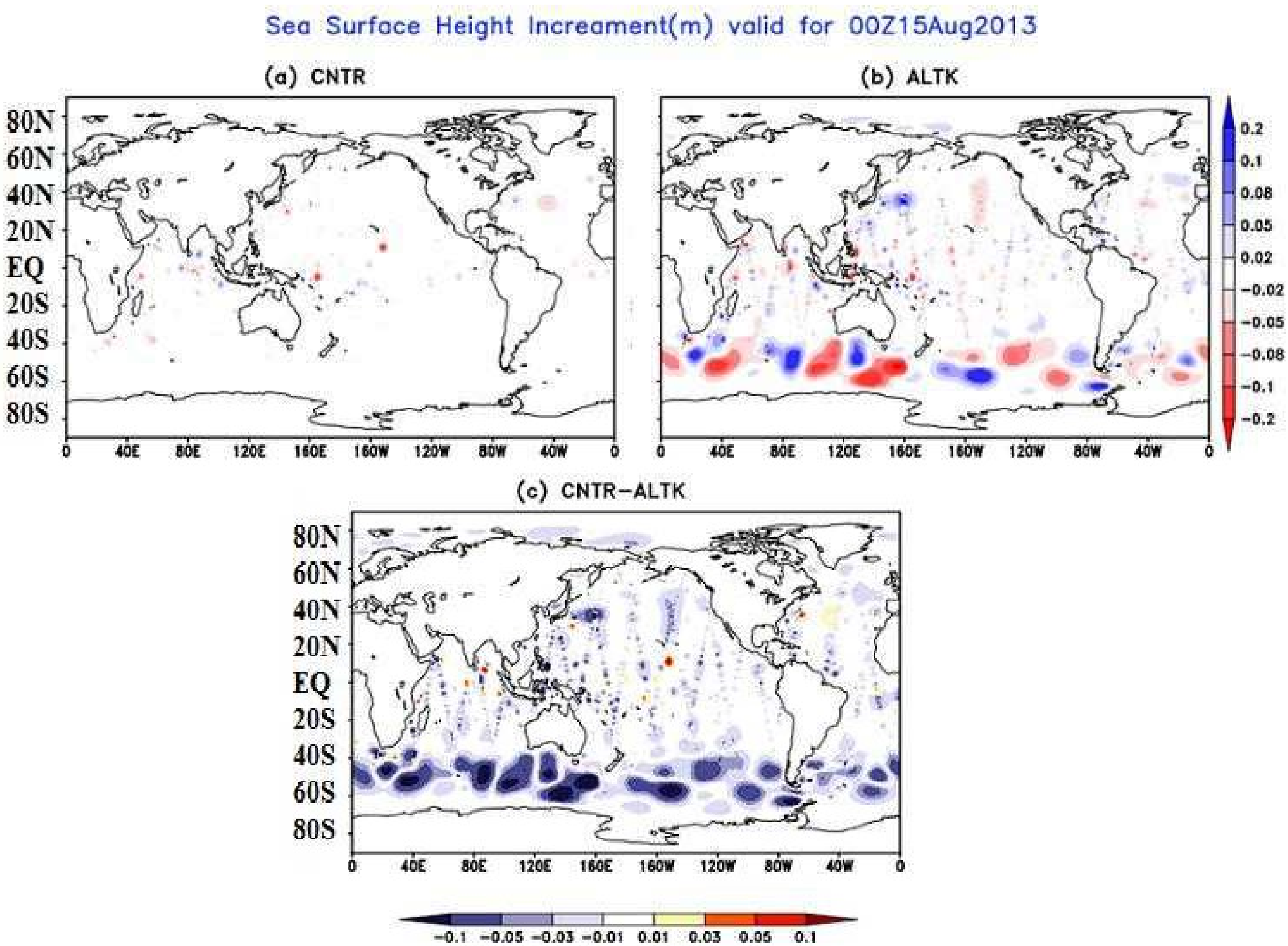
Assimilation – 01-31 August 2013



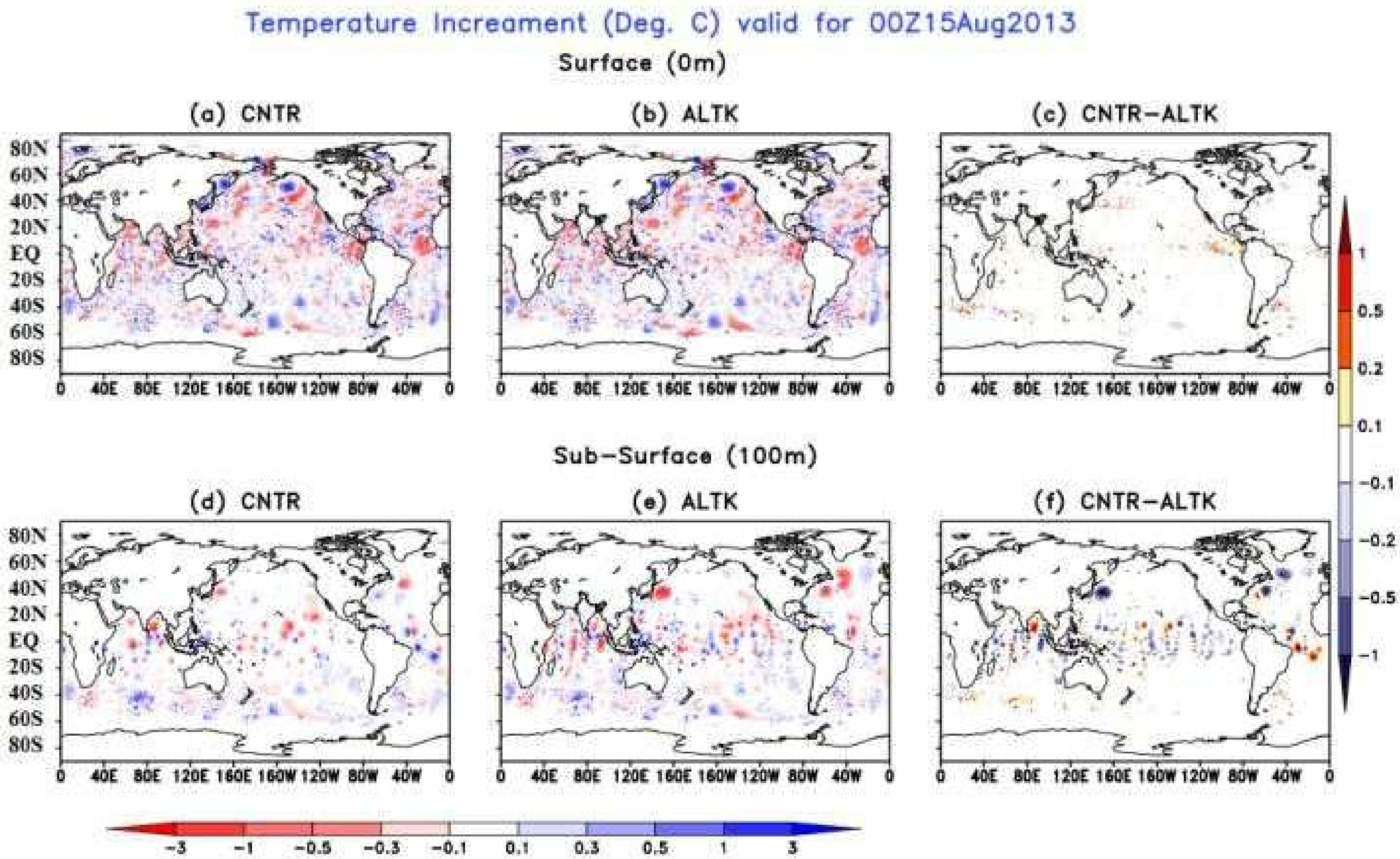


Scatter plot of Altika SLA with Jason-2 SLA for GLBO, INDO, ATLO, & PACO regions (a-d).

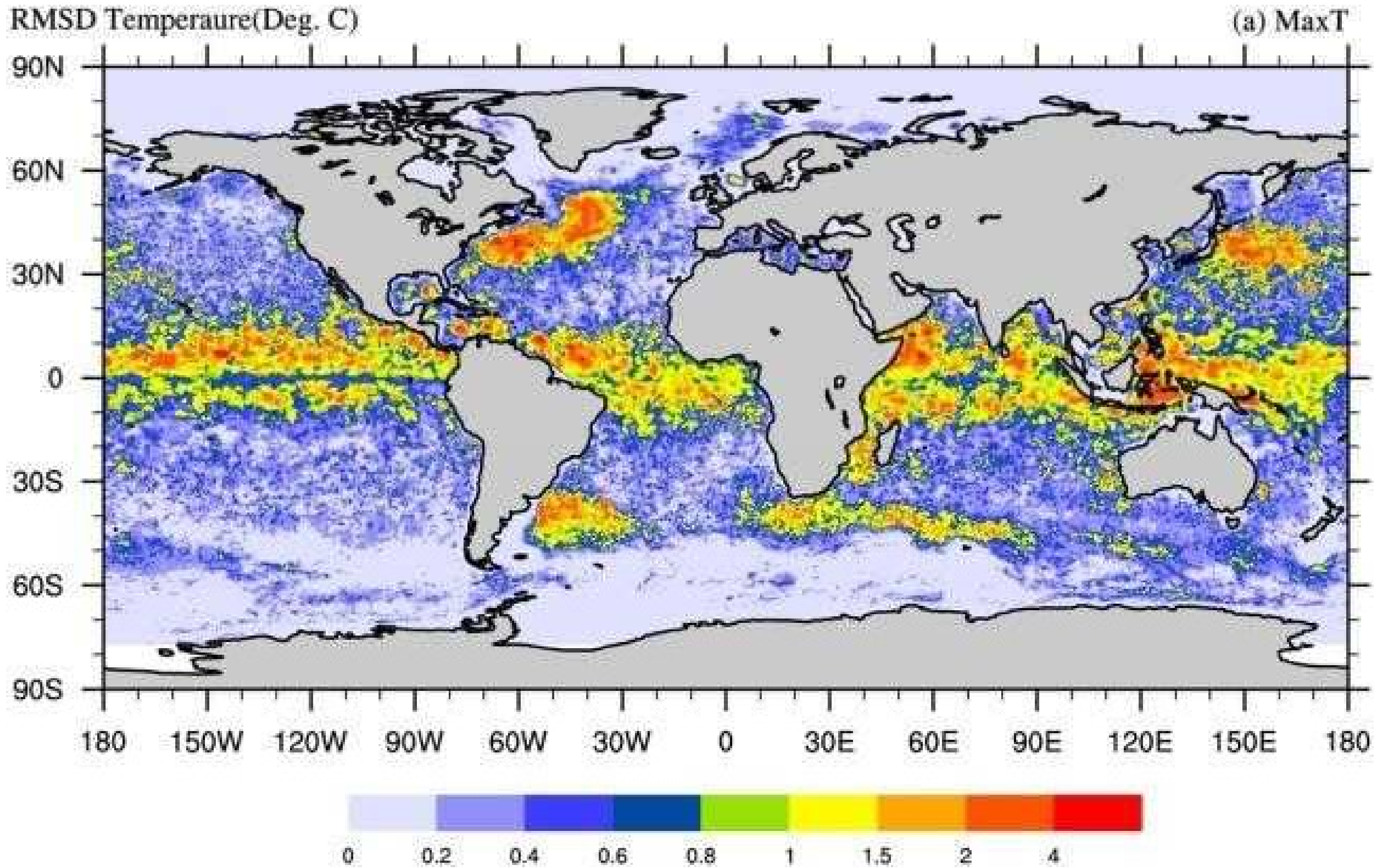
Impact of Sea Level Anomaly (SLA) observations



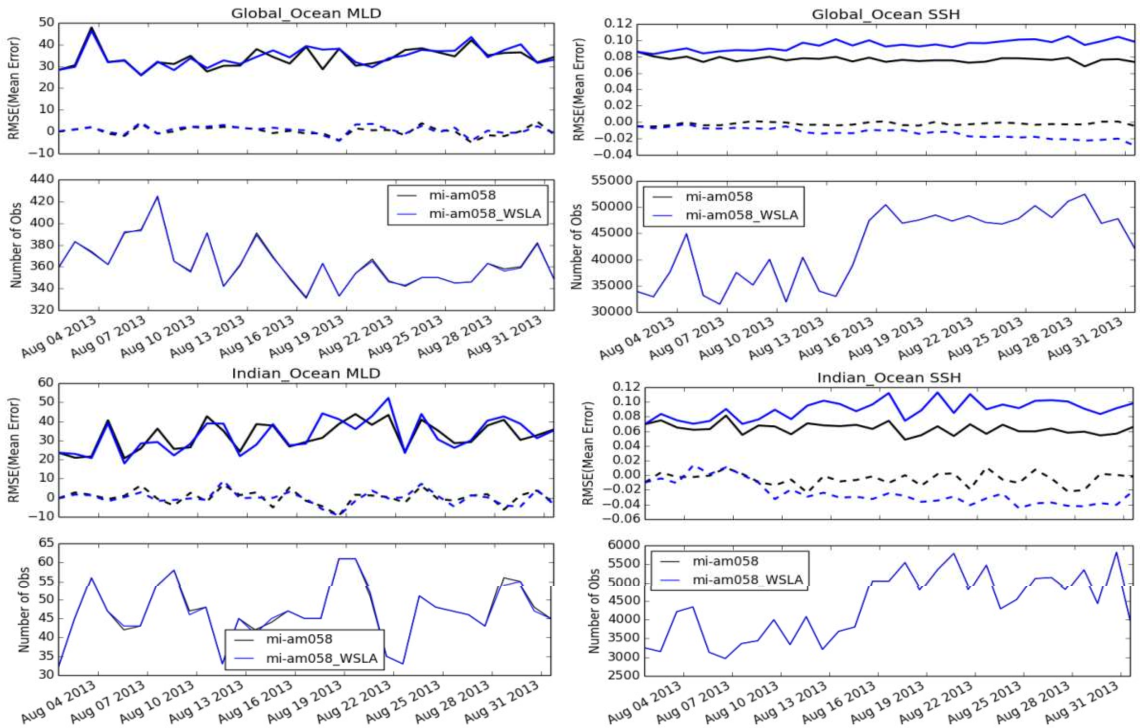
Impact of Sea Level Anomaly (SLA) observations



Impact of Sea Level Anomaly (SLA) observations



Impact of Sea Level Anomaly (SLA) observations



mi-am058: With SLA observations

mi-am058_WSLA: Without SLA

Variables	Mean error		RMSE		Std dev err		Avg N Obs
	SLA	Without SLA	SLA	Without SLA	SLA	Without SLA	
MLD	0.29	0.78	34.09	34.55	34.09	34.55	45.0
Salinity Profile	0.003	0.003	0.124	0.128	0.124	0.128	639
Temp. profile	0.019	0.016	0.655	0.66	0.654	0.66	767
SSH	-0.02	-0.014	0.076	0.095	0.076	0.094	5330
AVHRR SST	0.04	0.043	0.418	0.422	0.416	0.418	12931
In situ SST	-0.028	-0.023	0.479	0.478	0.479	0.477	4164

RMSE of MLD, Salinity, and temperature profiles are reduced after assimilation of SLA observations

Thanks