LETKF-ROMS: An Ocean Data Assimilation System

ARYA PAUL
INCOIS
Hyderabad

Workshop on Data Assimilation in Weather and Climate models, May 06-17 2024

INDIAN NATIONAL CENTRE FOR OCEAN INFORMATION SERVICES

HYDERABAD

OUTLINE

PART A

1) QUICK RECAP
2) HOW TO ESTIMATE OBSERVATION ERROR ?
3) ENSEMBLE DATA ASSIMILATION (LETKF) IN OCEAN MODEL
4) IS IT GOOD TO ASSIMILATE ALL OBSERVATIONS ?

PART B

SOME TIPS ON HOW TO IMPLEMENT LETKF

QUICK RECAP

A SIMPLE SCALAR PROBLEM

Let's estimate the temperature of this room.

$$x^a = x^b + PH^T\left(HPH^T + R\right)^{-1}\left(y - Hx^b\right)$$

Assume

$$R = \sigma_0^2\,;\; P = \sigma_b^2$$

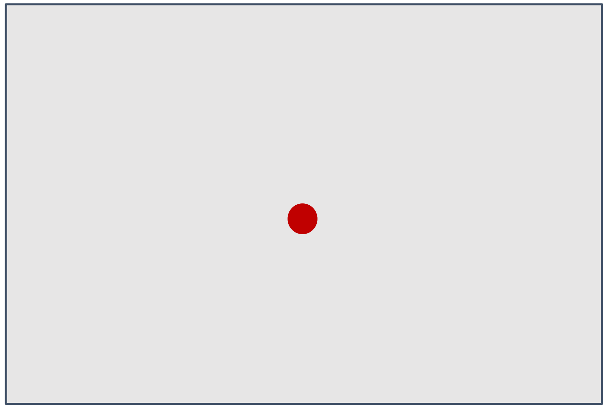
$$x^a = x^b + \sigma_b^2\left(\sigma_b^2 + \sigma_o^2\right)^{-1}\left(y_0 - x^b\right)$$

$$\Rightarrow x^a = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_b^2}x^b + \frac{\sigma_b^2}{\sigma_0^2 + \sigma_b^2}y_0$$

$$\sigma_b >> \sigma_0 \;\Rightarrow\; x^a \approx y_0$$

$$\sigma_0 >> \sigma_b \;\Rightarrow\; x^a \approx x^b$$

The model predicts the temp at the center of this room to be 31 deg whereas the thermometer shows 30 deg. What is the analysis ?

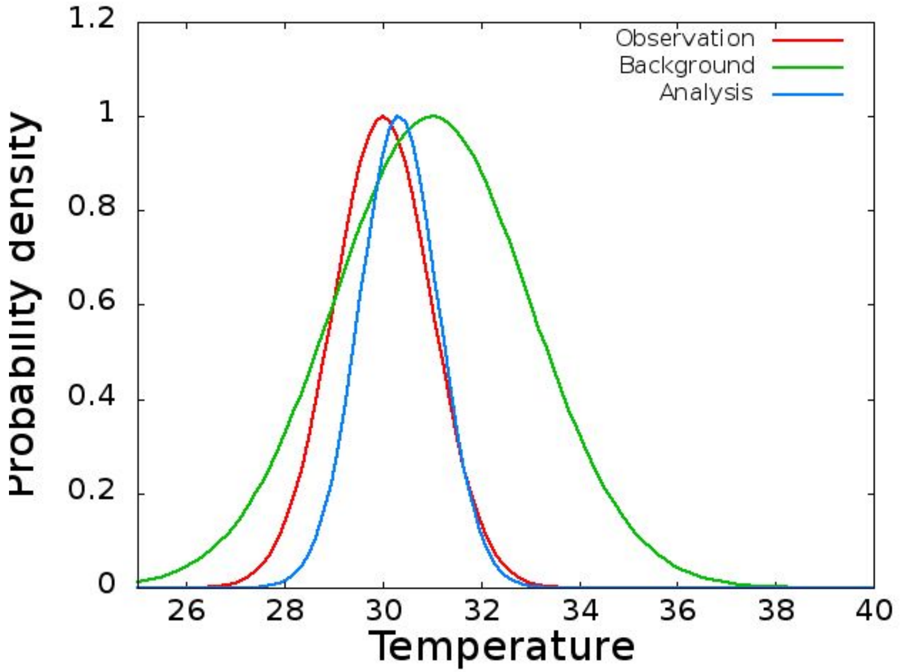


$$x^b = 31$$

$$y_0 = 30$$

$$H = 1$$

COVARIANCE INFLATION

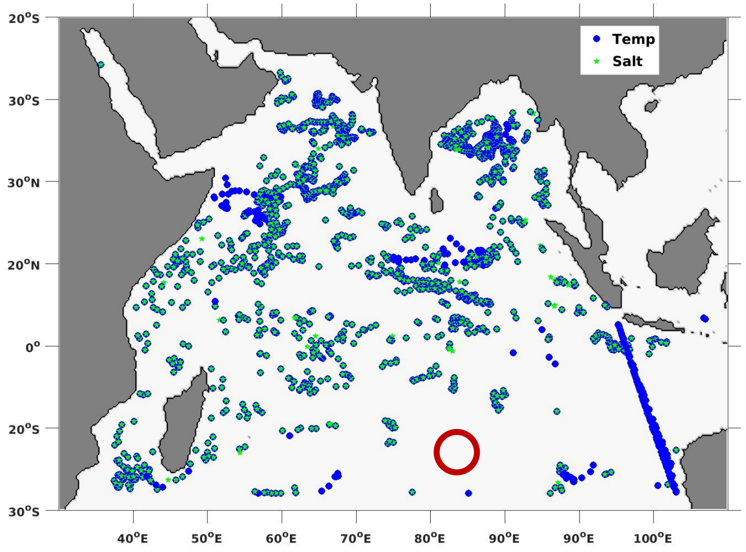


$$x^a = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_b^2} x^b + \frac{\sigma_b^2}{\sigma_0^2 + \sigma_b^2} y_0$$

$$x^b = 31.0, \sigma_b^2 = 2$$

$$y_0 = 30.0, \sigma_0^2 = 1$$

$$x^a = 30.33, \sigma_a^2 = 0.8$$



COVARIANCE INFLATION IS NECESSARY !!!

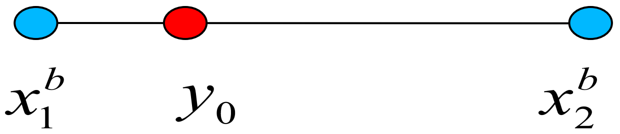
WHEN IS INFLATION NOT GOOD ?

ANS : SPARSE OBSERVATION. WHY ?

WHAT IS THE ROLE OF P ?

Suppose you model the temperature of two ends of a room but observe the temperature somewhere in between.

$$x^a = x^b + PH^T\left(HPH^T + R\right)^{-1}\left(y - Hx^b\right)$$



Suppose we observe a point in between two grid points.

$$Hx^b = \alpha x_1^b + (1-\alpha)x_2^b; \qquad 0 \le \alpha \le 1$$

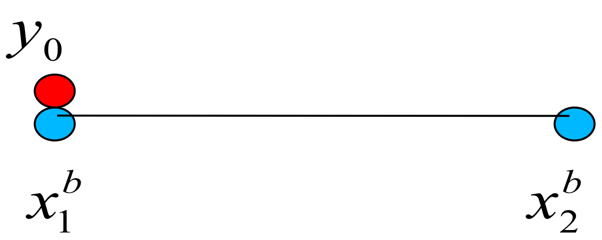
Assume

$$P = \begin{bmatrix} b_{11} & b_{12} \\ b_{21} & b_{22} \end{bmatrix} = \begin{bmatrix} \sigma_b^2 & \mu\sigma_b^2 \\ \mu\sigma_b^2 & \sigma_b^2 \end{bmatrix}; \qquad R = \sigma_0^2$$

$$\begin{pmatrix} x_1^a \\ x_2^a \end{pmatrix} = \begin{pmatrix} x_1^b \\ x_2^b \end{pmatrix} + \sigma_b^2 \begin{pmatrix} \alpha + \mu(1-\alpha) \\ \mu\alpha + (1-\alpha) \end{pmatrix} \frac{y_0 - \left[\alpha x_1^b + (1-\alpha)x_2^b\right]}{\left[\alpha^2 + 2\alpha(1-\alpha)\mu + (1-\alpha)^2\right]\sigma_b^2 + \sigma_0^2}$$

WHAT IS THE ROLE OF P ?

Case 1: No cross-correlation between two grid points, $\mu = 0$ and $\alpha = 1$



$$\begin{pmatrix} x_1^a \\ x_2^a \end{pmatrix} = \begin{pmatrix} x_1^b \\ x_2^b \end{pmatrix} + \sigma_b^2 \begin{pmatrix} \alpha + \mu(1-\alpha) \\ \mu\alpha + (1-\alpha) \end{pmatrix} \frac{y_0 - \left[\alpha x_1^b + (1-\alpha) x_2^b\right]}{\left[\alpha^2 + 2\alpha(1-\alpha)\mu + (1-\alpha)^2\right]\sigma_b^2 + \sigma_0^2}$$

$$\Rightarrow \begin{pmatrix} x_1^a \\ x_2^a \end{pmatrix} = \begin{pmatrix} x_1^b \\ x_2^b \end{pmatrix} + \sigma_b^2 \begin{pmatrix} 1 \\ 0 \end{pmatrix} \frac{y_0 - x_1^b}{\sigma_b^2 + \sigma_0^2}$$

$$x_1^a = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_b^2} x_1^b + \frac{\sigma_b^2}{\sigma_0^2 + \sigma_b^2} y_0$$

$$x_2^a = x_2^b$$

The analysis at grid point 2 is equal to the background. Observation had no effect on grid point 2.

WHAT IS THE ROLE OF P ?

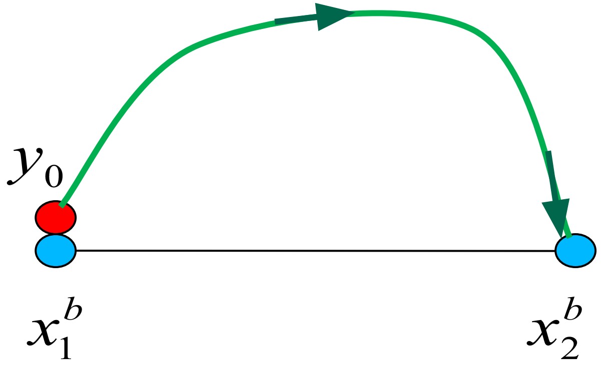
Case 2: $\alpha = 1, \quad \mu \neq 0$

$$\begin{pmatrix} x_1^a \\ x_2^a \end{pmatrix} = \begin{pmatrix} x_1^b \\ x_2^b \end{pmatrix} + \sigma_b^2 \begin{pmatrix} \alpha + \mu(1-\alpha) \\ \mu\alpha + (1-\alpha) \end{pmatrix} \frac{y_0 - \left[\alpha x_1^b + (1-\alpha) x_2^b\right]}{\left[\alpha^2 + 2\alpha(1-\alpha)\mu + (1-\alpha)^2\right]\sigma_b^2 + \sigma_0^2}$$

$$\Rightarrow \begin{pmatrix} x_1^a \\ x_2^a \end{pmatrix} = \begin{pmatrix} x_1^b \\ x_2^b \end{pmatrix} + \sigma_b^2 \begin{pmatrix} 1 \\ \mu \end{pmatrix} \frac{y_0 - x_1^b}{\sigma_b^2 + \sigma_0^2}$$

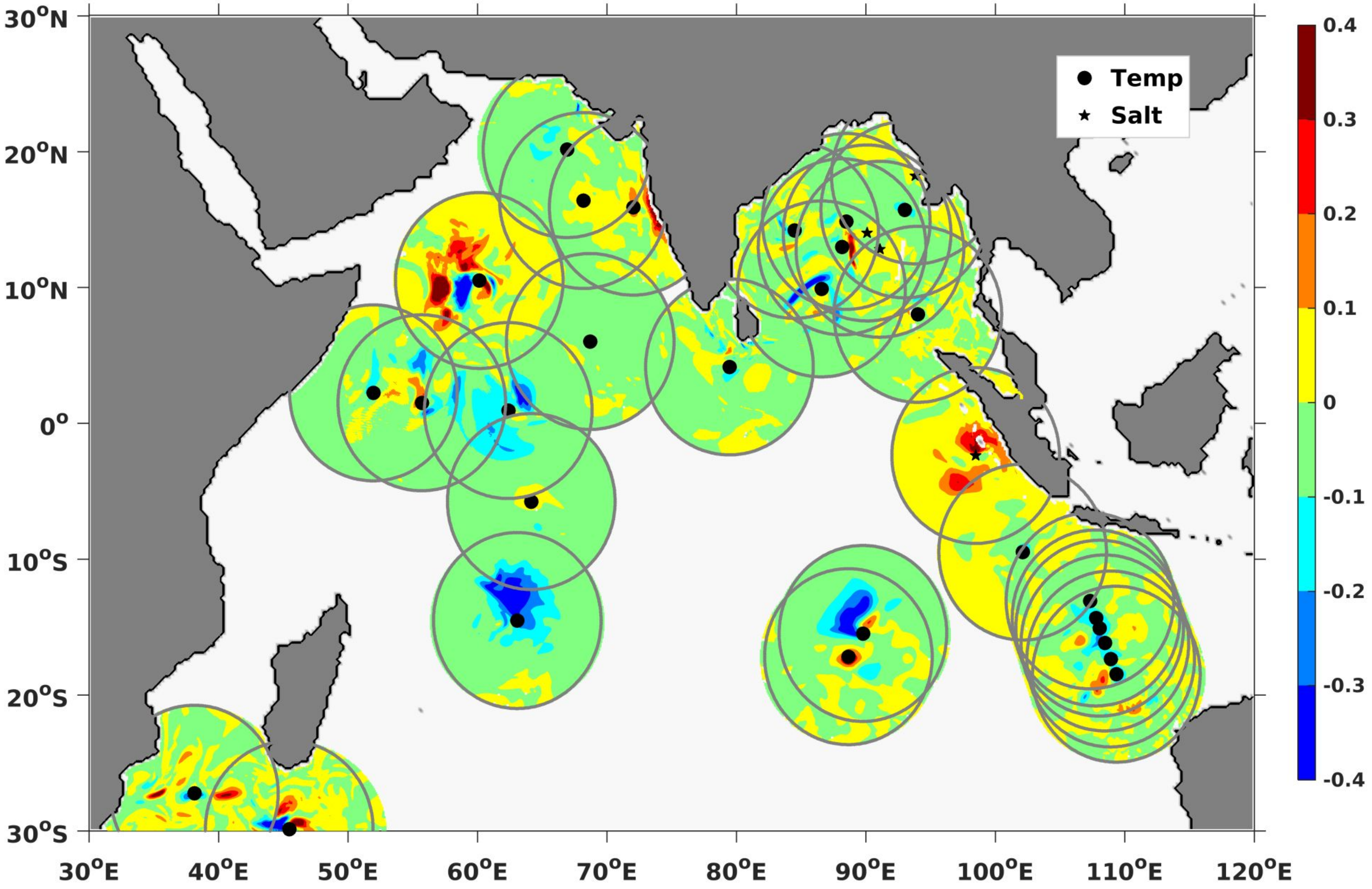
$$x_1^a = \frac{\sigma_0^2}{\sigma_0^2 + \sigma_b^2} x_1^b + \frac{\sigma_b^2}{\sigma_0^2 + \sigma_b^2} y_0$$

$$x_2^a = x_2^b + \mu\sigma_b^2 \frac{y_0 - x_1^b}{\sigma_0^2 + \sigma_b^2}$$



Now the solution at grid point 2 is influenced by the observation. The role of Background error covariance is to spread information from one grid point to the other.

IDEA OF LOCALIZATION



PLOT OF ANALYSIS - BACKGROUND TEMPERATURE

HOW TO ESTIMATE OBSERVATION ERROR ?

HOW TO ESTIMATE REPRESENTATION ERROR ?

Observation Error = Measurement Error + Representation Error

(From Instruments)

Received: 23 August 2018 | Revised: 26 July 2019 | Accepted: 20 August 2019 | Published on: 13 November 2019

DOI: 10.1002/qj.3649

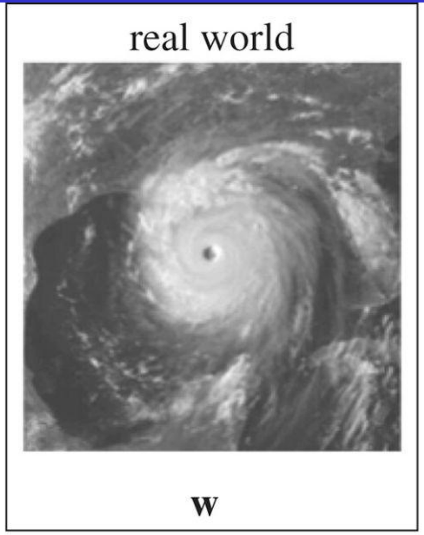
RESEARCH ARTICLE

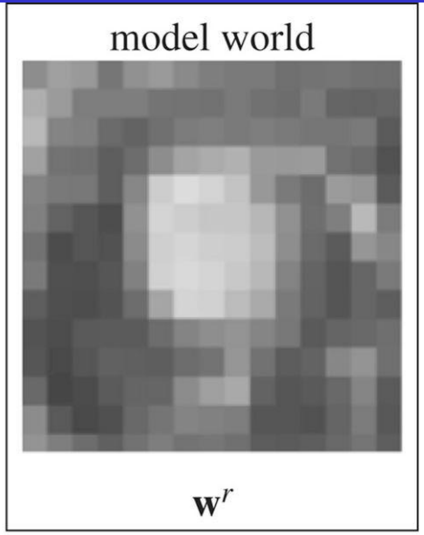
Quarterly Journal of the Royal Meteorological Society RMetS

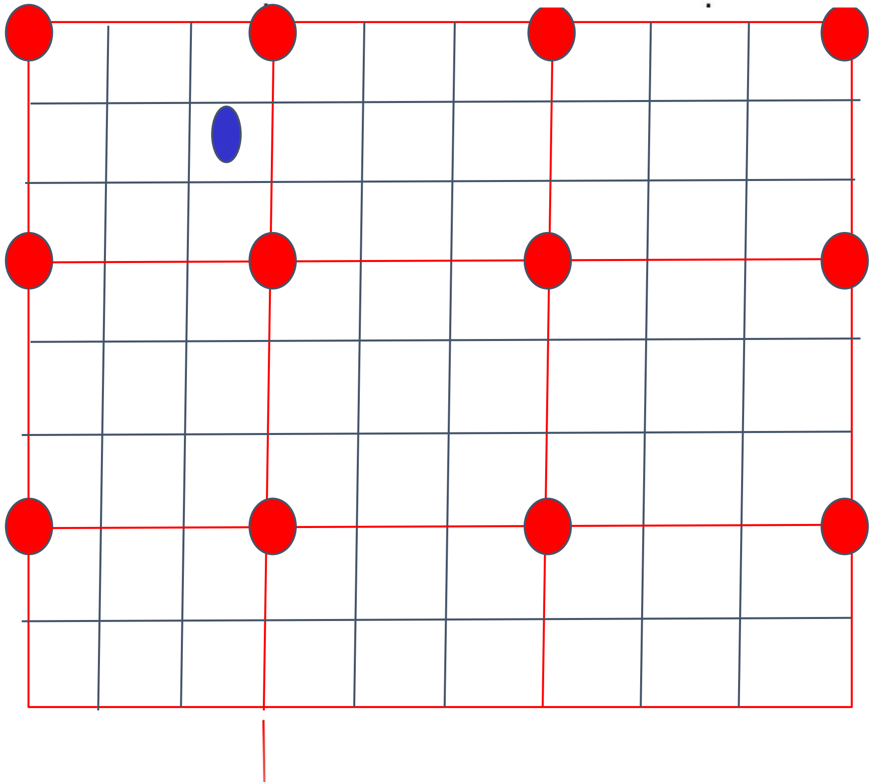
Impact of dynamical representational errors on an Indian Ocean ensemble data assimilation system

Sivareddy Sanikommu[1,2,3] | Deep Sankar Banerjee[1] | Balaji Baduru[1,4] | Biswamoy Paul[1] | Arya Paul[1] | Kunal Chakraborty[1] | Ibrahim Hoteit[2,3]

Observation Errors







How to estimate Representation Error RE ?

$$OE = Y - X^T$$

$$OE = Y - Y^T + Y^T - HX^T$$
$$= IE + RE$$

Y^T = True observation
X^T = True model state
Y = Given observation

Step 1: Take the high-resolution ($n \times n$) assimilation-free model state, Y, which plays the role of the true state of the ocean in the observation space. We implicitly assume that the chosen high-resolution assimilation-free model resolves the variability of the Indian Ocean at all scales.

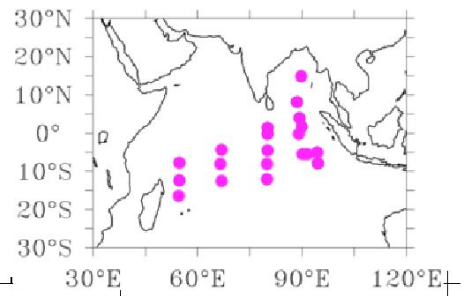
Step 2: Choose a lower-resolution grid ($m \times m$), such that m/n is an integer. Compute the spatial mean of Y on this lower-resolution grid to filter out the subgrid-scale variability and estimate the ocean state (Y_c) at the lower-resolution grid (typical re-gridding).

Step 3: Apply a simple bilinear interpolation operator P to map Y_c on the high-resolution observational space of Y.

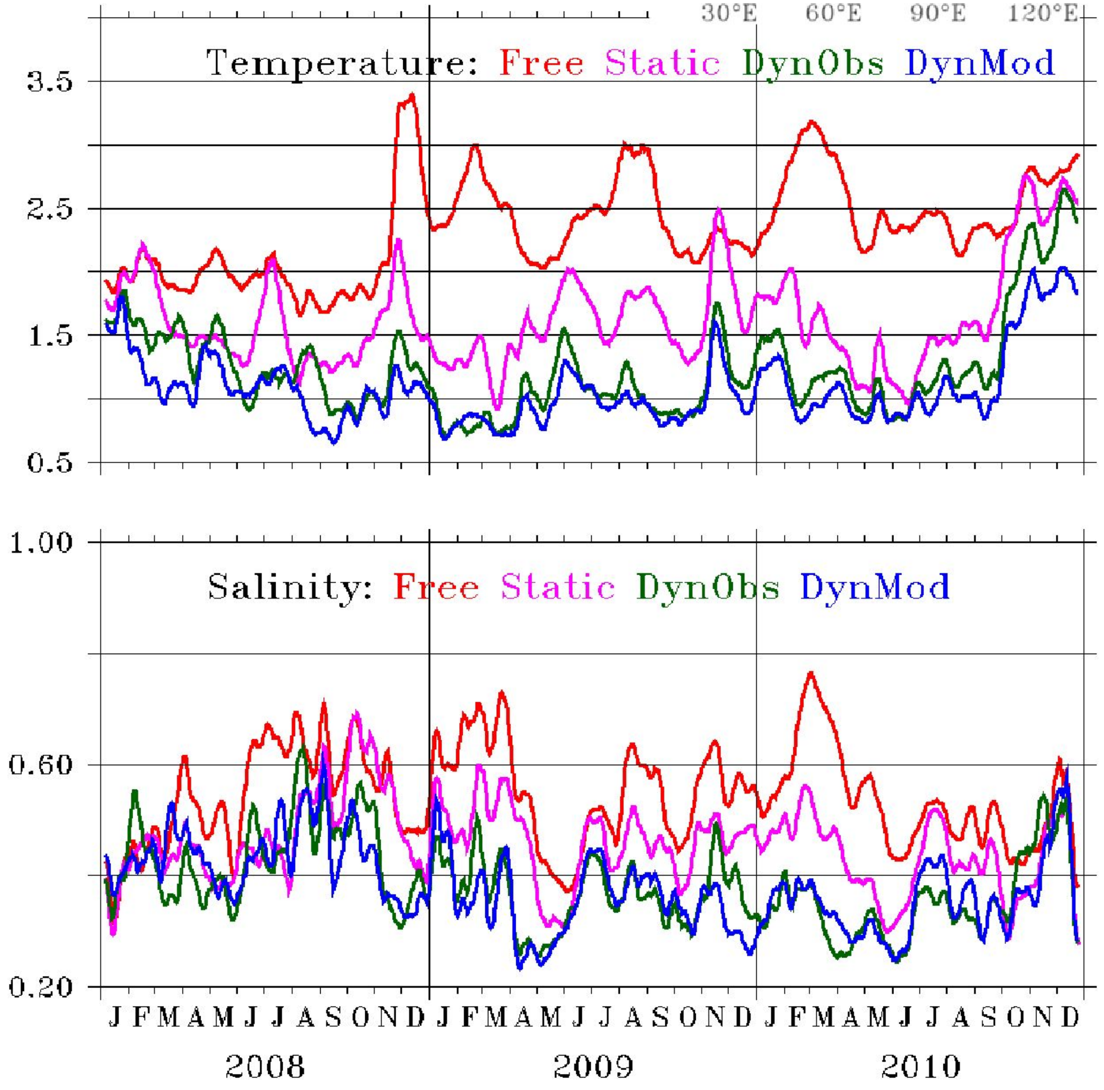
Step 4: Take the root-mean-square (RMS) of $[Y-P(Y_c)]$ using the $(m/n)^2$ points corresponding to each lower-resolution grid point to estimate RE at the low-resolution $m \times m$ grid.

1) https://doi.org/10.1002/qj.3130
2) https://doi.org/10.1002/qj.3649

Another method to estimate RE



RMSE



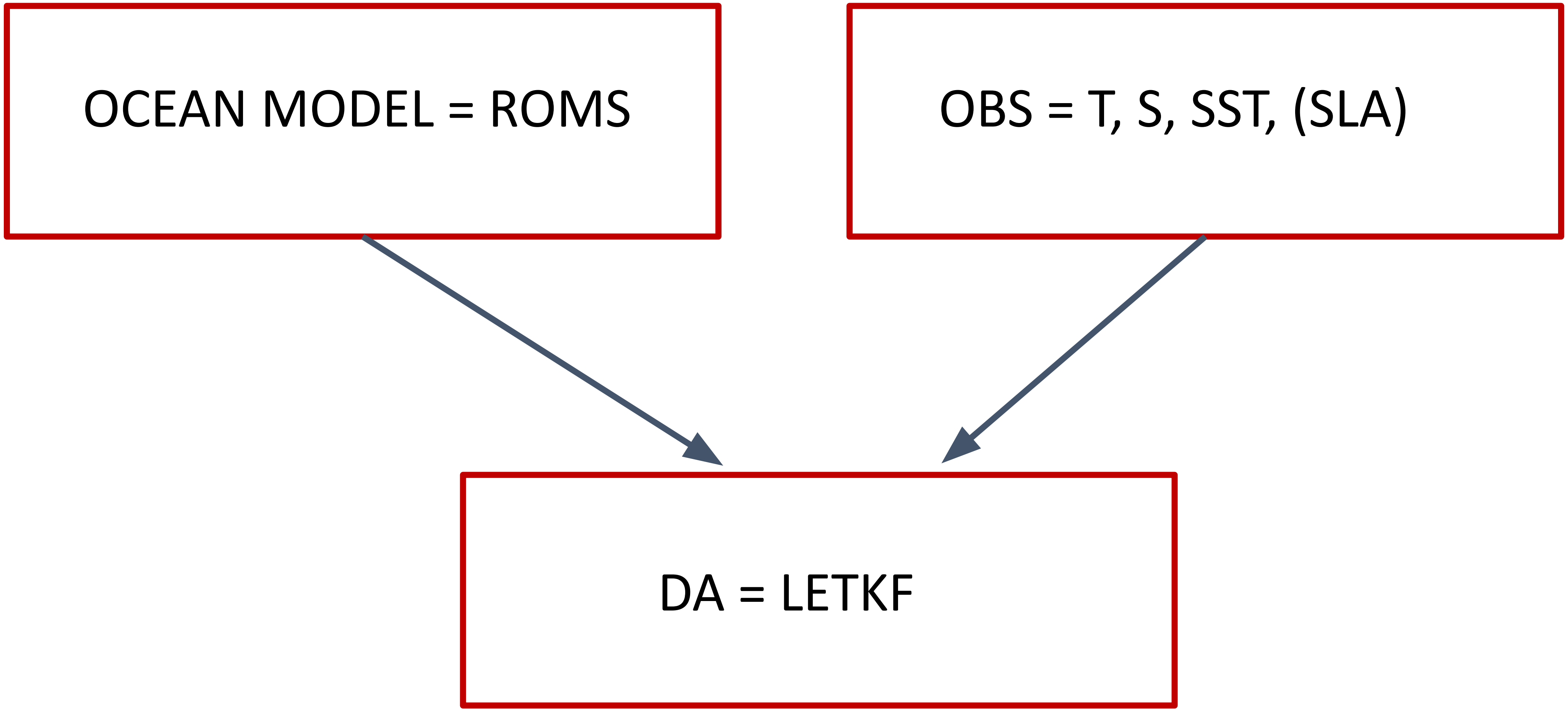
This method requires only vertical gradient of tracer from observations. No models needed.

RE = SF X (tracer_grad - min_tracer_grad)/max_tracer_grad

Behringer, D.W., Ji, M. and Leetmaa, A. (1998) An improved coupled model for ENSO prediction and implications for ocean initialization. Part I: The ocean data assimilation system. Monthly Weather Review, 126(4), 1013–1021.

ENSEMBLE DA IN OCEAN MODEL - A PRACTICAL APPLICATION

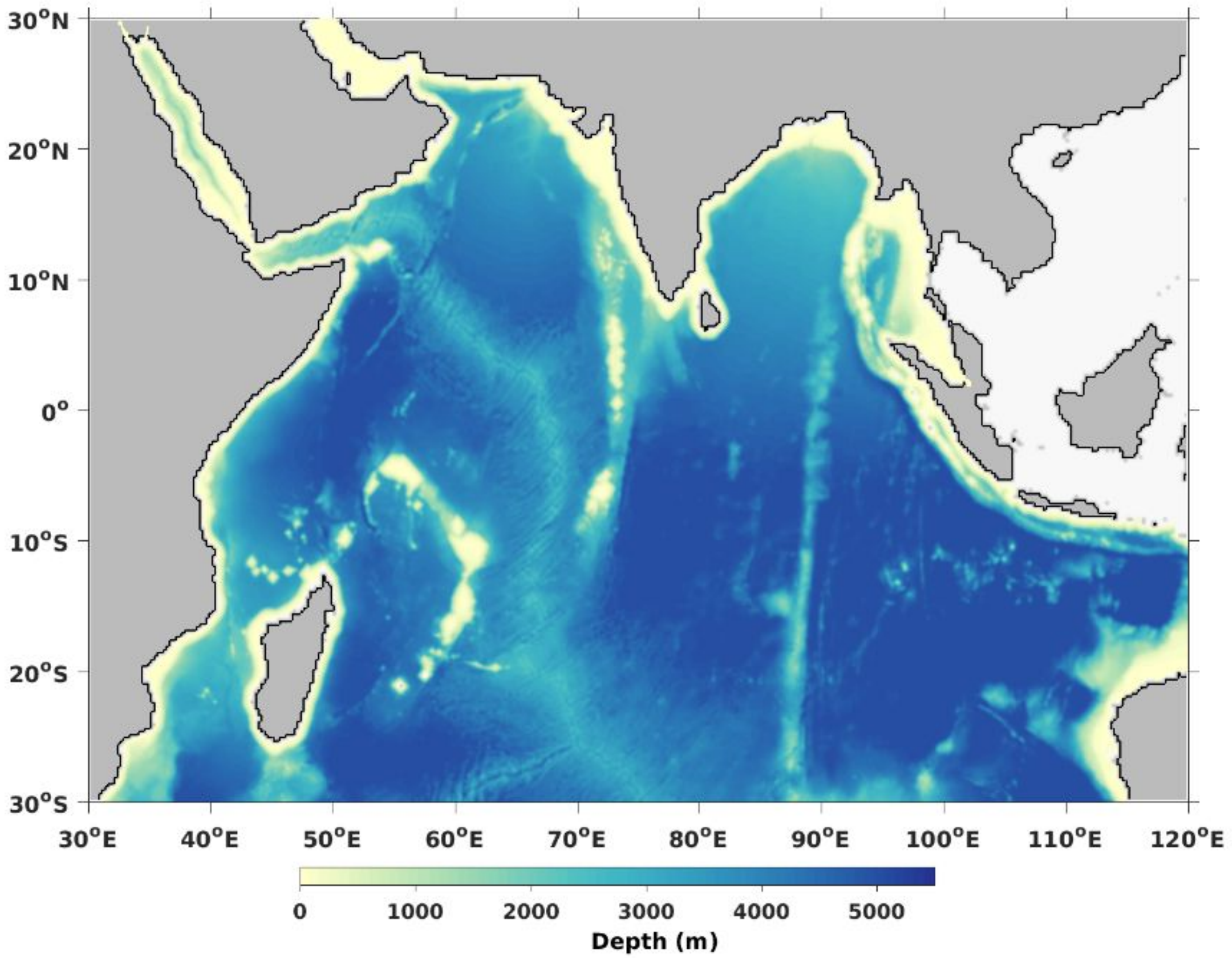
OCEAN DA



Assimilation Scheme :: **Local Ensemble Transform Kalman Filter (LETKF)**
No. of Ensembles :: 80

MODEL Specifications

Model Domain



Model :: Regional Ocean Modeling System (ROMS)

DOMAIN:
30ºE to 120ºE ; 30ºS to 30ºN

RESOLUTION:
1/12º (~ 9km, Horizontal)
40 sigma levels (Vertical)

BOUNDARY CONDITIONS:
Derived from INCOIS-GODAS.

ATMOSPHERIC FLUX:
NCMRWF flux from GFS model.

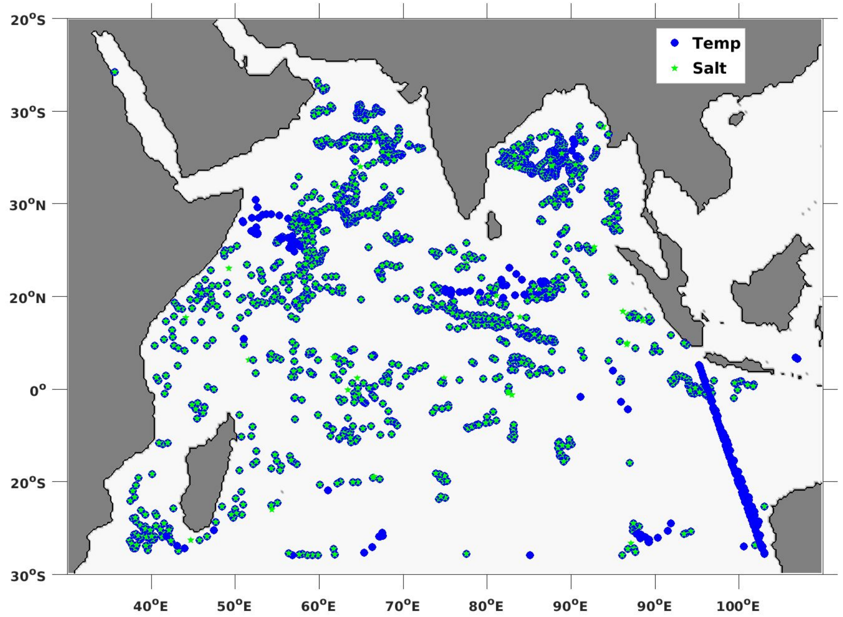
Observations

Assimilated Variables

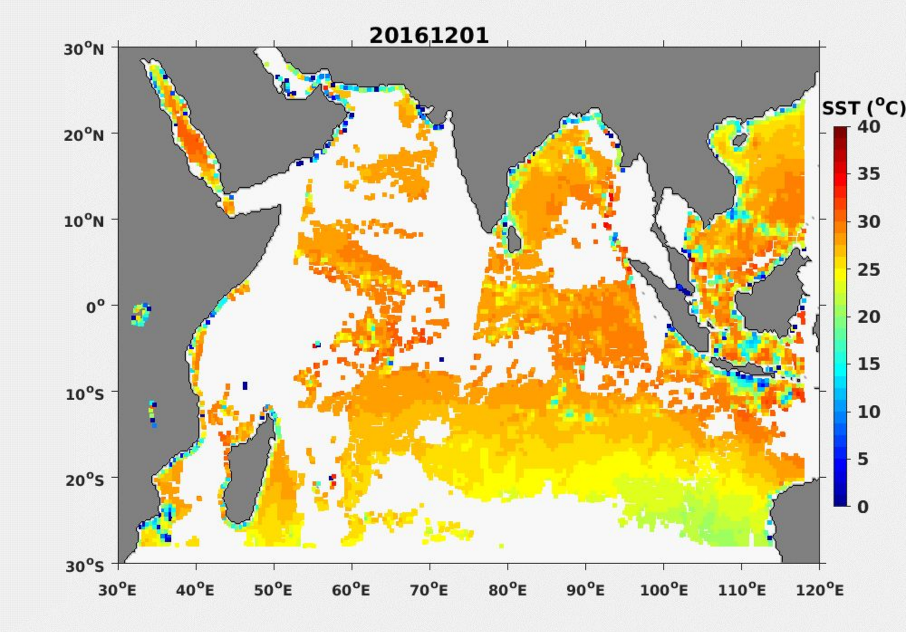
1. In-situ Temperature
2. Salinity Profiles
3. Sea Surface Temperature

Independent Variables

1. Sea Level Anomaly
2. Sea Surface Salinity
3. U,V Currents



Spatial Distribution of Assimilated Observations (Temp and Salinity) for Aug 2016- Aug 2017



Assimilated satellite track SST over Indian Ocean for Dec 2016

Local Ensemble Transform Kalman Filter (LETKF)

Physica D: Nonlinear Phenomena
Volume 230, Issues 1–2, June 2007, Pages 112-126

Efficient data assimilation for spatiotemporal chaos: A local ensemble transform Kalman filter

Brian R. Hunt [a], Eric J. Kostelich [b], Istvan Szunyogh [c]

https://doi.org/10.1016/j.physd.2006.11.008

LOCAL ENSEMBLE TRANSFORM KALMAN FILTER (LETKF)

LETKF also minimizes the same cost function

$$J(x) = (x - \bar{x}_b)^T P_b^{-1} (x - \bar{x}_b) + [y^0 - H(x)]^T R^{-1} [y^0 - H(x)]$$

but does so in a **local** space.

It projects (**transforms**) the model state into a local space spanned by the ensemble members.

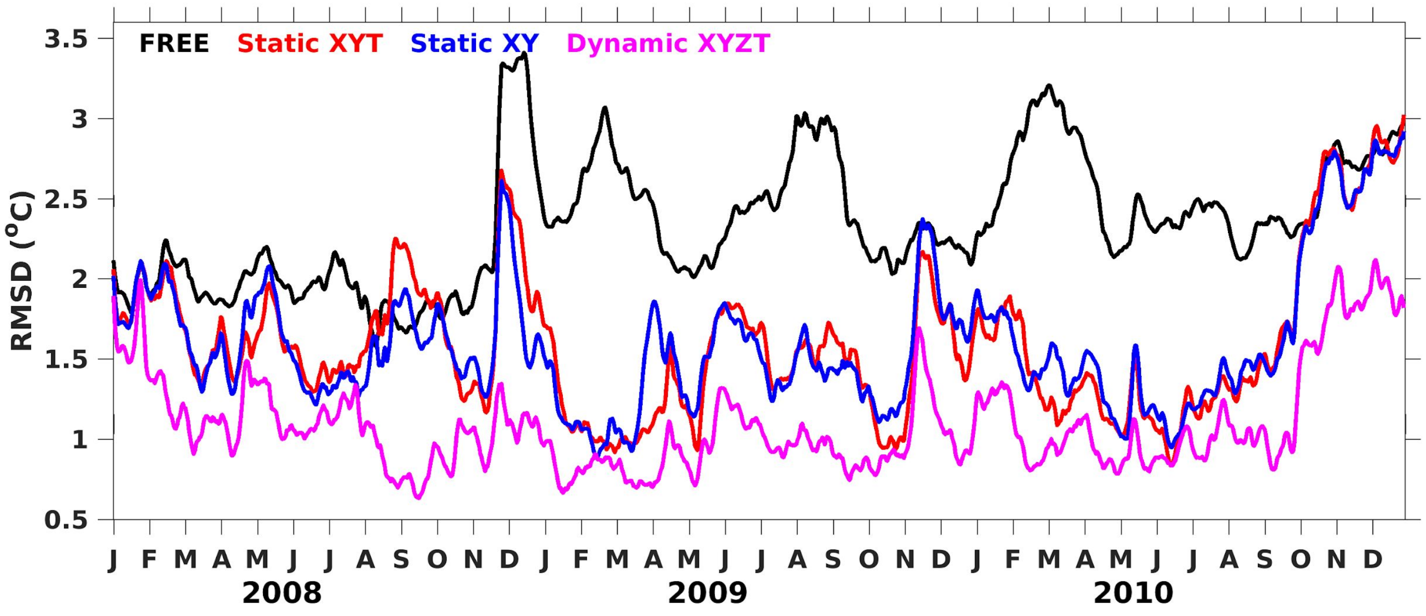
Salient features of our LETKF system

(a) It has 80 ensemble members forced by 80 different atmospheric fluxes.
(b) Use of two different mixing schemes. Done to arrest filter divergence.
(c) Some model parameters are varied across ensemble members. Done to arrest filter divergence.
(d) Spatio-temporal RE used.
(e) Assimilation window is 5 days.
(f) Analysis is available from 2016 Aug onwards. It is named as **R**egional **A**nalysis of **IN**dian Ocean (**RAIN**).

Novelty in the system

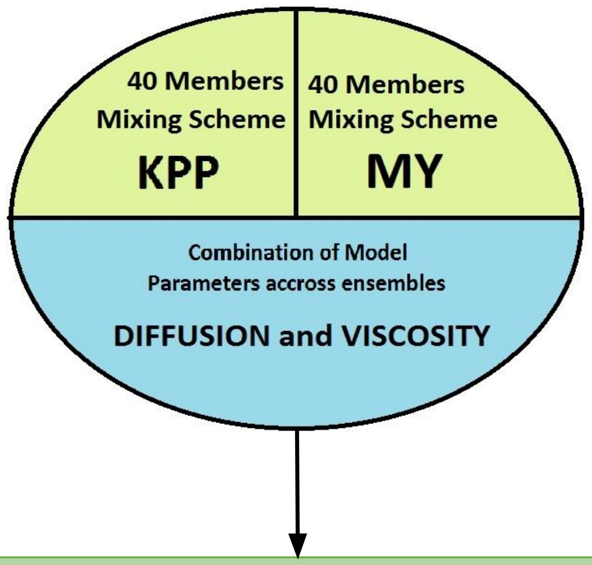
Introduction of spatio-temporal Representational Error

Time Series of RMSD in SST



Impact of dynamical representational errors on an Indian Ocean ensemble data assimilation system, Siva Reddy et al,
Quarterly Journal of Royal Meteorological Society, 2019. DOI: 10.1002/qj.3649

Introduction of two mixing schemes across ensembles

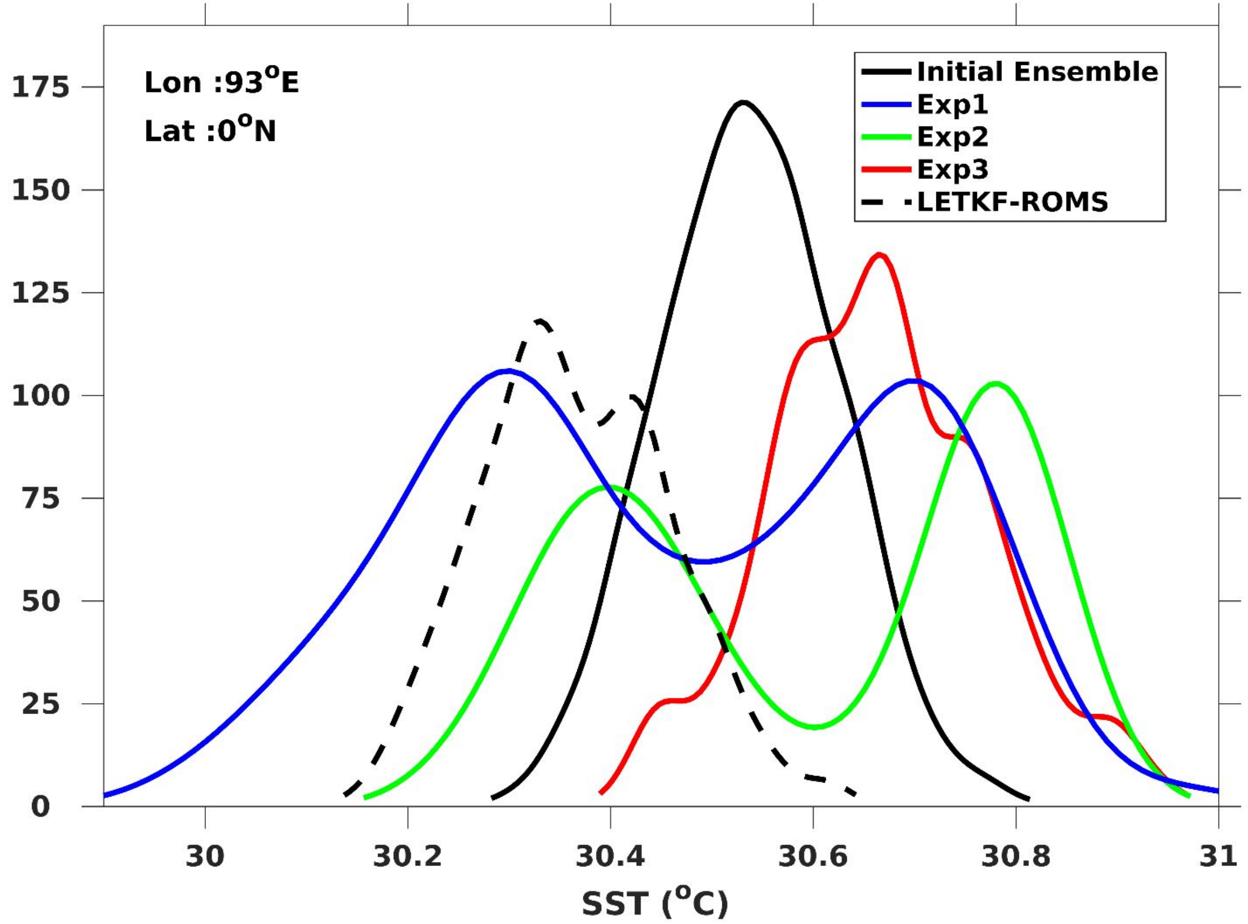


Ensemble combination with use of different mixing schemes and model parameters like diffusion and viscosity coefficients

Ensemble based regional ocean data assimilation system for the Indian Ocean: Implementation and evaluation, Balaji et al, Ocean Modeling, 2019,
https://doi.org/10.1016/j.ocemod.2019.101470

Bimodal ensemble

SST distribution across the ensemble members



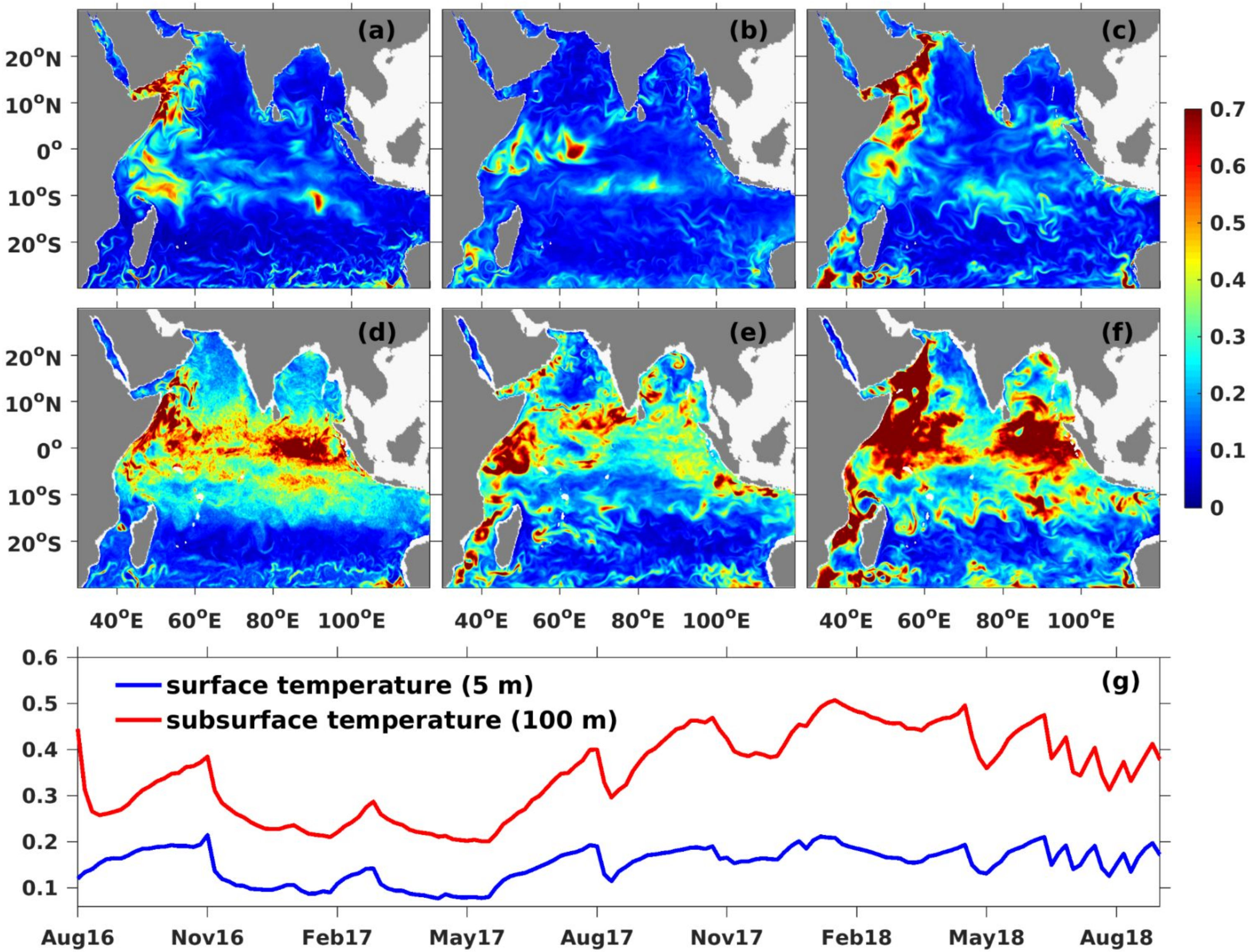
Exp 1: Same as LETKF-ROMS (no assm); ensemble flux

Exp 2: Same as LETKF-ROMS (no assm); identical flux

Exp 3: Same as exp 1; all ensemble members respond to KPP

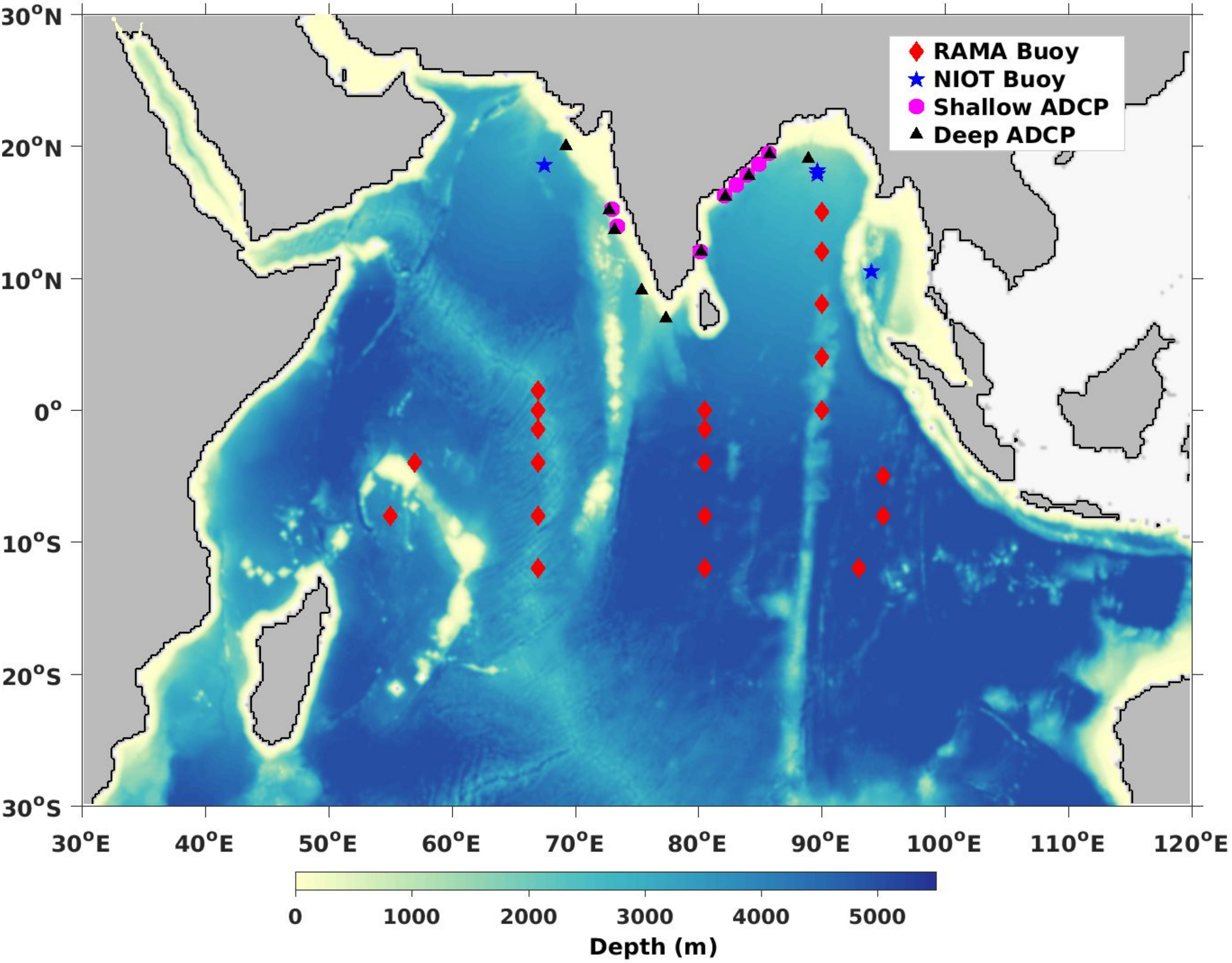
Initial ensemble (black) and after 600 days of run of three control experiments (exp1 — blue, exp2 —green and exp3 — red) and LETKF-ROMS (dashed black)

Ensemble spread

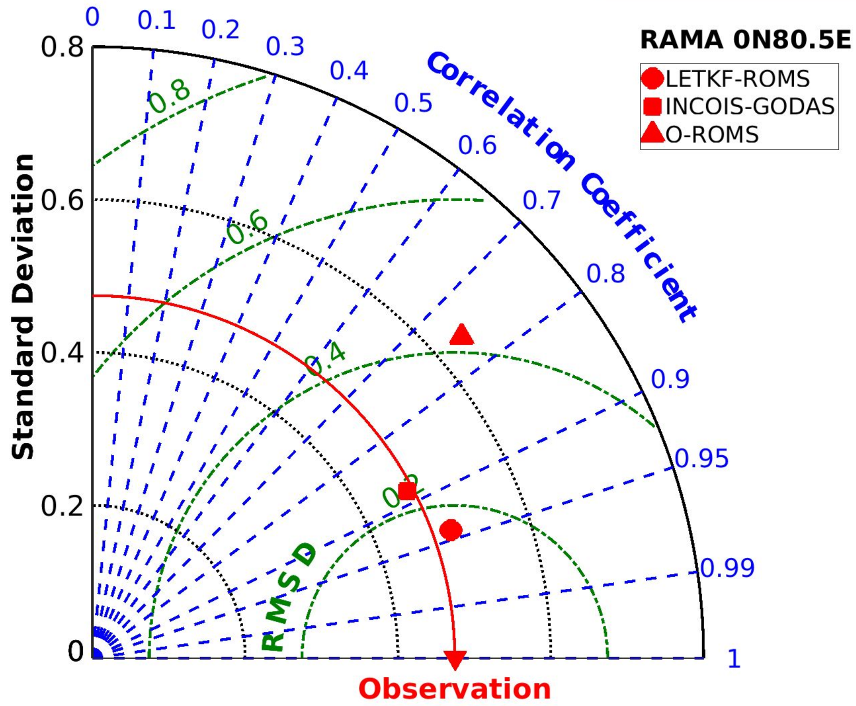


Spatial ensemble spread on 1st September 2016, 1st March 2017 and 1st September 2017 of the surface (5 m) temperature ((a), (b) and (c) respectively) and subsurface (100 m) temperature ((d), (e) and (f) respectively).
(g) Domain-averaged time series of spread in temperature at 5 m (blue) and 100 m (red) depth.

Analyses



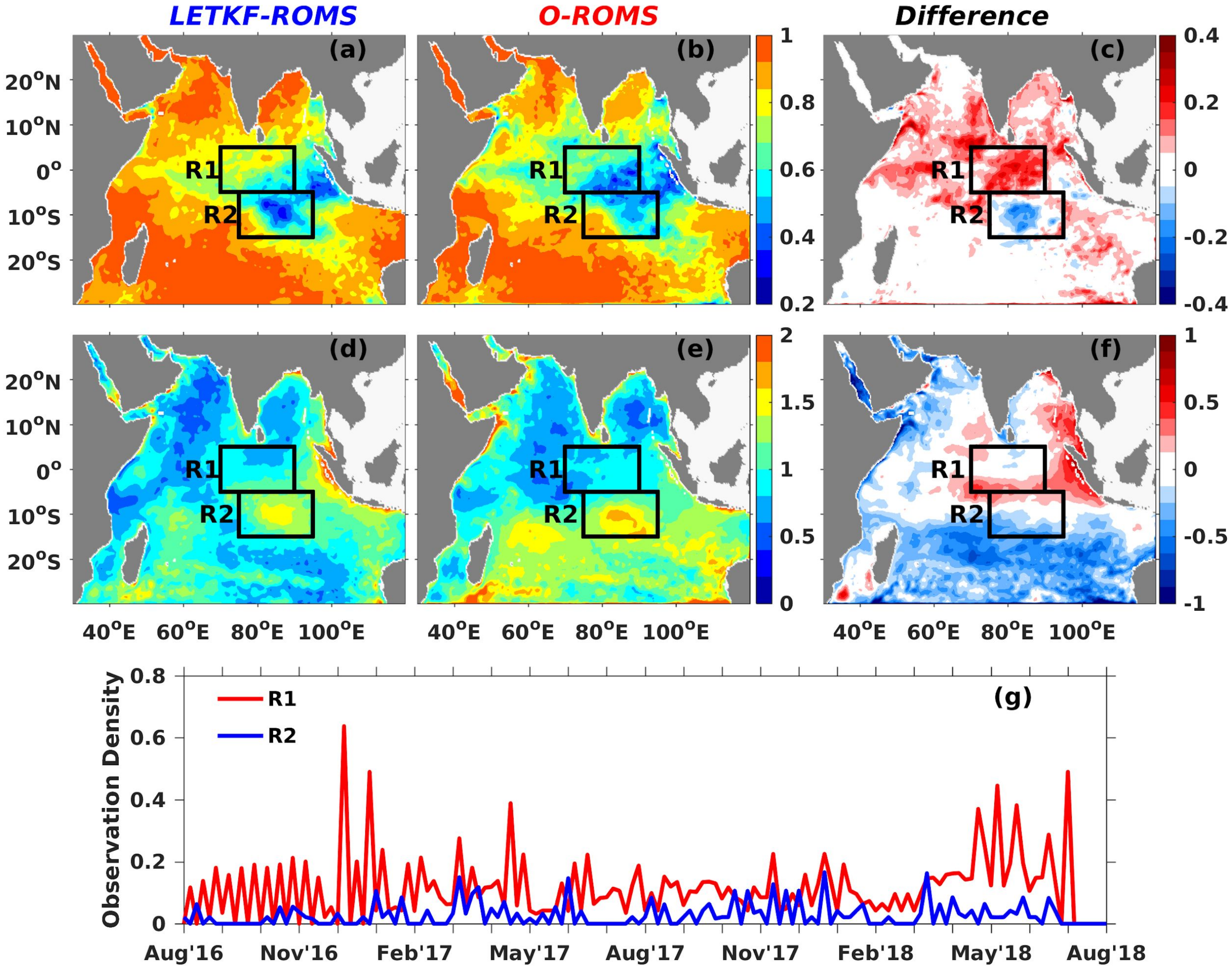
Location of in-situ observations (RAMA, NIOT, ADCP) used for comparison and validation of the analysis



Taylor Diagram of SST from LETKF-ROMS, INCOIS-GODAS and O-ROMS with respect to RAMA

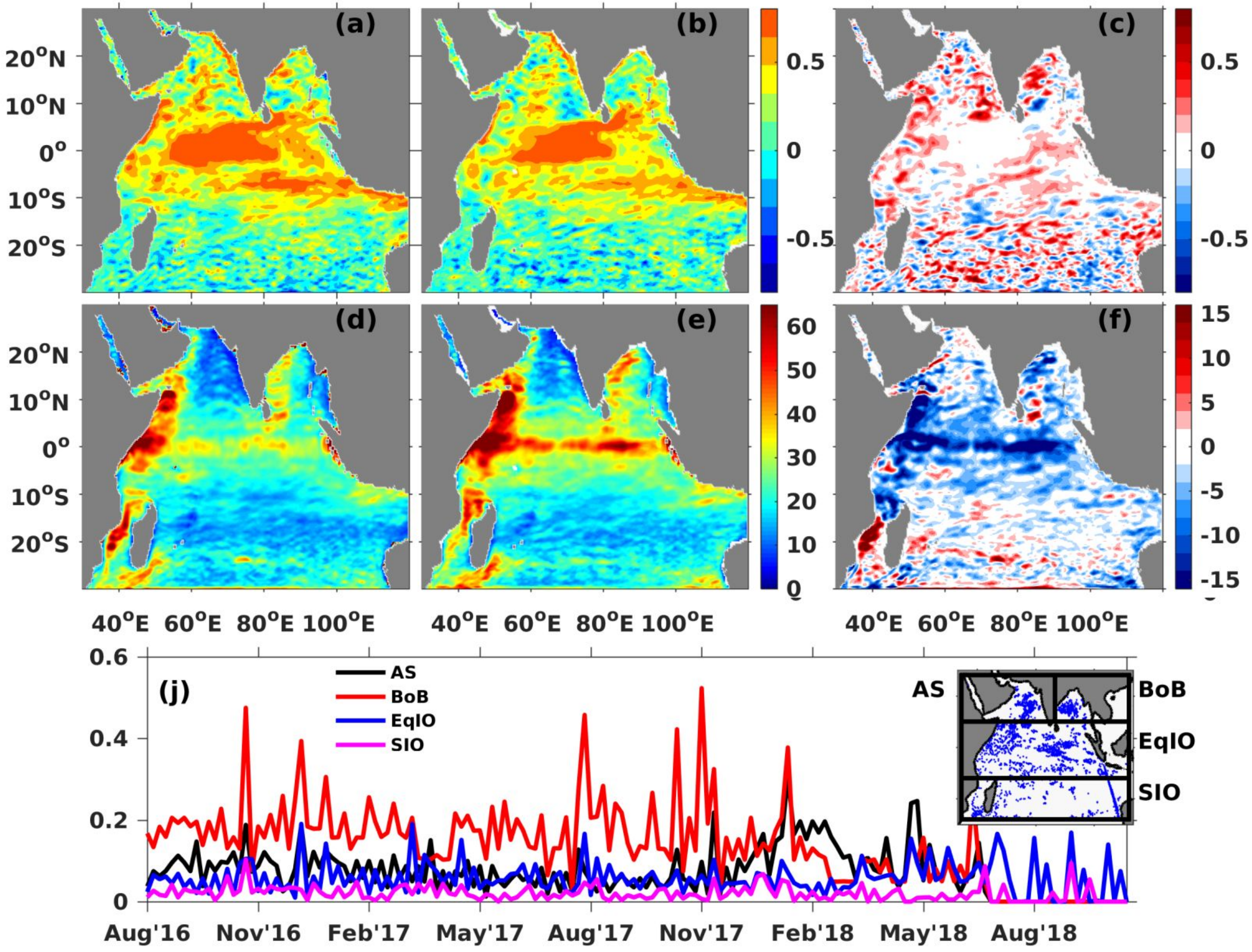
Comparisons at all locations is available in Technical Report. (Balaji et al., 2018)
http://www.incois.gov.in/documents/TechnicalReports/ESSO-INCOIS-MDG-TR-03 (2018).pdf

SST Analysis



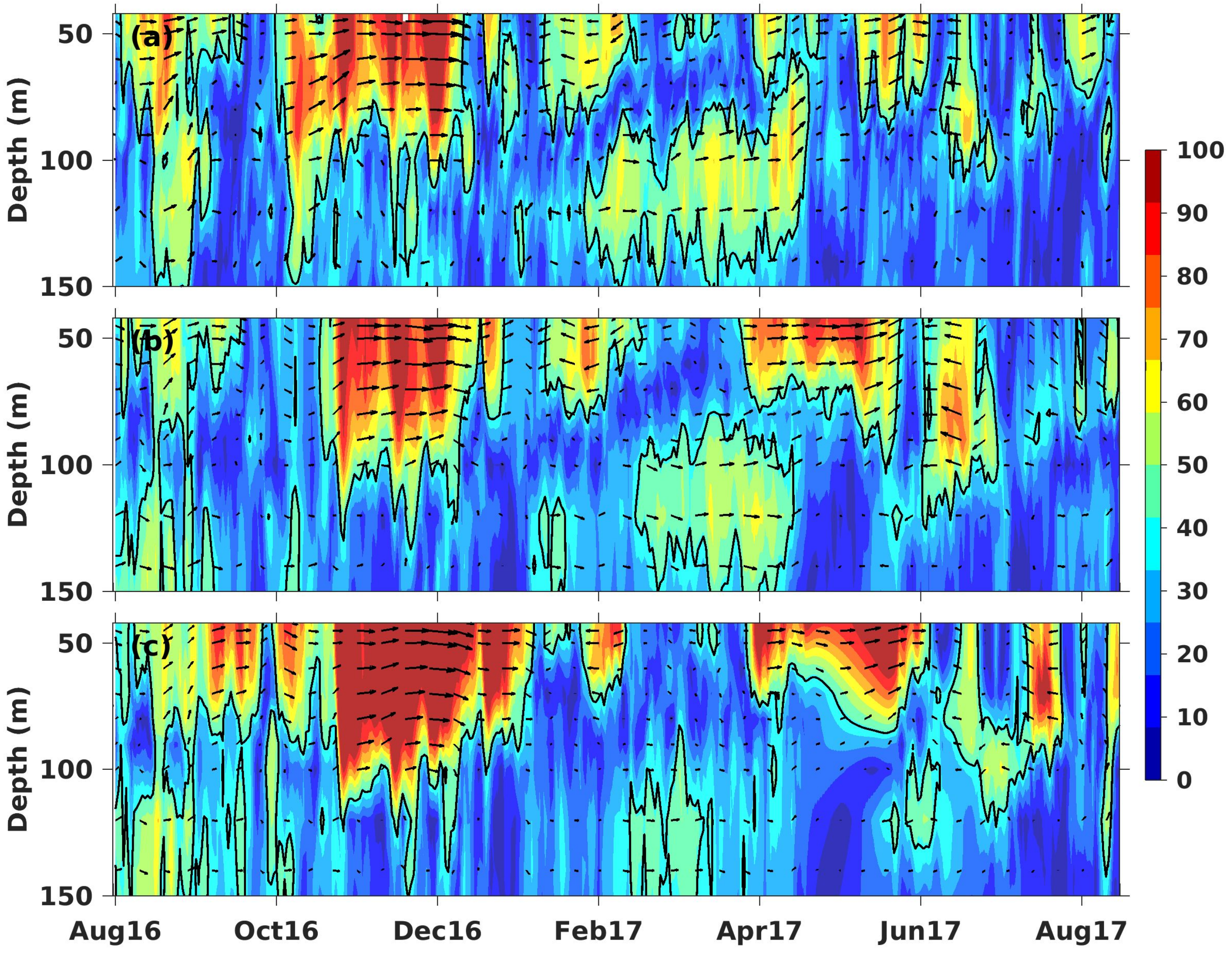
Spatial Correlation and RMSD of SST against AVHRR

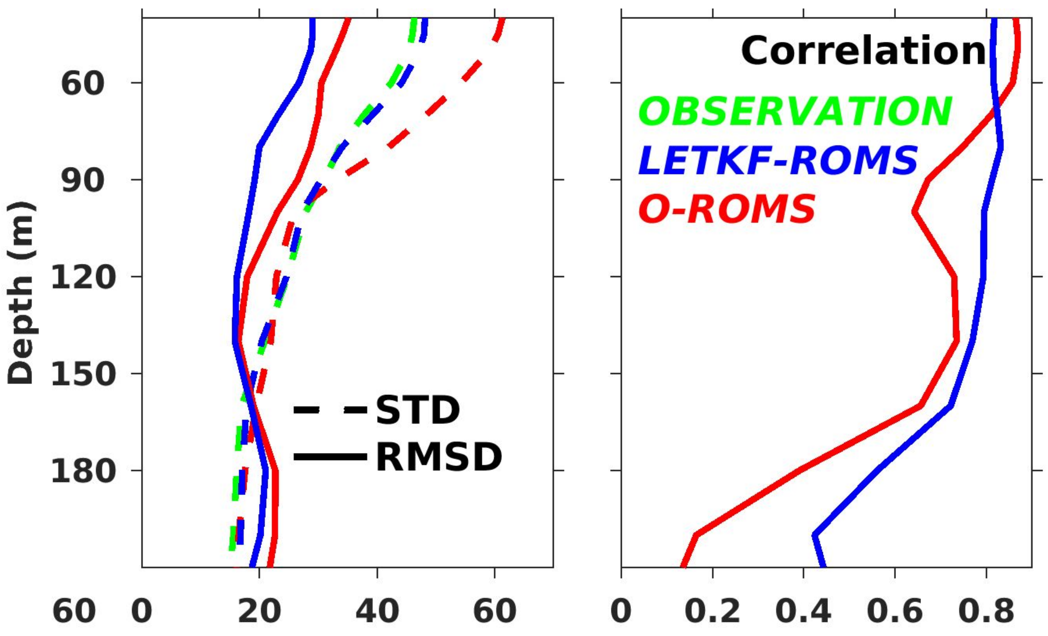
Zonal Current Analysis



Spatial Correlation, RMSD of Zonal Current against OSCAR

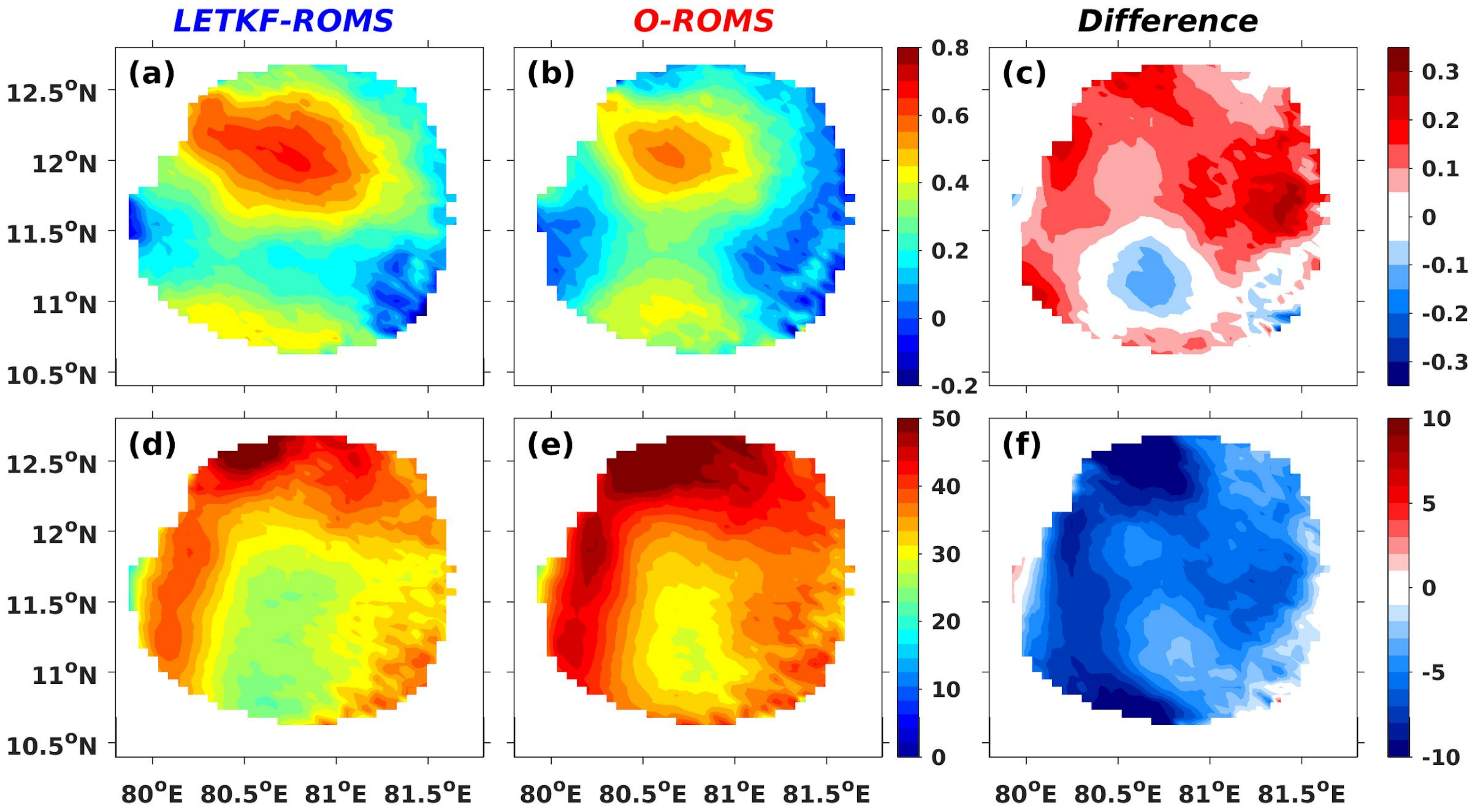
Zonal Current Analysis

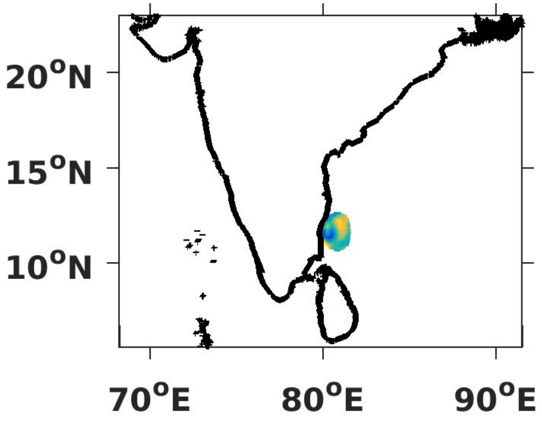




Time-depth evolution of currents at Equator, 80.5 ^{O} E from (a) ADCP, (b) LETKF-ROMS and (c) O-ROMS.

Current Analysis





Spatial Correlation and RMSD of Meridional Currents against HF Radar observations on East Coast of India

IS IT ALWAYS GOOD TO ASSIMILATE ALL AVAILABLE OBSERVATIONS ?

Ensemble Forecast Sensitivity to Observation (EFSO)

Do all observations improve the analysis ? NO

How do we determine the beneficial observations ? OSE & EFSO

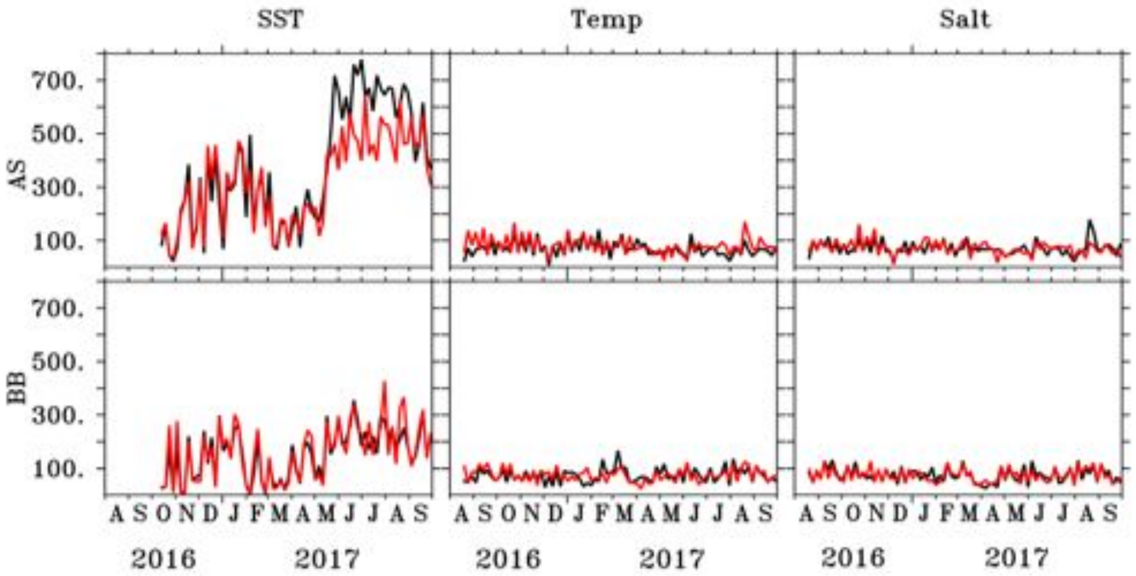
- Observing System Experiment (OSE) is an integral part of operational forecasting system to segregate the **Bad** (detrimental) observations from the **Good** (beneficial) observations in order to improve forecast skill.
- In an OSE, a continuous assimilation/forecast cycle is run assimilating as well as denying a set of real observations, and the results from these simulations are compared to assess the impacts of those observations on NWP forecasts. Being a data denial experiment, OSE is a resource-consuming experiment.
- Ensemble Forecast Sensitivity to Observations (EFSO) can evaluate the impact of each and every assimilated observation simultaneously. It is a **faster and less resource-consuming** alternative to traditional OSE.
- EFSO was not implemented for ocean earlier because of the unavailability of a proper weight matrix that helps evaluate impact of each observation.

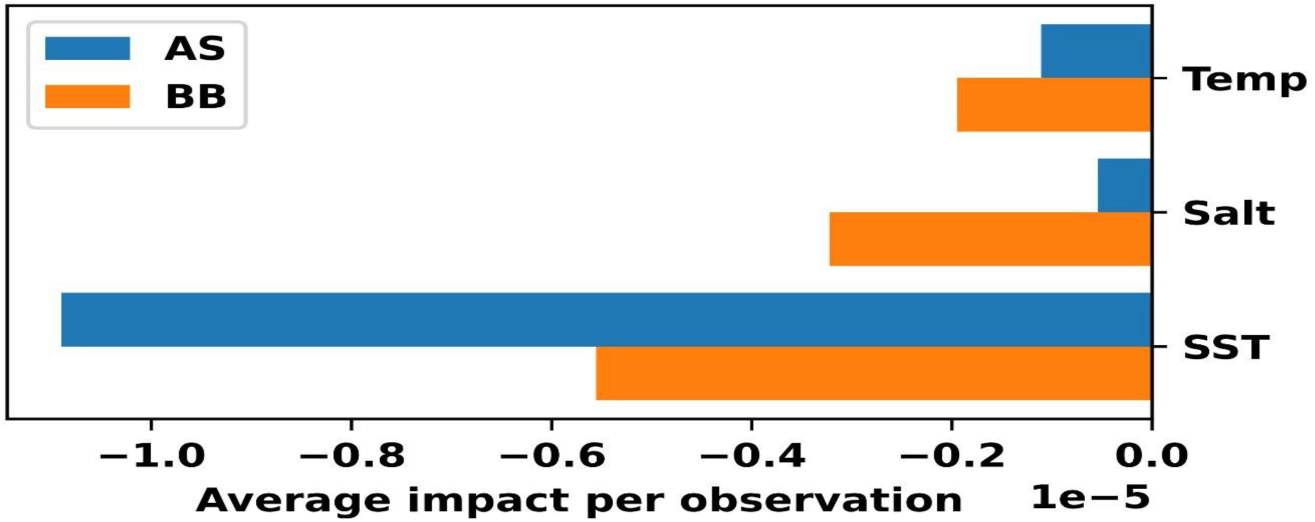
Ensemble Forecast Sensitivity to Observation (EFSO)

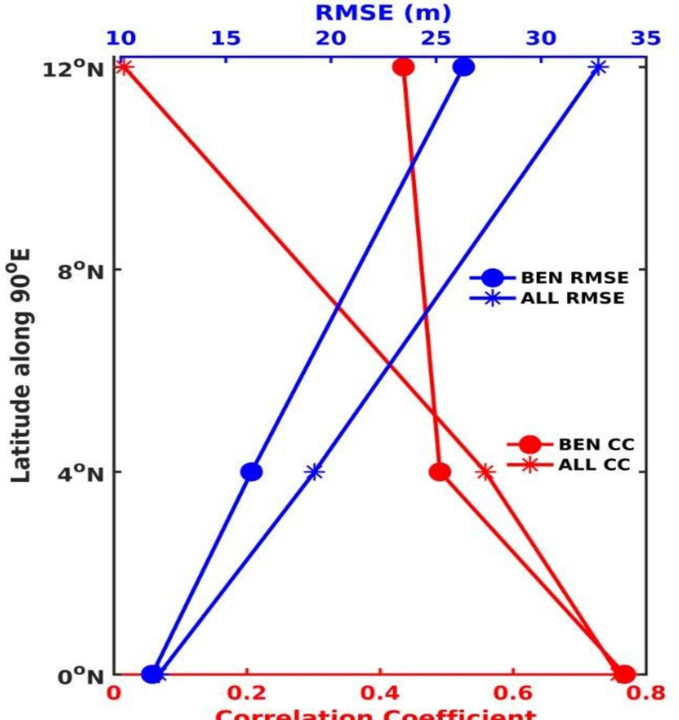
- Essentially EFSO measures the error reduction in a forecast when an observation is assimilated compared to that when it is not assimilated.
- Forecast Skill Metric, $\boldsymbol{\delta e} = \mathbf{e_A{}^T * C * e_A{}^T - e_N{}^T * C * e_N{}^T}$

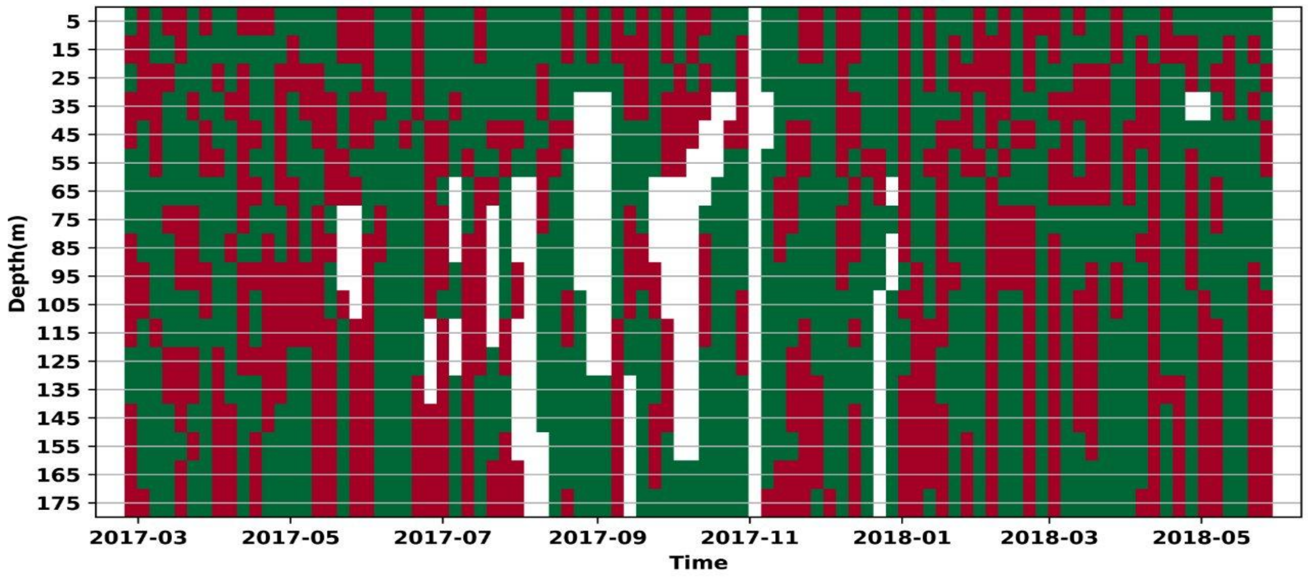
- The weight matrix (squared norm) **C** is a diagonal matrix containing the magnitude of the baroclinic vector at each model grid points defined by $\left(\left\|\frac{\nabla\rho\times\nabla p}{\rho^2}\right\|\right)$ diagonal elements.
- $\mathbf{e_A}$ and $\mathbf{e_N}$ are the errors in the forecast when an observation is **assimilated** and **not assimilated** respectively.
- The forecast is verified against a reference analysis generated by 80 ensemble LETKF-ROMS.
- We segregate the impact of each observation (**T, S,** and **SST**) at every model grid point for every vertical layer at each assimilation step.
- $\boldsymbol{\delta e} < \mathbf{0}$ indicates **improving forecast** skill and hence a **beneficial** observation.

Ensemble Forecast Sensitivity to Observation (EFSO) - Results







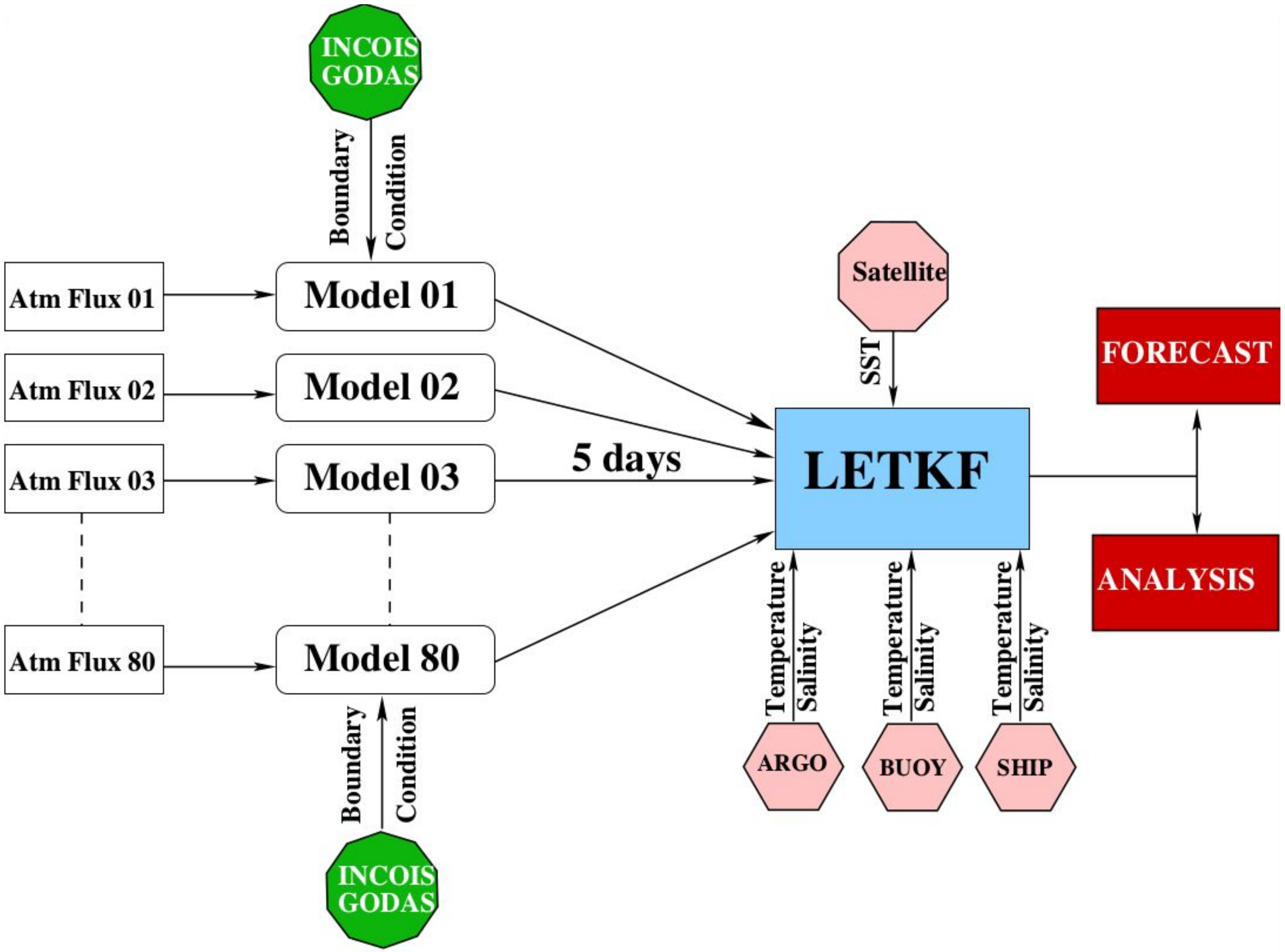


PENDING !!!

WHAT MORE ABOUT DA ? OSSE - MISSING !!!

PART B

Pictorial Illustration of the system



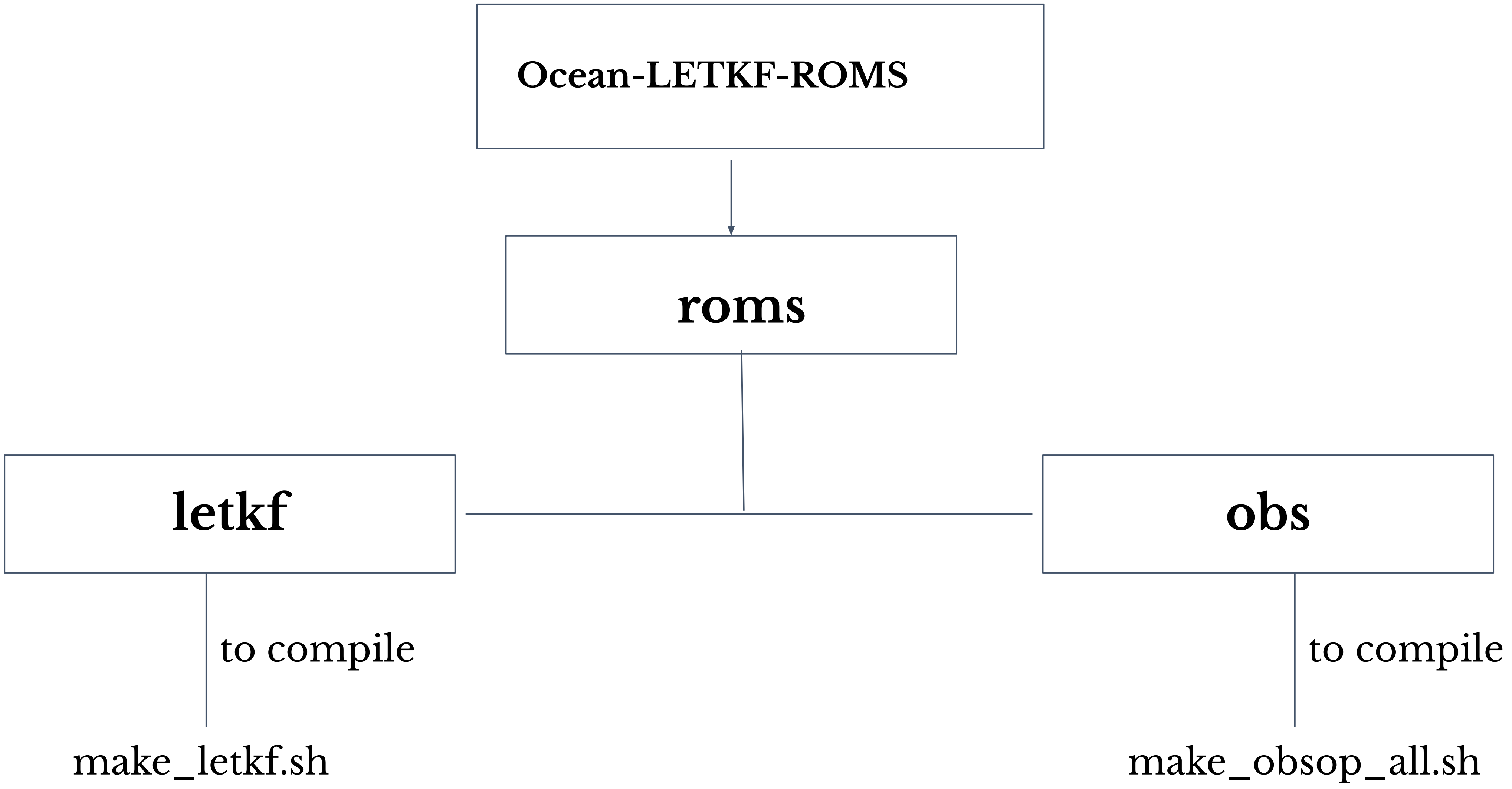
WITHIN LETKF-CODE

INCOIS

Ocean-LETKF-ROMS

- **config**
  1) machine.sh
  2) machine_fortran.sh
  3) machine_netcdf.sh
- **common**
  1) common.f90
  2) common_letkf.f90
  3) common_mpi.f90
- **roms**
  1) common folder
  2) letkf folder
  3) obs folder

WITHIN LETKF-CODE



Important Programs within LETKF

params_letkf.f90 → No. of ensembles, Cov Inflation etc

params_model.f90 → grid specifications of the underlying model

letkf.f90 → Main program of letkf.

letkf_obs.f90 → This module reads all observation and stores it in appropriate format.

letkf_tools.f90 → This module performs the main loop of the data assimilation

letkf_local.f90 → This module performs localization.

Main Program ----- letkf.f90

```
call initialize_mpi
call set_common_roms
call set_common_mpi_roms

call set_letkf_obs

call read_ens_mpi

call write_ensmspr_mpi

call das_letkf

call write_ens_mpi

call write_ensmspr_mpi
```

ROCOTO

Rocoto is a workflow management system used to execute different tasks with minimal manual intervention

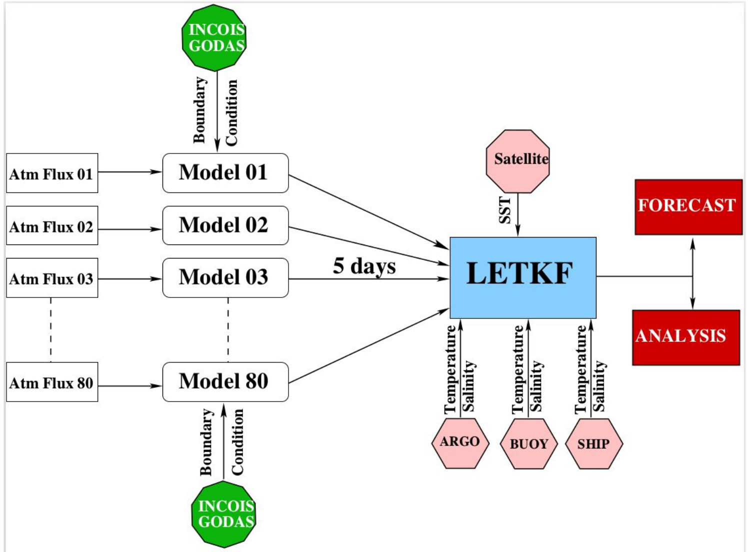
How to run it ?

rocotorun -w xml -d logfile

xml file containing the workflow management

logfile containing the status of the workflow

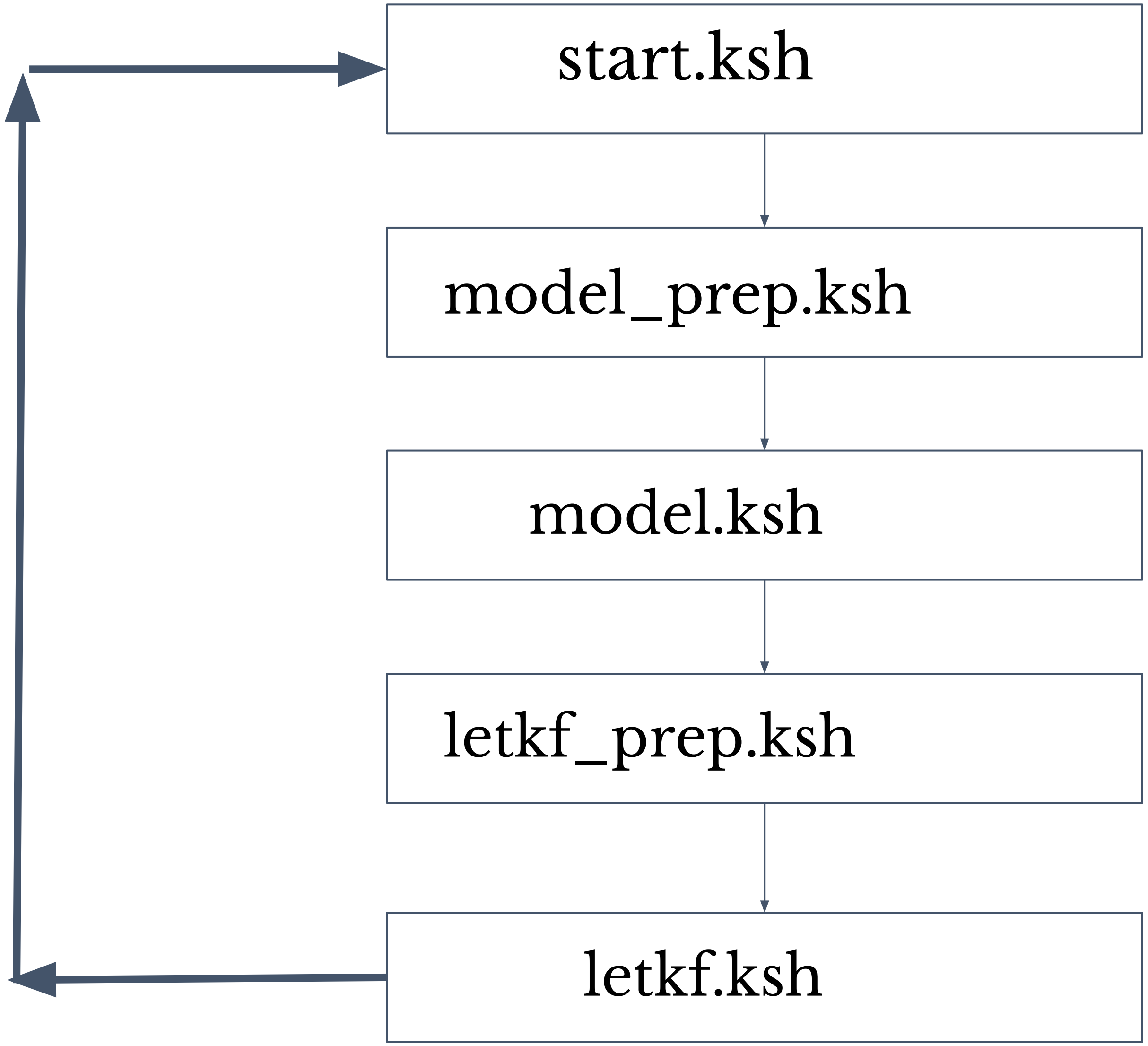
How does it work ?



Visit
http://christopherwharrop.github.io/rocoto/

HOW TO RUN LETKF ?

ROCOTO WORKFLOW



References

- Balaji, B., Deepsankar, B., Paul, B., Sanikommu Sivareddy, P.A. Francis, Abhisek Chatterjee, and Arya Paul (2018). LETKF-ROMS: An improved predictability system for the Indian Ocean. Technical Report, ESSO-INCOIS-MDG-TR- 03, available at http://moeseprints.incois.gov.in/id/eprint/4572.

- Balaji, B., Biswamoy Paul, Deep Sankar, B., Sivareddy, S., Arya Paul. 2019. Ensemble based Regional ocean data assimilation system for the Indian Ocean: Implementation and Evaluation. Ocean Modelling.

- Sivareddy, S., Deep Sankar, B., Balaji, B., Biswamoy Paul, Arya Paul, Kunal Chakraborty, Ibrahim Hoteit. 2019. Impact of Dynamical Representational Errors on an Indian Ocean Ensemble Data Assimilation System. Quarterly Journal of the Royal Meteorological Society.

- Biswamoy Paul, Balaji B, Arya Paul 2024. A study of forecast sensitivity to observations in the Bay of Bengal using LETKF, Frontiers in Marine Science

THANK YOU