Indian National Centre
for Ocean
Information Services

An operational Objective Analysis system at INCOIS for generation of Argo Value Added Products

TVS Udaya Bhaskar, M Ravichandran, R Devender
Modelling and Observation Group (MOG)

Indian National Centre for Ocean Information Services
Post Box No 21, IDA Jeedimetla, Gajularamaram,
Hyderabad – 500 055, India

April 2007

Technical Report No. INCOIS/MOG-TR-2/07

Table of Contents

1. Abstract

The system of objective analysis used at Indian National Centre for Ocean Information Services is described. It is a integral part of the Argo data processing system, and designed to operate with minimum of manual intervention. The analysis method, based mainly on the method of McCreary and Kessler is a method of estimating Gaussian weight for all the observation used for estimating the value at grid location. The errors are determined from a comparison of the observation with the estimated value. The analysis system is very flexible, and has been used to analyse many different types of variables.

2. Introduction

A fundamental problem in the geophysical sciences is how to use data collected at finite number of locations, and eventually at different times, to estimate value at any point of the space or space-time continuum. The ultimate aim of such estimate can be a simple visualization of the observed field, diagnosis of physical process of the use of the estimated field as input for numerical models, among many other applications. A wide set of techniques have been developed for both diagnostic and prognostic studies/analyses. Since the development of computers in the last decades made possible the automatic implementation of these techniques, they have been referred to as spatial objective analysis. To put it in a nut shell Objective Analysis (OA) is the process of transforming data from observations at irregularly spaced points into data at points of regularly arranged grid.

2.1 Objective Analysis definition

The graphic (figure 1) depicts the basic problem of OA, namely that we have irregularly spaced observations that must provide values for points on a regularly spaced grid. (Red dots represent observations and blue dots are grid points.) OA in general is the process of interpolating observed values onto the grid points used by the model/analysis in order to define the initial conditions of the atmosphere/ocean.

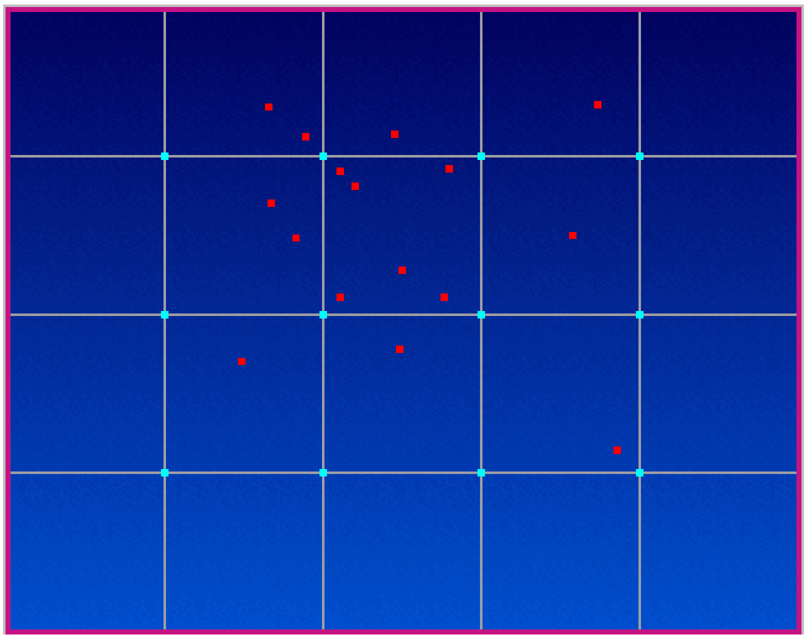


Fig 1 Irregularly spaced observations

Why isn't this just a simple exercise in mathematical interpolation? There are several answers to this question.

1. We can use our knowledge of oceanic behavior to infer additional information from the data available in the area. For example, we can use balance relationships such as geostrophy or mass continuity to introduce dynamical consistency into the analysis. If we use one type of data to improve the analysis of another, then the analysis is said to be multivariate.

2. We can adjust the analysis procedure to filter out scales of motion that can't be forecast by the model being used. For example, small mesoscale circulations represented in the observations may need to be smoothed out in an analysis for a global model.

3. We can make use of a first guess field or background field provided by an earlier forecast from the same model. The blending of the background fields and the observations in the objective analysis process is especially important in data sparse areas. It allows us to avoid extrapolation of observation values into regions distant from the observation sites. The background field can also provide detail (such as frontal locations that exist between observations).

Using a background field also helps to introduce dynamical consistency between the analysis and the model. In other words, that part of the analysis that comes from the background field is already consistent with the physical (dynamic) relationships implied by the equations used in the model.

4. We can also make use of our knowledge of the probable errors associated with each observation. We can weight the reliability of each type of observation based on past records of accuracy.

2.2 Analysis Equation

Now we will examine a fundamental OA equation in worded form in order to illustrate the basic principles that contribute to a numerical oceanography analysis.

In simplest terms, the OA equation attempts to determine the value of a particular oceanographic variable at a particular grid point (at a particular valid time).

In words, the analysis equation can be expressed as shown below (figure 2).

The analysis value at the grid point = The background (or first guess) value at the grid point + A correction

where

$$\text{correction} = \sum_{\text{sum}} \left(\text{weights} \times \text{new information (based on observations)} \right)$$

in other words

correction = a weighted average of a set of observations taken from sites surrounding the grid point

Fig 2 Objective Analysis equation

2.3 Importance of Background Field

In the simplest kind of OA scheme, the background values would not be used and the analysis would be based solely on new observations. In this case the equation would become:

$$\text{The analysis value at the grid point} = \sum_{\text{sum}} \left(\text{weights} \times \text{new observations} \right)$$

The observations themselves would be interpolated to the grid point by calculating a weighted average of the data. (One type of weight, for example, is proportional to the distance of the data from the grid point. The farther an observation is from the grid point, the less weight it gets.)

If a grid point has no nearby observations, the simple scheme described here is in trouble!

3. Kessler and McCreary Methodology

3.1 Data

The data used in this analysis consisted of all the CTD data measured by Argo floats in the Indian Ocean region (30º E – 120º E and 30º S – 30º N). The profiles' data were obtained from the INCOIS web site (http://www.incois.gov.in/Incois/argo/argo_webGIS_intro.jsp#) which are made available by USGODAE and IFREMER. Argo floats measure T/S from surface to 2000 m depth every 5/10 days. All profiles were subjected to real time quality control checks like density inversion test, spike test and gradient test (see Wong *et al.*, 2004).

3.2 Methodology

The gridding was carried out in two steps. First the temperature data Tn (x_n, y_n, z_n, t_n) from each profile n were linearly interpolated to standard depths (1 m from surface to 2000 m) there by creating a modified data set T'_n (x_n, y_n, z_n, t_n). This interpolation was done only when two sample in a profile are with in a selected vertical distances which increased from 5 m in the surface to 100 between 500 – 2000 m. Second, in a separate computation at each Levitus standard depth Z_0 (0, 10, 20, 30, 50, 75, 100, 125, 150, 200, 250, 300, 400, 500, 600, 700, 800, 900, 1000, 1200, 1400, 1600, 1800, 2000 m) the temperature T'_n were mapped from irregular grid locations (x_n, y_n, z_n) to regular grid (x_0, y_0, z_0) locations with a grid spacing of 1º X 1º . Specifically the value of the gridded temperature T'' at each grid point (x_0, y_0, t_0) was estimated by the operation

$$T''(x_0, y_0, t_0) = \frac{\sum_{n=1}^{Np} T'_n W_n}{\sum_{n=1}^{Np} W_n} \qquad \textbf{(1)}$$

where N_p is the total number of profiles with in the influence region of a particular grid point. The Guassian weight function W_n is given by

$$W_n(x_n, y_n, t_n) = exp\left\{-\left[\left(\frac{x_n - x_0}{X}\right)^2 + \left(\frac{y_n - y_0}{Y}\right)^2 + \left(\frac{t_n - t_0}{\tau}\right)^2\right]\right\} \quad (2)$$

This operation is similar to a single iteration of objective mapping as used by Levitus (1982). Visualising the three dimensional grid (x_0, y_0, z_0) with data points $T'_n(x_n, y_n, z_n)$ scattered irregularly through it, the mapping operation appears as a ellipsoid moving from grid point to grid point averaging the points that fall with in that ellipse. Each data point falls with in summation of several grid points, weighted according to the distance. In the regions of very sparse sampling, a single data point may be the only information for one or several grid points. If no data points fell with in the ellipsoid at a grid point, then that was left blank.

4. General Derivation of Objective Analysis Equation

In this section, we provide a derivation of the **Gauss-Markov** theorem which is the foundation for the method of Objective Analysis which is used in Oceanography and Meteorology for the estimation of fields based upon incomplete and noisy observations.

4.1 The derivation

We want to derive an optimal estimate, $\hat{x}$, of a field, x, as a linear combination of observation data, θ,

$$\hat{x} = A\theta \tag{1}$$

Our goal is to derive the form of the matrix, A, so that the expected mean square difference between the estimated field and the actual field (x) is minimized,

$$E[\varepsilon\varepsilon^T] = E[(\hat{x} - x)(\hat{x} - x)^T] = \textbf{minimum} \tag{2}$$

If we put (1) into (2) and expand, we get

$$E[\varepsilon\varepsilon^T] = E[A\theta\theta^T A^T - x\theta^T A^T - A\theta x^T + xx^T] \tag{3}$$

If we let C_x be the autocorrelation of the field ($E[x\ x^T]$), C_θ be the autocorrelation of the observations ($E[\theta\theta^T]$), and $C_{x\theta}$ be the cross correlation between the field and the observations ($E[x\theta^T]$), then we can write the above as

$$C_\varepsilon = AC_\theta A^T - C_{x\theta}A^T - AC_{x\theta}^T + C_x \tag{4}$$

The next step requires the application of the following matrix identity (proved in the appendix),

$$(A - B\ C^{-1})\ C\ (A - B\ C^{-1})^T - B\ C^{-1}\ B^T = A\ C\ A^T - B\ A^T - (B\ A^T)^T \tag{5}$$

using A in (4) for A in (5), and $C_{x\theta}$ for B as well as C_θ for C, we can reduce (x) to

$$C_\varepsilon = (A - C_{x\theta}C_\theta^{-1})C_\theta(A - C_{x\theta}C_\theta^{-1})^T - C_{x\theta}C_\theta^{-1}C_{x\theta}^T + C_x \tag{6}$$

(note we have also used the fact that $C_\theta = C_\theta^T$).

The matrices C_θ, is an autocorrelation matrix therefore both it and C_θ^{-1} are nonnegative definite (see appendix), therefore

$$(A - C_{x\theta}C_\theta^{-1})C_\theta(A - C_{x\theta}C_\theta^{-1})^T \tag{7}$$

and

$$C_{x\theta}C_\theta^{-1}C_{x\theta}^T \tag{8}$$

are both matrices with positive diagonal elements. This means that the diagonal elements of C_ε are therefore minimized when it is true that,

$$A - C_{x\theta}C_\theta^{-1} = 0 \tag{9}$$

Therefore we have,

$$A = C_{x\theta}C_\theta^{-1} \tag{10}$$

This is the estimator that we are seeking.

$$\hat{x} = C_{x\theta}C_\theta^{-1}\theta \tag{11}$$

Further, we can write down what the expected error for the estimator as,

$$C_\varepsilon = C_x - C_{x\theta}C_\theta^{-1}C_{x\theta}^T \tag{12}$$

Equations (11) and (12) constitute the **Gauss-Markov** estimator for the linear minimum means square estimate of a random variable.

4.2 Linear Observations

Upon reflection of equations (11) and (12) a problem arises: the determination of C_θ and $C_{x\theta}$ require having the true field values, *x*, but all one can actually observe are the measurement data θ.

In order to account for this important distinction we need to make some assumptions about how the measurements are related to the actual state of the system. We will assume that the observations are a linear function of the actual state plus random noise,

$$\theta_s = Hx_s + v_s \tag{13}$$

where *H* is a known matrix that maps the data to the observations and *v* is the random measurement error or noise. We introduce the index *s* here in order to be explicit that we are indicating the values at some space-time location *s*, *where the observations are*, which is not necessarily the location where the estimate $\hat{x}$ of the field is being made.

With this, we can write

$$C_{x\theta} = E[x(Hx_s + v)^T] = C_{xs}H^T + C_{xv} \tag{14}$$

Applying (13) to the definition of C_θ, gives

$$C_\theta = E[(Hx_s + v)(Hx_s + v)^T] = HC_xB^T + C_{sv}^T B^T + BC_{sv} + C_v \tag{15}$$

If we suppose that *the actual state and the noise are uncorrelated*, then the terms, C_{xv} and C_{sv} are each zero.

So now we have

$$\hat{x} = C_{xs}H^T[HC_sH^T + C_v]^{-1}\theta \tag{16}$$

and,

$$C_\varepsilon = C_x - C_{xs}H^T[HC_sH^T + C_v]^{-1}(C_{xs}H^T)^T \tag{17}$$

If we consider the case where the measurements θ are the same quantity as what we are estimating x (i.e. we are using density data to estimate the density field, as opposed to using salinity and temperature to estimate the density field), then H is just the identity matrix, so our estimator is,

$$\hat{x} = C_{xs}[C_s + C_v]^{-1}\theta \tag{18}$$

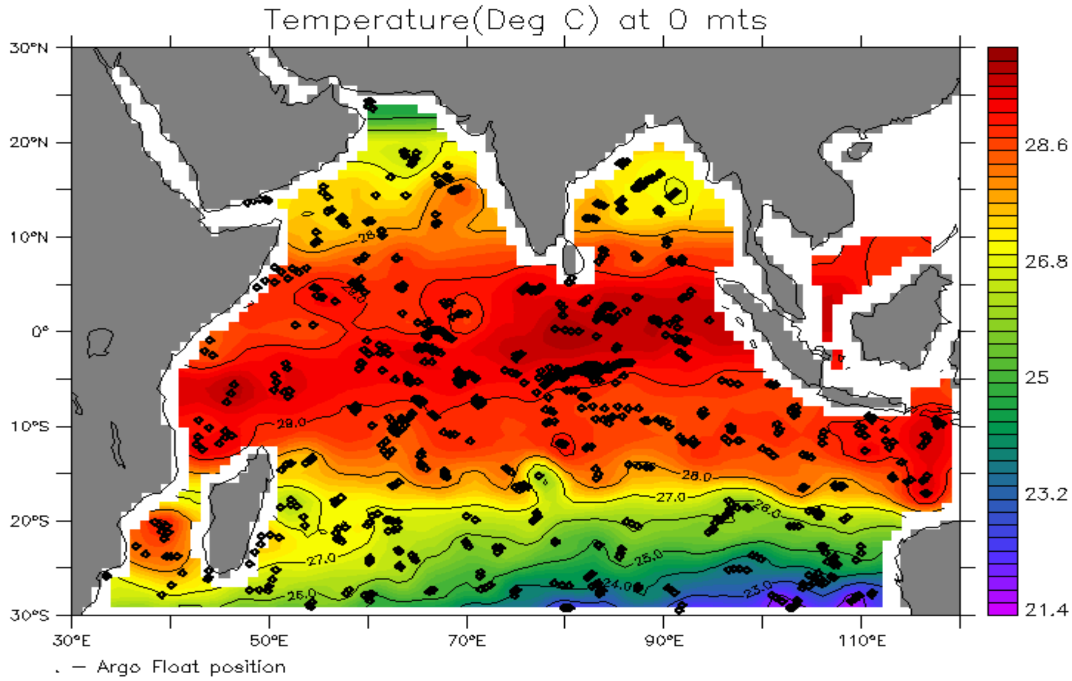
and,

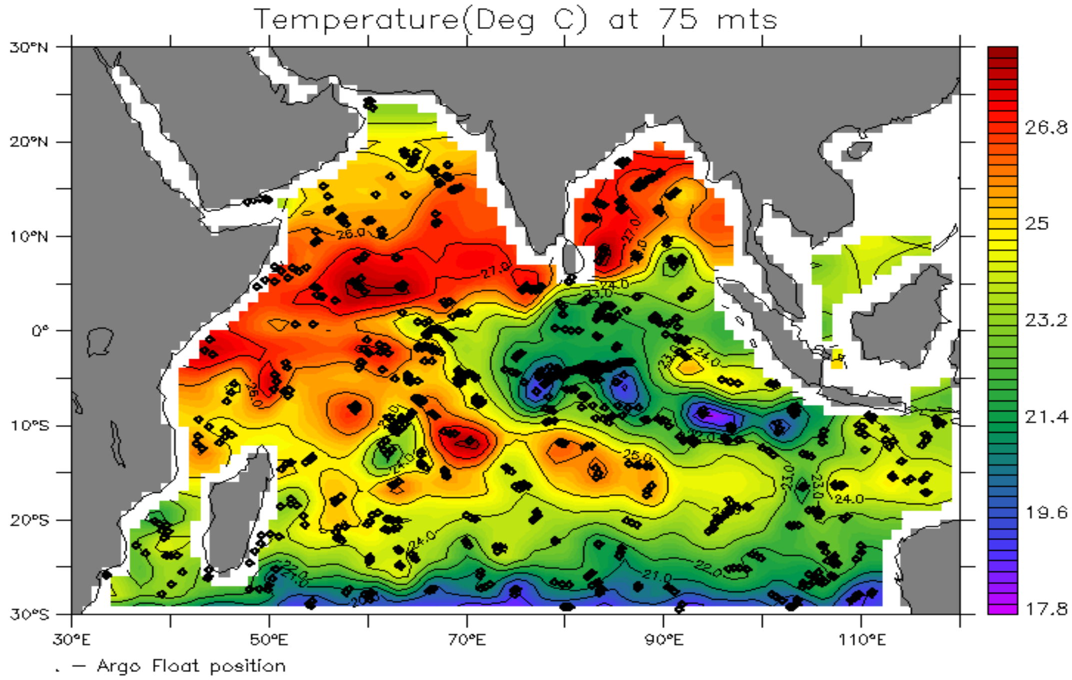
$$C_\varepsilon = C_x - C_{xs}[C_s + C_v]^{-1}C_{xs}^T \tag{19}$$

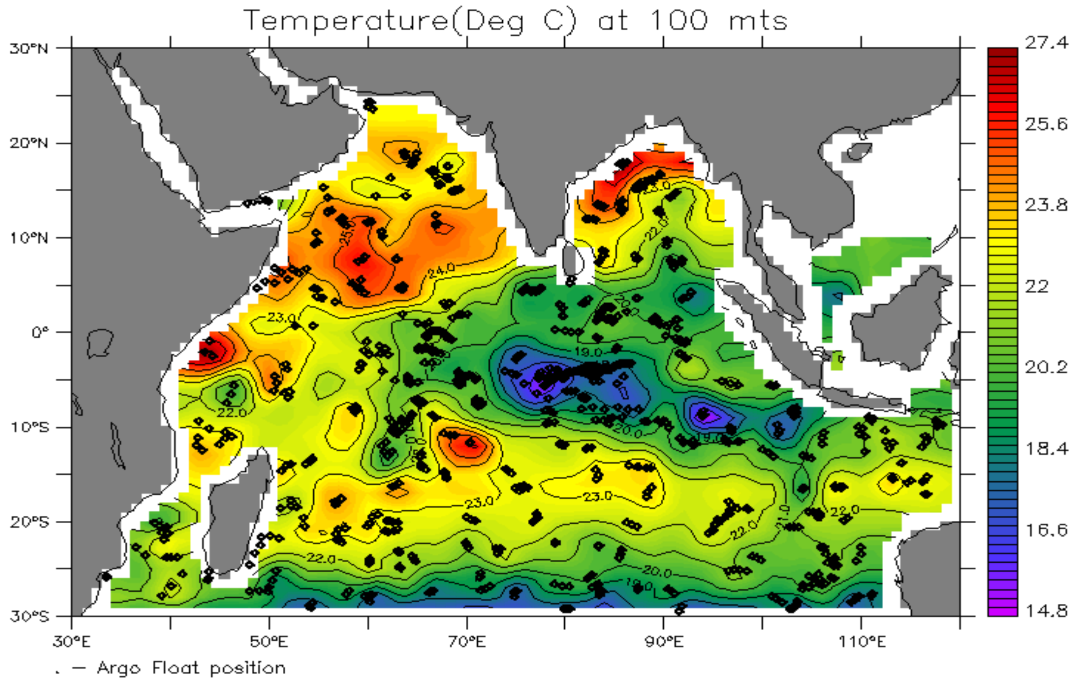
If the noise is white noise, then C_v is a diagonal matrix and we see that the effect of not having the true state correlations, but estimates of it based upon the observations, is to increase the diagonal elements of the matrix to be inverted by the measurement noise variance.

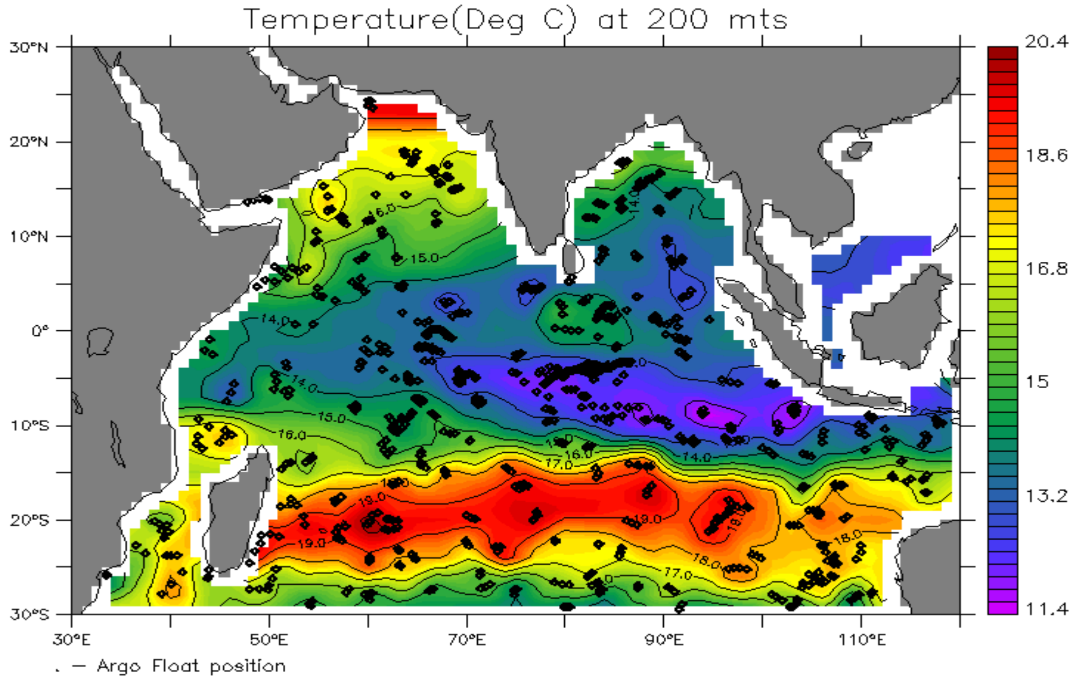
5. Monthly Argo Value Added Products

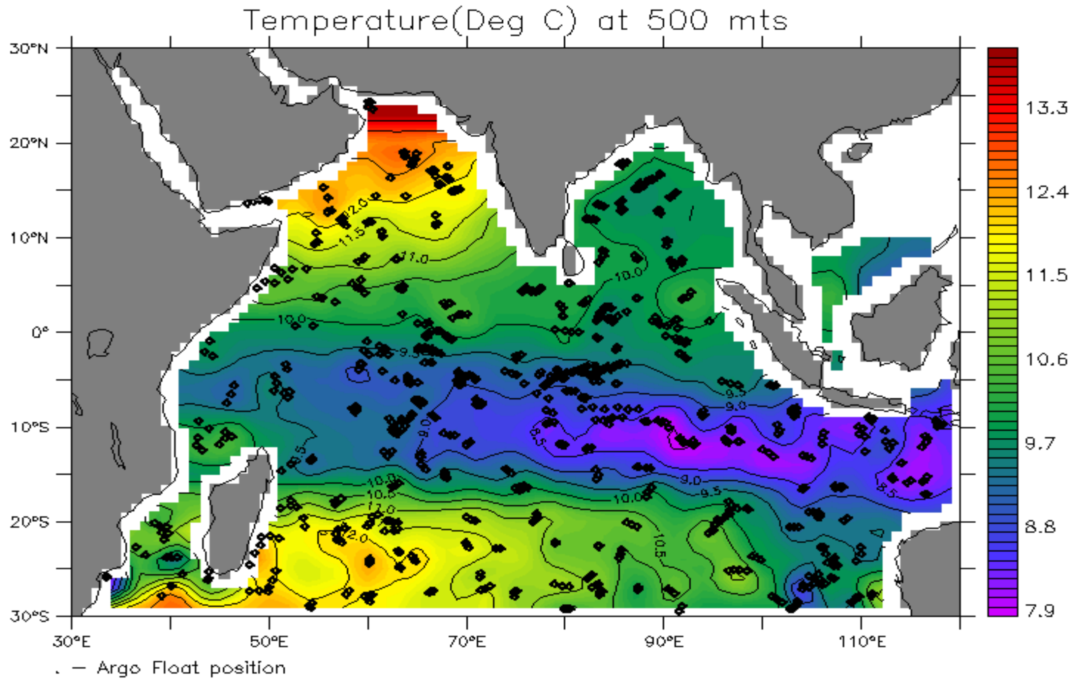
5.1 Temperature products

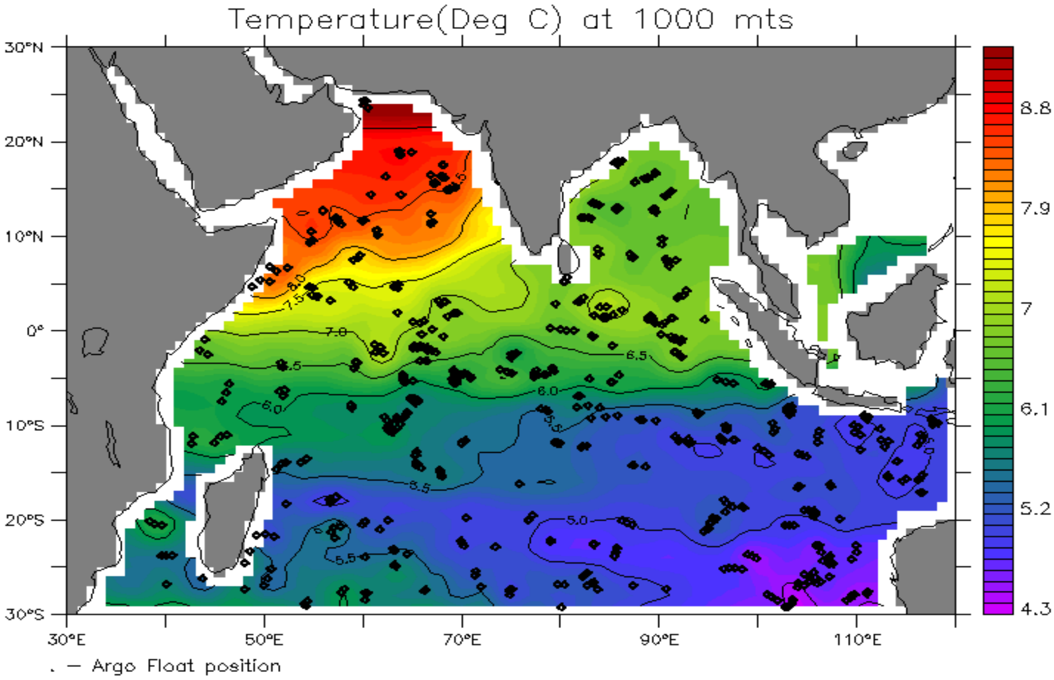



Temperature(Deg C) at 75 mts
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E
26.8
25
23.2
21.4
19.6
17.8
. — Argo Float position

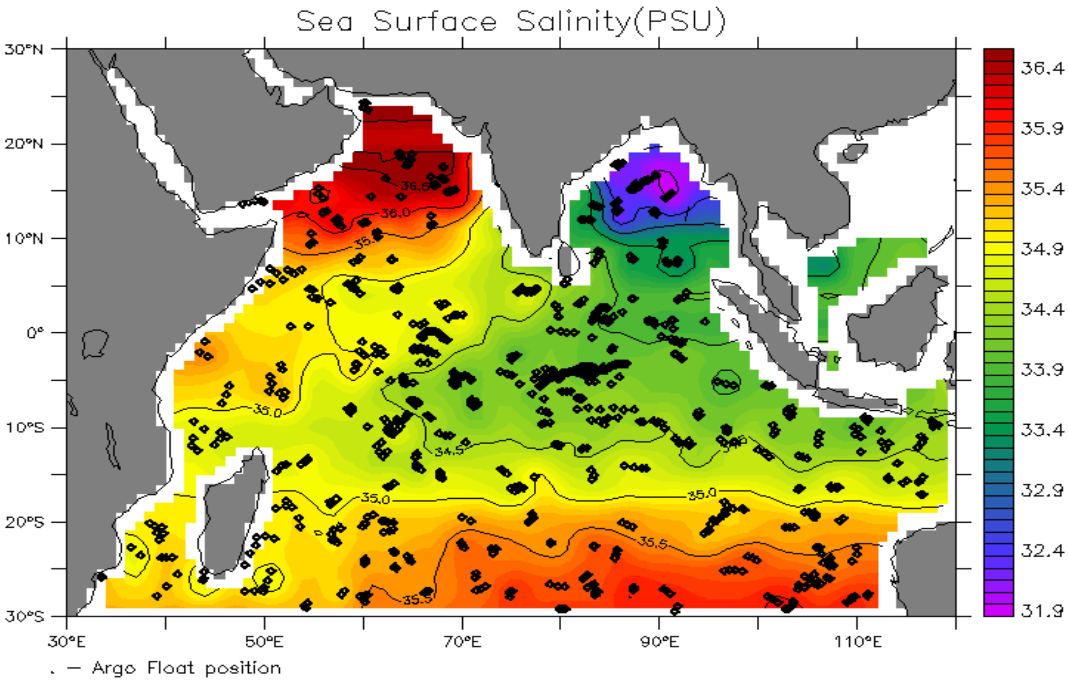

Temperature(Deg C) at 100 mts
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E
27.4
25.6
23.8
22
20.2
18.4
16.6
14.8
. — Argo Float position

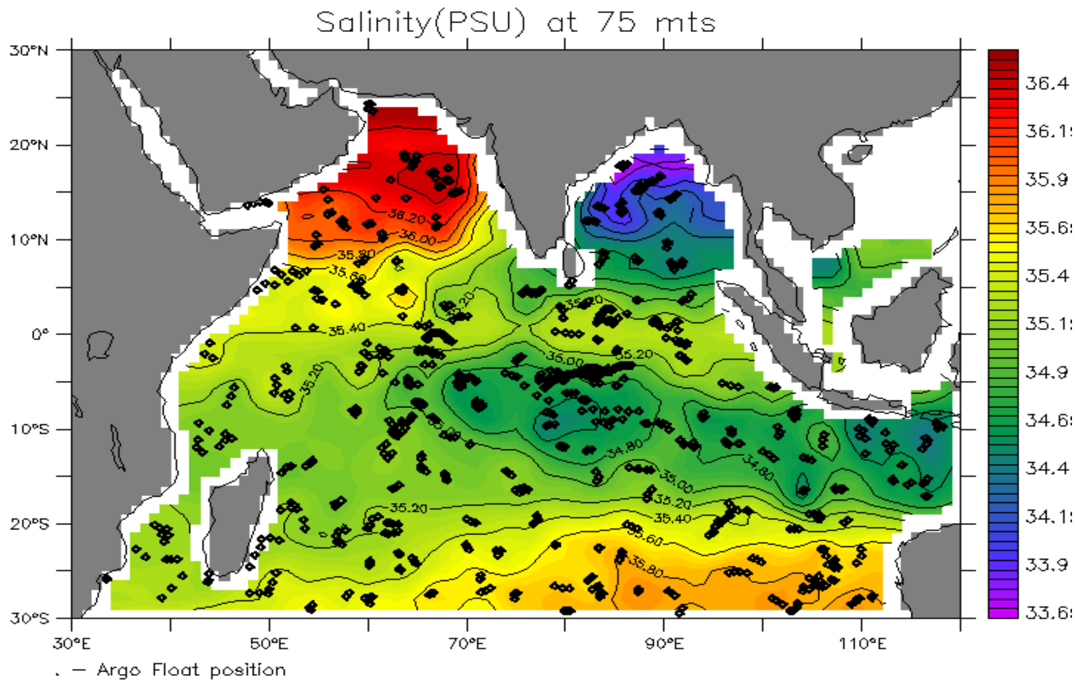

Temperature(Deg C) at 200 mts
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E
. — Argo Float position
20.4
18.6
16.8
15
13.2
11.4

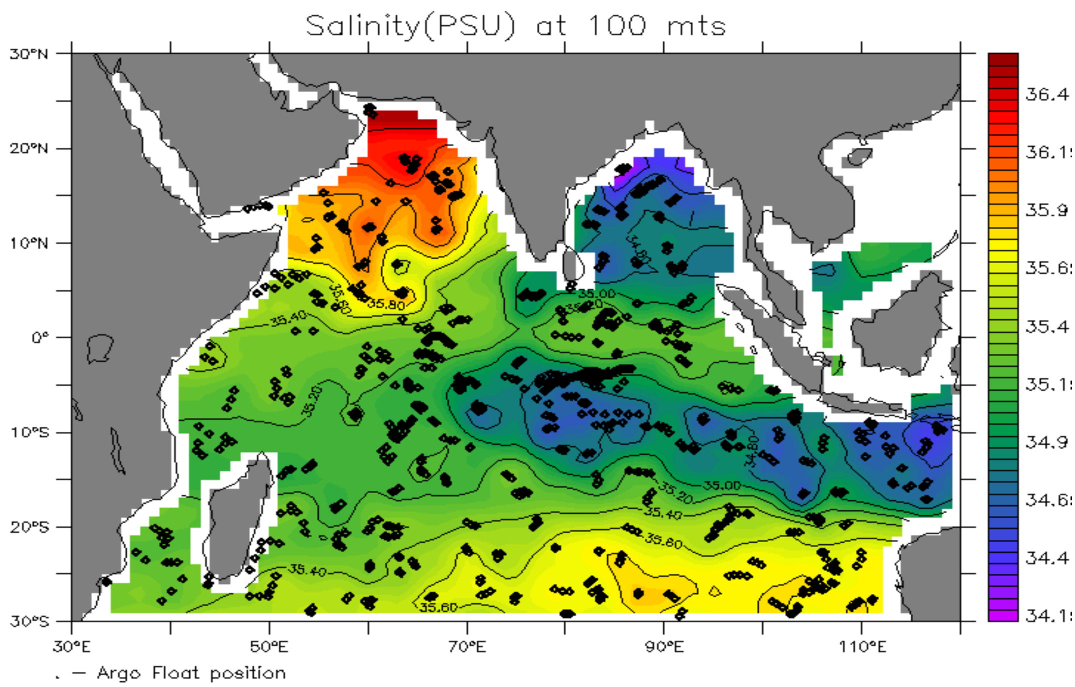

Temperature(Deg C) at 500 mts
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E
. — Argo Float position
13.3
12.4
11.5
10.6
9.7
8.8
7.9

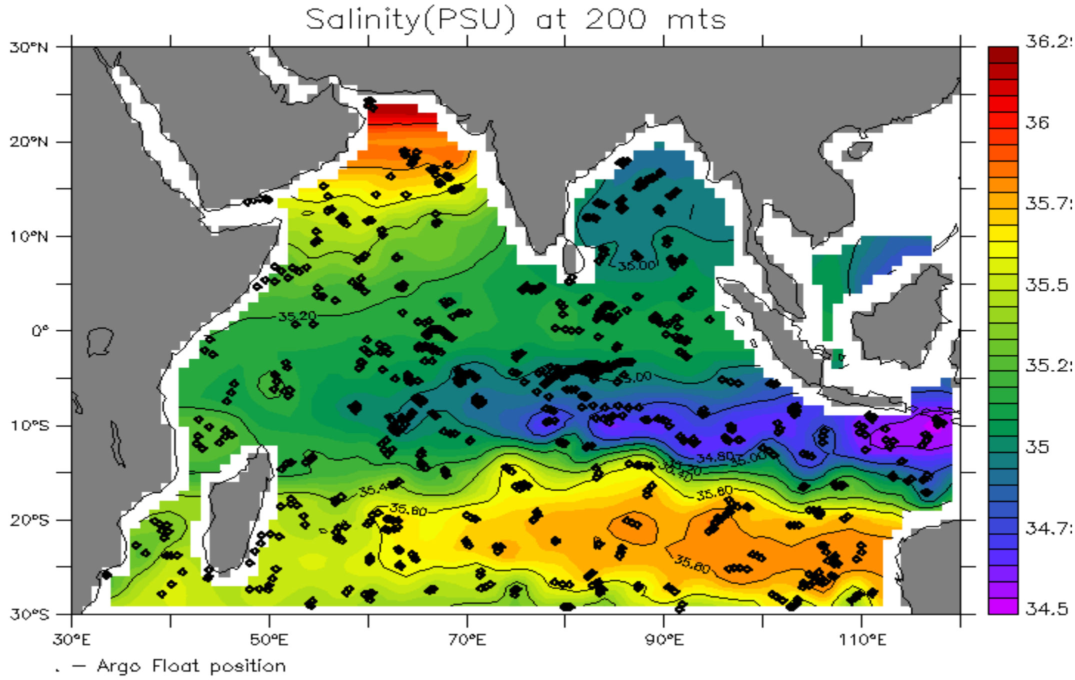


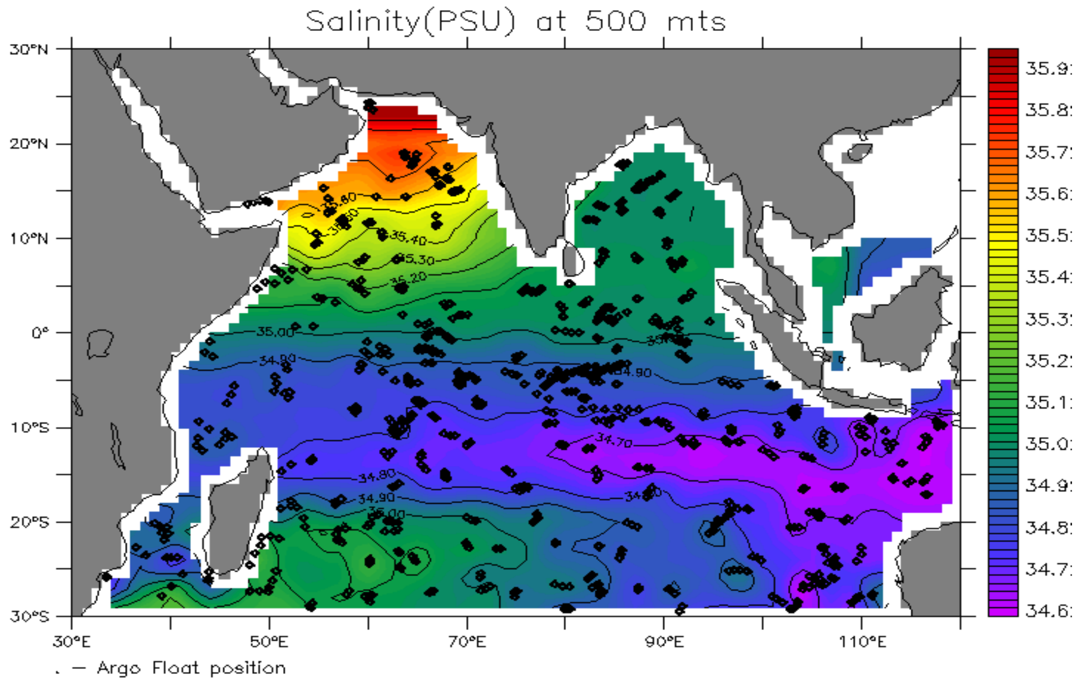
5.2 Salinity Plots

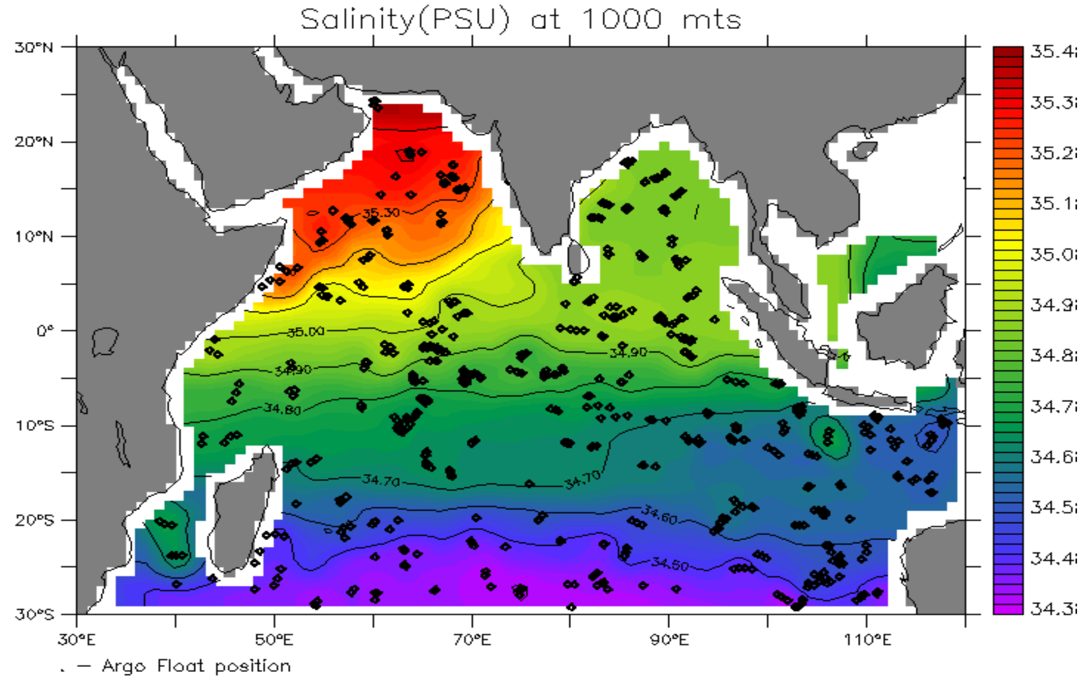



Salinity(PSU) at 75 mts
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E
. — Argo Float position
36.4
35.9
35.4
34.9
34.4
33.9

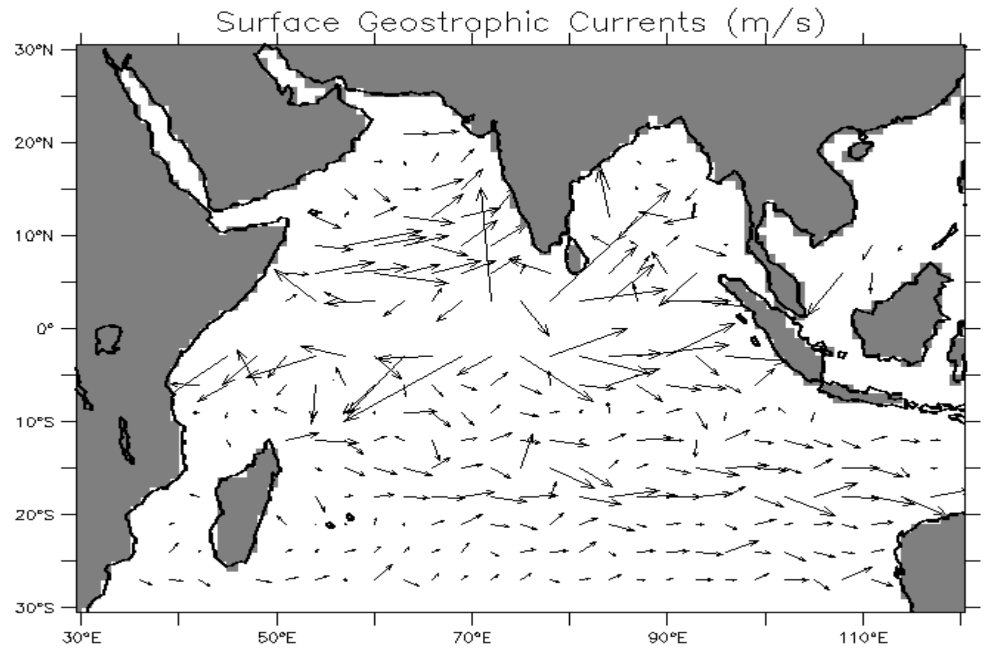

Salinity(PSU) at 100 mts
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E
. — Argo Float position
36.4
35.9
35.4
34.9
34.4

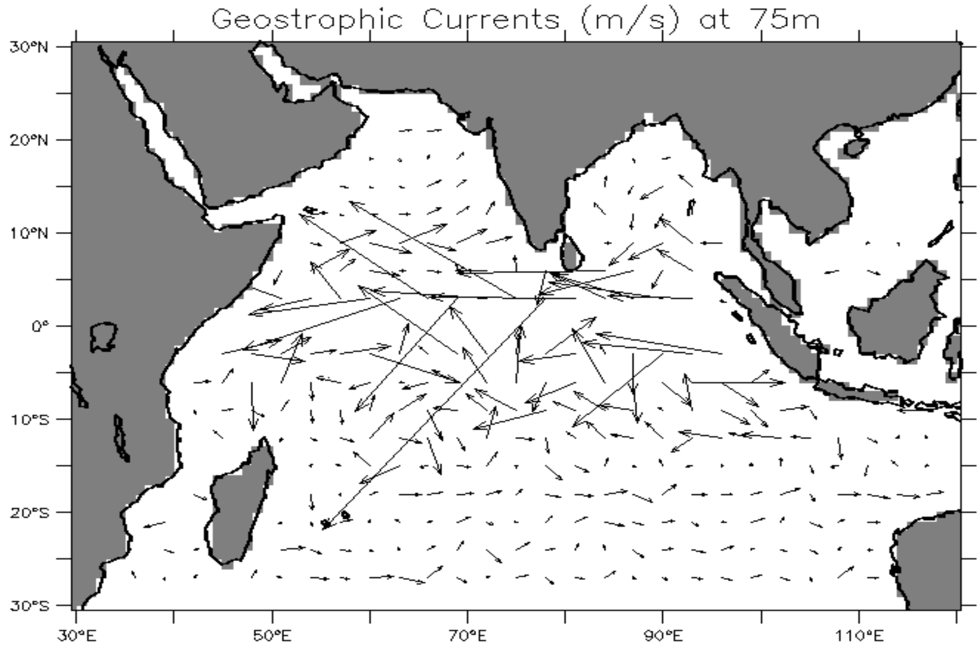

Salinity(PSU) at 200 mts
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E
. — Argo Float position
36.2
36
35.7
35.5
35.2
35
34.7
34.5

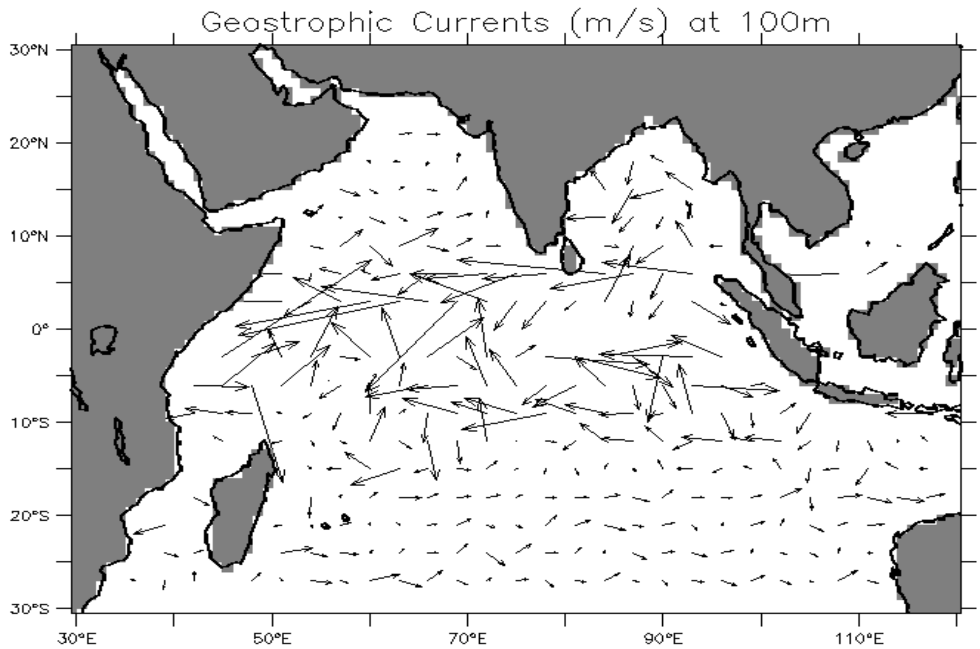

Salinity(PSU) at 500 mts
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E
. — Argo Float position
35.9
35.8
35.7
35.6
35.5
35.4
35.3
35.2
35.1
35.0
34.9
34.8
34.7
34.6

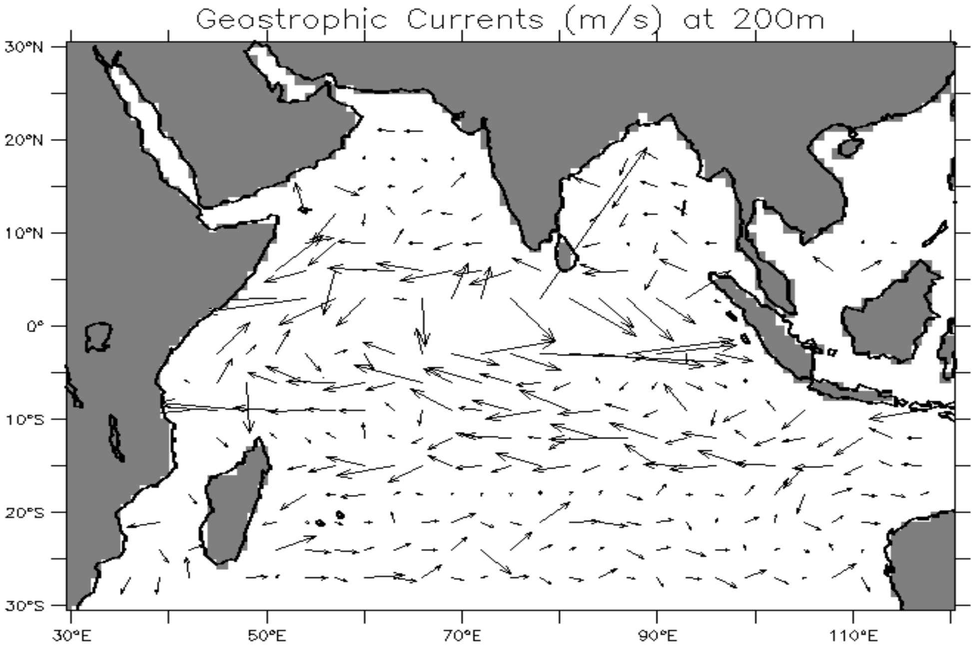


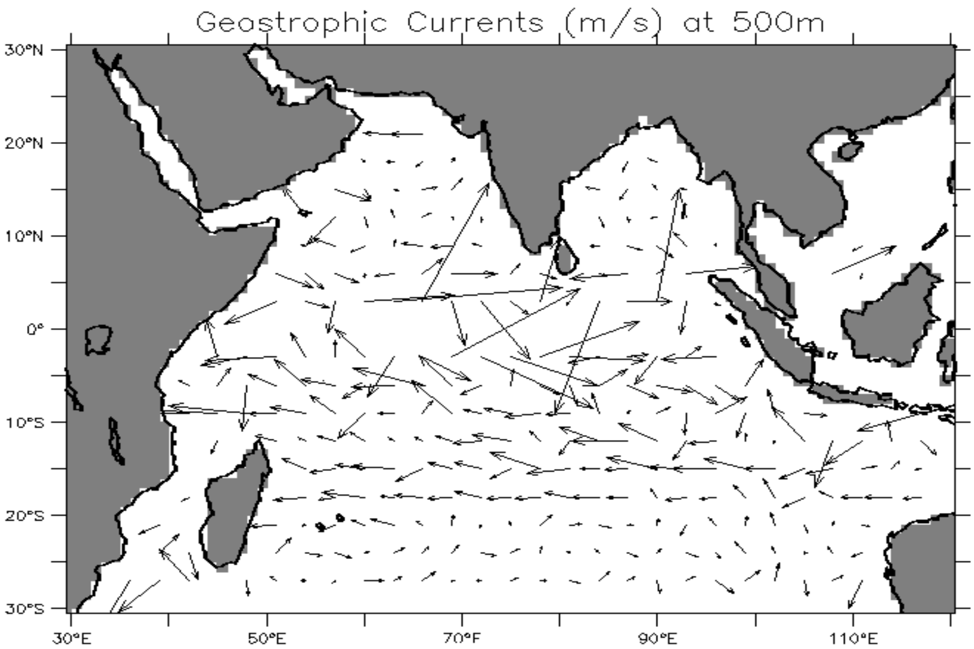
5.3 Geostrophic Currents

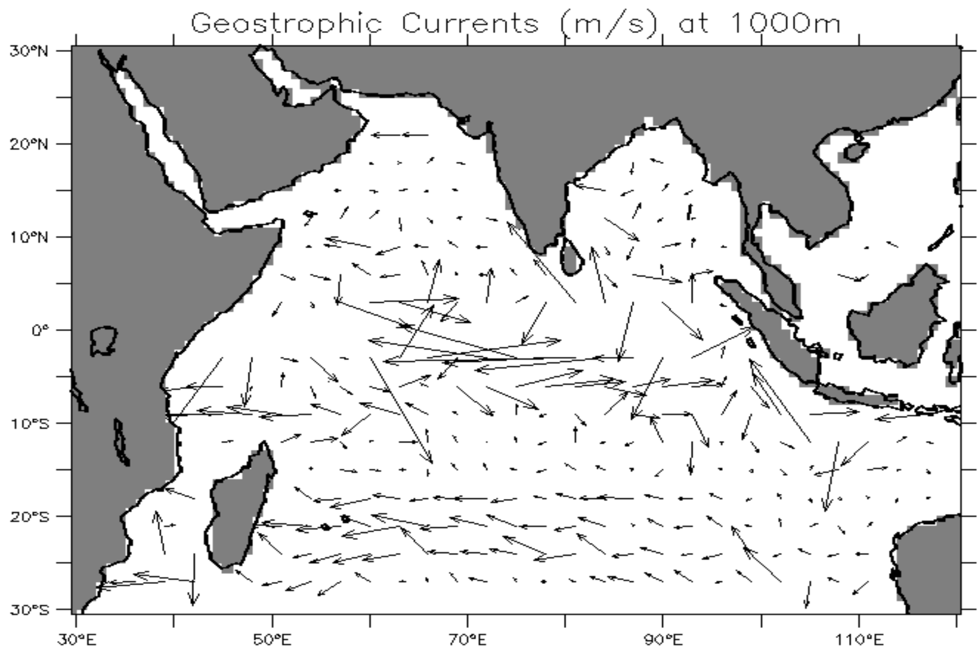



Geostrophic Currents (m/s) at 75m
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E

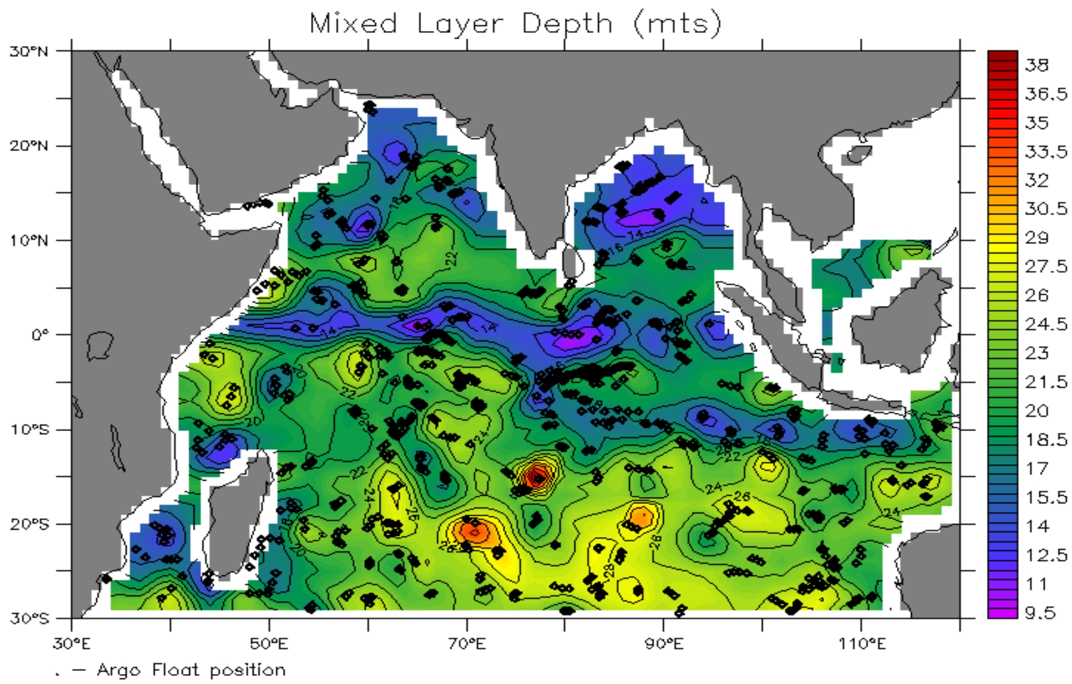

Geostrophic Currents (m/s) at 100m
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E

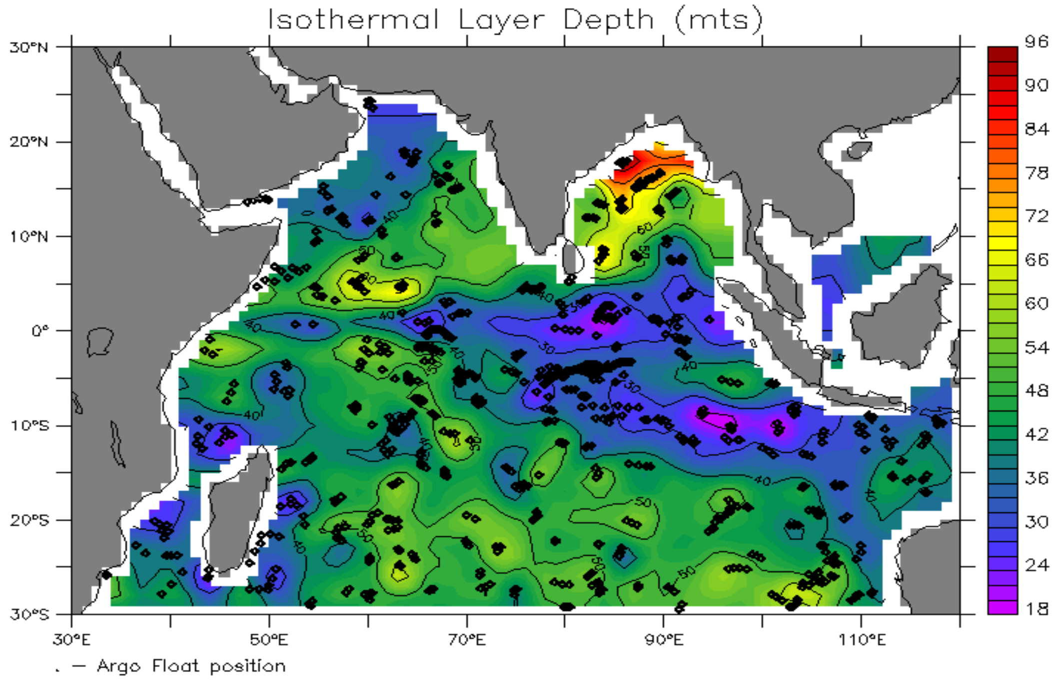

Geostrophic Currents (m/s) at 200m
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E


Geostrophic Currents (m/s) at 500m
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°F
90°E
110°E

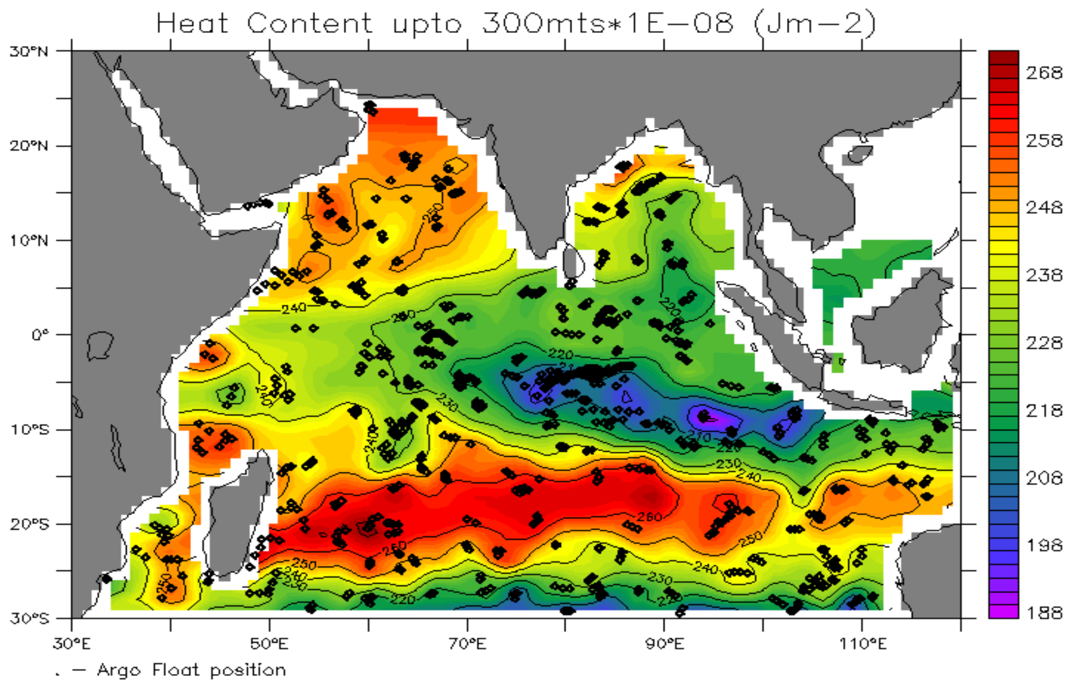


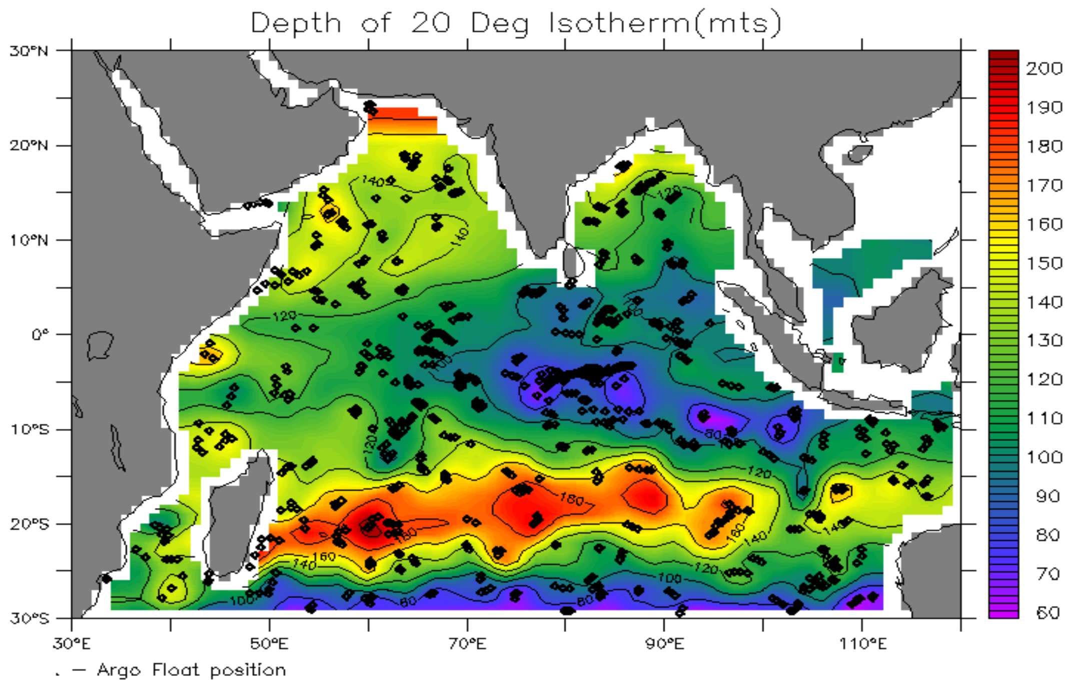
5.4 Mixed Layer Depth and Isothermal Layer Depth



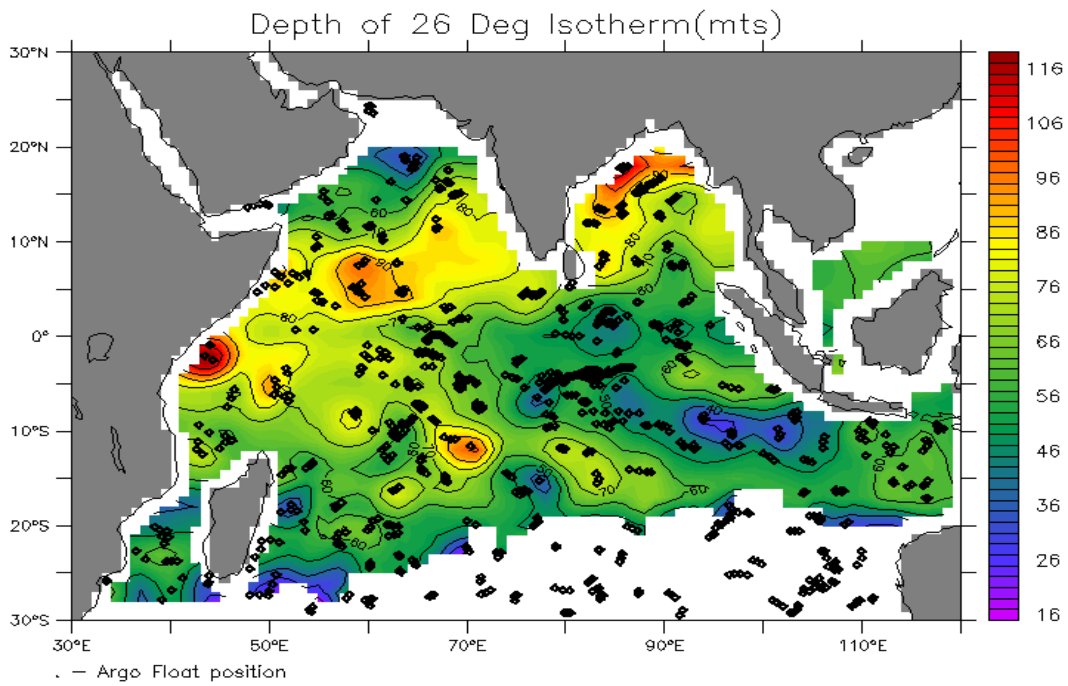


5.5 Heat Content (300 m), D20, D26, Dynamic Height

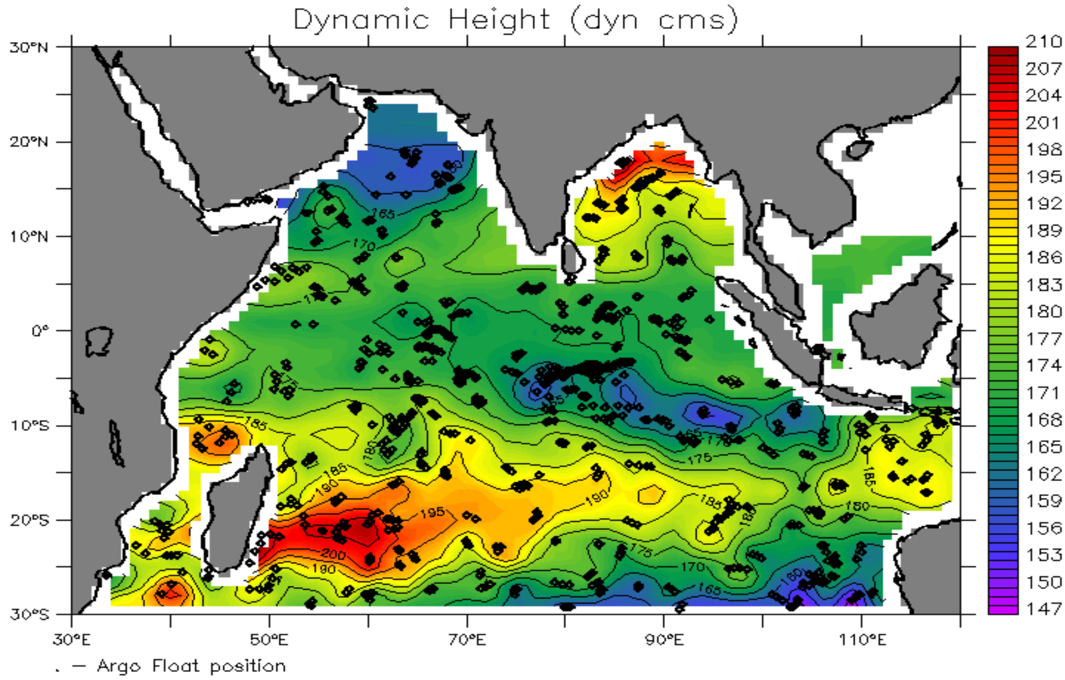




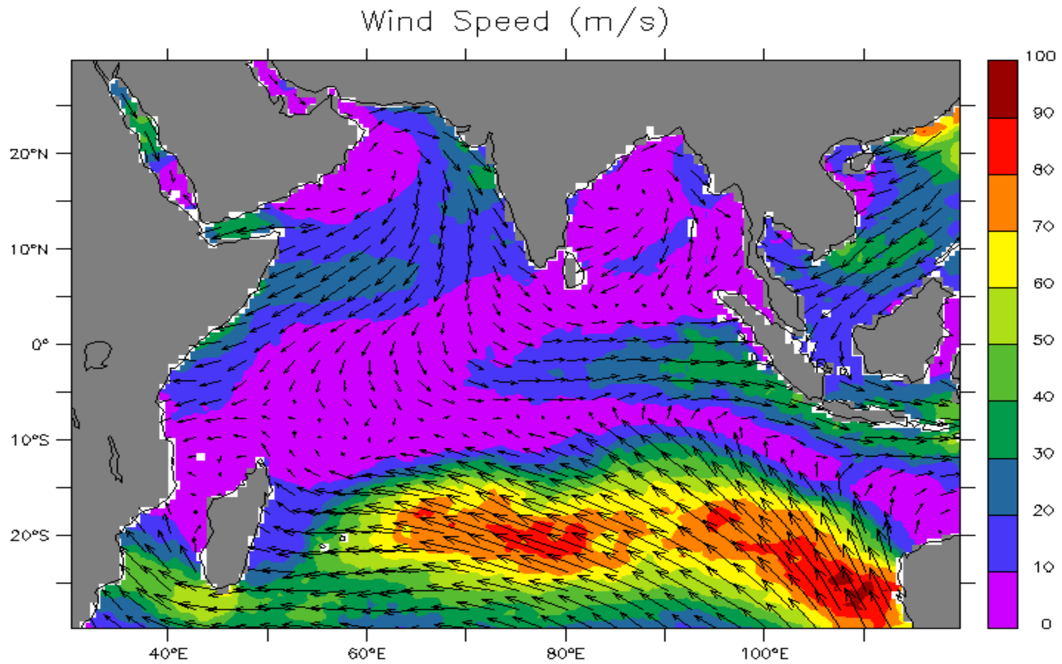
Depth of 20 Deg Isotherm(mts)
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E
. — Argo Float position
200
190
180
170
160
150
140
130
120
110
100
90
80
70
60

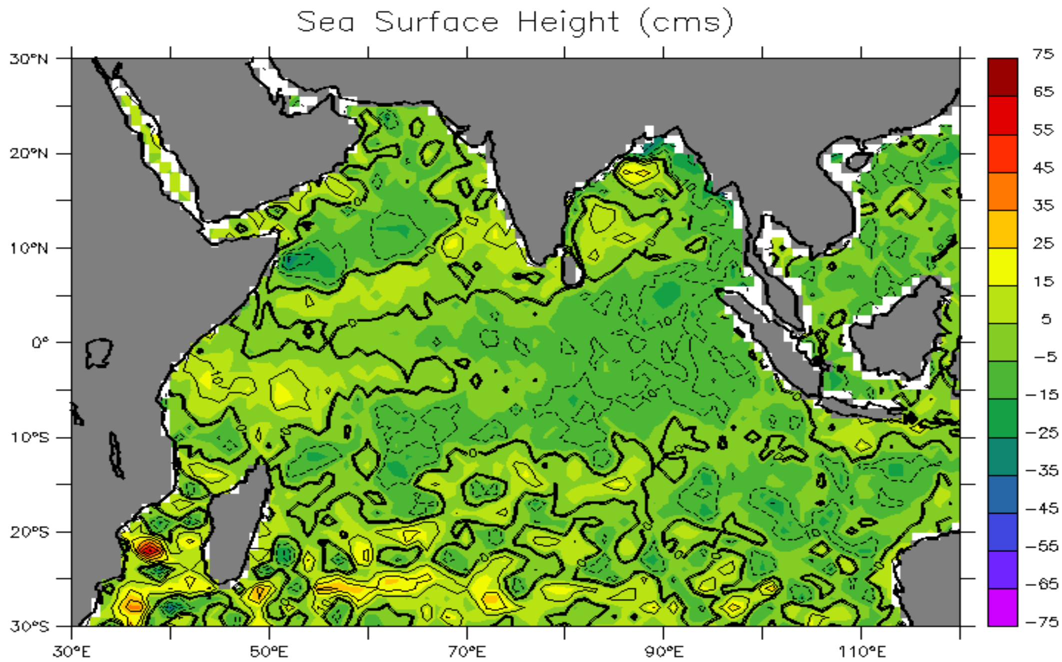


Depth of 26 Deg Isotherm(mts)
30°N
20°N
10°N
0°
10°S
20°S
30°S
30°E
50°E
70°E
90°E
110°E
. — Argo Float position
116
106
96
86
76
66
56
46
36
26
16



5.6 Wind and Sea Surface Height Anomaly





6. Acknowledgements

We express our gratitude to Director, Indian National Centre for Ocean Information Services (INCOIS) for his constant encouragement. Argo data were collected and made freely available by the International Argo Project and the national programmes that contribute to it (http://www.argo.ucsd.edu, http://argo.jcommops.org). The authors wish to acknowledge use of the Ferret program, a product of NOAA's Pacific Marine Environmental Laboratory, for analysis and graphics in this paper.

7. References

- Bretherton, F.P., R.E. Davis, and C.B. Fandry, 1976; *A technique for Objective Analysis and Design of Oceanographic Experiments applied to MODE-73*, Deep Sea Research, V 23, pp. 559-582
- Carter, E.F. and A.R. Robinson, 1987; *Analysis Models for the Estimation of Oceanic Fields*, Jour. of Atmospheric and Oceanic Technology, V. 4, No. 1, pp. 49-74
- Gandin, L.S., 1965; *Objective Analysis of Meteorological Fields*, Isreal Program for Scientific Translations, Jerusaleum, 242 pages
- Liebelt, P.B., 1967; *An Introduction to Optimal Estimation*, Addison-Wesley, Reading Mass., 273 pages
- Kessler and McCreary, 1993: The Annual Wind-driven Rossby Wave in the Subthermocline Equatorial Pacific, Journal of Physical Oceanography 23, 1192 -1207.
- Wong, A., R. Keeley, T. Carval and the Argo Data Management Team. 2004. Argo quality control manual, ver. 2.0b, Report, 23 pp.

8. Appendix 1

The derivation of the **Gauss-Markov** theorem depended upon the matrix identity,

$$(A - BC^{-1})C(A - BC^{-1})^T - BC^{-1}B^T \equiv ACA^T - BA^T - (BA^T)^T \quad (1)$$

This identity is not very intuitive and so we will provide the proof here. This proof depends upon one assumption being true, that $C = C^T$.

Let us define,

$$X \equiv (A - BC^{-1})C(A - BC^{-1})^T - BC^{-1}B^T \quad (2)$$

and,

$$Y \equiv ACA^T - BA^T - (BA^T)^T \quad (3)$$

If we can establish that $X \equiv Y$, then we have our proof.

We start by expanding X,

$$\begin{aligned} X &\equiv (A - B\,C^{-1})\,C\,(A - B\,C^{-1})^T - B\,C^{-1}\,B^T \\ &= (A - B\,C^{-1})\,C\,(A^T - C^{-T}B^T) - B\,C^{-1}\,B^T \end{aligned} \quad (4)$$

and since we have assumed that $C = C^T$, then $C^{-1} = C^{-T}$, so

$$\begin{aligned} X &= (A - B\,C^{-1})\,C\,(A^T - C^{-1}B^T) - B\,C^{-1}\,B^T \\ &= (A\,C - B\,)\,(A^T - C^{-1}B^T) - B\,C^{-1}\,B^T \\ &= ACA^T - BA^T - AB^T + BC^{-1}B^T - BC^{-1}B^T \\ &= ACA^T - BA^T - (BA^T)^T \equiv Y \end{aligned} \quad (5)$$

9. Appendix 2

In this note we prove that a covariance matrix is non-negative definite.

Consider the product, *H*, that is the product of an arbitrary vector, *a*, and the covariance vector, *x*:

$$H = a^T E[(x - \mu)(x - \mu)^T]a \tag{1}$$

We can re-arrange the above by moving the constant vector *a* inside expectation operator so that we have,

$$H = E[a^T (x - \mu)(x - \mu)^T a] \tag{2}$$

If we define

$$Y \equiv (x - \mu)^T a \tag{3}$$

(which is a random variable because x is), then (2) is

$$H = E[\, Y^T \, Y] \tag{4}$$

This ``squared'' quantity is clearly never negative, so that we can conclude that the covariance matrix $E[(x-\mu)(x-\mu)^T]$ is non-negative definite.