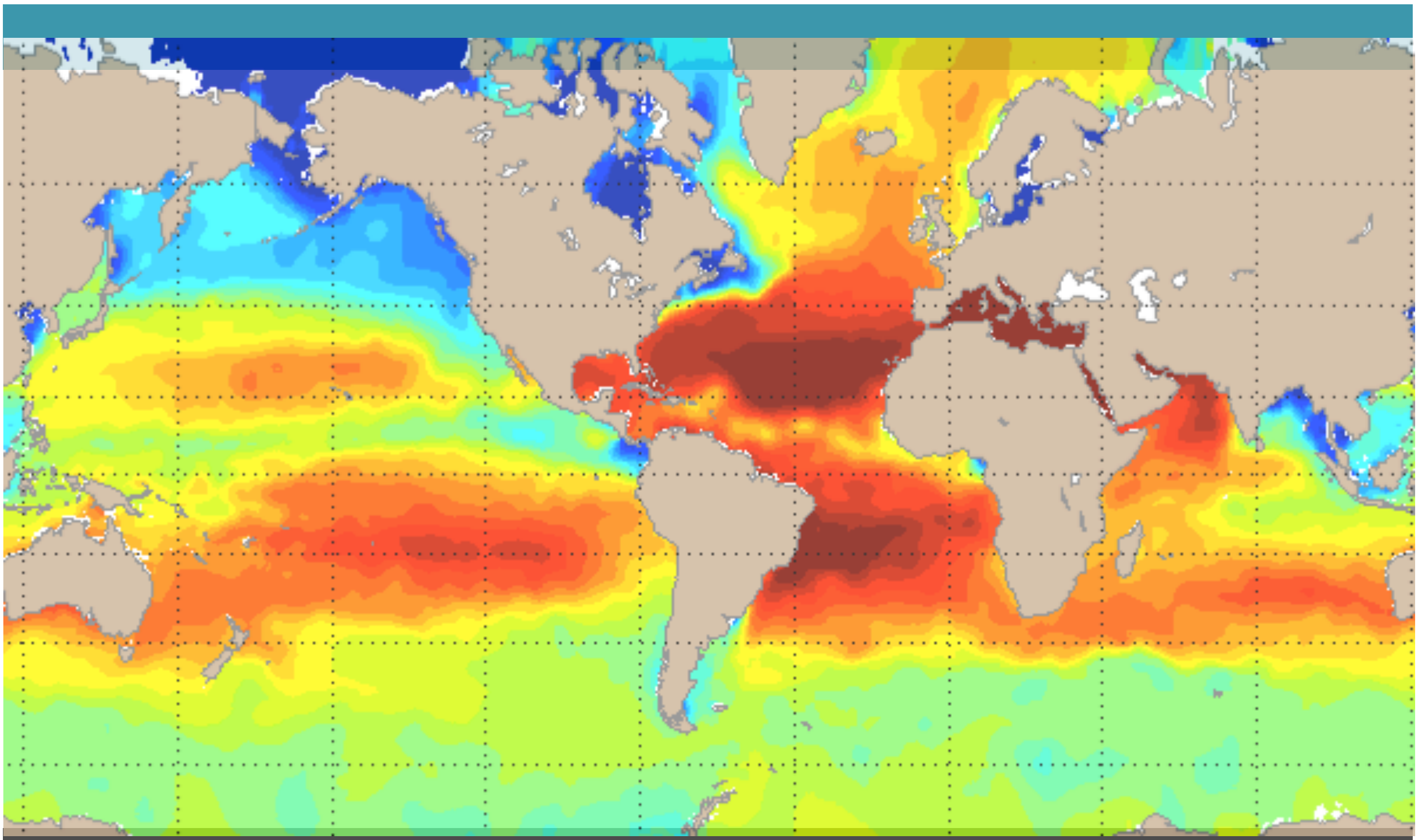


ISAS-Tool Version 6: Method and configuration

F. Gaillard

Rapport LPO 12-02

Laboratoire de Physique de Océans, UMR 6523

History

Auteur	Mise à jour	Date
F. Gaillard	*Création du document – V4 beta*	*03/02/2007*
R. Charraudeau	*V4.00 – Version française*	*23/11/2007*
F. Gaillard	*V4.01 – Version française*	*11/02/2008*
F. Gaillard	*V4.1b - English version*	*19/03/2008*
F. Gaillard	*Minor corrections*	*25/09/2008*
F. Gaillard	V5.1	18/06/2009
F. Gaillard	V5.2b	11/01/2010
F. Gaillard	V5.3: Split method and configuration/program	15/12/2010
F. Gaillard	V6: update method and describe CA2 configuration based on ARV11	19/01/2012

Content

1 Introduction

ISAS (In Situ Analysis System) is an analysis tool for the temperature and salinity fields. Originally designed for the synthesis of ARGO dataset, it has been tested for the first time on the POMME area in the North-East Atlantic in 2000, it was later extended to the Atlantic and the Global ocean as the Argo array was setting up. It is developed and maintained at LPO (Laboratoire de Physique des Océans) within the **Argo** Observing Service (SO-ARGO) where it is used for research purposes on ocean variability. **ISAS** is made available to the **Coriolis** datacenter for exploitation in operational mode. It can accommodate a wide range of in situ measurements if they are provided in the standard NetCdf format distributed by the Coriolis datacenter (http://www.coriolis.eu.org/) . It is based on optimal interpolation and the estimated quantity is the anomaly on depth levels relative to a reference climatology. ISAS is uni-variate, which means that temperature and salinity variables are estimated independently.

This document describes the statistical method used to produce the estimate and the specific choices performed to implement the method. The practical implementation and use of the tool is described in the **user's manual** corresponding to the appropriate ISAS version.

2 Estimation method

ISAS uses estimation theory for mapping a scalar field on a regular grid from sparse and irregular data (Bretherton et al.,1976) the first implementation is described in Gaillard et al. (2008). We use here the unified terminology recommended for data assimilation (Ide et al, 1997).

The interpolated field, represented by the state vector x, is constructed as the departure from the value of a reference field at the grid points . This reference is derived from previous knowledge (climatology or forecast). Only the unpredicted part of the observation vector, or departure from the reference field at the data points, called innovation, is used:

$$d = y^o - x^f \qquad \textbf{Eq. 1}$$

The analyzed field x^a is a linear least square estimator, obtained as the linear combination of the innovation that minimizes the statistical error. A covariance matrix of error is associated to this solution ($\mathbf{P}^a$).

$$\begin{aligned} x^a &= x^f + \mathbf{K}^{OI} d \\ \mathbf{P}^a &= \mathbf{P} - \mathbf{K}^{OI}\mathbf{C}_{ao}^T \end{aligned} \qquad \textbf{Eq. 2}$$

In the objective analysis formalism, the gain matrix $\mathbf{K}^{OI}$ is built from the matrices that express the covariance of the field, from grid point to data point and from data point to data point and the observation noise covariance matrix.

$$\mathbf{K}^{OI} = \mathbf{C}_{ao}\left(\mathbf{C}_{oo} + \mathbf{R}\right)^{-1} \qquad \textbf{Eq. 3}$$

The error on the estimation is given by the diagonal of this matrix, usually presented as a percentage of the a priori variance.

This solution makes implicit use of an observation or mapping matrix $\mathbf{H}$, such that : $y^o = \mathbf{H}x + \varepsilon$, which by analogy with the Ide et al. (1997) formalism can be expressed as: $\mathbf{H}^T = \mathbf{P}^{-1}\mathbf{C}_{ao}$.

It should be noticed that this formalism provides at the same time an estimate of the misfit between observations and analysis, also called analysis residuals:

$$\begin{aligned} \delta &= y^o - \mathbf{H}x^a \\ \delta &= \mathbf{R}\left(\mathbf{C}_{oo} + \mathbf{R}\right)^{-1} d \end{aligned} \qquad \textbf{Eq. 4}$$

We make use of these residuals in order to detect erroneous data, either outliers, biases or drifts. The advantage of such method is that the residuals are computed with the correct mapping matrix. Moreover, it is not necessary to perform an analysis at each data point to obtain the residuals, they are obtained at once for the whole data set.

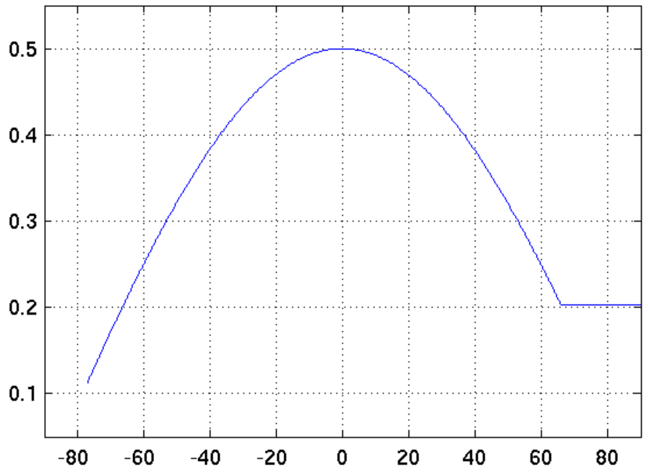
The definition of the variables is as follows:

- d: Innovation vector
- y^o: Observation vector
- x^f:Climatology or forecast at observation location (vector)
- x^a: Value of the analyzed field at the grid points (vector)
- ε: Observation error (measurement noise + representativity error)
- δ: Data misfit or residual (vector)
- $\mathbf{K}^{OI}$: Optimal estimation matrix, similar to the Kalman gain matrix
- $\mathbf{P}$: A priori Covariance matrix of the anomaly field at grid points
- $\mathbf{P}^a$: Covariance matrix of the error on the analyzed field
- $\mathbf{R}$: Covariance matrix of the error on observations (cumulates measurement error and representation error)
- $\mathbf{C}_{ao}$: Covariance matrix of the anomaly field between grid points and observation points
  $\mathbf{C}_{oo}$: Covariance matrix of the anomaly field at observation points.

3 Grid and bathymetry

The horizontal grid is ½ degree Mercator from 77°S to 66.5°N, where it is thus isotropic with a resolution that increases with latitude, from 66.5°N to the North pole, the latitude step is fixed. The vertical resolution increases from 20 m at 2000m to 5 meter in the upper layer. Near surface two levels have been added at 0 and 3 m. The grid properties are summarized Figure 1.

The bathymetry is an interpolation over our grid of the file `etopo2bedmap.nc` produced by MERCATOR from the 2 minutes bathymetry file of the NGDG `Bathy_Etopo2.nc`. The interpolation is done using the median of the 4 surrounding points.



Levels (m):	Step
[0 3]	3
[5:100]	5
[110:800]	10
[820:2000]	20

Figure 1 : Horizontal and vertical resolution of the analysis grid

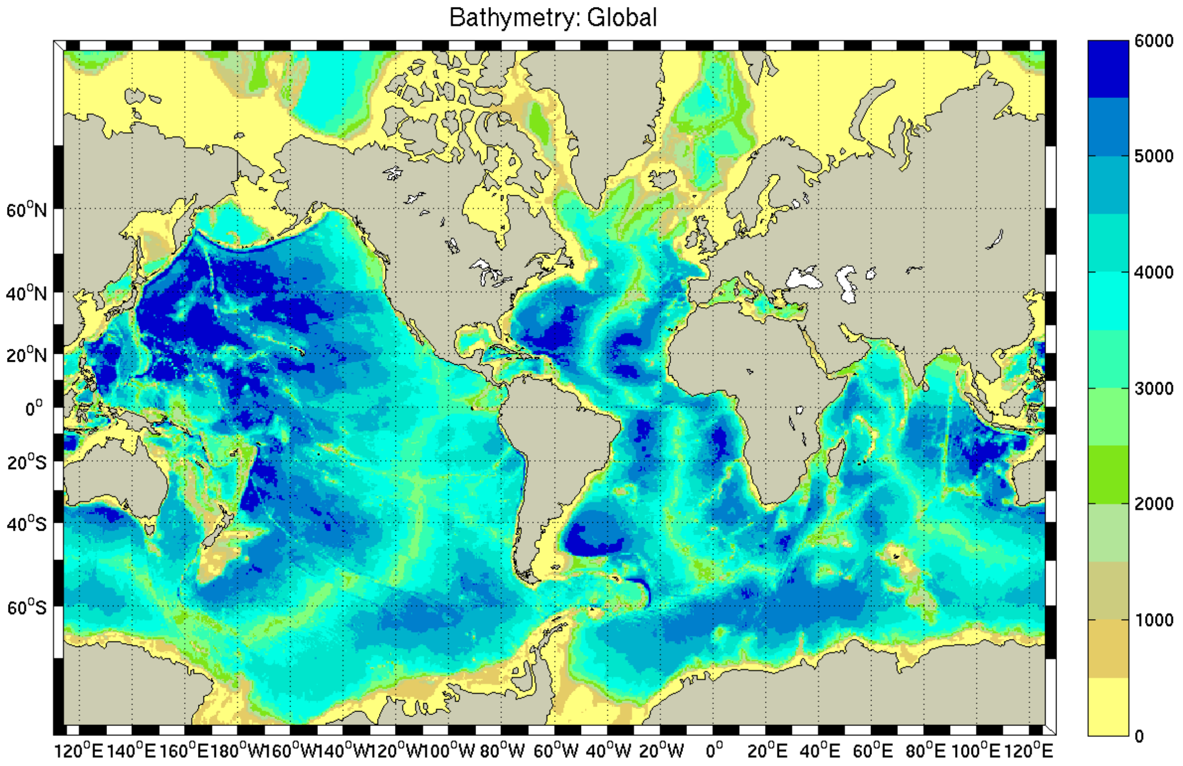


Figure 2 : Bathymetry interpolated on the analysis grid (global view).

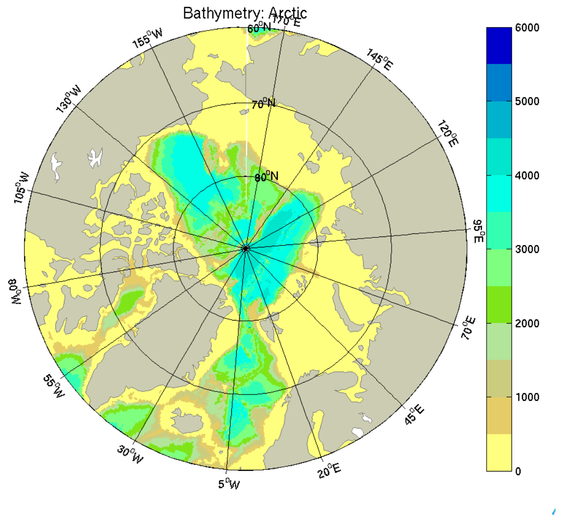


Figure 3 : Bathymetry interpolated on the analysis grid (polar view).

4 Reference climatology

The climatology is an important component of the a priori statistical information used by ISAS to compute the estimate. The mean ocean state defines the background or reference state (x^f) and the variance is necessary to compute the elements of the covariance matrices that appear in equations 2-4.

Two types of climatologies can be used with ISAS: a climatology derived from the NODC WOA atlas or a climatology constructed from a previous ISAS analysis. In each case gridded fields of the following variables must be provided on the ISAS grid:

- Monthly mean for temperature and salinity
- Variance for temperature and salinity relative to the monthly mean. At the moment this variance is assumed constant over the annual cycle.

4.1 Mean field from NODC climatology

In the case of NODC climatology, provided on a 1 degree grid and low vertical resolution, the data must be interpolated on the ISAS higher resolution grid but it is also necessary to extrapolate the NODC atlas from the ocean to the land. This interpolation is done in three steps:

1. Extend NODC atlas at constant latitude over land at NODC levels
2. Perform 2D horizontal bilinear interpolation onto ISAS grid at NODC levels
3. Apply land mask
4. Perform vertical linear interpolation on ISAS vertical levels

Monthly fields are provided by NODC only from 0-1500m. They are extended to 2000m using the annual mean. The transition is smoothed with a linear filter (1/4, 1/2, 1/4).

4.2 Mean field from ISAS climatology

An ISAS monthly mean climatology can be easily computed by averaging over several years the gridded fields from a previous analysis. The ARV11 climatology is based on the D2CA1S2 analysis over the period 2004-2010. The monthly climatology has been obtained by averaging all years for each month. A mean full depth annual field has also been constructed by combining the average of the monthly climatology in the upper 2000m with the WOA05 climatology at greater depth.

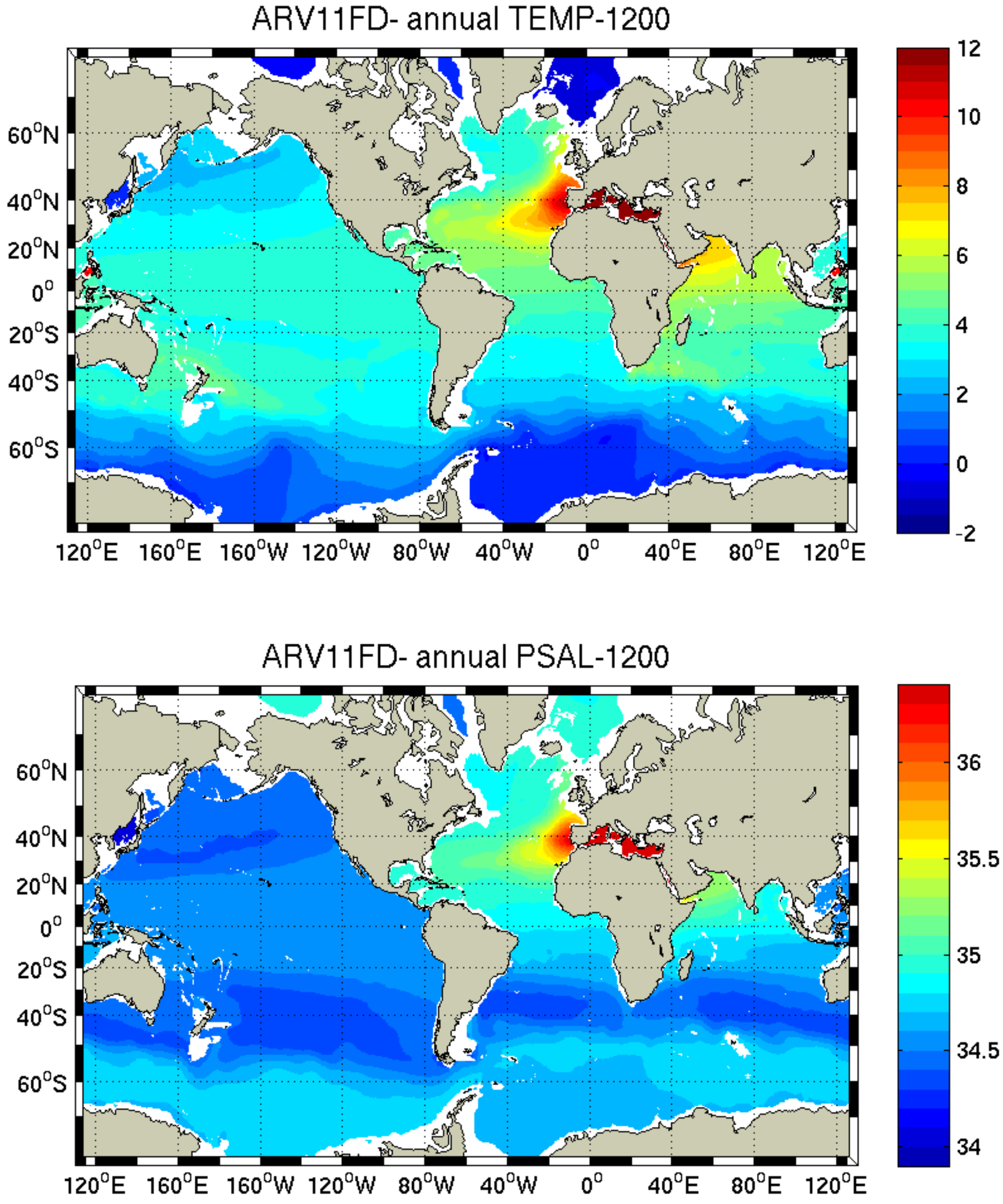


Figure 4 : Monthly climatology ARV11 in May used for the CA2 configuration. Upper pannel : temperature, lower pannel : salinity.

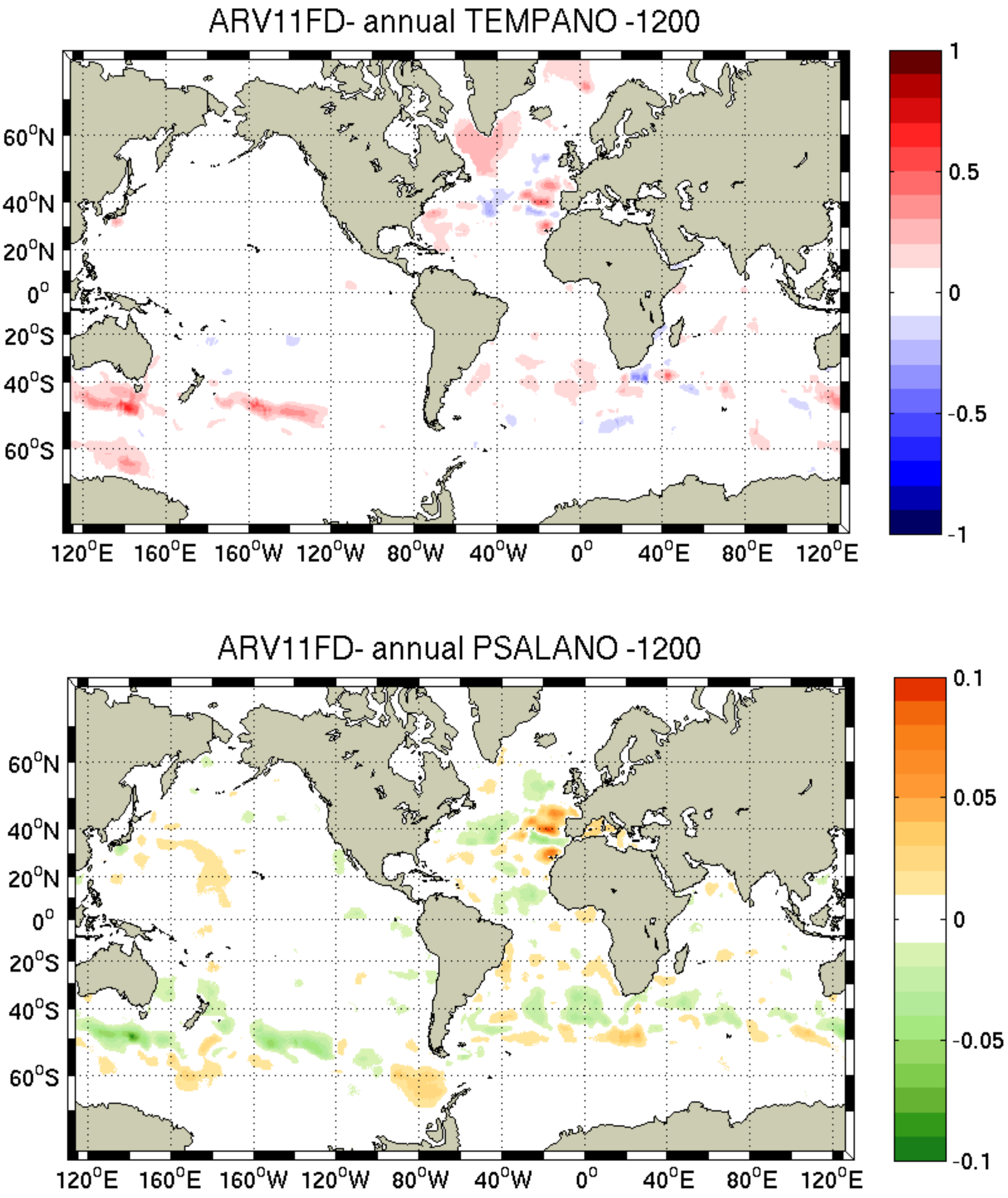


Figure 5 : Anomaly of the ARV11 annual 2004-2010 mean relative to WOA05. Upper pannel : temperrature, lower pannel : salinity.

5 Variances and Covariance scales

5.1 Covariances

Statistical information on the field and data noise are introduced through the covariance matrices that appear in equation 1. We assume that the covariance of the analyzed field can be specified by a structure function modeled as the sum of two Gaussian function, each function associated with two horizontal space scales and one time scale:

$$\boldsymbol{C(dx, dy, dt) = \sum_{i=1}^{2} \sigma_i^2 \exp -\left(\frac{dx^2}{2L_{ix}^2} + \frac{dy^2}{2L_{iy}^2} + \frac{dt^2}{2L_{it}^2}\right)} \qquad \textbf{Eq. 5}$$

where $dx, \ dy, dt$, are the space and time separations, $L_{ix}, \ L_{iy}, \ L_{it}$ the corresponding e-folding scales. The weight given to each ocean scale is controlled by the variances σ_i^2. The total variance is computed as the variance of the anomaly relative to the monthly reference field as explained in the previous section. It is considered as the sum of four terms:

$$\boldsymbol{\sigma^2 = \sigma_{L1}^2 + \sigma_{L2}^2 + \sigma_{UR}^2 + \sigma_{ME}^2} \qquad \textbf{Eq. 6}$$

σ_{L1}^2 and σ_{L2}^2 are the two terms appearing in eq. 5, their sum is the total field variance. The remaining sum is the total error variance: σ_{ME}^2 corresponds to the measurement errors and σ_{UR}^2 represents small scales unresolved by the analysis and considered as noise, also called representativity error.

A unique σ_{ME}^2 profile has been computed from the measurement errors of the standard database and substracted from the total variance to obtain the ocean variance σ_{ocean}^2 (first three terms of the sum). We express the variances associated to each scales as a function of the ocean variance by introducing normalized weights:

$$\sigma_{L1}^2 = \sigma_{ocean}^2 W_1; \sigma_{L2}^2 = \sigma_{ocean}^2 W_2; \sigma_{UR}^2 = \sigma_{ocean}^2 W_{UR}$$

$$W_1 + W_2 + W_{UR} = 1$$

5.2 Covariance scales

The free parameters of the system are the weights W_i that define the distribution of variance over the different scales. The error matrix combines the measurement error and the representativity error due to unresolved scales, it is assumed diagonal, although this is only a crude approximation since both errors are likely to be correlated for measurements obtained with the same instrument, or within the same area and time period. The first scale length L_1 is fixed over the ocean, separate values can be used in x and y. Most of the time it is set to 300km, the target Argo resolution. The second length L_2 is proportional to the Rossby Radius computed from the annual full depth climatology. This value is bounded by 300 km in the Y direction and 600 km in the X direction (Figure 6) for the highest limit and twice the grid size for the lowest limit.

5.3 Variances

The variance fields are computed from a data set prepared for a previous analysis, that have been interpolated on ISAS levels. In the case of ARV11 all data over the period 2004-2010 were used. The variance is computed at each grid point as the mean square anomaly relative to the monthly climatology for all data within a square of size 10 grid points. The annual mean standard deviation (STD) is the square root of the mean monthly variances. To take into account that some areas remain under-sampled (the southern ocean in particular) and that some erroneous data may remain in the data set, we impose an upper and a lower bound, and for data void areas, a default value is proposed. These values were computed as follows: At each level we computed the distribution, the lower bound is defined as the value of the first 0.6 percentile, the upper bound is the value at the 99.9 percentile from 0 to 20m and of the 99.5 percentile below. The default value is the most probable value. An exponential is fitted through the lower bound and default value profiles, for the maximum we applied a linear filter twice (¼ ½ ¼). The resulting profiles are shown Figure 7

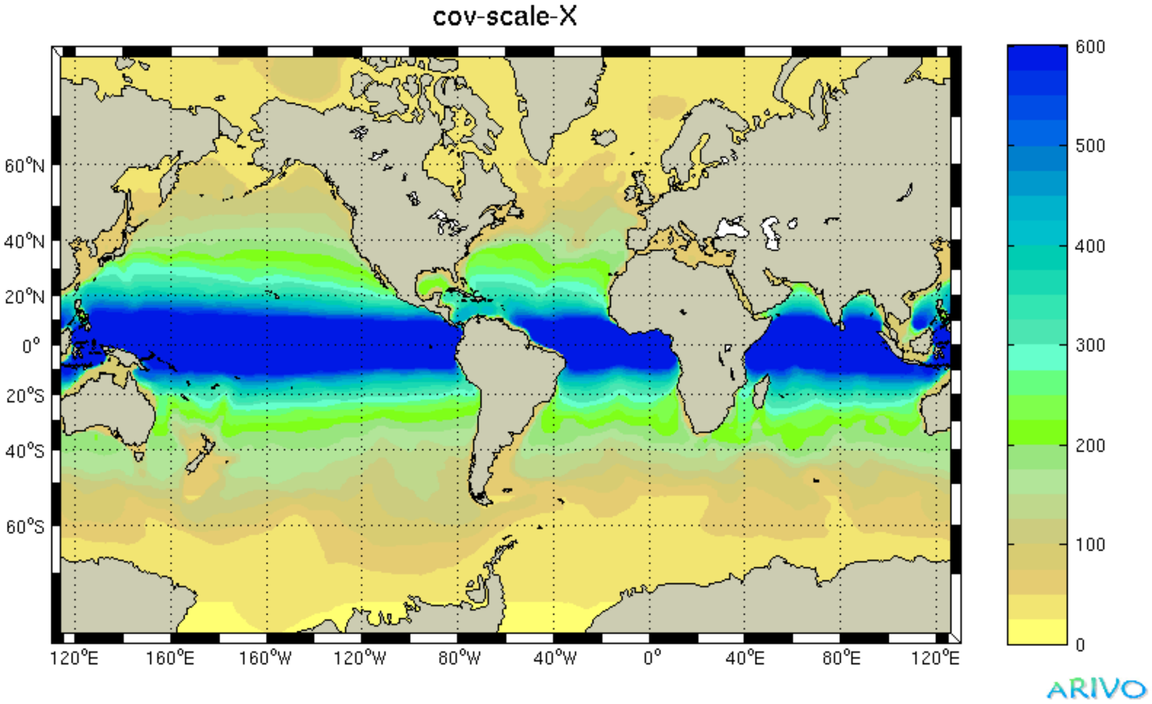


Figure 6 : Covariances scales L2 in the X direction (deduced from the Rossby Radius computed in the ARV11 full depth annual mean).

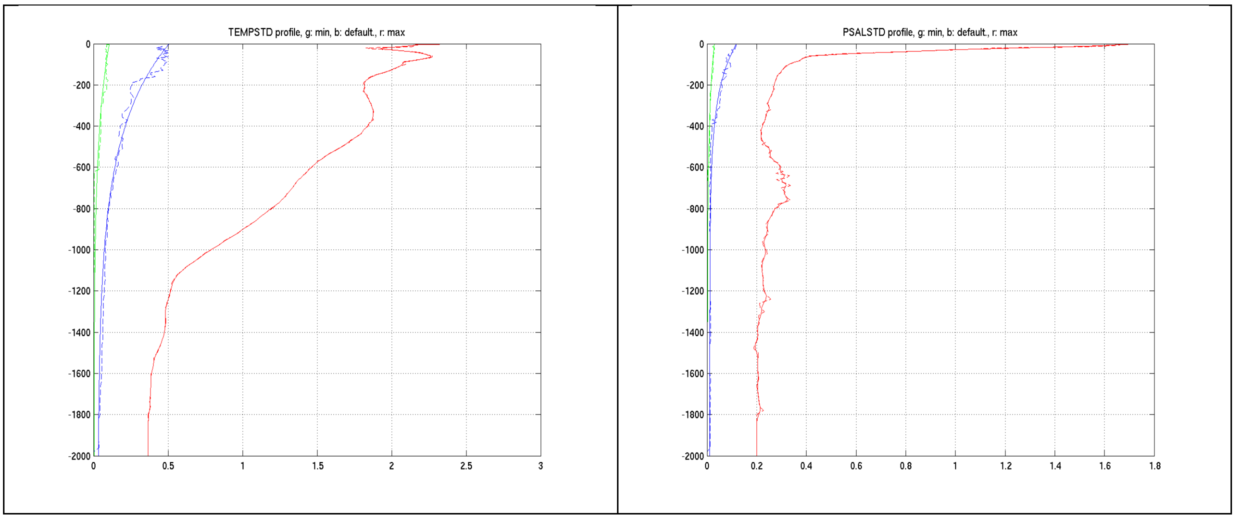


Figure 7 : Vertical profiles for the minimum (green), default (blue) and maximum (red) values for the standard deviation of temperature (left) and salinity (right).

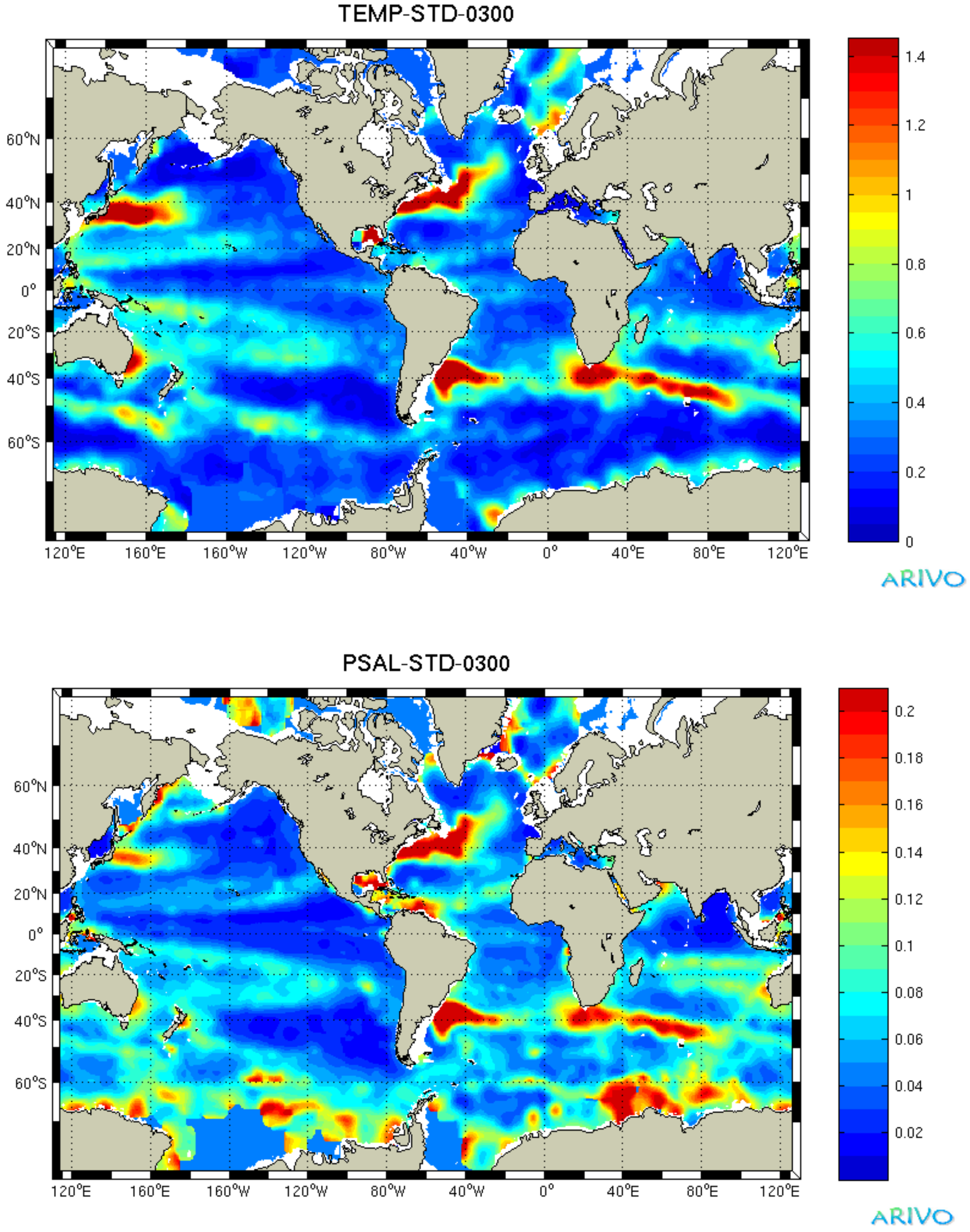


Figure 8 : Standard deviation of temperature and salinity at 300 m in CA2 (ARV11 based) configuration.

6 Areas and masks

For the practical implementation of the method, the global ocean has been divided in areas. Each area defines the group of points that are processed at once (the yellow area of Figure 9) . A mask associated to each area indicates which data are taken into account (the yellow and green area of Figure 9). Generally the 'useful' area include all points surrounding the 'analyzed' area, but some points can be masked to avoid mixing data from different basins.

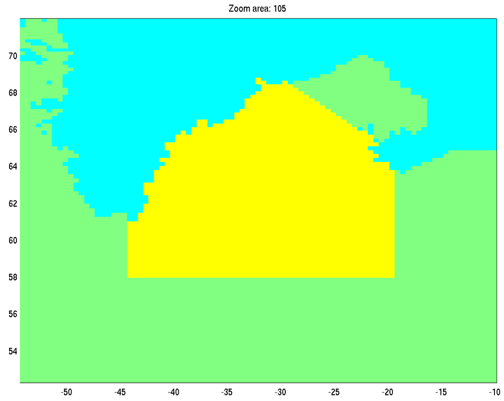


Figure 9 : Mask for area 105 : In yellow the 'analyzed' area, in green the 'useful' area.

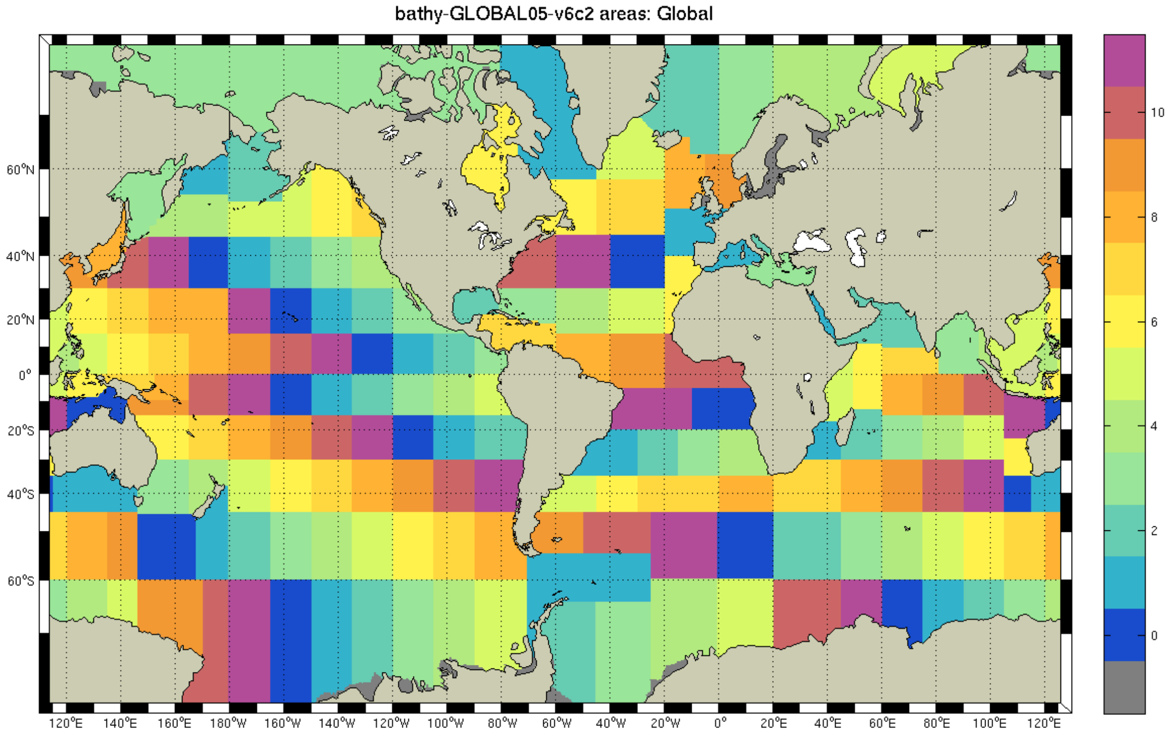
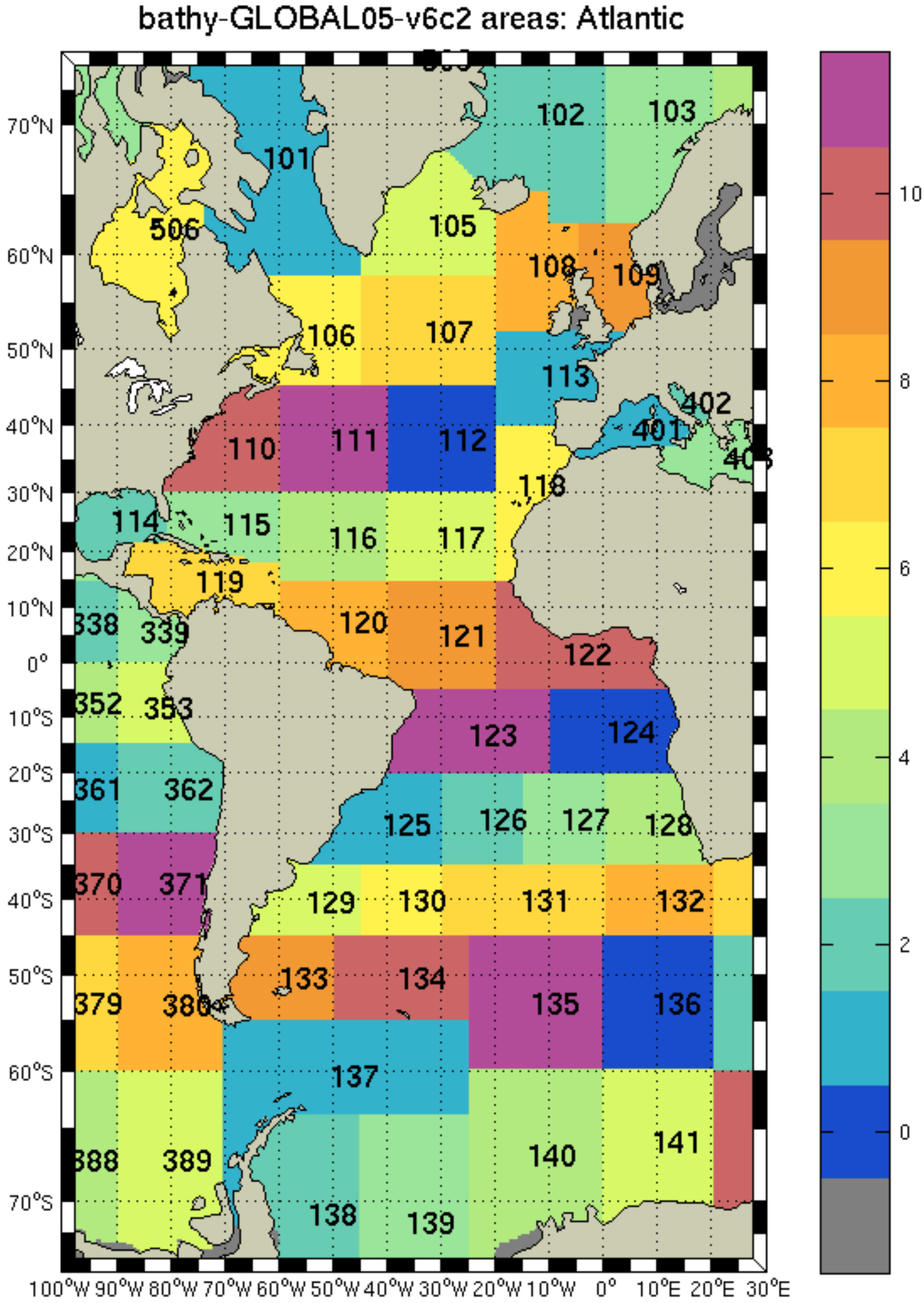


Figure 10 : Areas defined over the global ocean



bathy-GLOBAL05-v6c2 areas: Atlantic

Figure 11 : Area numbering, Atlantic

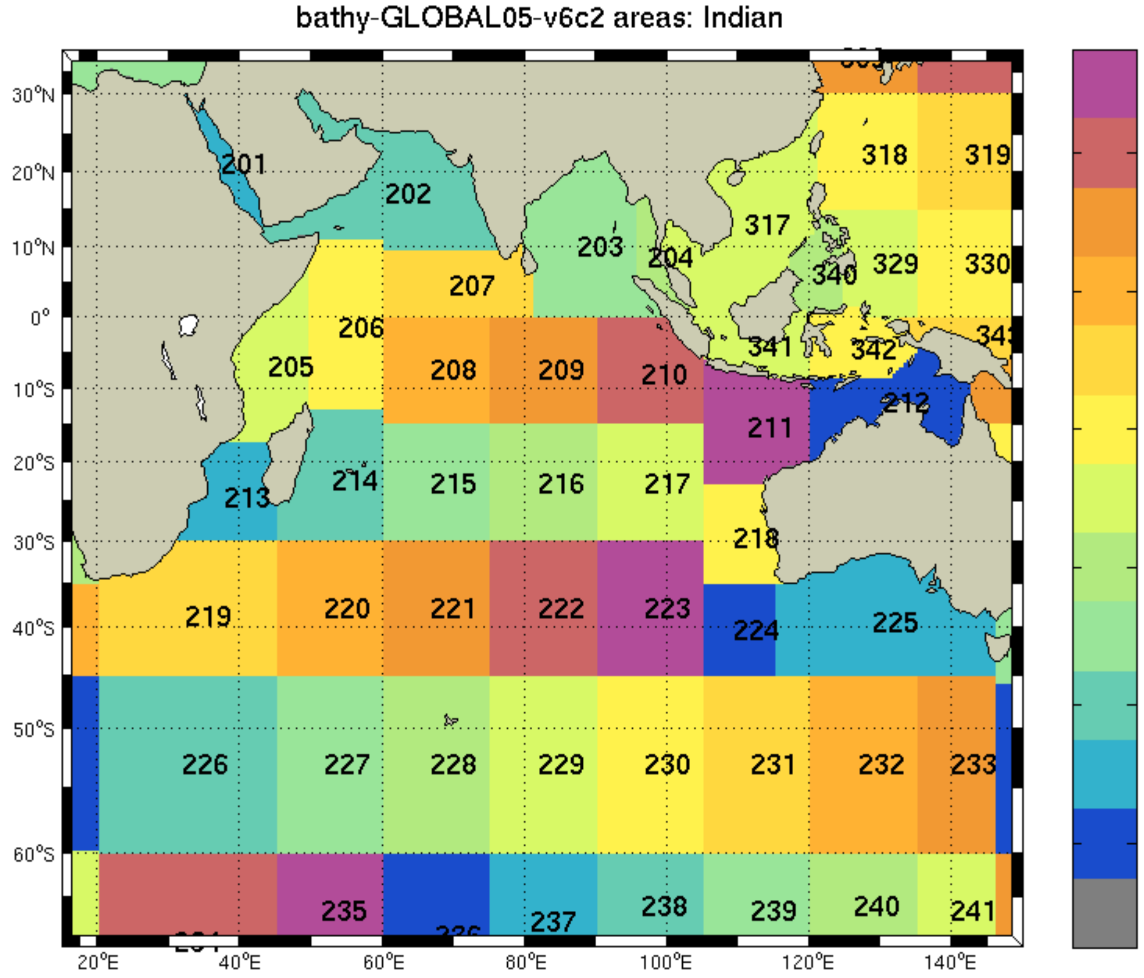


Figure 12 : Area numbering, Indian Ocean

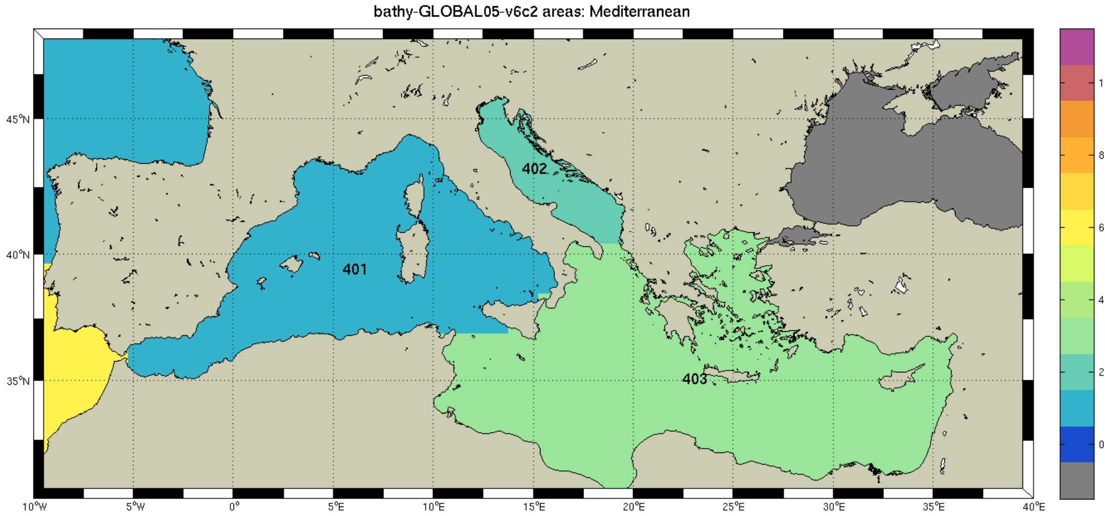


Figure 13 : Area numbering, Mediterranean

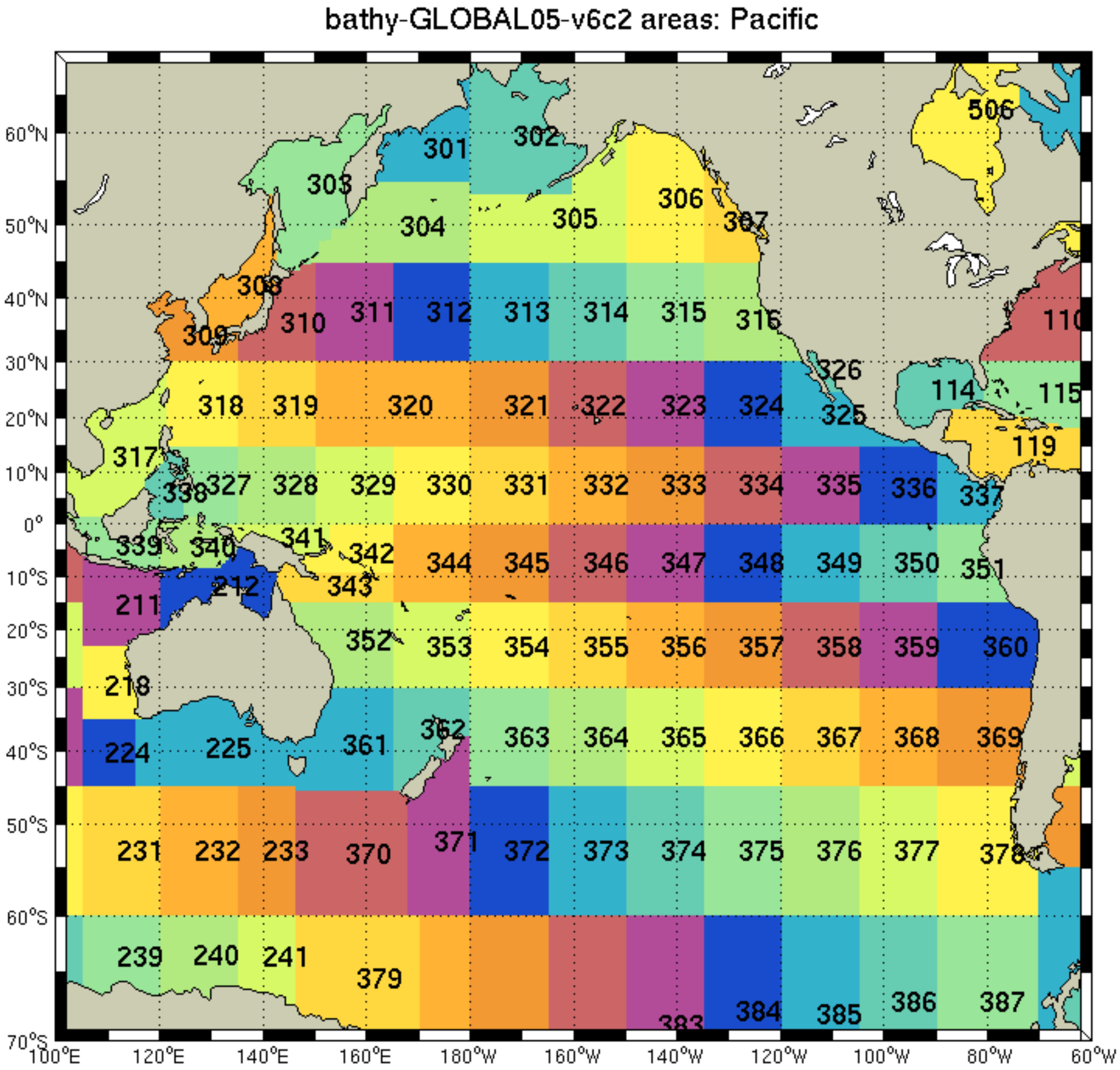


Figure 14 : Area numbering, Pacific

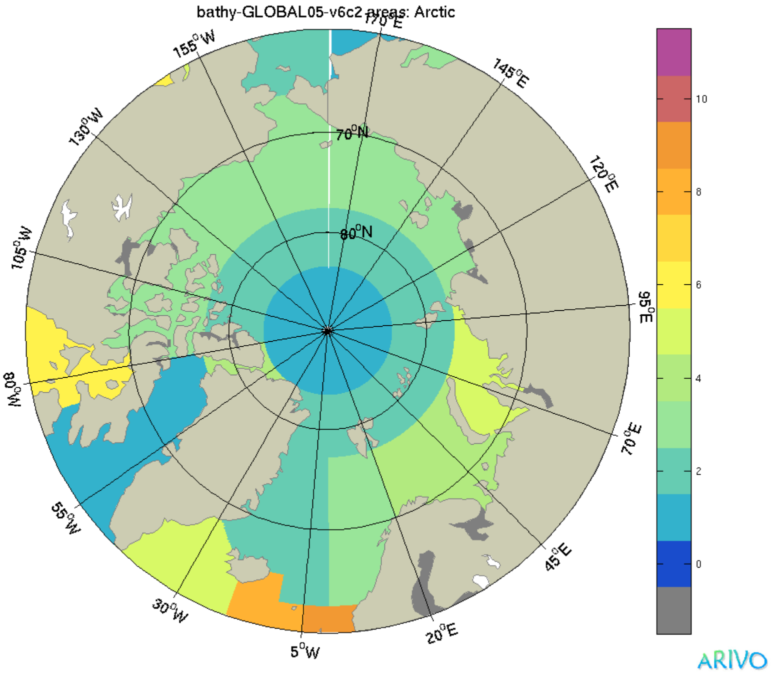


Figure 15 : Area numbering, Arctic

7 References

Gaillard, F., E. Autret, V.Thierry, P. Galaup, C. Coatanoan, and T. Loubrieu , 2009 : Quality control of large Argo data sets. JOAT, Vol. 26, No. 2. 337–351

Bretherton, F., R. Davis, and C. Fandry, (1976), A technique for objective analysis and design of oceanic experiments applied to Mode-73. *Deep Sea Research*, 23, 1B, 559--582.

Ide, K., P. Courtier, M. Ghil and A. C. Lorenc (1997), Unified Notation for Data Assimilation: Operationnal, Sequential and Variational. Journal of the Meteorological Society of Japan, 75, no 1B, 181-189.