

4. Interpolation onto a regular grid

4.1 Optimal interpolation method

The optimal interpolation method was used to compute climatological property distributions of the selected standard levels on a regular grid. The methodology we use in this study has been described previously in the oceanography literature, where it is variously referred to as Gauss-Markov, optimal or statistical interpolation, or objective analysis (Gandin, 1963; Bretherton et al., 1976). The technique is commonly used to interpolate irregularly sampled, noisy data onto regular grids for subsequent analysis.

The method employs a location-dependent background or first-guess field G . An analysed grid-point value F_o is the first guess evaluated at the grid point plus an interpolated analysis error. The latter is an interpolation to the grid-points of the differences between observation values and values of the background field at the observing points. Thus an analysed grid-point value F_o is:

$$F_o = \sum w_i (F_i - G_i) + G_o.$$

The weights w_i are those weights which minimise the ensemble average of the squared difference between the analysis value and the true value of the field signal. For any specific set of observing sites (i,j) and grid-point (o) locations the first-guess-minus-observation differences and the grid-point first-guess error are considered to be stochastic variables with a joint statistical distribution for which the covariances are known or can be computed or modelled.

The minimisation gives a set of linear equations for optimal weights \mathbf{w} :

$$(\mathbf{A} + \lambda \mathbf{I}) \mathbf{W} = \rho \quad (1)$$

where \mathbf{A} is the signal correlation matrix with elements $A_{ij} = \rho(\Delta x_{ij})$, \mathbf{I} is the identity matrix, and $\lambda = \sigma_n^2 / \sigma_f^2$, an Δx_{ij} represent the spatial separation between points i and j, $\rho_f = \rho(\Delta x_{io})$.

The method requires knowledge of variances of signal and noise and of the spatial autocorrelation function ρ for *increment* fields. The signal is defined as variability with scales larger than the smallest scales of interest. The noise is variability with smaller scales, plus random instrumental errors.

An advantage of the optimal interpolation method is that it also returns an estimate of the uncertainty (error variance). The relative error ε' depends on the observation locations, and on the levels of signal and noise variance:

$$\varepsilon'_o = (1 - \sum \sum \rho_{oi} A^{-1}_{ij} \rho_{oj}), \quad (2)$$

Another important advantage over empirical distance-weighting schemes is that the optimal interpolation method takes into account relative separations among the observing sites but not only the individual separations between the grid-point and the observing sites. The presence of the correlations among input increments in the weight determination algorithm controls for redundancy of information from sources whose increments are statistically related.

4.2. Data reduction

The optimal interpolation requires inversion of the covariance matrix, which becomes impractical for a large number of observational points. Usually, only the data points closest to

the grid node are taken to obtain a field estimate. It means, that most of the observed data are actually lost for the analysis. In this case the climatic estimate is based on a number of observations which may not be representative of the regional long-term mean conditions.

Chelton and Schlax (1991) analysing colour scanner satellite data introduced the notion of time averaging to the standard optimal interpolation method in recognition that some temporal averaging is desirable to reduce the aliasing of high-frequency variability in the signal. Time of observation is neglected in our interpolation and original data are subject to spatial averaging prior to optimal interpolation. The ocean was subdivided into 0.5-degree latitude zones between 80°S and 90°N. Each zone was in turn subdivided into boxes with the longitude size equal to 55 km for the zone mid-latitude. Thus the quantity that is estimated is not the signal at a particular estimation point but its average over a 55x55 km box. Data averaging was done within each box if at least 4 profiles were available. Similar averaging procedure was used by Levitus (1982) and in the later updates of the NOAA ocean climatology, where one-degree data averages served as input for the further analysis. However, unlike in NOAA climatologies we performed averaging on the potential density surfaces, referenced to the pressure of the respective standard level. Since diapycnal processes are assumed to be important in the near-surface layer, within the upper 100-meter layer averaging was performed on the isobaric (standard depth) surfaces.

The averaging results in a 10-fold reduction of the input data, from 1,059,535 original profiles to only 106 330 profiles (for all data mean averaging), with the total number of the box-averaged profiles being 21311. **Fig. 12** shows the distribution of the averaged and original profiles used finally for the optimal interpolation of temperature and salinity.

4.3. Modelling spatial lag correlation

The correlation structure for the increment field has a central and highly sensitive role in the optimal interpolation method. The optimal interpolation requires the knowledge of the spatial correlation function ρ , and the signal-to-noise ratio γ^{-1} . Unfortunately, because of a general data paucity the determination of the spatial correlations for the fields of oceanographic parameters is a difficult task, compared with the situation in meteorology, where time series observations are often routinely available. Satellite observations however provide sufficient information to determine statistical characteristics of oceanographic fields but only at the ocean surface (Kuragano and Kamachi, 2000). Repeat XBT sections were also used to infer the spatial correlation structure. Meyers et al (1991) give an example of space-time scale determination of sea surface temperature and depth of the 20°C isotherm in the Tropical Pacific Ocean.

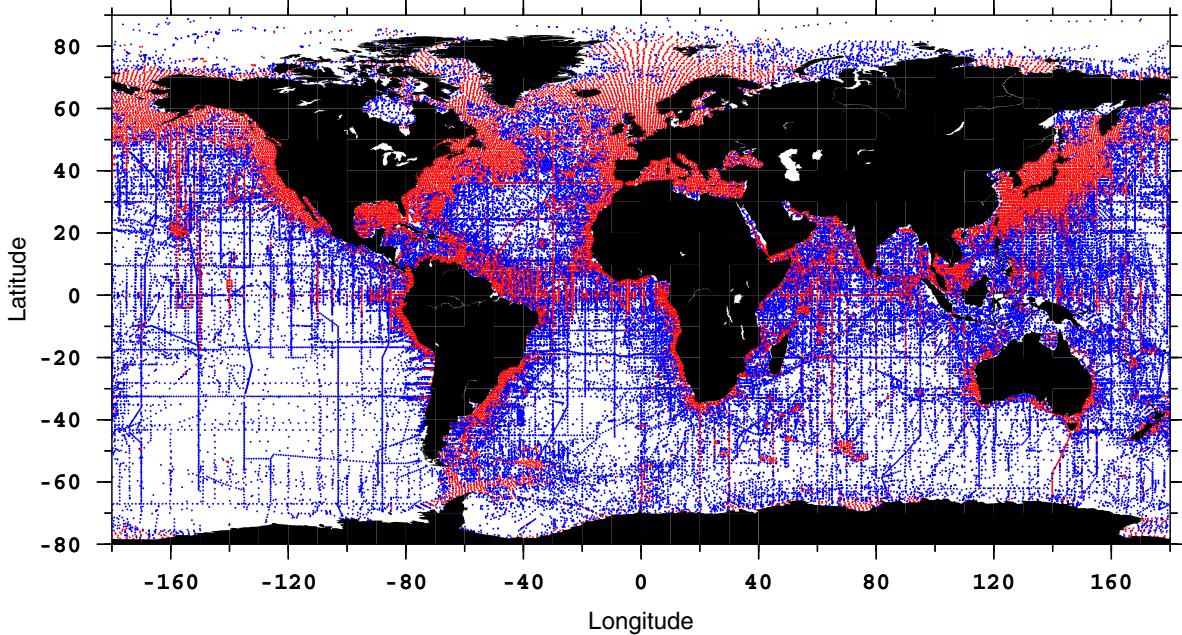


Fig.12: Observed (blue) and averaged (red) T,S-profiles used as input for the optimal interpolation.

In most applications of optimal interpolation the negative squared exponential (called often "Gaussian function") has been favoured as the shape of the correlation function, although many other functions have been considered. A "gaussian" model of the autocorrelation function was used for this study, which employs a decorrelation length scale R :

$$\rho(r) = \exp(-r^2/R^2) \quad (3)$$

We note, that our assumption of this covariance is highly arbitrary, and we choose the "Gaussian" form because of the relatively simple structure as well as the lack of a more appropriate choice. The parameter R (decorrelation scale or e-folding length scale) is a measure of the spatial scale of correlation. The shortcomings of the "Gaussian" function when applied to a typical geophysical data can be summarised as a propensity to overestimate the weights at short lags and underestimate them at large lags (McIntosh, 1990). This amounts to oversmoothing at small scales and may be countered by artificially lowering the value of λ , causing the interpolation to follow the data more faithfully.

It is well known, that many important features of the oceanic circulation and water mass distribution are closely connected to the bottom topography. Thus, most of the World Ocean boundary currents are situated above the continental slope, having a cross-current scale of $o(50\text{km})$, e.g. small compared with the typical size of the ocean basin. Similar, fresh water river plumes significantly alter property distributions in the coastal areas, but are effectively separated from the open ocean areas by sharp fronts. It is therefore desirable to preserve the "narrowness" of boundary currents in the analysed fields. Using large decorrelation scales would result in unrealistically wide boundary currents. We included the topographic information into the calculation of the characteristic length scale R . For each grid node the distance D between the node and the coast was determined, with the respective R determined from:

$$\begin{aligned} R &= R_{\max} \text{ for } D \geq D_{\max}, \\ R &= R_{\min} + (R_{\max} - R_{\min}) D/D_{\max} \text{ for } 0 < D < D_{\max} \end{aligned} \quad (4)$$

where R_{min} is the e-folding scale at the coast line ($D=0$), and R_{max} is the maximum e-folding length scale in the open ocean. The spatial correlation scale R_{max} is set as 450 km in our calculation and is intended to represent typical signal scales, D_{min} is set as 500 km. The spatial distribution of the decorrelation length scale is given in **Fig.13**.

4.4. Computational details

A set of 45 standard levels was selected for which the objectively analysed property fields were computed:

0, 10, 20, 30, 40, 50, 75, 100, 125, 150, 175, 200, 250, 300, 350, 400, 450, 500, 600, 700, 800, 1000, 1100, 1200, 1300, 1400, 1500, 1750, 2000, 2250, 2500, 2750, 3000, 3250, 3500, 3750, 4000, 4250, 4500, 4750, 5000, 5250, 5500, 5750, 6000 meters.

We note that the NOAA climatologies are available for 33 levels for the depth range 0-5500 meters.

For each grid point, data are initially selected in a large subdomain, whose radius is arbitrary set as 750 km (about 2 times larger than the space correlation scale). The first-guess field estimate is taken as the mean of the data in the subdomain. Next, data are selected for the objective mapping: inside a subdomain up to 150 nearest observations are retained

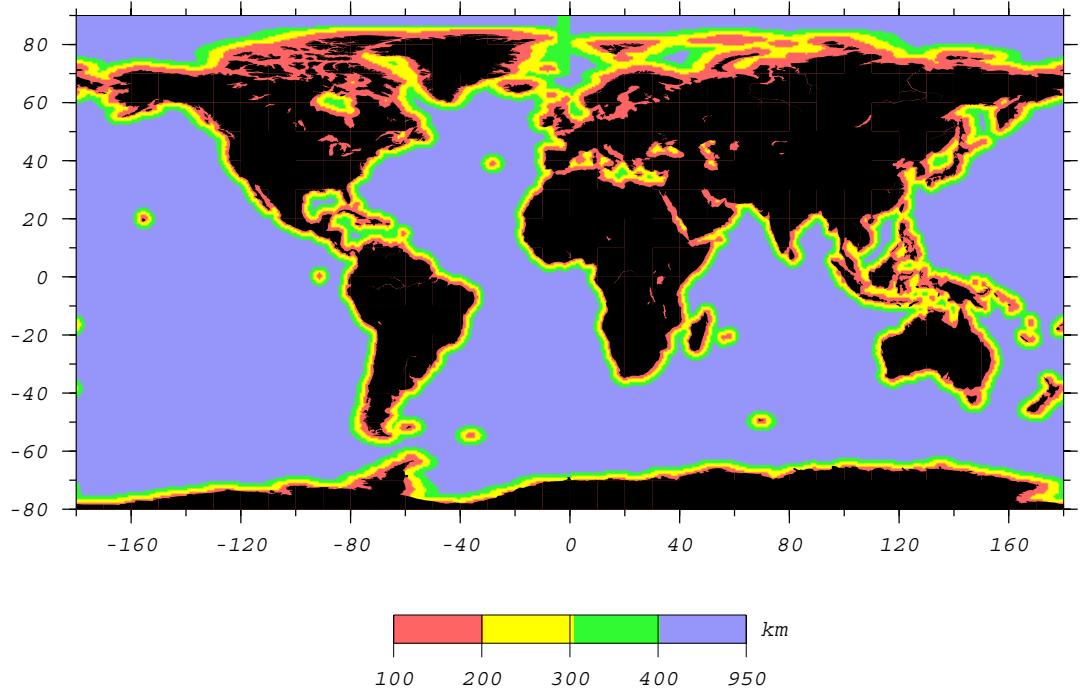


Fig. 13: Decorrelation length scale (km).

To avoid averaging of the data from basins essentially separated from each other optimal interpolation was done for a number of oceanic basins or coastal areas. These interpolation areas are shown in **Fig. 14**.

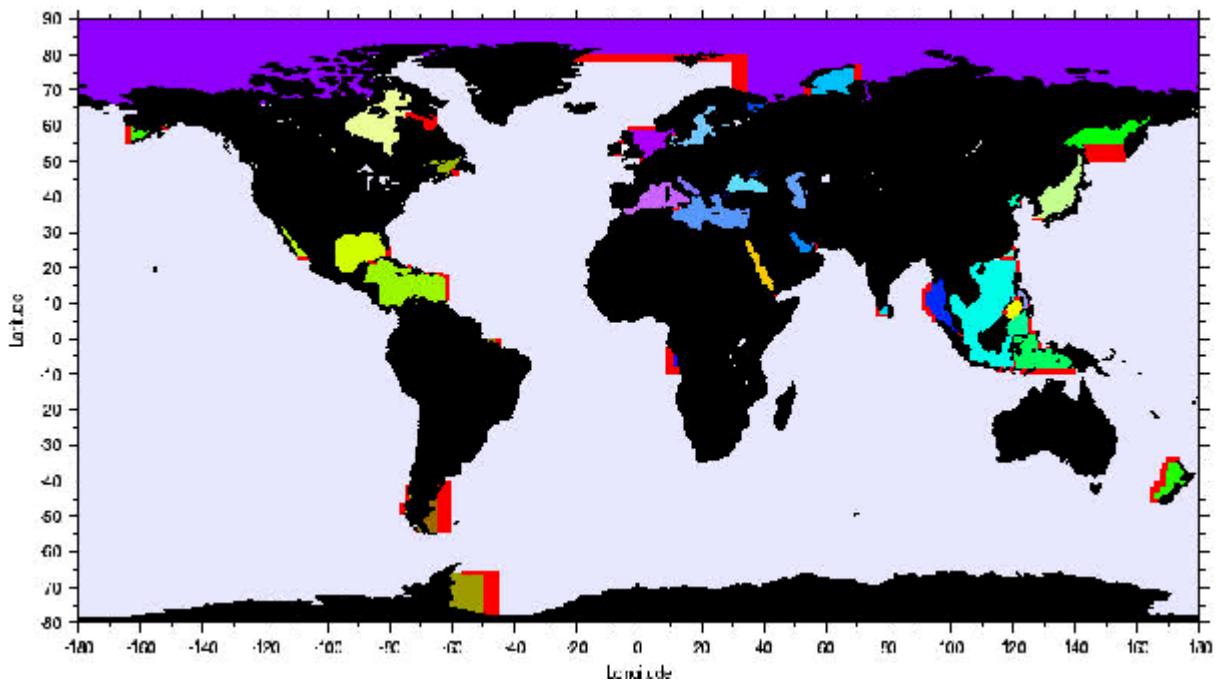


Fig. 14: Interpolation areas. Overlapping areas are shown in red.

Property distributions at the 200 meter level (**Fig. A1a-b**) and near the bottom (**Fig. A1c-d**) are given in the Appendix . Also shown are property standard deviation calculate on isopycnal surfaces (**Fig A2**).