

Using routine hydrographic sections for estimating the parameters needed for Optimal Statistical Interpolation. Application to the northern Alboran Sea*

MANUEL VARGAS-YÁÑEZ, FLORIAN GARIVIER, OLIVIER PIRRA
and M. GARCÍA-MARTÍNEZ

Instituto Español de Oceanografía, C.O. Málaga (Fuengirola), Puerto Pesquero de Fuengirola s/n,
29640 Fuengirola, Málaga, Spain. E-mail: manolo.vargas@ma.ieo.es

SUMMARY: Optimal Statistical Interpolation is widely applied in the analysis of oceanographic data. This technique requires knowing some statistics of the analysed fields such as the covariance function and the noise to signal ratio. The different parameters needed should be obtained from historical data sets, but, in contrast with the case of meteorology, these data sets are not frequently available for oceanographic purposes. Here we show that using routine hydrographic samplings can provide a good estimate of the statistics needed to perform an Optimal Statistical Interpolation. These data sets allow the covariance function to be estimated between two points with different horizontal and vertical coordinates taking into account the possible lack of homogeneity and isotropy which is rarely considered. This also allows us to accomplish a three-dimensional analysis of hydrographic data, which yields smaller analysis errors than a traditional two-dimensional analysis would.

Keywords: Optimal Statistical Interpolation, time series, routine sampling, Alboran Sea, covariance function.

RESUMEN: USO DE MUESTREOS HIDROLÓGICOS SISTEMÁTICOS PARA LA ESTIMACIÓN DE LOS ESTADÍSTICOS NECESARIOS EN TÉCNICAS DE INTERPOLACIÓN ÓPTIMA. APLICACIÓN AL SECTOR NORTE DEL MAR DE ALBORÁN. – La interpolación óptima es ampliamente aplicada al análisis de datos oceanográficos. Esta técnica requiere el conocimiento de algunos estadísticos de los campos analizados, como la función de covarianza o la razón ruido-síñal. Los distintos parámetros necesarios deberían obtenerse a partir de series históricas de datos, sin embargo, al contrario de lo que ocurre en meteorología, esto no es posible en la mayor parte de los casos cuando esta técnica es aplicada a datos oceanográficos. En este trabajo se muestra que el uso de muestreos hidrológicos rutinarios puede proporcionar una buena estimación de los estadísticos necesarios en interpolación óptima. Estos datos permiten la obtención de la función de covarianza entre dos puntos cualesquiera con diferentes coordenadas verticales y horizontales, tomando en consideración la posible falta de homogeneidad o isotropía. Esto permite realizar un análisis tridimensional de los datos que proporciona valores de los errores del análisis más bajos que los producidos por la interpolación bidimensional frecuentemente empleada.

Palabras clave: Interpolación Óptima, series temporales, muestreo sistemático, Mar de Alboran, función de covarianza.

INTRODUCTION

A significant component of our present knowledge about the water mass properties and its general circulation comes from hydrographic cruises and

from data obtained from ships of opportunity. Traditionally, these surveys were accomplished by means of bottle samples or XBTs and CTDs obtained at a set of predefined stations. More recently we have begun to obtain hydrographic data from profiling floats which drift freely with currents at a fixed depth. Studying these data requires contouring

*Received September 6, 2004. Accepted April 29, 2005.

different variables as well as computing derived fields such as geostrophic velocity, vorticity, etc., which involves computing certain variables by means of finite differences. For this purpose, the observed data are interpolated onto a regular grid. Several methods are used for interpolating oceanographic data (see Thiébaux and Pedder, 1987 for a review), including distance-weighting schemes and Optimal Statistical Interpolation (OSI hereafter). OSI was originally developed by meteorologists (Gandin, 1963), and first described in an oceanographic context by Bretherton *et al.* (1976), although it was referred to as objective analysis. It has several advantages over traditional distance-weighting methods as it takes into account the correlation between nearby observations, allows the analysis error to be estimated and considers the observational errors, smoothing the interpolated field. Another advantage of OSI is that univariate analysis can be easily extended to multivariate analysis as long as the cross correlation between different variables is known (Gomis *et al.*, 2001).

The OSI is optimal in a statistical sense (see Section 3), but requires knowing the covariance function of the analysed field as well as the ratio of the error variance to the signal variance. Another characteristic of this method is that it is applied to increment fields, the differences between the observed fields and a certain drift, background or first guess. As pointed out by Gomis *et al.* (2001), “Both the statistical mean field and correlations between increment pairs should in principle be derived from historical data sets”. During recent decades several oceanographic atlases have appeared, such as that by Levitus *et al.* (1994), or for the specific case of the Mediterranean sea, MEDATLAS (1997). These atlases provide temperature and salinity mean values which can be used for calculating increments from the observed values, but they do not allow us to obtain the covariance function or the noise to signal ratio. In almost all of the cases, these parameters are estimated from a single oceanographic cruise. Freeland and Gould (1976), estimated the longitudinal and the transverse velocity correlation function from neutrally buoyant floats, assuming isotropy, stationarity and homogeneity. Robinson *et al.* (1987), used an homogeneous, isotropic and stationary correlation function to study the mesoscale fields in the Eastern Mediterranean. Carter and Robinson (1987), obtained a non-isotropic and time dependent correlation function from both a XBT survey and from an

array of moored current meters, and Mariano and Brown (1992), considered the possible heterogeneous character of the correlation function dividing the interpolation domain into several spatial and time bins. In most of the cases using a single survey or short length time series implies that several assumptions have to be made in a subjective way and in some cases some parameters that are needed to define the covariance function are chosen in a trial and error form. Nevertheless, the advantages of this objective analysis method lie in the explicit form of the assumptions, the repeatability of the operations and the ability to obtain error fields, but as stated by Roemmich (1983), “with limited data, human judgement cannot be removed from the problem. From that standpoint, linear estimation is no different from or better than, say, hand-contouring”.

Repeated standard hydrographic sections can be used as observations similar to those from meteorological stations allowing us to obtain accurate mean or climatological values as well as correlations between pairs of stations. The purpose of this work is threefold: 1) We want to illustrate the potential use of these periodic surveys for calculating background fields, covariances between different stations and noise to signal ratios in a similar way to that used in meteorology. These results could also be extended to the recent use of profiling floats in the Argo project. 2) Estimating the statistics of different variables will allow us to check several hypotheses usually assumed in the OSI of oceanographic data, such as isotropy and homogeneity. 3) Finally, we address the problem of finding a covariance function that allows a fully three-dimensional OSI, combining the horizontal and vertical directions and improving previous attempts to find a covariance function for the interpolation of vertical hydrographic sections. Values of these functions and the different parameters involved are presented for the specific case of the Alboran Sea.

DATA

Ecomálaga is a multidisciplinary project aimed at studying environmental conditions within the continental shelf of the area surrounding Málaga Bay in the northern riparian of the Alboran Sea (Fig. 1). From October 1992 to the present, three transects perpendicular to the coast have been revisited four times a year. These transects extend from Cape Pino, Málaga and Vélez and each com-

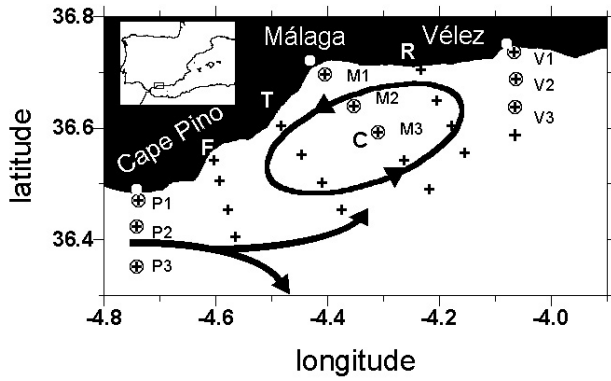


FIG. 1. – Position of the stations sampled in Ecomálaga project. Stations with circles and labelled with letters P, M and V and numbers 1, 2, 3 are those stations sampled since 1992 and used in estimating the correlation and covariance functions. The rest of the stations (crosses) are those sampled since 2000 and have not been used for statistics estimations; nevertheless, they have been used in Results for a comparison between 2D and 3D interpolation schemes. Arrows show a schematic of the circulation in the upper layer of this area.

prise three stations. These stations are identified by the letter, P for the Cape Pino transect, M for Málaga and V for Vélez, and by a number, 1 for the most inshore station, 2 for the intermediate one, and 3 for the offshore station. The three inshore stations are over a bottom depth of 25 m. Intermediate stations have a depth of 75 m in the cases of M2 and V2 and 110 for P2. The offshore stations have depths of 540, 200 and 300m for P3, M3 and V3 respectively. Here we analyse time series to a maximum depth of 200m. At each station a CTD cast is made. We will analyse temperature, salinity and specific volume anomaly (*svan* hereafter) from 5 to 200m at an interval of 5m. Time series analysed extend from October 1992 to October 2003 (included). There were a total of 45 cruises, but the time series analysed have some gaps and the final data set has a total number of 41 measurements for each station and depth. Since the year 2000, three more transects (F, T and R in Fig. 1) have been added, as well as two more stations on the Málaga transect and one more on the Vélez transect. Nevertheless, as these time series are not long enough, we will not consider them when estimating mean values and correlations.

OPTIMAL STATISTICAL INTERPOLATION

OSI was first developed by meteorologists (Gandin, 1963), but since then, many papers have applied it to oceanographic data (see for instance among many others Bretherton *et al.*, 1976;

Freeland and Gould, 1976; Roemmich, 1983; Robinson *et al.*, 1987; Carter and Robinson, 1987; Thiébaux and Pedder, 1987; Lorenc, 1991; Fukumori and Wunsch, 1992; Mariano and Brown, 1992; Pedder, 1993; Gomis *et al.*, 2001). Here we present a brief summary.

OSI is applied to increment fields (θ) which represent the difference between the observed variable (ϕ) at a space point \mathbf{r} and a drift or background value at the same point (μ). It is also assumed that the observations and therefore the increments have a certain error (ε), which represents not only instrumental error but also the part of the variance of the increment field which is associated with those wavenumbers and frequencies not resolved by the data. We consider a set of n observations which can be decomposed as $\phi_i = \mu_i + \theta_i + \varepsilon_i$. The problem addressed is to estimate the increment field at a grid point “g” as a linear combination of the n observed increments (which contain the error term):

$$\hat{\theta}_g = \sum_{i=1}^n w_{gi} \varphi_i \quad (1)$$

being $\varphi_i = \theta_i + \varepsilon_i$.

If w_g is a column vector containing the n weights in (1) and φ is a vector with the n observed increments, we can write (1) in matrix form $\hat{\theta}_g = w_g^T \varphi$, where $()^T$ denotes transposition and the hat denotes estimation. The analysis error will be the difference between the true increment and the estimated increment. OSI is optimal in the sense that it minimises the variance of the analysis error $\langle (\theta_g - \hat{\theta}_g)^2 \rangle$ where the brackets denote expectation or average over an infinite ensemble of independent realisations of the process.

The solution to the minimisation problem is $w_g = A^{-1} C_g$, $\hat{\theta}_g = C_g^T A^{-1} \varphi$ and the error variance is $\sigma_s^2 (1 - C_g^T A^{-1} C_g)$. Where σ_s^2 is the signal variance, $C_g^T = (C_{g1}, \dots, C_{gn})$ denotes the correlation between the grid point “g” and the n observation points and $A = C + \gamma I$. C is a $n \times n$ matrix where the C_{ij} element is the correlation between the observation points “i” and “j”, I is the identity matrix and γ is the noise to signal ratio $\sigma_\varepsilon^2 / \sigma_s^2$.

RESULTS

In most of the oceanographic applications where a fully three-dimensional (3D) data set is available, (a typical CTD cruise) there are several limitations to applying the OSI method. The first one is that the

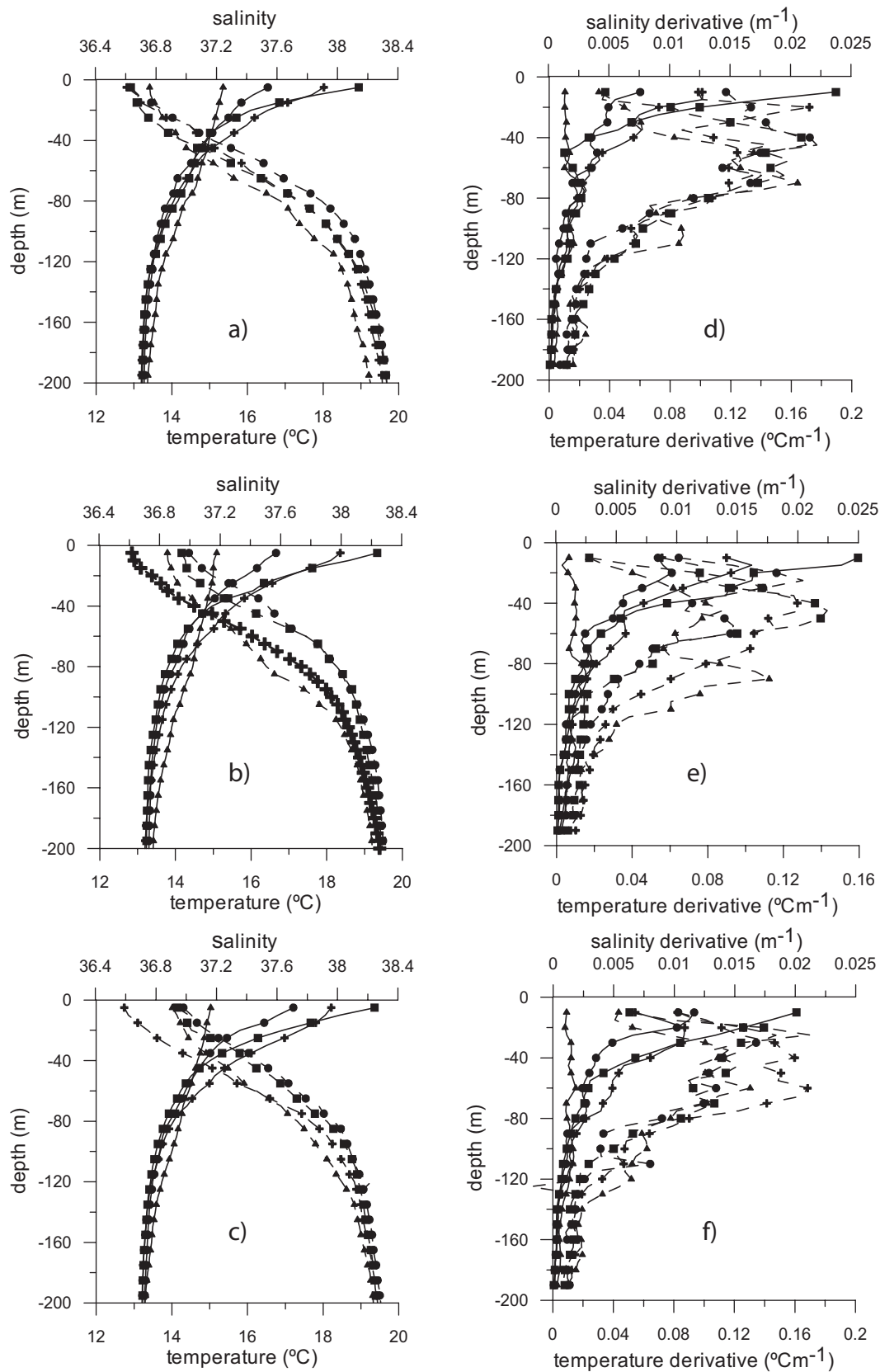


Fig. 2. – Temperature (solid line) and salinity (dashed lines) mean profiles at P3, M3 and V3 (a, b and c respectively). In all the cases we have obtained a profile for each season of the year. Squares correspond to summer (July-September), crosses to autumn (October-December), triangles to winter (January-March) and circles to spring (April-June). Figures 2d, e, f. Vertical derivatives of temperature and salinity. Line and symbol criteria are the same as in Figures a, b, c.

3D field is analysed by means of the 2D analysis of successive horizontal surfaces at different depths or pressure levels. This method does not take into account the information provided by neighbouring points at different horizontal levels. The reason for this simplification is that the correlation function is usually assumed to be isotropic and obviously this does not hold when both the horizontal and vertical directions are involved in the analysis. When a vertical hydrographic section is analysed this problem is solved by analysing and mapping the data for each single depth (1D analysis, Lavín, 1999), or by introducing a scale factor relating the horizontal and vertical distances (Roemmich, 1983). Nevertheless the covariance function used in these cases implies some assumptions that are not appropriate as we will show in the following sections.

Estimating the background or drift field

In many cases there are no climatological data sets for estimating the background mean fields or the needed covariance functions. The mean field is sometimes estimated by fitting a low-order polynomial to the data. In our case we can get the mean values by averaging all the available data for each station and depth. The average takes into account the possible existence of seasonal cycles and different averages are accomplished for each season of the year. Once these climatological profiles are obtained, we derive the increments by subtracting from each profile the climatological profile corresponding to its location and the corresponding season of the year. Figures 2a, b and c show the seasonal temperature and salinity profiles for the P3, M3 and V3 stations.

Horizontal correlations

The mean values show some onshore-offshore and west to east gradients mainly due to the distribution of Atlantic and Mediterranean waters and to the sloping of isotherms and isohalines due to the Atlantic current flowing to the south of our study area (not shown). It is usually assumed that the mean field over a horizontal section captures most of the anisotropy of the field and so the increment correlations depend only on the distance. The procedure in meteorology is to compute the correlation between the increments at different pairs of stations and to represent these correlations as a function of the distance. Figures 3a-f show the correlation

between all the pairs of stations available at 5, 25, 50, 75, 100 and 200 m. In all the figures we include the correlation for temperature (triangles), salinity (circles) and *svan* (open squares). The number of stations available for calculating correlations decreases with depth. This increases the errors and makes the decrease in correlation with distance more difficult to see clearly for the 100 and 200m cases. From 5 to 75 m this behaviour is clear. We propose two different functions for modelling the decrease of correlation with distance, a gaussian function and an exponential one:

$$\rho(r) = e^{-\frac{r^2}{2L_x^2}} \quad \rho(r) = e^{-\frac{r}{L_x}},$$

where r is the distance between two points for a fixed horizontal surface. Theoretically the correlation for $r=0$ is 1. Nevertheless the correlations estimated do not tend to unity as the distance approaches zero, or in our case, to the minimum separation between stations. This is due to the observational noise. The noise at different locations is independent and thus cancels when the covariance is estimated. On the other hand, the covariance estimate is divided by the variance of the increment field, and in this case the variance of the noise is retained. Instead of the functions above we fit functions of the form

$$\rho(r) = R_0 e^{-\frac{r^2}{2L_x^2}} \quad \text{and} \quad \rho(r) = R_0 e^{-\frac{r}{L_x}}$$

(Thiébaux and Pedder, 1987). In both cases we estimate two parameters from the fit, R_0 and the decay length scale L_x . The fit parameter R_0 provides an estimation of

$$\frac{\sigma_s^2}{\sigma_s^2 + \sigma_\varepsilon^2}$$

and the noise to signal ratio can be calculated as

$$\gamma = \frac{1}{R_0} - 1.$$

If the fit is made for the covariance function then we could estimate the decay length scale and the observed increment variance $\sigma^2 = \sigma_s^2 + \sigma_\varepsilon^2$. We used both the gaussian and the exponential function to model the correlation and covariance dependence with distance. Using the mean squared error (*mse*) as criterion, the exponential function provides a better fit to the sampled correlation and covariance functions, nevertheless, the differences

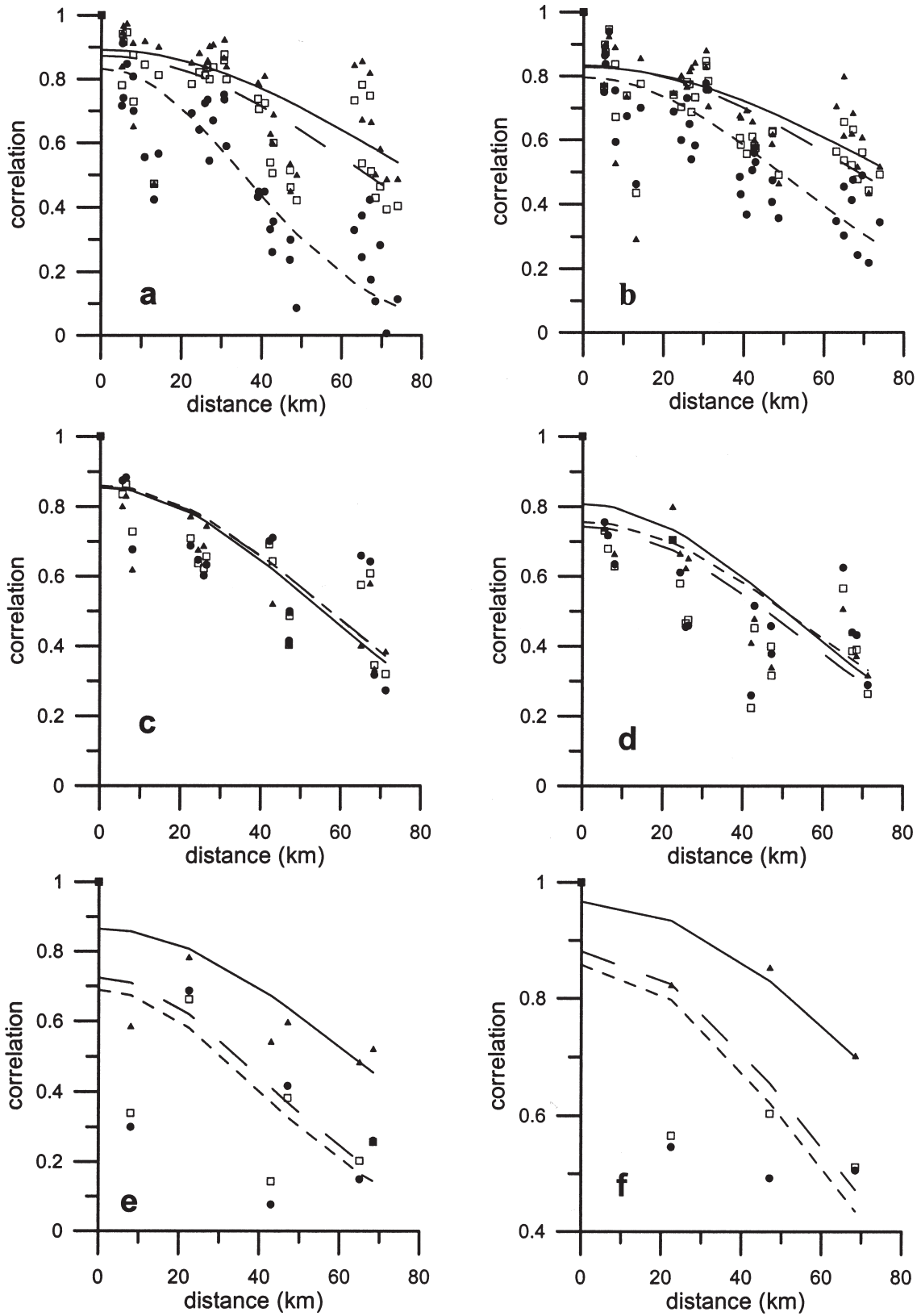


Fig. 3. – Correlation versus distance for 5, 25, 50, 75, 100 and 200 db (a, b, c, d, e and f respectively). In all the figures we include the correlation for temperature (triangles), salinity (circles) and *svan* (open squares) and the fit using a function of the form

$$R_0 e^{-\frac{r^2}{2L_x^2}}. \text{The solid line is the fit for temperature, the dashed line for salinity and broken line for } svan.$$

are not very large (*mse* for the exponential fit is typically 10% lower than the *mse* for the gaussian one). The exponential model also provides a lower noise to signal ratio. On the other hand, the exponential function derivative is discontinuous at the origin. This is not a major problem for univariate analysis, but would not be a suitable correlation function in the case of multivariate analysis where some cross-correlation functions are related through their first derivatives. For this reason we will use the gaussian function in the interpolations performed in Section 4. We have also included the gaussian fit in Figure 3. Nevertheless, we provide the fit parameters for both correlation functions in Table 1 and in Figure 4. They show clearly the decrease of the variance with depth for temperature and *svan*. To find the functional dependence of the total variance with depth we fit exponential functions to the temperature and *svan* variances. The results are

$$\sigma^2(z) = 1.74e^{-\frac{z}{40}} \text{ } ^\circ\text{C}^2$$

for temperature and

$$\sigma^2(z) = 2.64e^{-\frac{z}{42}} (10^{-12} \text{ m}^3 \text{ kg}^{-1})^2$$

for *svan*, where z is the depth in meters. For the case of salinity, the variance seems to increase from the surface to a depth of around 40 m, and then it decreases with depth to the maximum depth sampled. The variance is a measure of the range of variability of the variable. It is reasonable to accept that a variable will change in a wider range in those areas affected by stronger gradients. Upward and downward movements of the water masses will cause larger deviations from the mean in those areas affected by strong vertical gradients. In the same way changes in the slope of the isohalines due to changes in the strength or position of the main currents will have a similar effect. Figures 2d, e and f show the vertical derivative of temperature and salinity at the three offshore stations. In all the cases, the maximum temperature vertical gradients are at the sea surface decreasing with depth. The only exception is in winter when the maximum derivative could be in the subsurface, although this is not clear on the plot. The reason is obviously that the temperature is dominated by the seasonal thermocline developed from spring to autumn and broken in winter when the permanent thermocline associated with the Atlantic and Mediterranean

TABLE 1. – Different parameters representing the covariance and correlation function for temperature, salinity and *svan* at different depths. In all the cases the first value is obtained by fitting an exponential function to the covariances or correlations represented versus distance while the second value corresponds to the gaussian fit (see text, Section 4.2).

temperature/depth	σ^2 (°C ²)	R_0	L (km)	γ
5	2.24/2.08	0.96/0.89	144/74	0.05/0.12
25	1.48/1.37	0.89/0.83	148/76	0.12/0.21
50	0.41/0.37	0.94/0.85	80/54	0.06/0.17
75	0.18/0.16	0.91/0.81	71/52	0.1/0.24
100	0.08/0.08	0.92/0.87	103/60	0.09/0.16
200	0.02/0.02	0.99/0.97	212*/85	0.01/0.03
average		0.94/0.87	126/67	0.07/0.16
salinity/depth	σ^2	R_0	L (km)	γ
5	0.07/0.06	1.05*/0.83	35/35	-0.05*/0.2
25	0.09/0.08	0.92/0.8	68/51	0.09/0.26
50	0.09/0.08	0.95/0.86	84/55	0.06/0.16
75	0.05/0.04	0.86/0.76	76/56	0.2/0.32
100	0.02/0.013	0.82/0.69	40/39	0.2/0.45
200	0.003/0.002	0.93/0.86	89/59	0.08/0.17
average		0.94/0.8	65/49	0.09/0.26
svan/depth	σ^2 (10 ⁻¹² m ³ /kg) ²	R_0	L (km)	γ
5	2.2/1.96	0.96/0.87	107/63	0.04/0.15
25	1.9/1.69	0.91/0.83	118/68	0.1/0.2
50	1.0/0.93	0.94/0.86	84/55	0.06/0.17
75	0.5/0.44	0.86/0.74	67/52	0.16/0.35
100	0.2/0.15	0.84/0.72	45/41	0.19/0.38
200	0.03/0.026	0.94/0.88	100/61	0.06/0.14
average		0.92/0.82	87/57	0.10/0.23

* this value, as well as the $R_0 > 1$, makes no sense and is an artefact of the fit.

waters has a more important role. For salinity, the seasonal effects which probably exist (see Vargas-Yáñez *et al.*, 2005), are of second order importance with respect to the permanent halocline which governs the vertical gradients. For all seasons and for all three stations the maximum vertical gradient of salinity occurs around 46m depth. This is consistent with the variations of salinity variance with depth. In this case we model this functional dependence as

$$\sigma^2(z) = 0.11e^{-\frac{|z-46|}{40}}$$

We also modelled the decrease of the variance with depth using a gaussian function but the results were considerably worse.

For the other two parameters involved in the correlation and covariance function, L_x and γ , no clear depth dependence is found and it is more likely that these parameters do not vary systematically with depth. As the time series grows in length, either a clear pattern should evolve, or else the amplitude of these random fluctuations should decrease.

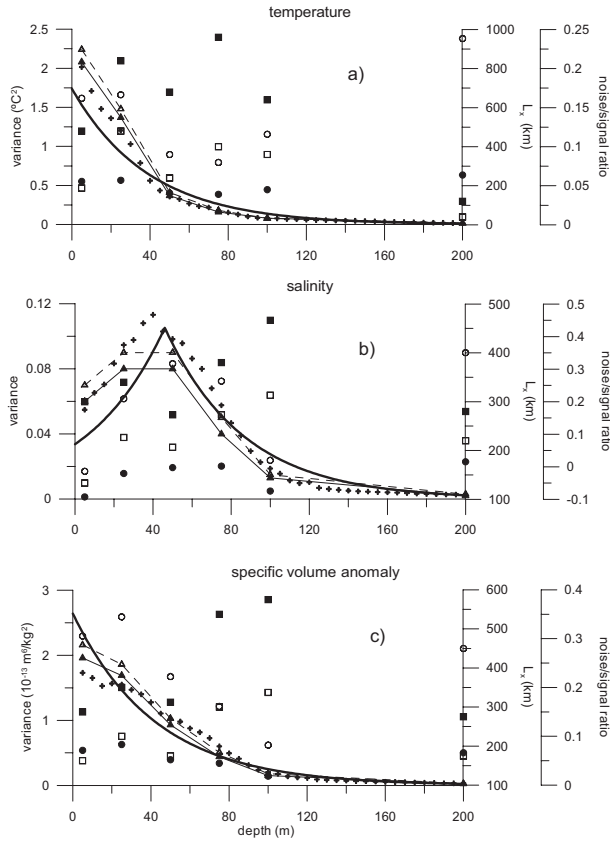


FIG. 4. – Variance obtained for different depths (5, 25, 50, 75, 100 and 200 db): We obtained the covariance as a function of the distance (in the same way that we did for correlation in Fig. 3) and then we fit both a gaussian

$(\sigma^2 e^{-\frac{r^2}{2L_x^2}})$ and an exponential function $(\sigma^2 e^{-\frac{r}{L_x}})$.

Triangles represent the parameter of the fit σ^2 for the different depths analysed. The same fit was applied to the correlation function and from it we obtained R_θ (see the text). Squares are the noise to signal ratio for each depth, calculated from R_θ , and the decay length scale is represented by circles. Filled symbols correspond to the gaussian fit and open symbols to the exponential one. The thick solid line is the theoretical depth dependence of the variance presented in Section 4.2, and crosses represent the variance as a function of depth obtained from the covariance estimation for a fixed station and different depths and then setting $z_1=z_2$ (see text, Section 4.3). Panels a, b and c show the same calculations for temperature, salinity and *svan* respectively.

Vertical correlations

As already stated, when we deal with a vertical section or we wish to analyse a 3D field, we have two alternatives: One is to analyse the different depth levels separately and the other is to consider the covariance between two points with position vectors \mathbf{r}_1 and \mathbf{r}_2 such as

$$\gamma(\mathbf{r}_1, \mathbf{r}_2) = \sigma^2 e^{-\frac{r^2}{2L_x^2} - \frac{\Delta z^2}{2L_z^2}}$$

(Roemmich, 1983), where r is the horizontal distance between the two points (we assume isotropy in

the horizontal) and Δz is the vertical distance. The decay length scales L_x and L_z account for the anisotropy between the horizontal and vertical directions.

It is easy to see from the results in Section 4.2 that this function does not represent the true statistics of the oceanographic variables analysed above. First, let's consider that for $\mathbf{r}_1=\mathbf{r}_2$, $r=\Delta z=0$, this expression provides the same variance regardless of the depth. For two pairs of points separated by the same horizontal distance (r), but located at different depths, this function does not take into account the dependence of the variance on depth. If we consider pairs of points with the same horizontal coordinate ($r=0$) but different depths, this expression indicates that the covariance between the points depends on the vertical separation but not on the depth of the two points. Figure 5 shows the covariance between pairs of points with the same horizontal coordinate (we fix the station) as a function of the depth difference of the two points. We have repeated these calculations changing the depth of the initial point, in other words, the curve labelled '5 m' represents the covariance between a point located at 5 m depth and all the points with the same horizontal coordinate and depths from 5 to 200 m as a function of the depth difference. The curve labelled with 25 m is the covariance between a point at 25 m depth and all the points with depths ranging between 25 and 200 m, expressed as a function of the depth difference, etc. Figures 5a, b and c show that it is not realistic to assume that the covariance function depends only on the depth difference. According to Figure 5 we could say that for two points with vertical coordinates z_1 and z_2 , it depends on z_1 and Δz , or in a more simple form that it depends on both z_1 and z_2 , and the homogeneity hypothesis is clearly not acceptable for the vertical dependence.

Let's consider a certain station (horizontal coordinate fixed). We estimate from the increment time series the covariance for all the possible pairs of values z_1 and z_2 . We repeat these calculations for the three offshore stations where a more extensive description of the covariance depth dependence is possible. Results for P3, M3 and V3 are very similar confirming that the covariance function is homogeneous in the horizontal direction, so we averaged results for the three stations. Figure 6 shows the isolines for constant covariance. It is important to note that the covariances for $z_1=z_2$ represent the variance as a function of depth. The difference between this and the calculations in Figure 4 is that here we have

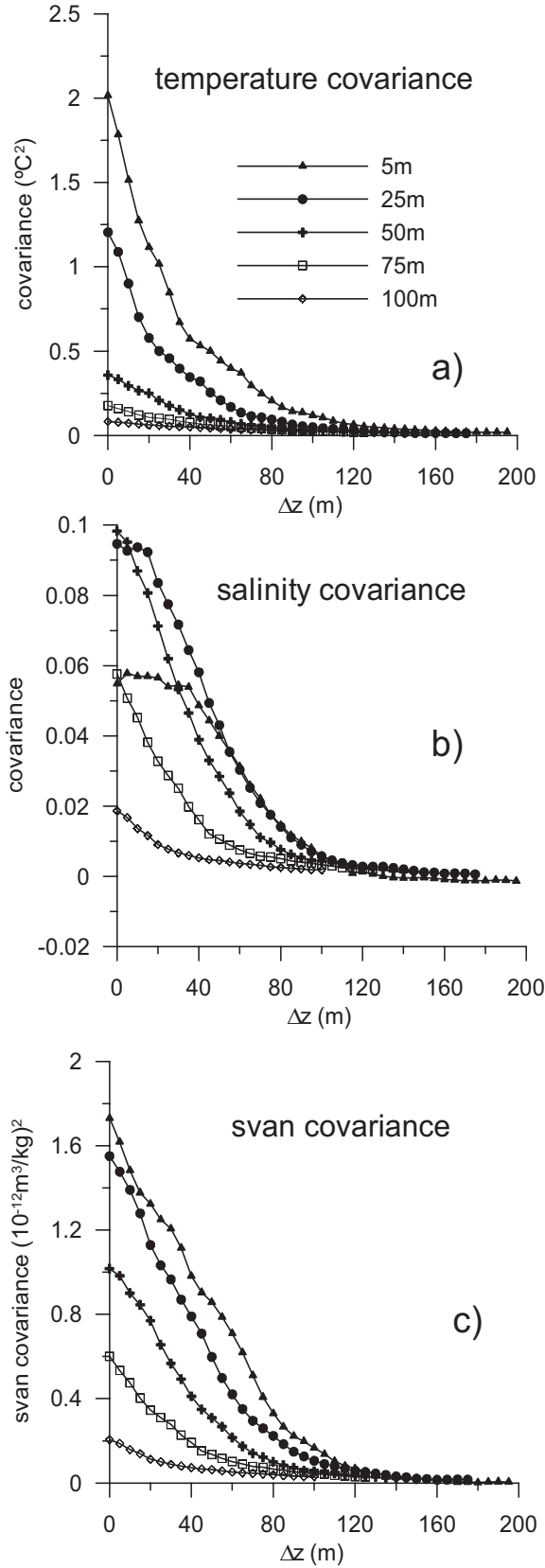


FIG. 5. – Covariance as a function of the depth difference between pairs of points with the same horizontal coordinate. Different curves correspond to different depths of the initial point.

a continuous description of the water column instead of 6 discrete levels, but we only consider the results from the most offshore stations instead of using all the available stations at a given depth level. The new dependence of variance with depth has been included in Figure 4 as crosses and these results are consistent with the previous ones.

Visual inspection of Figure 6 shows that the covariance is roughly constant on ellipses decreasing as the semi-axis increases. That is, we treat the covariance as if it takes the form

$$f(\alpha_1 z_1^2 + \alpha_2 z_2^2 + \alpha_3 z_1 z_2),$$

where f is a monotonic decreasing function. We propose that the covariance decreases as a gaussian function, $\text{cov}(z_1, z_2) = be^{-(\alpha_1 z_1^2 + \alpha_2 z_2^2 + \alpha_3 z_1 z_2)}$ or as an exponential one,

$$\text{cov}(z_1, z_2) = \beta e^{-\sqrt{(\alpha_1 z_1^2 + \alpha_2 z_2^2 + \alpha_3 z_1 z_2)}}$$

For the gaussian model an intuitive interpretation is easy. In the case that $a_1 = a_2$ we can write

$$z_1 = \bar{z} - \frac{\Delta z}{2} \quad \text{and} \quad z_2 = \bar{z} + \frac{\Delta z}{2}$$

where we assume, without losing generality, that we are in the case

$$z_1 < z_2, \quad \bar{z} = \frac{z_1 + z_2}{2}$$

and $\Delta z = z_2 - z_1$. In this case the covariance function can be w

$$\text{cov}(z_1, z_2) = \left[\beta e^{-\frac{\bar{z}^2}{2C_1^2}} \right] \cdot e^{-\frac{\Delta z^2}{2C_2^2}}$$

being $C_1^2 = \frac{1}{2(2\alpha_1 + \alpha_3)}$

and $C_2^2 = \frac{1}{\alpha_1 - \alpha_3/2}$

Now β can be interpreted as the variance at the sea surface and the covariance function is made of two factors, the first one representing the variance dependence with depth, where depth is now the mean value of z_1 and z_2 , and a second term representing the decrease of covariance with the vertical distance between the two points.

According to Figure 6b, it is clear that the maximum variance for salinity is at the subsurface. Averaging P3, M3 and V3, resulted in an average

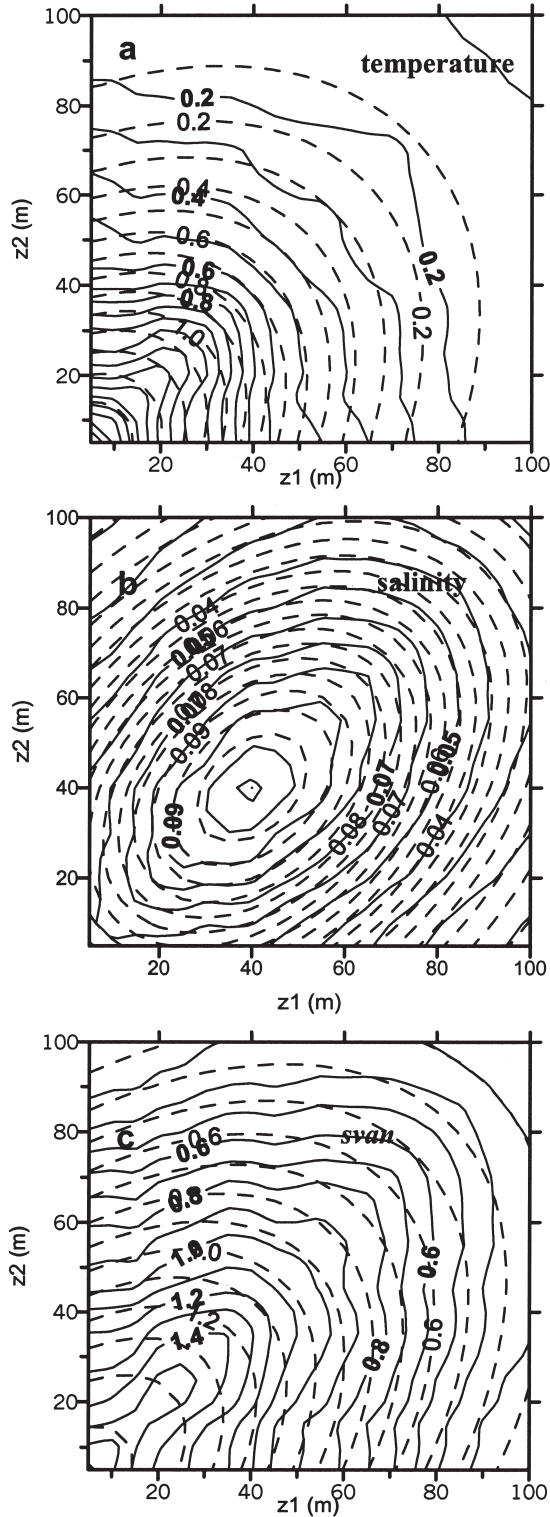


FIG. 6. – Solid lines are isolines of constant covariance for pairs of points with the same horizontal coordinate and different depths. Calculations were made for the most offshore stations (P3, M3 and V3) and then averaged. Dashed lines are the fit using a function

$$\text{cov}(z_1, z_2) = \beta e^{-[\alpha_1(z_1 - z_1^0)^2 + \alpha_2(z_2 - z_2^0)^2 + \alpha_3(z_1 - z_1^0)(z_2 - z_2^0)]}$$

Values of the different parameters can be seen in Table 2. *svan* values are multiplied by 10^{12} .

maximum at 40 m depth. Thus, we propose for salinity a dependence with depth in the form:

$$\text{cov}(z_1, z_2) = \beta e^{-[\alpha_1(z_1 - z_1^0)^2 + \alpha_2(z_2 - z_2^0)^2 + \alpha_3(z_1 - z_1^0)(z_2 - z_2^0)]}$$

for the gaussian fit or

$$\text{cov}(z_1, z_2) = \beta e^{-\sqrt{\alpha_1(z_1 - z_1^0)^2 + \alpha_2(z_2 - z_2^0)^2 + \alpha_3(z_1 - z_1^0)(z_2 - z_2^0)}}$$

for the exponential one, where $z_1^0 = z_2^0 = 40$. This is consistent with previous results where we found a maximum variance for salinity at 46 m, the depth corresponding to the maximum vertical salinity gradients (Fig. 2). Table 2 shows the parameters for both the exponential and gaussian fits.

In the case of an exponential fit an intuitive interpretation is not so easy, nevertheless, as $\alpha_1 = \alpha_2$ for the three variables this allows us to write the covariance as a function of the mean depth and the depth increment

$$\text{cov}(z_1, z_2) = \beta e^{-\sqrt{(2\alpha_1 + \alpha_3)\bar{z}^2 + (\alpha_1 - \frac{\alpha_3}{2})\frac{\Delta z^2}{2}}}$$

We can check if this result is consistent with previous ones. For instance in the case that $z_1 = z_2$ then we get the dependence of variance with depth

$$\beta e^{-\frac{\bar{z}}{C}}$$

being

$$C = \frac{1}{\sqrt{2\alpha_1 + \alpha_3}}$$

Using the parameters in Table 2 we obtained a variance dependence on depth of the form

$$\sigma^2(z) = 2e^{-\frac{\bar{z}}{32} \circ C^2}$$

for temperature,

$$\sigma^2(z) = 0.15e^{-\frac{|\bar{z} - 40|}{30}}$$

for salinity and

$$\sigma^2(z) = 2.6e^{-\frac{\bar{z}}{45}} (10^{-12} m^3 / kg)^2$$

for *svan*. These results are consistent with the results reported in Section 4.2. In the case of a gaussian fit the depth dependence of the variance would be

$$\sigma^2(z) = 1.5e^{-\frac{\bar{z}^2}{2(32)^2} \circ C^2}$$

TABLE 2. – Parameters of the fit for the covariance between two spatial points at a fixed horizontal position and different depths (see text, Section 4.3). For each coefficient the first value corresponds to the exponential fit and the second one to the gaussian one.

	temperature (°C ²)	salinity	svan 10 ⁻¹² (m ³ /kg) ²
σ^2	2.0/1.5	0.15/0.1	2.6/1.6
α_1	0.0008/0.0004	0.001/0.0004	0.0007/0.0002
α_2	0.0008/0.0004	0.001/0.0004	0.0007/0.0002
α_3	-0.0006/-0.0003	-0.0009/-0.0003	-0.0009/-0.0002

for temperature,

$$\sigma^2(z) = 0.1e^{-\frac{(\bar{z}-40)^2}{2(32)^2}}$$

for salinity and

$$\sigma^2(z) = 1.6e^{-\frac{\bar{z}^2}{2(50)^2}} (10^{-12} m^3/kg)^2$$

for *svan*. These results also show a reasonable agreement with the results reported in Section 4.2. According to the *mse* criterion the exponential model offers only a slightly better description of the vertical correlations. On the other hand the gaussian function offers a more intuitive approach. We show the parameters of both fits in Table 2. In Figure 6 we have included the gaussian fit and this will be the function used in the 3D interpolation in Section 4.5.

Vertical sections

When interpolating oceanographic data onto a vertical section, a standard approach is to consider the covariance between two points separated by a horizontal distance r and a vertical distance Δz as

$$\gamma(r, \Delta z) = \sigma^2 e^{-\frac{r^2}{2L_x^2} - \frac{\Delta z^2}{2L_z^2}}$$

(see for instance Roemmich, 1983). As shown in 4.2 and 4.3 this function assumes homogeneity in the vertical direction which is not true. We have obtained separately a covariance function for the horizontal case, taking into account the dependence of the variance on depth. We also considered the possible dependence on depth of the noise to signal ratio and the decay length scale but we found that a constant value was appropriate. We have also found a function for the covariance between two points with the same horizontal coordinate but different depths, which takes into account the non-homogeneity of this function. The most intuitive approach to the problem of finding the covariance between pairs of points with different horizontal and vertical

coordinates is to combine both results into one single function:

$$\text{cov}(r, z_1, z_2) = \beta e^{-[\alpha_1(z_1 - z_1^0)^2 + \alpha_2(z_2 - z_2^0)^2 + \alpha_3(z_1 - z_1^0)(z_2 - z_2^0)]} \cdot e^{-\frac{r^2}{2L_x^2}} \quad (2)$$

where, according to the results discussed in Sections 4.2 and 4.3, we use a gaussian model to represent the horizontal and vertical dependence of the covariance function.

A first requirement of this function is that the depth dependence of the variance that we have obtained by analysing the horizontal covariance at different depth levels and the one we obtained for the vertical direction is the same. This was demonstrated in the previous section. The following question is that we have no evidence that the covariance function in this general case can be factorised in this way. Figure 7 shows the covariance function for the temperature, salinity and *svan* between different pairs of stations and for all the possible pairs of depths (solid line) and the theoretical value obtained using (2). Differences between estimated and theoretical values are around 18%-30% of the estimated values. Using the exponential fit for the vertical part of the correlation function the differences between the observed and the theoretical values are around 18%-28%, only slightly lower than those mentioned in the previous section. Based on these results and Figure 7, we can accept that Expression (2) roughly describes the spatial dependence of the covariance function.

For the 3D CTD data set obtained from a single survey, we can define a 3D grid and accomplish a fully three-dimensional interpolation, estimating the increment field at each grid point as a linear combination of all the observations. In this way, the interpolated value not only takes into account the nearby observations at the same horizontal level, but also nearby observations at different depths. We have considered the 24 temperature, salinity and *svan* profiles from the July 2003 Ecomálaga survey (see Fig. 1 for the position of the casts). We defined a 3D grid from 4.75°W to 4°W and from 36.3°N to 36.75°N and from 5 to 100 db with a horizontal interval of 0.05° (both in latitude and longitude) and a vertical interval of 5db. We accomplish the interpolation onto the 3D grid simultaneously. Then, to represent the results we selected the interpolated data corresponding to certain depth levels, or latitude or longitude in order to contour the results at

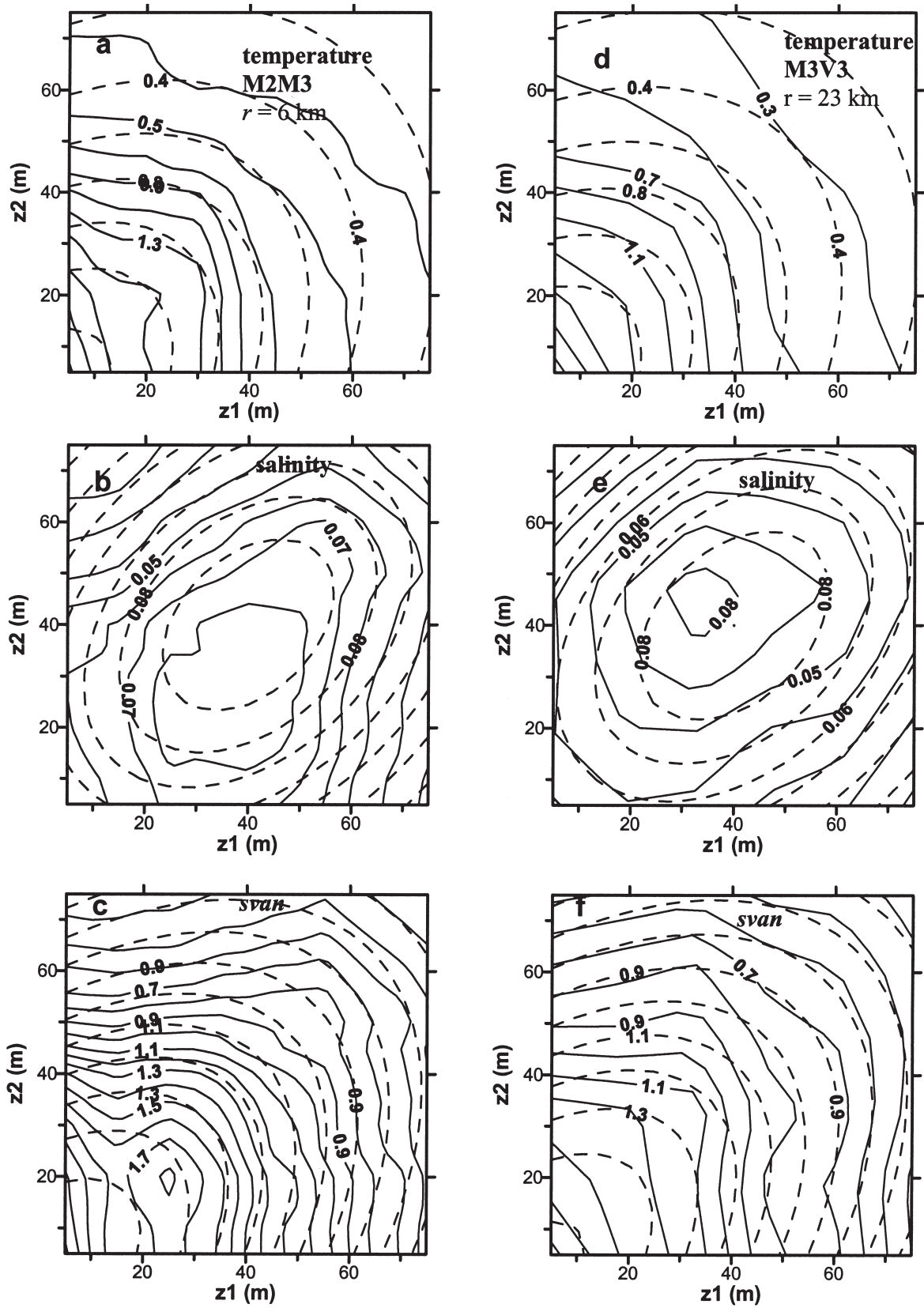


FIG. 7. – Isolines of constant covariance (solid line) for pairs of points with different vertical and horizontal coordinates (a, b, c correspond to stations M2 and M3, $r=6$ km and Figures d, e, f to M3, V3). Dashed lines correspond to the theoretical values predicted by the expression (2, Section 4.4) and parameters in Tables 1 and 2.

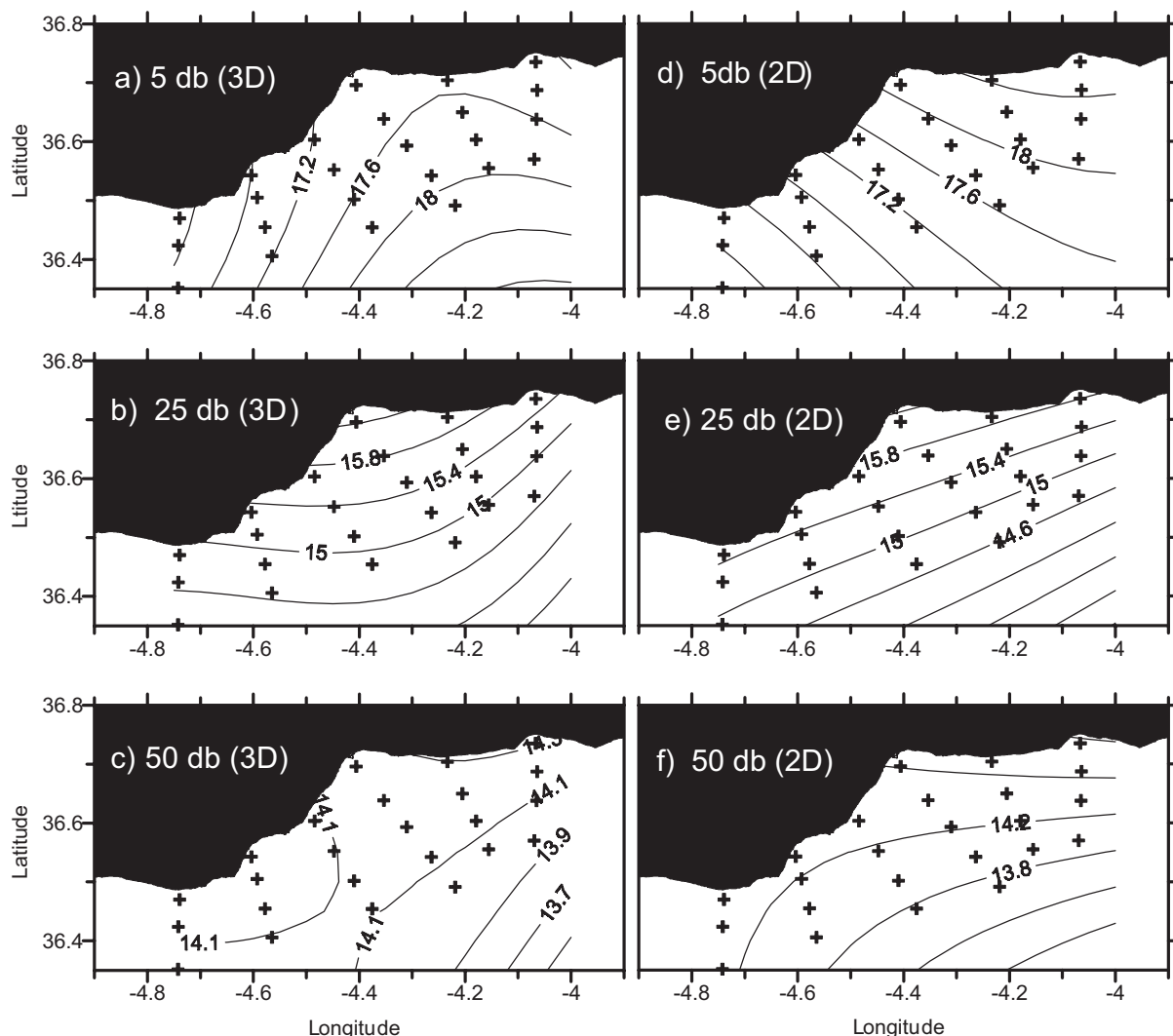


FIG. 8. – Temperature interpolated fields from the Ecomálaga 0703 survey. Cast positions are indicated with crosses. Figures a, b and c represent three horizontal sections at 5, 25 and 50 db from the three dimensional interpolation. Figures d, e, f are the same sections from a two-dimensional interpolation at those levels.

horizontal or vertical sections. Figure 8 shows the temperature field at three selected vertical levels (5, 25 and 50 db) obtained through the 3D interpolation (Fig. 8a, b, c), and the results from a 2D interpolation at the same levels (Fig. 8d, e, f). It is clear that the 2D interpolation produces smoother fields. For instance, the 2D interpolation is not able to capture the curvature change of the isotherms at the 5 and 25 db interpolations. Some sub-mesoscale features are also filtered out in the 50 db case. Figure 9 presents the error variance for the interpolations corresponding to Figure 8. Note that errors are larger in the 2D case than in the 3D OSI. In Figure 9 it is clear that errors increase considerably in the extrapolated areas, where the interpolated field should not be considered and it is usually masked. In the interpolated area errors are much lower, reaching values around 0.015°C^2 for the 3D case at 5db and 0.03°C^2

for the 2D analysis. At 25 and 50 db the differences increase, with the minimum errors around 0.01°C^2 and 0.007°C^2 at 25 and 50 db in the 3D analysis and 0.03°C^2 at 25 and 50 db in the 2D case. For the case of salinity the 3D/2D error variances have minimum values around 0.004/0.007, 0.004/0.007 and 0.004/0.005 at 5, 25 and 50 db respectively. For the *svan* and the same depths the minimum values are 0.005/0.01, 0.005/0.01 and 0.006/0.013 ($10^{-12}\text{m}^3/\text{kg}$)

SUMMARY AND DISCUSSION

We have shown that using routine hydrographic sections can provide a useful method for estimating both climatological fields as well as correlations, covariances and noise to signal ratios between pairs of stations. The estimation of these functions can

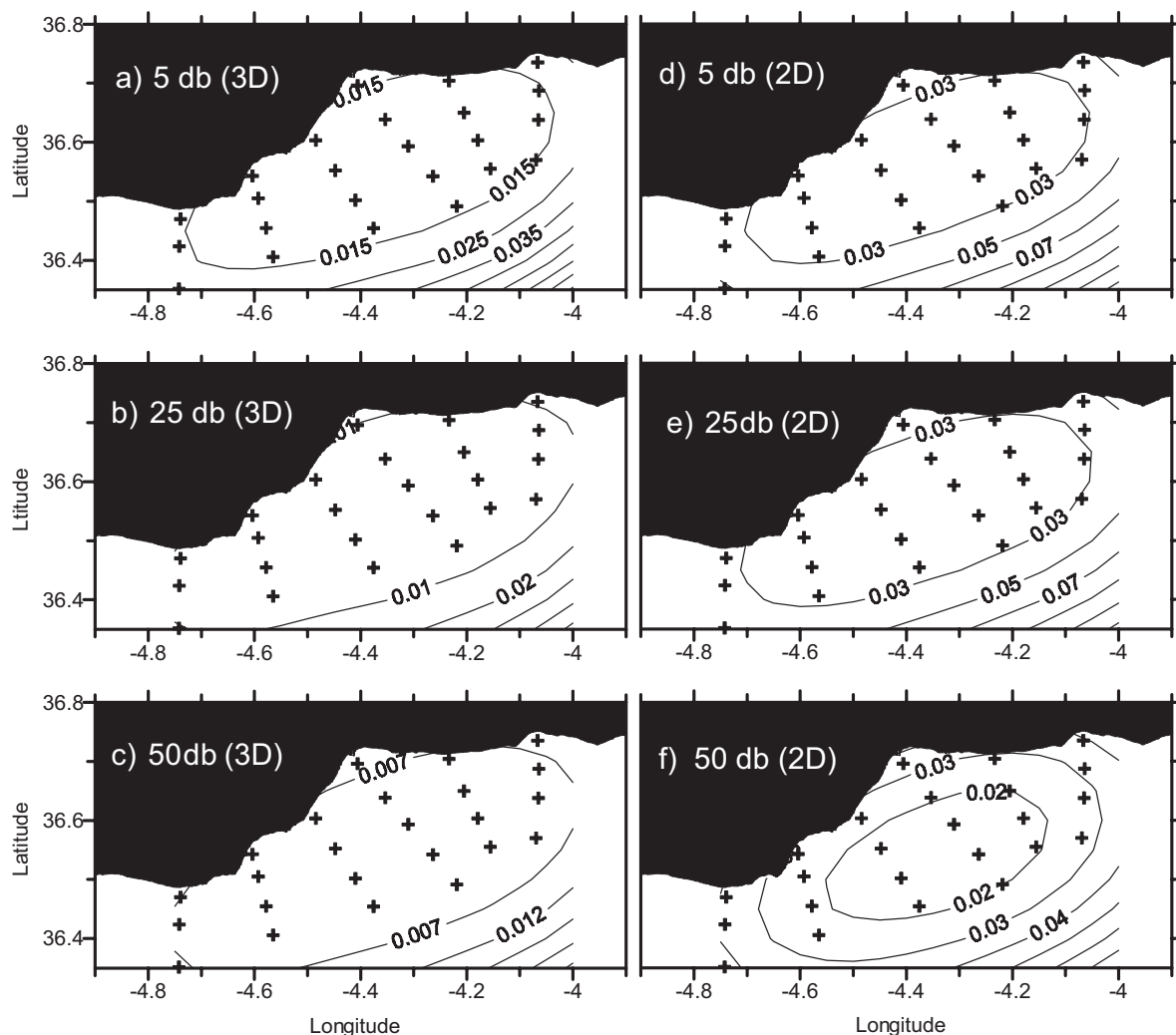


FIG. 9. – Error variance for the interpolated fields in Figure 8.

help to perform an interpolation of hydrographic data which is optimal in a statistical sense, reducing the number of subjective assumptions.

For the particular case of the northern riparian of the Alboran Sea we have obtained the climatological 3D fields of temperature, salinity and *svan* as well as the covariance and correlation functions between any couple of points with different horizontal and vertical coordinates. The climatological or background fields estimated in this work are only useful for this specific area of the Alboran Sea. In other parts of the Mediterranean, the background field should be obtained from the available climatologies. Nevertheless, the covariance function and the noise to signal ratios obtained here could be used in other areas, as the dynamics underlying the statistics of the hydrographic fields are likely to be similar.

We have proved that the variance depends on the depth but not on the horizontal coordinate. The horizontal decaying distance and the noise to signal

ratio are independent of the depth. We have also shown that the covariance between pairs of points at different depths but with the same horizontal coordinate does not depend on the horizontal coordinate. These results support the idea that the statistics of the analysed fields are homogeneous in the horizontal but not in the vertical. We have assumed isotropy between the two horizontal directions (although this point has not been checked) but obviously horizontal and vertical directions are not isotropic. We have estimated a covariance function for any two points and shown that it depends on the horizontal distance and the particular values of depth of both points. The knowledge of this function allows us to interpolate vertical sections as well as make a fully three dimensional interpolation of hydrographic data. Applying these results and comparing them with traditional 2D interpolation methods show that the 3D analysis allows us to capture some features that are filtered in the 2D analysis and reduces the analysis errors.

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 Res.* 1, 23: 559-582.
- Carter, E.F. and A.R. Robinson. – 1987. Analysis models for the estimation of oceanic fields. *J. Atmospheric Ocean Tech.*, 4: 49-74.
- Fukumori, I. and C. Wunsch. – 1991. Efficient representation of the North Atlantic hydrographic and chemical distributions. *Prog. Oceanogr.*, 27: 111-195.
- Freeland, H.J. and W.J. Gould. – 1976. Objective analysis of meso-scale ocean circulation features. *Deep-Sea Res.*, 23: 915-923.
- Gandin, L.S. – 1963. Objective analysis of meteorological fields. Transl. from Russian by Israel Program for scientific translations, 1965 (NTIS No. TT65-50007).
- Gomis, D., S. Ruiz and M.A. Pedder. – 2001. Diagnostic analysis of the 3D ageostrophic circulation from a multivariate spatial interpolation of CTD and ADCP. *Deep-Sea Res.* 1, 48: 269-295.
- Lavín, A. – 1999. *Fluxes, trends and decadal changes in the subtropical North Atlantic*. Ph.D. Thesis, Universidad de Cantabria.
- Levitus S., R. Burgett and T.P. Boyer. – 1994. World Ocean Atlas 1994 Volume 3: Salinity. NOAA Atlas NESDIS 3. U.S. Department of Commerce, Washington, D.C. 99 pp.
- Levitus S. and T.P. Boyer. – 1994. World Ocean Atlas 1994 Volume 4: Temperature. NOAA Atlas NESDIS 4. U.S. Department of Commerce, Washington, D.C. 117 pp.
- Lorenc, A.C. – 1981. A global three-dimensional multivariate statistical interpolation scheme. *Mon. Wea. Rev.*, 109: 701-721.
- Mariano, A. J. and O. B. Brown. – 1992. Efficient objective analysis of dynamically heterogeneous and nonstationary fields via the parameter matrix, *Deep-Sea Res.*, 39, N° 7/8: 1255-1271.
- MEDATLAS, 1997. Mediterranean Hydrological Atlas on CD-ROM. IFREMER (Ed.), published by IFREMER/DIDIT/IDT on behalf of the MEDATLAS consortium under contract MAS2-CT93-0074.
- Pedder, M. A. – 1993. Interpolation and filtering of spatial observations using successive corrections and gaussian filters. *Mon. Wea. Rev.*, 121: 2889-2902.
- Robinson, A.R., A. Hecht, N. Pinardi, J. Bishop, W.G. Leslie, Z. Rosentrub, A.J. Mariano and S. Brenner. – 1987. Small synoptic/mesoscale eddies and energetic variability of the eastern levantine basin. *Nature*, 327: 131-134.
- Roemmich, D. – 1983. Optimal Estimation of Hydrographic station data and derived fields. *J. Phys. Oceanogr.*, 13: 1544-1549.
- Thiébaux, H.J. and M.A. Pedder. – 1987. *Spatial objective analysis with application in atmospheric science*. Academic Press, London.
- Vargas-Yáñez, M., A. Chasles, A. Berthelelot, T. Ramírez, D. Cortés, A. García, A. Carpena, J.M. Serna, M. Sebastian, J. Mercado, J. Cortés and J.P. Álvarez. – 2005. Proyecto Ecomálaga 1992-2001. Parte I: Oceanografía Física. *Inf. Técn. Inst. Esp. Oceanogr.* 183: 1-73.

Scient. ed.: C. Millot

