SACLANTCEN MEMORANDUM
serial no.: SM-217

SACLANT UNDERSEA RESEARCH CENTRE

MEMORANDUM

Test and evaluation of objective mapping applied to oceanographic data from GIN Sea '86

G. Peggion

March 1989

Page count for SM-217
(excluding covers)

Pages	Total
i–vi	6
1–16	16
@1–@8	8
	30

SACLANT Undersea Research Centre
Viale San Bartolomeo 400
19026 San Bartolomeo (SP), Italy

tel: 0187 540 111
telex: 271148 SACENT I

NORTH ATLANTIC TREATY ORGANIZATION

Report no. changed (Mar 2006): SM-217-UU

Test and evaluation of objective mapping applied to oceanographic data from GIN Sea '86

G. Peggion

The content of this document pertains to work performed under Project 04 of the SACLANTCEN Programme of Work. The document has been approved for release by The Director, SACLANTCEN.

Issued by:
Underwater Research Division

J. Hovem
Division Chief

intentionally blank page

Test and evaluation of objective mapping applied to oceanographic data from GIN Sea '86

G. Peggion

Executive Summary: Sonar performance, and hence the capability to detect and track submarines, requires a knowledge of the environmental conditions to serve as input data for analysis and modelling applications. Since most of the oceanographic observations are made at irregularly distributed locations, it is necessary to assign estimates of physical variables to points on a regular grid. An ideal procedure has to filter the disturbances inherent in the data, minimize observational errors, and provide the 'best' solution in regions lacking data. The estimated field must satisfy two requirements. First, it must be close to the observations, to an accuracy defined by mathematical conditions, regardless of the physical nature of the records. Second, it must obey dynamical/statistical relationships which are known to be satisfied by the 'real' ocean. Briefly, it is at least necessary that the regularly distributed field preserve the dominant spatial and/or temporal features of the records.

This research is a preliminary investigation to develop a new procedure to provide an interpolated field which preserves the dominant spatial features contained in the observational data. A later report will present the results applied to the observations collected at sea during the GIN Sea program of SACLANTCEN. The purpose is to test different interpolation techniques and suggest new criteria to evaluate *a priori* the mathematical accuracy of interpolated solutions. The method develops as follows. A test function is mapped on an irregular grid. This distribution of artificial data is evaluated on a regular grid using different interpolation techniques. The estimated field and the values of the known function are compared at the points of the regular grid by computing standard statistical distributions. For any given irregular and regular point set, the accuracy of the solution is tested as a function of the length scale associated with the input data. By decomposing observed data into basic functions, it is possible to preserve the spatial features of the records and minimize the interpolation errors.

In addition, this study contributes to improving the processing and graphical interpretation of data, by determining the inherent errors in such procedures.

intentionally blank page

Test and evaluation of objective mapping applied to oceanographic data from GIN Sea '86

G. Peggion

Abstract: This study is an investigation to produce an objective analysis procedure to assimilate oceanographic data from the GIN Sea program of SACLANTCEN into numerical model initialization. More specifically, this study suggests a preprocessing analysis to define the most appropriate interpolation scheme. Here, interpolation is considered accurate if the resulting field distribution preserves the dominant spatial features contained in the data set. The method develops as follows. A known continuous function is given and mapped on an irregular grid. This artificial distribution is evaluated on a regular grid using different interpolation algorithms. The interpolated solutions are compared with the values of the analytical function at the same mesh points, by computing standard statistical distributions.

Application of interpolation routine to oceanographic data illustrates differences and similarities between interpolated distributions preserving synoptic and mesoscale features, and indicates the limits of the solutions for a valid scientific analysis of the observed field.

Keywords: data assimilation ○ GEOINT ○ GIN Sea ○ interpolation ○ KRIGPAK ○ objective analysis ○ oceanography ○ ZGRID

Contents

1
Introduction

In most oceanographic applications, estimates of a field must be assigned to points of a regular net from widely-spaced, irregularly-distributed reporting points. An ideal procedure would filter the noise inherent in the data without appreciable alteration of the field, and provide an 'exact' solution where no samplings are available. The problem is generally considered successfully solved when the solution minimizes disturbances, observational errors, and small scale fluctuations.

Essentially, there are two methods for producing a map distribution of values. One method may be called subjective, the other objective. When subjective techniques are used, the solutions are dependent on personal interpretations; even long-term, experienced analysts are often unable to supply assumptions and hypotheses for a rigorous formulation of the procedure. Thus, it may be necessary to formulate the problem *objectively*. The field produced by an objective method must satisfy two basic requirements. First, it must be close to observations with an accuracy which ideally is the accuracy of the observations themselves, but practically limited by mathematical conditions. This requirement must be included in any reliable method for processing irregularly distributed data, and be independent of the physical variables at the reporting points. Second, the quality and type of the observations must constrain the solution to dynamical and/or statistical relationships which are known to be satisfied by the 'real' field. In any case, it is necessary to define in a *subjective* way, when and how the appropriate dynamical/statistical information can be introduced in the procedure. To satisfy these two requirements, a regularly distributed oceanic field must at least preserve the dominant spatial features of the records, according to a model formulation describing the physics of the phenomena under investigation.

Considering the complexity of the problem it is a good policy to proceed by steps. This study is an investigation to provide an objective analysis procedure to apply to the observations collected during the sea trials of SACLANTCEN scientific programs. More specifically, the purpose is to introduce criteria for evaluating the mathematical accuracy of different interpolation routines and software packages already available at SACLANTCEN. The solutions are analyzed and compared, with particular attention to the horizontal spatial structure of the records. Finally, the interpolation routines are applied to oceanographic data. Although no physical constraints are imposed on the solutions, the experiments indicate the most appropriate algorithms for several scientific investigations.

One of the most common applications of data assimilation is the display of data sets in some type of contouring map. Map-drawing software packages generally include reliable methods of processing irregular data to obtain accurate estimates on a regular grid. From this point of view, graphical representations become objective analysis problems which fulfil only the mathematical requirement previously mentioned. In recent years, the significant development in computer technology for generating images and the high level of accuracy in map displays have made graphic representation a powerful tool not only for the final display of scientific results, but also for the analysis and scientific interpretation of large amounts of data. Two errors are associated with graphic displays: the error inherent to the countouring routines, and the error caused by the interpolation itself. While the distortions and inaccuracies generated by map drawing are negligible, interpolation schemes may introduce severe misinterpretations and misinformation, which too often are underestimated or forgotten. Thus, it becomes necessary to define criteria for an *a priori* estimate of the accuracy of the interpolated fields. This must be done case by case, because the solutions depend so heavily on the scattered data distribution, the smoothness of the values, and the resolution of regular mesh, that it is not possible to present results and conclusions of general applications.

Document overview Section 2 illustrates the methodology and discusses the criteria for the comparison. The results are presented and compared in Sect. 3. Different interpolation routines are applied to oceanographic data in Sect. 4. Finally, Sect. 5. summarizes and discusses this research.

2
Methodology and preliminary processing

The experiments have been kept as simple as possible so that the basic problems could be understood easily, yet retain some connections with the final oceanographic products that are saught. Thus, the irregular data set is specified at the locations of the GIN Sea '86 cruise stations (Fig. 1). However, for a correct evaluation and estimate of the different solutions, values on the irregular grid are assigned from a known continuous function. This artificial distribution is evaluated on a regular grid with different interpolation methods, and compared with the values of the analytical function at the same mesh points. The method can be thought of as an operator (say $\mathcal{A}$) who maps regularly distributed data into scattered data and a second operator (the interpolation itself, say $\mathcal{B}$) who transforms the irregularly into regular data distributed values. If the two regular data sets coincide, the interpolation would be the inverse of the operator $\mathcal{A}$, and we would have found the optimal interpolation algorithm. However, even in this remote possibility the results would hold only for a restricted class of assigned data and point distributions. To evaluate the accuracy of the interpolated values, it is necessary to define *objective* criteria for making possible the comparison between different interpolation methods since most of the numerical procedures belong to self-contained software packages for which little technical explanations are given, and very few of them supply an adequate analysis of the error. To overcome the problem, the input and output values on the regularly distributed mesh point are compared by computing statistical distributions such as standard deviation and cross-correlations.

An important feature to be defined is the extension of the regular grid. Here, the regular mesh is given by the rectangle bounded by the maximum and minimum values of longitude and latitude of the station locations, as illustrated by the marked box in Fig. 1. This domain contains a region of high concentration of data points (in the Farœ–Shetland Islands regions) and several large zones mostly in the northern portion, in which few or no data at all are available. With these spatial features the algorithms can be tested in their interpolation and extrapolation performance as well. Generally, interpolation schemes also serve as schemes for extrapolation; in both cases they must model the function in between/beyond the known points by some plausible form. However, interpolation is 'usually' safe, whereas extrapolation may be hazardous.

The regular mesh resolution also needs to be assigned. As is well known, given a finite number of values, only features of length scale larger than 2Δ (where Δ is the length scale of the station grid spacing) can be reproduced. In our experiments, the regular mesh is defined on a 20×12 grid on the longitude and latitude axes, respectively. In the region of our interest, this is equivalent to a resolution of about 40 and 60 km, respectively. This grid spacing is adequate for the analysis of meso-to-large-scale oceanic features. Because of the limited north–south extension of the geographical region, the numerical domain is referred to a cartesian coordinate system (i.e. distances on the longitudinal axes are constant with latitude). Moreover, the previous spatial features are transformed in their correspondent dimensionless variables, and the domain is henceforth referred as to the rectangle $R_0 = \{(x, y)|0 \leq x \leq 1, 0 \leq y \leq 1\}$, where the dimensionless variables x and y are for longitude and latitude, respectively.

The choice of the testing functions must be carefully done to extract the maximum information with a restricted number of experiments. As it is known, the functions $\phi_{n,m}(x, y) = \exp^{i(n\pi x + m\pi y)}$ form a set of orthonormal and complete functions in the nondimensionalized rectangle $R : \{(x, y) : |x| \leq 1, |y| \leq 1\}$; i.e. any given function $f(x, y)$ such that $\int_R f^2 \, dx \, dy < \infty$ can be written in its modal Fourier representation

$$f(x, y) = \sum_{n,m} a_{n,m} \phi_{n,m}(x, y), \qquad a_{n,m} = \int_R f \; \phi_{n,m} \, dx \, dy.$$

Because of their properties, the functions $\phi_{n,m}$ are a good testing tool for our problem. They allow us to verify the quality of the solutions as a function of the associated wavelength, and to generalize the results to a wide class of functions by superposition. Without a loss of generality only functions of the type $\sin(n\pi x)\sin(m\pi y)$ in the domain R_0 are henceforth considered.

Three different experiments are performed for different values of the parameters n and m, which are representative of different wavelengths. The first case ($n = 1$, $m = 1$) relates to long waves, i.e. whose length scale is comparable to the total extension of the domain. The case ($n = 5$, $m = 1$) represents waves of intermediate length scale range, i.e. whose wavelength is resolved by many grid points, and a few wavelenghts are reproduced in the domain. Finally, the case ($n = 10$, $m = 6$) is an extreme case, representing the shortest wavelength associated with the regular grid. The analytical functions relative to each case (henceforth, the exact solutions) are illustrated in Fig. 2.

3
The interpolating routines

In this section the results from interpolation routines and software packages available at SACLANTCEN are presented. Among the algorithms with complete documentation and of common application, only those with good mathematical accuracy and computational efficiency have been considered. For instance, the Numerical Algorithm Group (NAG) Library furnishes two-dimensional interpolation routines. However, a correct application of those routines requires such a large amount of working space (of order N^4, where N is the number of the original data points) that they are computationally inefficient. All of the applied interpolating routines are briefly discussed, separately. There are some basic common features. For example, all of them may suppress extrapolation in areas that are too far from data points, but in the following experiments the search radius is selected so that interpolation is done over the whole domain. Moreover, all the routines have the possibility of further smoothing. However, this aspect of the problem is not considered here, and the solutions presented in the following sections have the 'default' smoothing factor.

3.1. ZGRID

This routine belongs to a software display package developed by Taylor, Richards and Halstead (1971). Arbitrarily placed data points may be interpolated onto a rectangular grid using two-dimensional Laplace or spline interpolation or a combination of the two. The choice of the interpolating scheme is expressed by the algorithm intrinsic parameter `CAY`. An over-relaxation method is used to perform the interpolation. The original data points are initially moved to the nearest grid point and then shifted back to their original position as the shape of the surface is constructed. The relaxation process is usually continued until the estimate of the maximum error is less than one percent of the surface value, or until 100 iterations are reached.

This routine has been applied in two extreme situations: the case of a mere laplacian interpolation (relative to `CAY` $= 0$, and henceforth referred to `ZGRID(0)`, and the case in which the spline interpolation predominates (henceforth referred to `ZGRID(100)`. Figures 3 and 4 illustrate the solutions of the performed interpolations. Both cases behave satisfactorily for the longest waves. When the laplacian interpolation is used, the resulting surfaces tend to have rather sharp peaks at the data point, but no spurious peaks appear in regions devoid of data. On the contrary, when spline interpolation dominates, the interpolated surfaces pass more smoothly through the data, but steep peaks may appear in areas lacking data.

3.2. UNIRAS

The Universal Raster System (UNIRAS) is a software graphics package. Among all the options and tools for accurate and sophisticated image displays, it contains two interpolation packages for the pre-processing of the data mapping. One (GEOINT) performs interpolation/extrapolation for irregularly scattered data points onto a regular grid; the other (KRIGPAK) develops a more complete method and also offers an estimate of the standard error per grid point. In the following paragraphs routines from these packages are applied and analyzed.

3.2.1. GEOINT

This package contains four basic interpolation procedures. Two of them (GINTP2 and GINTP3) consider local polynomial fitting, but have a limited range of applications because they do not handle fault lines. (A fault line divides the surface in two regions, one largely independent of the other. In this case interpolation is performed using data only from the same side of the fault line.)

The other two routines (GINTPF and GINPT1) can handle fault lines and are indicated to be superior to the previous ones in most of the gridding applications: GINTPF because of its accuracy, GINPT1 because of its computational efficiency (UNIRAS, 1986). Both routines perform interpolation using double linear interpolation, slope calculation, quadratic interpolation, and a final smoothing. To speed the process GINPT1 initially sorts the data points into grid.

Because of the similarities of all these routines, only the solutions from GINTPF are displayed (Fig. 5). As expected the routine performs quite well in the proximity of data points, but the effects of the smoothing operations are noticable. The routine applies a two-dimensional filter designed to reduce the surface curvature at each data point in regions of low sampling concentration, yet to retain the accuracy of the interpolation in regions of high concentration. This is particularly evident in the first case (Fig. 5a): in the upper part of the domain the surface is considerably flattened with respect to the exact solution. On the other hand, the smoothing and the control of the surface gradient give positive effects in the third case (Fig. 5c). Although the shortest waves are poorly represented even in regions of high data concentration, the solution presents evidence of small scale variability even in zones lacking data, without generating sharp peaks. The most accurate solution is for the second case (Fig. 5b) for which grid spacing, wavelength, and smoothing radius of influence all have equivalent length scale.

3.2.2. KRIGPAK

Unlike most interpolation methods, Kriging is a two-stage process. In the first stage the user determines the spatial structure of his data, providing a model for a 'semi-variogramm'. That is, at every point the distances between sampling locations are grouped, and the average difference of the data values per distance group is computed at various directions. In general a semi-variogramm is created on the principle that close points do not differ much in value and the difference in value increases with distance up to a certain distance. Points beyond this distance have no further influence. In such a way, the user defines a search radius for the contribution of data to the estimating values at the grid points.

Tests have been made to verify the dependence of KRIGPAK on the semi-variogramm definition. The experiments indicate that the solutions may be very sensitive to the choice of the semi-variogramm model. Figure 6 shows solutions obtained for different sets of the parameters that define the standard semi-variogramm model as given in UNIRAS (1986). These parameters are the maximum distance of influence (henceforth referred as to r) and the maximum value difference (henceforth referred as to s). It appears evident that as the maximum distance of influence is much greater than the wavelength of the solution, or as the maximum value difference increases outside the range of the function itself, the interpolated surface is dramatically smoothed and most of its basic characteristic features cannot be recovered anymore.

The second stage is the interpolation process itself. The interpolated values are chosen to minimize the standard error, which is not dependent on the individual samplings, but rather on the average behaviour in the zone of influence.

Two among all the routines of KRIGPAK are applied here. The first (`GKRIG2`) assumes that there is no significant trend in value within the radius of influence of a sample; the second (`GUK2D`) takes this trend into consideration. However, for the particular cases of our experiments, there are no substantial differences between the solutions, and, therefore, only the results from `GKRIG2` are displayed (Fig. 7). The solutions are given for the semi-variogramm paramters $r = 1$ and $s = 1$. This model is associated with the smallest influence distance (1 grid square), and a maximum difference value connected with half wavelength. It follows that the longest waves are represented with a high level of accuracy in all the domain. As the wavelength decreases, the solution remains accurate in regions of high point concentration, and in regions where extrapolation dominates, the surface is quite smooth without steep peaks. For the shortest waves the solution becomes less accurate, but the interpolated surface retains evidence of small-scale features even in zones lacking data points.

3.3. COMPARISON

So far, different interpolation methods and software packages have been analyzed separately. In the following section, the solutions are compared with each other. Comparison is made by defining in a *subjective* way an *objective* criterion. As anticipated in Sect. 1, we rely on some simple standard statistical functions. For each of the interpolated surfaces and associated exact solutions we compute

$$\begin{aligned}
&\text{mean value :} && M = \frac{1}{\mathrm{NX} \times \mathrm{NY}} \sum_{i=1}^{\mathrm{NX}} \sum_{j=1}^{\mathrm{NY}} f(x_i, y_j) \\
&\text{variance :} && c = \sum_{i=1}^{\mathrm{NX}} \sum_{j=1}^{\mathrm{NY}} (f(x_i, y_j) - M)^2 \\
&\text{standard deviation :} && \sigma = \sqrt{\frac{c}{(\mathrm{NX} \times \mathrm{NY} - 1)}} \\
&\text{local error :} && \epsilon^+ = \max_{i,j}(f(x_i, y_j) - f_{\mathrm{ex}}(x_i, y_j)) \\
& && \epsilon^- = \min_{i,j}(f(x_i, y_j) - f_{\mathrm{ex}}(x_i, y_j)) \\
&\text{correlation coefficient :} && R = \frac{1}{\sqrt{c\ c_{\mathrm{ex}}}} \sum_{i=1}^{\mathrm{NX}} \sum_{j=1}^{\mathrm{NY}} (f(x_i, y_j) - M)(f_{\mathrm{ex}}(x_i, y_j) - M_{\mathrm{ex}}).
\end{aligned} \tag{3.1}$$

where the subscript ex indicates the exact solution, NX and NY the number of mesh points in the x- and y-directions, respectively. Values of these funtions are given in Tables 1–4.

It follows that for the longest waves all the interpolated surfaces are quite accurate. The highest correlation coefficient is for the KRIGPAK routines, which also present the minimum excursion of local errors. Good also are the solutions from ZGRID using either laplacian or spline interpolations. Almost unexpectedly, GEOINT routines are behaving poorly. Although they preserve the mean value, their correlation coefficients are only about 60% of the correlation coefficients obtained by the other solutions. At the intermediate wavelength GEOINT and ZGRID(100) give the most accurate solutions, while the performance of KRIGPAK is less impressive. Finally, none of the routines shows a high level of accuracy for the shortest waves.

In regions lacking data, the routines may calculate quite different extrapolated surfaces. In general, ZGRID(100) generates high peaks and sharp gradients, while the routine ZGRID(0) gives smooth extrapolated surfaces with flat gradients. The GEOINT package appears to extend the axial symmetry of the data also into the extrapolation regions. However, it is not possible to determine if this is an intrinsic feature of the routines, or merely due to a lucky choice of the exact functions and data point distribution. Finally, KRIGPAK preserves the gradient fairly well into the

Table 1

Values of statistical functions for each of the adopted routines and the exact function ($n = 1$, $m = 1$: $\sin(\pi x)\sin(\pi y)$)

Subroutine	Mean Value	Variance	Standard Deviation	Local error ϵ^+	ϵ^-
exact solutions	0.34	22.89	0.31	–	–
ZGRID(0)	0.48	11.88	0.22	−0.09	0.58
ZGRID(100)	0.42	16.57	0.26	−0.02	0.47
GINTPF	0.34	17.61	0.33	−0.12	0.61
GINTP1	0.34	16.74	0.26	−0.16	0.59
GUK2D(a)	0.37	19.38	0.30	−0.30	0.36
GUK2D(b)	0.37	13.67	0.23	−0.30	0.50
GKRIG2(a)	0.41	15.39	0.25	−0.11	0.30

Table 2

Values of statistical functions for each of the adopted routines and the exact function ($n = 5$, $m = 1$: $\sin(5\pi x)\sin(\pi y)$)

Subroutine	Mean Value	Variance	Standard Deviation	Local error ϵ^+	ϵ^-
exact solutions	0.06	51.20	0.46	–	–
ZGRID(0)	0.04	37.73	0.39	−0.81	0.43
ZGRID(100)	0.06	52.61	0.46	−0.74	0.88
GINTPF	0.06	49.07	0.45	−0.58	0.62
GINTP1	0.04	50.04	0.45	−0.74	0.75
GUK2D(a)	0.01	44.99	0.43	−1.32	0.76
GUK2D(b)	0.02	58.19	0.49	−1.42	1.04
GUK2D(c)	0.04	74.45	0.55	−1.51	1.29
GKRIG2(a)	0.06	42.80	0.42	−1.19	0.73

Table 3

Values of statistical functions for each of the adopted routines and the exact function ($n = 10$, $m = 6$: $\sin(10\pi x)\sin(6\pi y)$)

Subroutine	Mean Value	Variance	Standard Deviation	Local error ϵ^+	ϵ^-
exact solutions	0.0	52.25	0.46	–	–
ZGRID(0)	−0.02	24.15	0.31	−1.23	1.26
ZGRID(100)	−0.08	134.56	0.75	−3.44	2.05
GINTPF	0.0	29.06	0.34	−1.22	1.43
GINTP1	0.0	28.52	0.34	−1.34	1.33
GUK2D(a)	0.01	33.47	0.34	−1.42	1.49
GUK2D(b)	0.01	12.80	0.23	−1.26	1.34
GKRIG2(a)	−0.01	26.07	0.33	−1.41	1.37

Table 4

The cross correlation coefficient between exact and interpolated surfaces as function of the different wavelengths

Subroutine	$n = 1$, $m = 1$	$n = 5$, $m = 1$	$m = 10$, $n = 6$
ZGRID(0)	0.83	0.85	0.30
ZGRID(100)	0.93	0.83	0.13
GINTPF	0.51	0.88	0.33
GINTP1	0.58	0.84	0.25
GUK2D(a)	0.95	0.70	0.23
GUK2D(b)	0.82	0.18	0.02
GKRIG2(a)	0.96	0.73	0.27
GKRIG2(b)	0.80	0.30	0.04

long wave range, but has a tendency to smooth and flatter the extrapolated surfaces in the other length scales.

Looking at these results on the global range of the wavelength, we may attempt the following conclusions. KRIGPAK routines results are quite accurate for long waves, i.e. when the semi-variogramm model is such that data values which are close together in space (relative to r) are also close in values (relative to s). Outside this limited range, KRIGPAK behaves poorly. Unfortunately, tests have indicated that KRIGPAK algorithms may break down for small values of r and/or s. Then it does not seem possible to obtain better estimates for intermediate and short waves. On the contrary, the GEOINT package is not too accurate for the long wave, but particularily efficient in the other wavelength ranges. Even with the shortest wave, GEOINT routines furnish the best correlation coefficient (relative to the other solutions). The experiments also confirm the dependence of ZGRID algorithms on the adopted interpolation method. The possibility of steep peaks and the growth of local error increase as the spline interpolation dominates, but laplacian interpolation has the tendency of smoothing, perhaps too much, the solutions. By adjusting the parameter CAY properly, a trade-off between smoothness and avoidance of spurious peaks could be obtained. Taylor, Richards and Halstead suggest that values CAY = 5 or CAY = 10 usually give a good surface representation.

4

An application to oceanographic data

The experiments presented in the previous sections assumed *a priori* knowledge of the solutions and defined the accuracy of the algorithms as a function of the dominant wavelength contained in the records. The following experiments apply the interpolation routines to the temperature values at 25 m depth of the GIN '86 cruise and comment on the scientific validity of the interpolated temperature distributions. Data have been extracted using the environmental oceanographic database developed at SACLANTCEN for the Applied Oceanography Group. Among the scientific questions addressed by the GIN Sea project, the observational program sought to determine the configuration and variability of the synoptic and mesoscale features of the region. Therefore, even though signals in all the wavelength range are included in the records, the dominant ones are in the meso-to-large length scale. Using these oceanographic terms, the long waves defined in the previous sections are associated with synoptic features, the intermediate with mesoscale variability, and the short with small-scale fluctuations that are not of concern in this investigation.

The solutions interpolated on the regular grid defined in Sect. 2 are illustrated in Fig. 8. The disagreements and similarities among the different solutions are evident. A clarifying example is given by the 8.5° isotherm. For all the interpolated functions, this isotherm encloses a pool and an eddy of cold water, both in proximity to the hydrographic stations located in the northern side of the GIN '86 region. However, the shape and the size of these features are quite different. GEOINT algorithms suggest a narrow tongue of cold water extending in the north–south direction, with the maximum of the southern penetration at around 62°N. Eastward and southward of the tongue there are two additional small cold eddies enclosed by the same isotherm. The routine `ZGRID(10)` predicts a pool of cold water which had absorbed the two small isolated eddies generated by the GEOINT package. The resulting tongue is wider eastward and extends southward up to 61°N. The extrapolated surface in the upper left corner of the domain is flat with a smooth gradient, but in the upper right corner, extrapolation leads to a solution with a sharp gradient. Finally, KRIGPAK routines generate the widest pool of cold water which covers almost the whole northern part of the basin. Both small eddies predicted by the GEOINT package have disappeared. The eastern eddy has been absorbed into the pool of cold water, the southern swept away by the interpolation. In that region the resulting temperature distribution increases southward between the 8.5° and the 9° isotherms.

Since it is implicitly assumed that the records adequately describe the dominant features of the field, then in regions of high point concentration all interpolated solutions are valid to an accuracy as indicated in the previous sections. More difficult is the evaluation of the validity of the extrapolated surfaces. The accuracy of an interpolated solution preserving large scale features might be extended to the extrapolated surface, because synoptic features are resolved by many points of both irregular and regular grids. On the other hand, extrapolated solutions in the mesoscale range are of low reliability, because the regions lacking data have extensions greater than the spatial dimensions of the features under investigation. For a valid scientific interpretation of the records, solutions dominated by mesoscale features must obey physical/dynamical constraints in regions of low data point concentrations.

Although mathematical tools alone do not indicate which one of the solutions is the closest to the real field, knowledge of the dominant wavelength contained in the records must restrict the choice of the numerical interpolation algorithm for scientific and numerical applications. The ZGRID and GEOINT packages are accurate for objective mapping of the mesoscale variability of the observed fields. For numerical purposes they can be applied successfully to initialization of eddy-resolving models, which are concerned with the evolutionary state of the environmental conditions. KRIGPAK results are accurate in reproducing features in the long wave range, and its applications are to be found in climatological processing and initialization of numerical models concerned with seasonal/annual variabilities. The use of KRIGPAK routines might be advisable also in processing noisy data in order to filter observational errors without appreciable alteration of the field distribution.

5

Conclusions

The tests presented in the previous sections evaluate the performance of several numerical interpolation algorithms and verify the accuracy of the interpolated solutions. A solution is considered accurate if it preserves the dominant spatial features contained in the data set. Tests have been done for prescribed distributions of the irregular and regular points. The scattered data distribution automatically defines a wavelength which is implicitly assumed to be associated with the dominant features under investigation. To avoid aliasing phenomena the regular grid should never be too fine with respect to the data point distribution, and it must be consistent with the model formulation describing the physics of the observed field. Experiments have been performed with three different testing functions representing long, intermediate, and short (with respect to the extension of the domain) waves.

Among the routines examined in this study, the experiments indicate that GEOINT and ZGRID are the most accurate algorithms in the intermediate wavelength range, although these software packages tend to generate sharp peaks in regions lacking data. On the other hand, KRIGPAK preserves the large-scale features contained in the data field, but smoothes and flattens the solution. KRIGPAK numerical procedures strongly depend on the definition of the semi-variogramm model. For a correct use of these routines, evaluations *a priori* of the field seem necessary, and the proper solution may be obtained by tuning the parameters r and s in the semi-variogramm function.

Aside from the limitations introduced by the definition of the spatial distributions, the interpolated solution also depends on the smoothness of the data field. Although it is implicitly assumed that data are not contaminated by noise or observational errors, the observed field usually contains signals in all the wavelengths. Application of one of those interpolation subroutines may emphasize or eliminate some of these features. Therefore, precise estimates of the dominant features of the field distributions and a correct choice of both regular grid and interpolation algorithm are necessary for an accurate interpolated solution. If these prerequisites are verified, no additional smoothing should be required. Smoothing procedures may minimize small-scale fluctuations of no interest, but may greatly affect the structure of the field. They are never advisable if the physics of the field distribution must be preserved, but may be required if the quality of the graphical representation needs to be increased.

Application of the routines to oceanographic data has made evident the similarities and differences between interpolated solutions preserving synoptic and mesoscale features. The definiton of the length scale of the dominant features and the choice of the most accurate corresponding algorithm are necessary, but do not provide sufficient conditions for a correct scientific analysis and interpretation of the data. In regions of high point concentration, the interpolated solutions are valid and accurate, but extrapolated surfaces may not respect the physics of the observed field.

References

Taylor, J., Richards, P. and Halstead, R. Computer routines for surfaces generation and display. Ottawa, Canada, Department of Energy, Mines and Resources, Marine Science Branch, 1971. (Manuscript Report Series, 16)

UNIRAS Reference Manuals, Version 5. Lyngby, Denmark, UNIRAS, 1986.

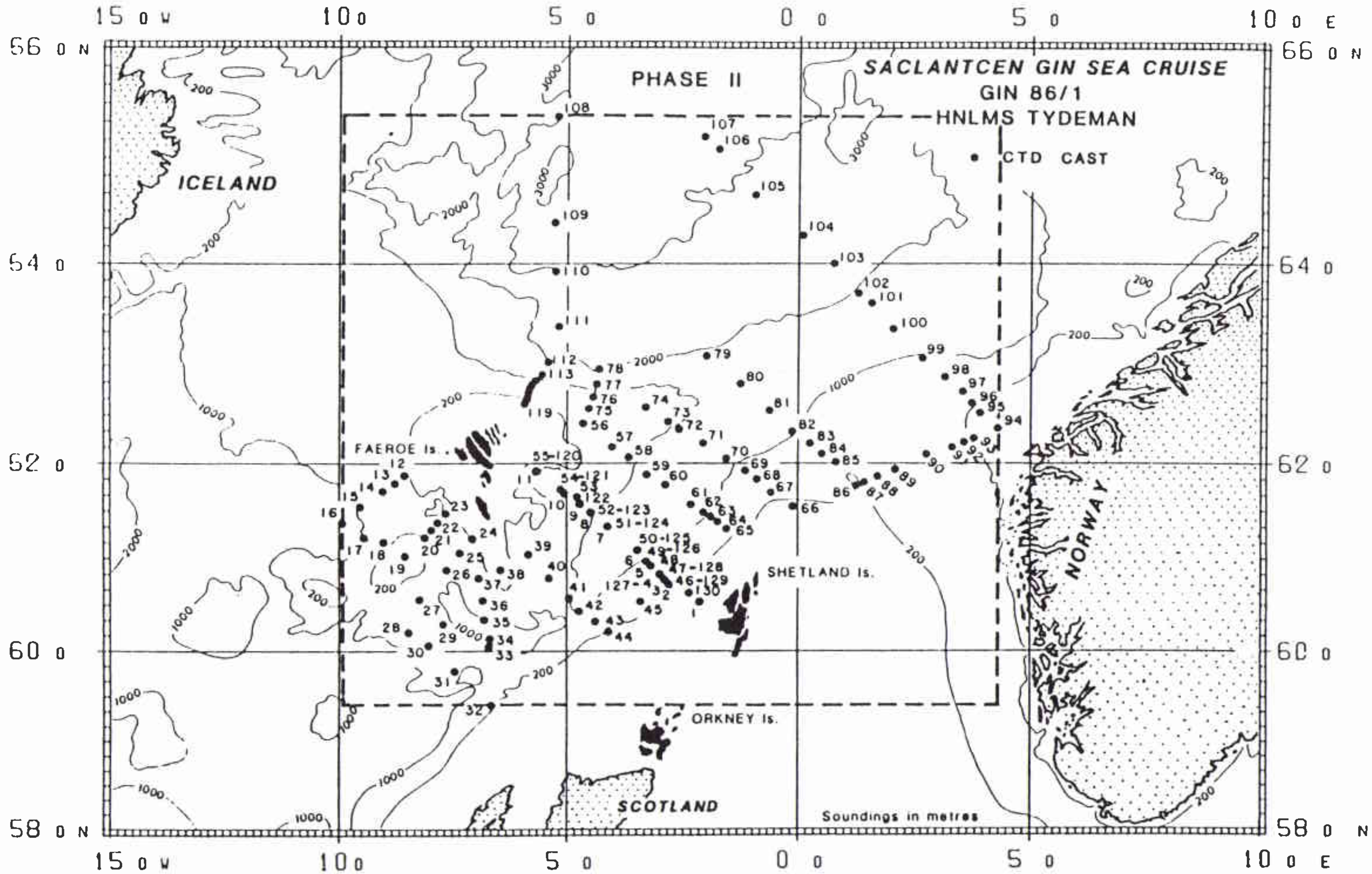

Fig. 1. GIN '86 cruise data distribution. The dots represent CTD casts taken during the June '86, and the numbering is according to the sampling. The box marks the area where interpolation/extrapolation is done.

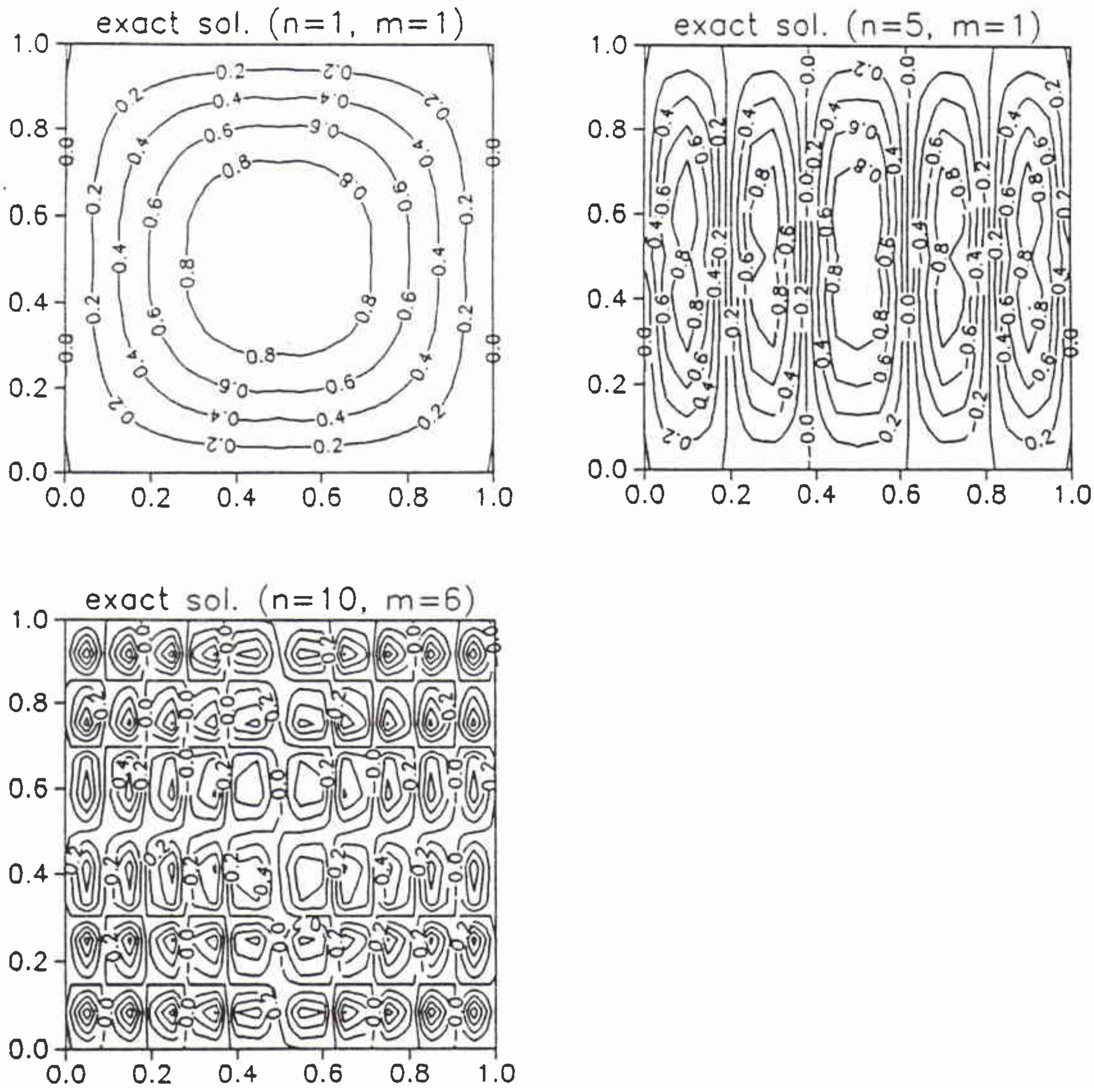

Fig. 2. **The exact solution of the numerical experiments.**

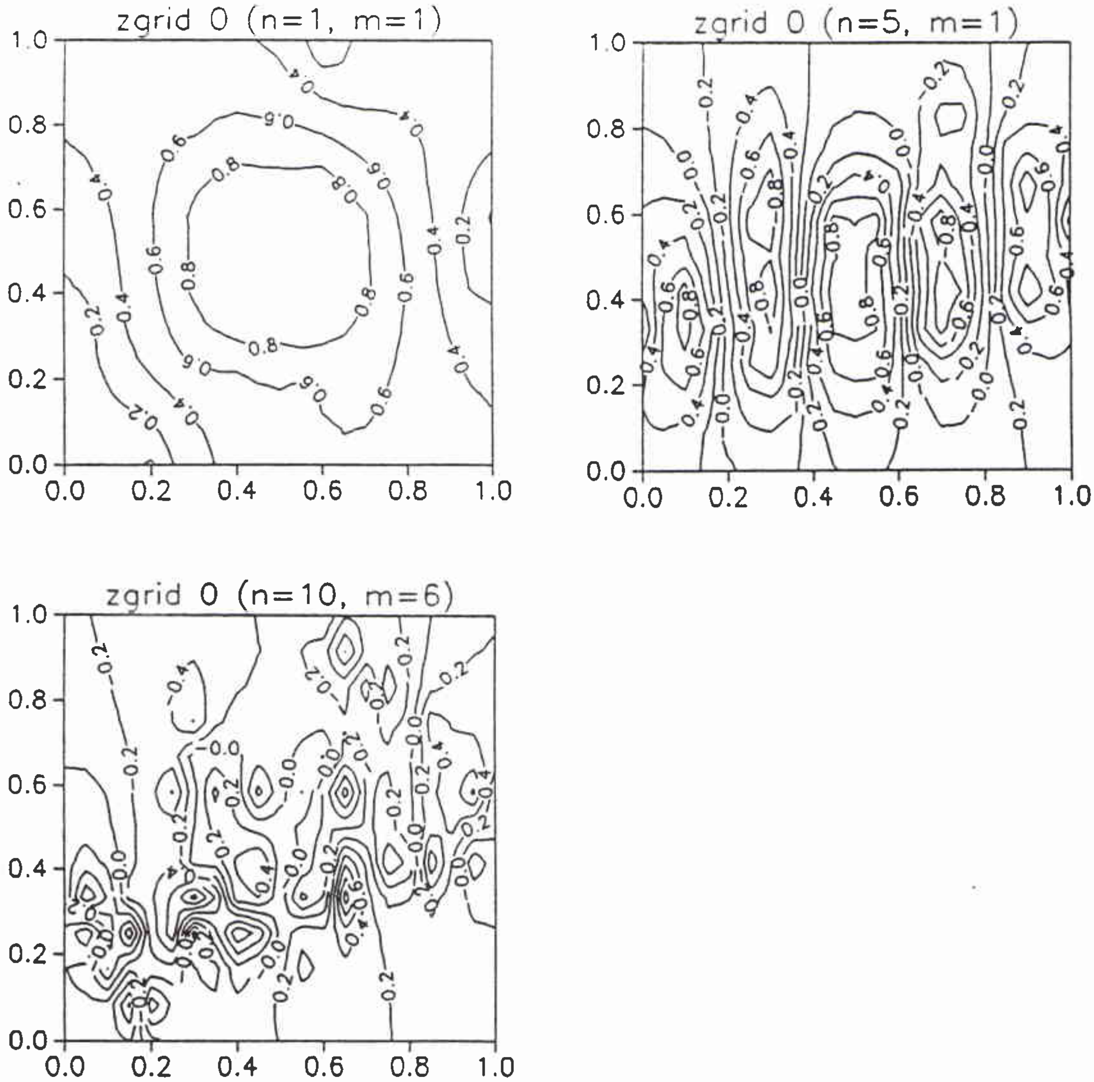

Fig. 3. The interpolated solutions from ZGRID(0).

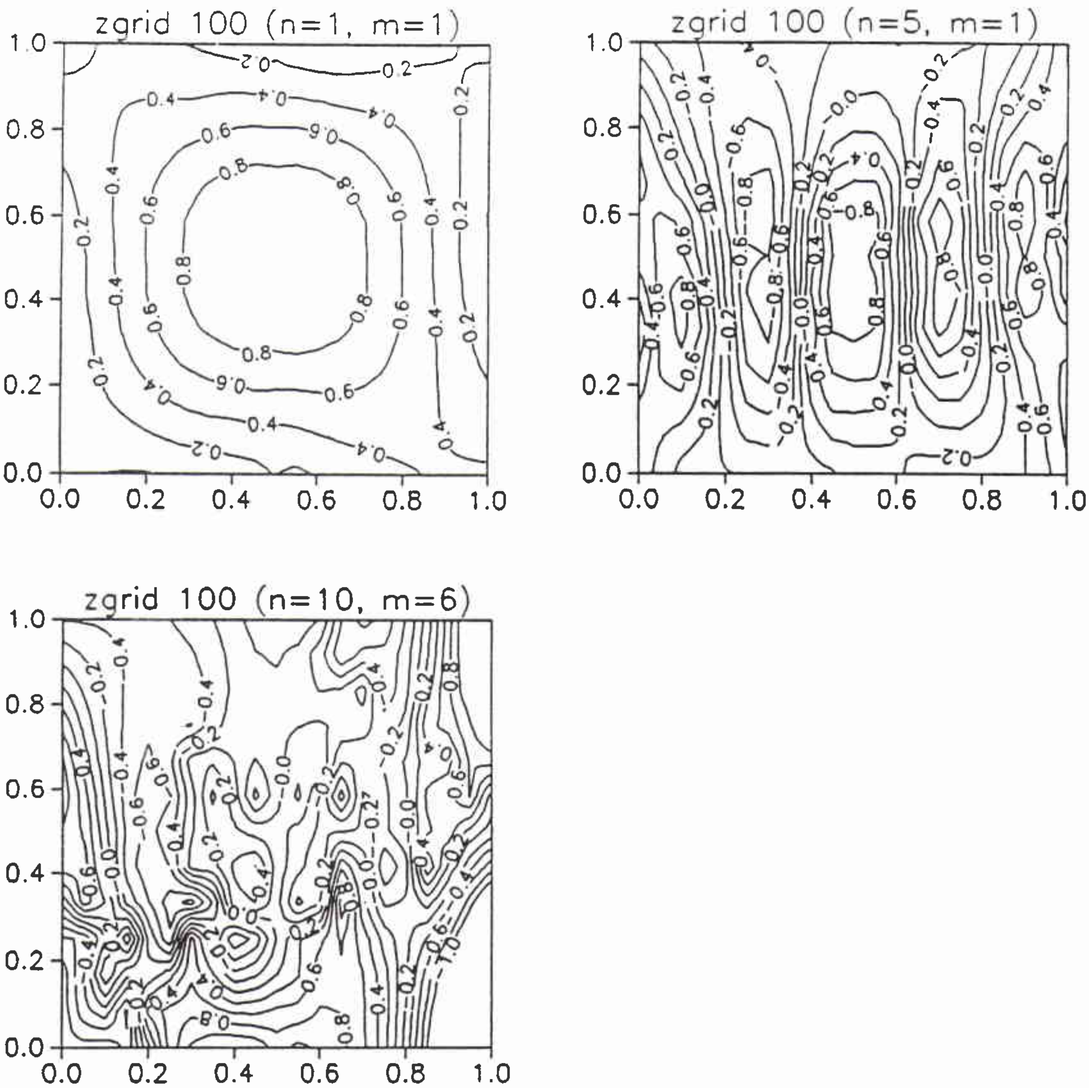

Fig. 4. The interpolated solutions from ZGRID(100).

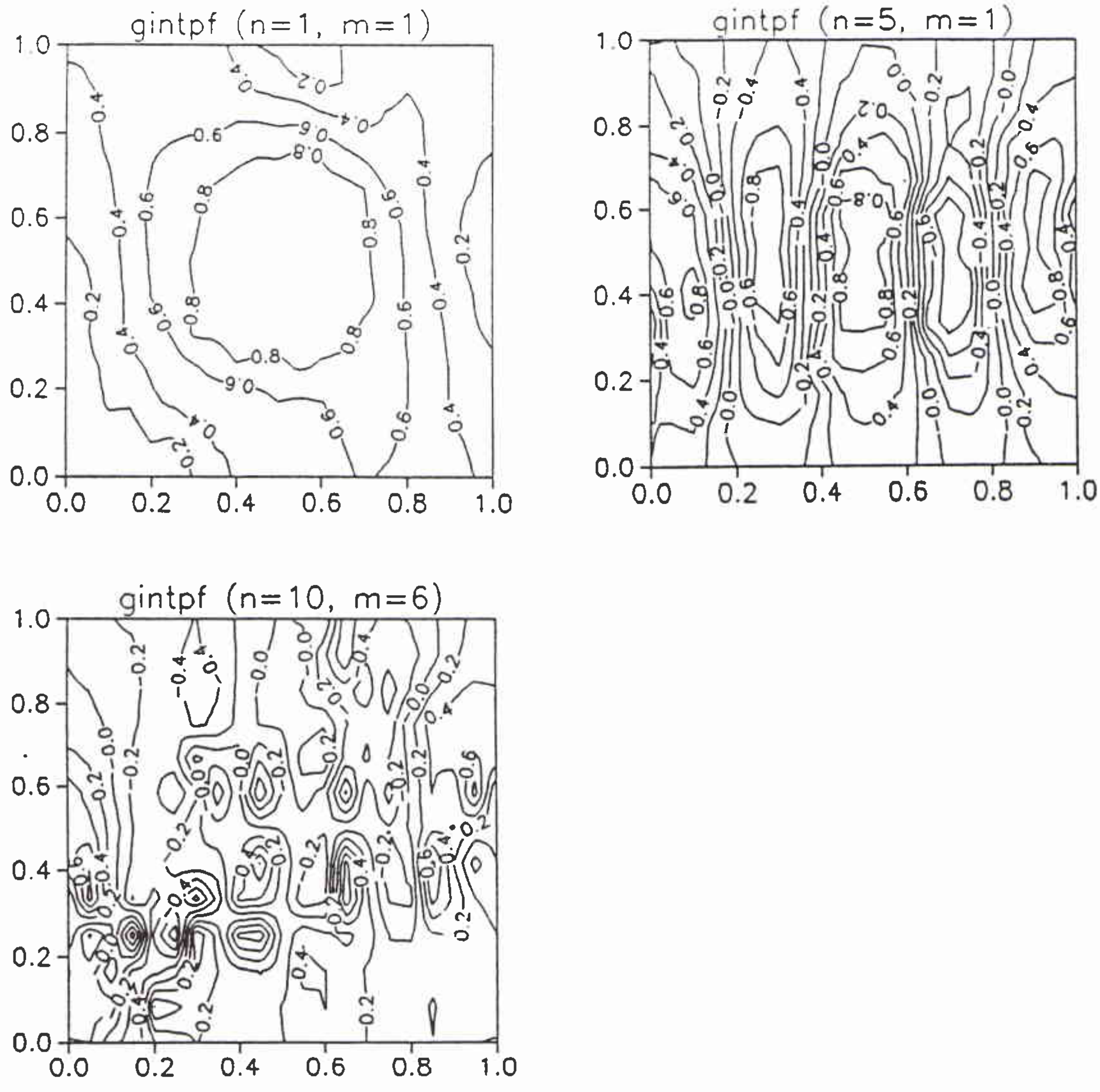

Fig. 5. **The interpolated solutions from GEOINT routine GINTPF.**

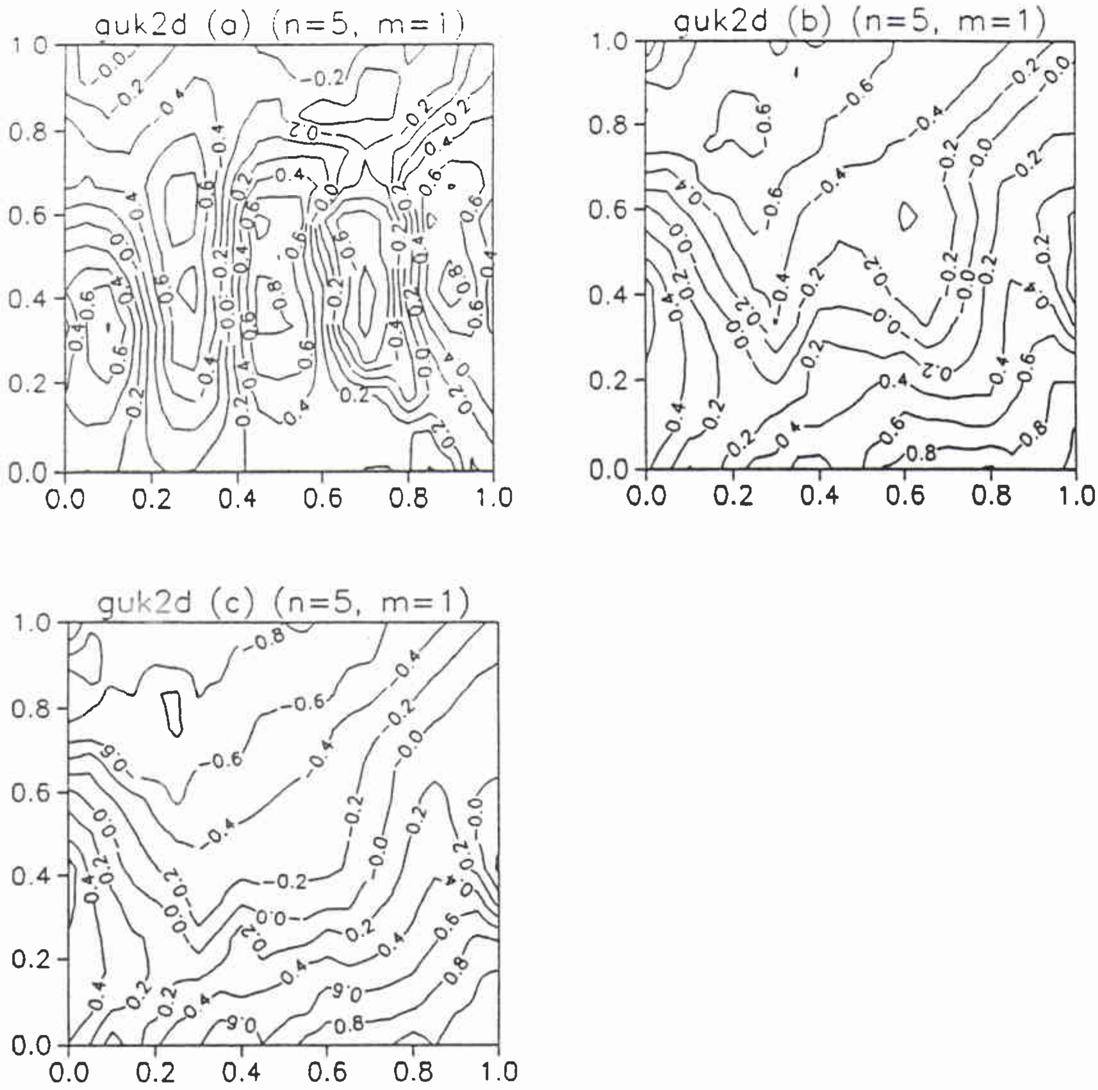

Fig. 6. The interpolated solutions from KRIGPAK routine GUK2D as function of the semivariogramm model. (a) $r = 1$, $s = 1$; (b) $r = 15$, $s = 2.5$; (c) $r = 1$, $s = 10$.

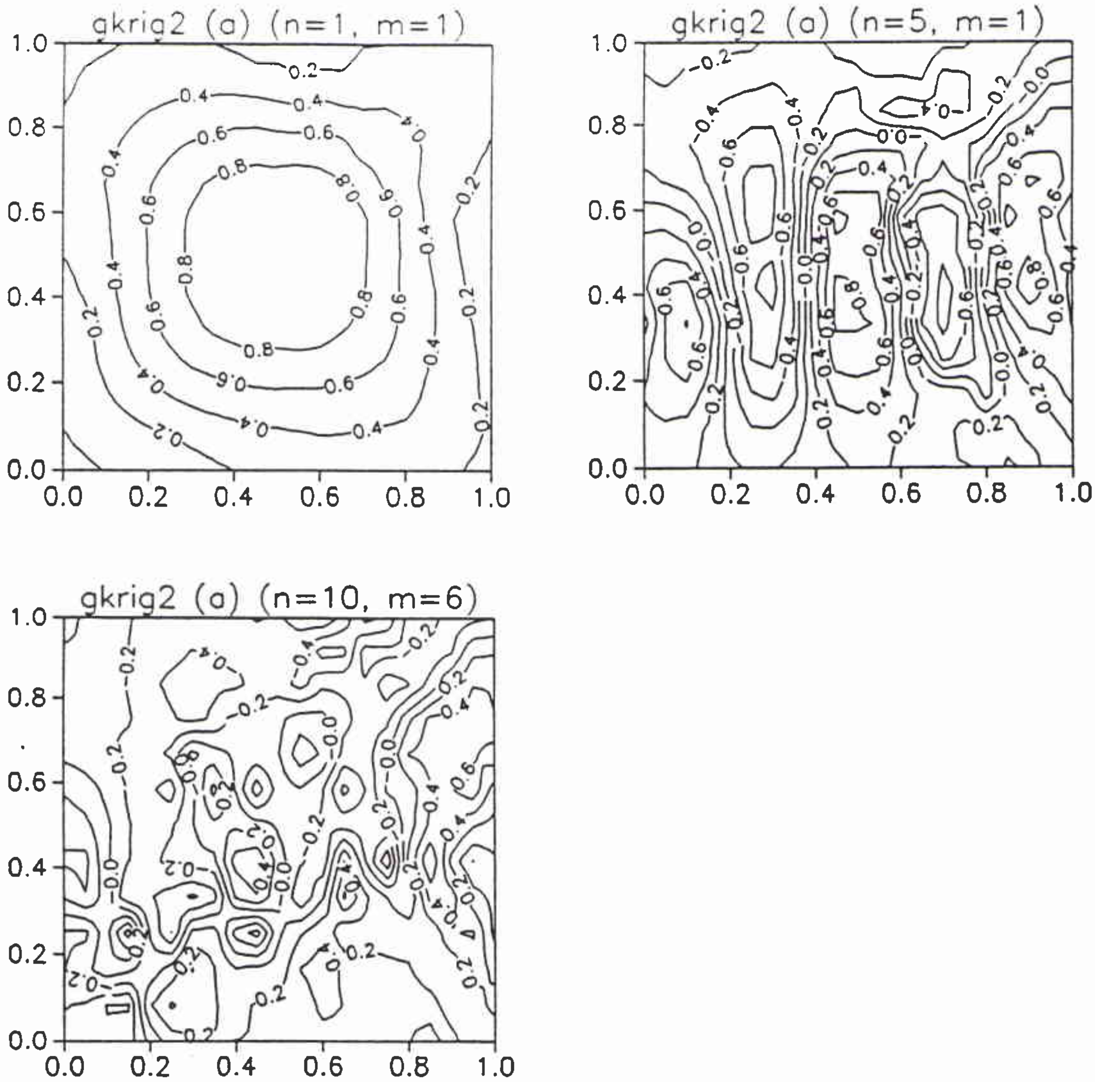

Fig. 7. The interpolated solutions from KRIGPAK routine GKRIG2.

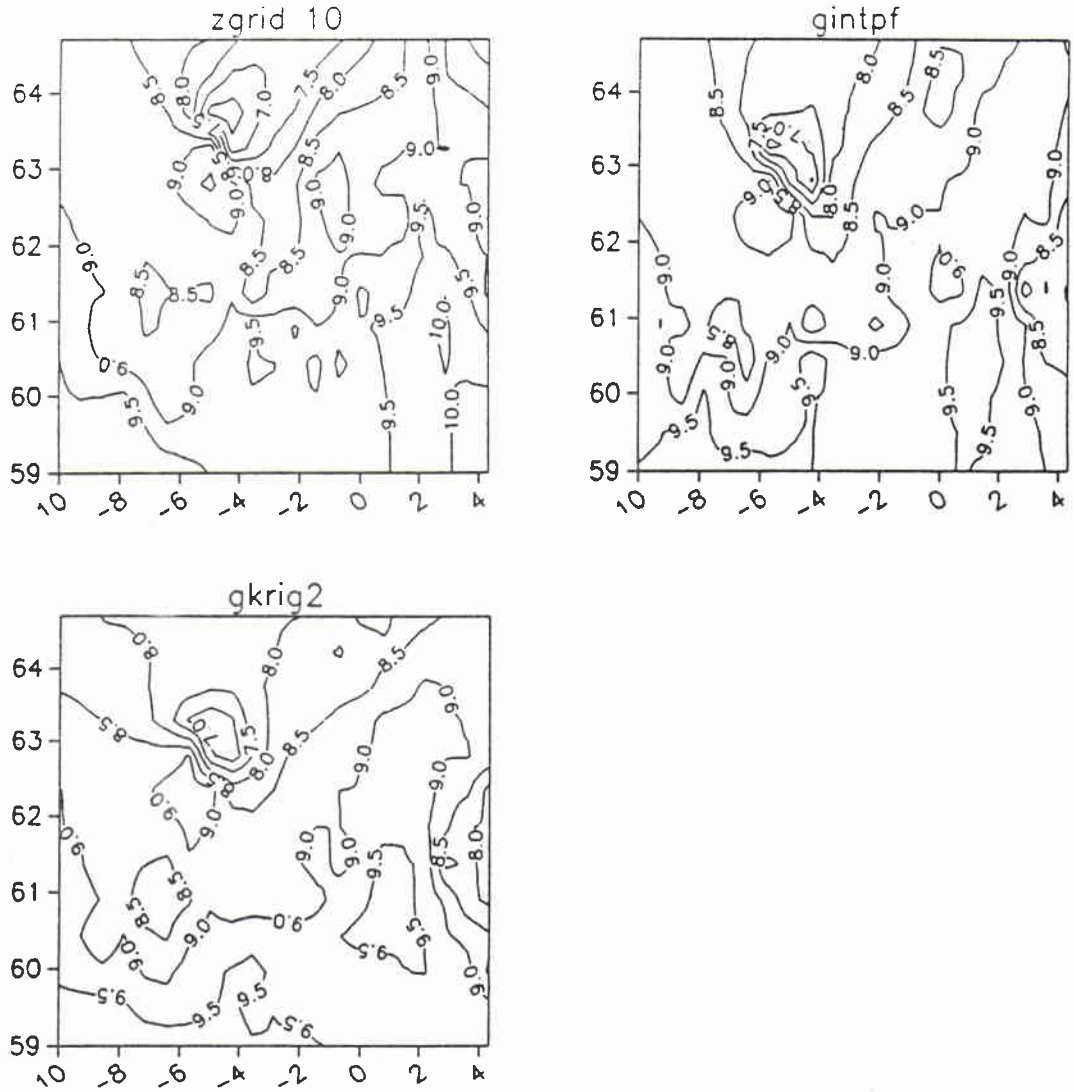

Fig. 8. The interpolated temperature distributions from temperature values at 25 m depth taken during GIN '86 cruise.

Report no. changed (Mar 2006): SM-217-UU

Initial Distribution for SM-217

Ministries of Defence

JSPHQ Belgium	2
DND Canada	10
CHOD Denmark	8
MOD France	8
MOD Germany	15
MOD Greece	11
MOD Italy	10
MOD Netherlands	12
CHOD Norway	10
MOD Portugal	5
MOD Spain	2
MOD Turkey	5
MOD UK	20
SECDEF US	60

NATO Authorities

Defence Planning Committee	3
NAMILCOM	2
SACLANT	3
SACLANTREPEUR	1
CINCWESTLANT/ COMOCEANLANT	1
COMSTRIKFLTANT	1
CINCIBERLANT	1
CINCEASTLANT	1
COMSUBACLANT	1
COMMAIREASTLANT	1
SACEUR	2
CINCNORTH	1
CINCSOUTH	1
COMNAVSOUTH	1
COMSTRIKFORSOUTH	1
COMEDCENT	1
COMMARAIRMED	1
CINCHAN	3

SCNR for SACLANTCEN

SCNR Belgium	1
SCNR Canada	1
SCNR Denmark	1
SCNR Germany	1
SCNR Greece	1
SCNR Italy	1
SCNR Netherlands	1
SCNR Norway	1
SCNR Portugal	1
SCNR Turkey	1
SCNR UK	1
SCNR US	2
SECGEN Rep. SCNR	1
NAMILCOM Rep. SCNR	1

National Liaison Officers

NLO Canada	1
NLO Denmark	1
NLO Germany	1
NLO Italy	1
NLO UK	1
NLO US	1

NLR to SACLANT

NLR Belgium	1
NLR Canada	1
NLR Denmark	1
NLR Germany	1
NLR Greece	1
NLR Italy	1
NLR Netherlands	1
NLR Norway	1
NLR Portugal	1
NLR Turkey	1
NLR UK	1

Total external distribution	236
SACLANTCEN Library	10
Stock	34
Total number of copies	280