Perspectives on mapping the MARGO reconstructions by variogram analysis/kriging and objective analysis

Christian Schäfer-Neth[*], André Paul, Stefan Mulitza

DFG Research Center Ocean Margins and Department of Geosciences, University of Bremen, D-28334 Bremen, Germany

Abstract

Paleo-data are not always useful in their original scattered distribution: For many numerical modeling issues and for display and comparison, gridded versions that provide meaningful estimates for under-sampled regions are a must. We constructed a test data set with a spatial resolution identical to the Multi-proxy Approach for the Reconstruction of the Glacial Ocean (MARGO) samples from the World Ocean Atlas temperatures and assess the performance of (i) variogram estimation and kriging and (ii) the Levitus objective analysis in reconstructing the original data. The two methods complement each other with respect to the facility of application and the quality of the results. Kriging requires a careful parameter adjustment but delivers the smallest deviation from the original data (1.22 °C in the global average), whereas the Levitus analysis provides a fast and efficient tool for checking the samples from different proxy data against each other during the compilation of the final MARGO database, at the expense of a slightly higher error (1.56 °C).

© 2004 Elsevier Ltd. All rights reserved.

1. Introduction

Every major advance in the reconstruction of paleooceanographic data, such as by the CLIMAP (1981) and GLAMAP (Sarnthein et al., 2003) projects, has triggered numerous studies that employed the new data in the form of regularly gridded fields. Common applications are the forcing of ocean and atmosphere models and the validation of coupled climate models, but gridding is already necessary for displaying and comparing data. However, interpolating the scattered paleo data to a regular grid is no straightforward task, and unsuitable methods may easily generate artifacts that will mislead any further studies. By compiling not only species counts but in addition other available temperature proxies, and by improving the spatial sampling density, Multi-proxy Approach for the Reconstruction of the Glacial Ocean (MARGO) marks a new step towards a better reconstruction of the glacial sea surface. Again, it is desirable to produce regular grids from the new seasonal data. Using a test data set based on modern temperatures but with a spatial resolution identical to the MARGO data set, we compared two different interpolation methods: variogram analysis and kriging (Deutsch and Journel, 1992), and the objective analysis by Levitus (1982). According to the results of this comparison, variogram analysis and kriging is best suited for the final gridding of the new MARGO temperature reconstructions, which underlines a remark by Wunsch (1996, Chapter 5.4), that "Kriging […] deserves more oceanographic attention".

2. Test data set

For comparing the different interpolation techniques, we adopted the following strategy:

We constructed three test data sets (annual mean, January–March and July–September) from the unanalyzed World Ocean Atlas 1998 temperature data (WOA,

*Corresponding author. Tel.: +49-421-218-7188; fax: +49-421-218-7040.

E-mail address: csn@uni-bremen.de (C. Schäfer-Neth).

1998) for 10 m depth. Except for an $1° \times 1°$ averaging, these temperatures have not been treated by any interpolation, specifically, they are not biased due to some inherent a priori ''information'' from any of the methods we discuss here. Apart from the high southern latitudes during austral winter, these data provide an almost complete global coverage which facilitates a comparison with the gridding results. Furthermore, the analyzed 10 m WOA data have been used to calibrate the transfer techniques throughout the groups participating in MARGO.

The WOA temperatures were averaged to every even longitude and latitude, that is, the $2° \times 2°$ grid used by the CLIMAP Project Members (1981), that we chose as the target for the present interpolation exercise. The choice of a $2°$ spacing is a compromise; in the areas of densest sampling, a gridding to $1°$ or even $0.5°$ would be feasible, but where data are sparse, $5°$ would be a more reasonable distance. At this stage of a preliminary test for a gridding of the actual MARGO data, the $2°$ grid served as a test bed, which can be easily refined or coarsened by changing a few parameters of the employed programs, depending on future requirements and data availability. To construct an input data set for the present study, we binned the original MARGO core locations into $2°$ longitude/latitude squares and sampled the $2°$ WOA data set at these points (Fig. 1). In this way, the test data and the original MARGO data are of identical resolution.

After interpolating from the test points back to the $2° \times 2°$ grid, we computed the absolute differences between the original and the gridded fields as a measure of interpolation skill, used to assess the interpolation techniques.

3. Interpolation/gridding methods

An interpolation of paleooceanographic data from the core locations to a regular grid faces three main problems: (i) There are areas with dense spatial sampling in contrast to others that are practically void of data. In the case of the MARGO data set, examples for both situations can be found in the northern North Atlantic Ocean and the subtropical South Pacific Ocean, respectively (Fig. 1). These extremes in sampling density require an interpolation method that is able to fill in the large gaps, but does not discard the small-scale information present in the densely sampled regions. Ideally, when estimating the values in the under-sampled areas, the interpolation scheme should use the information on small-scale variability and the spatial correlations of the samples that is available from the well-sampled areas. The use of such information would facilitate extrapolations beyond the areas sampled by MARGO, e.g., into the polar regions. (ii) The spatial

temperature variability changes itself quite considerably between the different ocean basins. For example, it is very high at oceanic fronts but very low in the centers of the subtropical gyres. These regional changes should be accounted for in the mapping process. If there is other related knowledge available, such as information on the ocean currents, it might be sensible to include this information as well. (iii) The reconstructions at individual core sites may differ greatly with respect to their error margins. If these can be quantified, the gridding method should use this information for weighting the samples and computing the possible errors of the gridded data.

If the gridding is intended as an interface between geological samples and ocean circulation models, the more straightforward interpolation methods such as nearest neighbor interpolation, splines, and triangulation are not well suited for a number of reasons. Triangulation interpolates with linear distance weights within triangles that are constructed between the sample locations, and the gridded data set emerges as a surface of connected triangular planar tiles. Thus, the value at a given point is determined only by the values of the three surrounding vertices, and potentially useful information from farther samples is discarded. In case of highly uneven sample spacings this is clearly not desirable as small-scale characteristics of the interpolated field are abruptly removed or distorted at the transition from dense to sparse sampling. Another problem arises from the piecewise planarity of the gridded field because it places the horizontal gradients at the edges of the interpolating triangles (Taylor et al., 2004). These locations can be very different from the real-world gradients, and an ocean circulation model driven with a field containing such patterns would develop unrealistic fronts and currents. Spline interpolation avoids these sharp wrinkles, but suffers from a tendency to generate local bulges or depressions around detached sample locations, thereby generating unrealistic current loops in ocean models. Nearest neighbor schemes with an inverse distance weighting often induce pronounced gradients halfway between the sample points. As with triangulation, the location of these gradients is dictated entirely by the arrangement of the samples and not by the physical processes that underly the sampled values. Besides being insufficient for forcing ocean-only models, interpolated fields that contain these artifacts cannot be drawn on when results from coupled ocean–atmosphere models are to be assessed. Furthermore, these methods are poor extrapolators because the extrapolated field is dominated by the spatial derivatives of the data set at the very boundary of the sampled area, which results in a high probability of overshoots.

This situation can be greatly improved if the gridding scheme propagates information from the sample locations to the grid points in a more sophisticated manner,
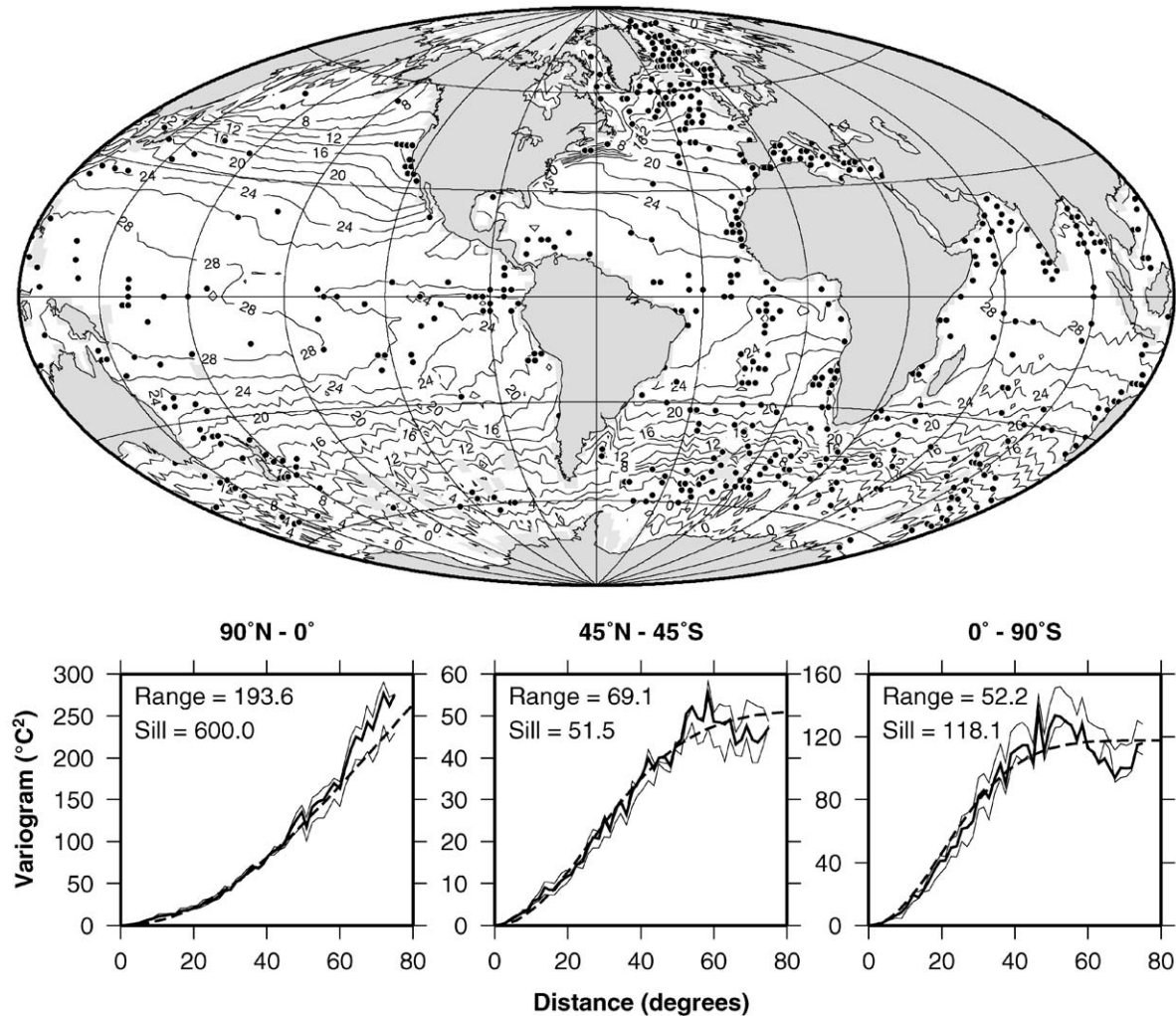
Fig. 1. Test data set. Top: Unanalyzed annual mean WOA (1998) 10 m temperatures (contour interval 2 °C). The locations of the input test data points (dots) were obtained by binning the MARGO core sites into 2° × 2° boxes. The geographical grid spacing is 30° and the central meridian is located at 60°W. The total number of data points is 530 (507 for the January–March data and 482 for July–September, not shown here). Light grey shading indicates gaps in the WOA data set. Bottom: Empirical (heavy solid) and modeled (dashed) modified Gaussian variograms for the three latitude belts. For comparison, the empirical variograms for the winter and summer data are indicated by the thin lines. Note the different scales of the vertical axes. The sill of the variogram for the northern belt cannot be identified from the empirical variograms. It was set to the indicated value based on a comparison of the empirical variograms's inflection points.

implicitly accounting for the underlying physical processes that determined the sample data. This is even more desirable if the interpolation is applied to paleo-data that are representative for a background climatic state different from today. One possibility to incorporate the physical constraints would be to assimilate the sample data into a coupled atmosphere–ocean model. In our associated paper (Paul and Schäfer-Neth, 2004) we discuss an example of this approach, based on the same test data set. Here, we compare variogram analysis and kriging (VAK hereafter) and the objective analysis devised by Levitus (1982, LOA). We demonstrate that these approaches are able to cope with uneven sampling density, and to reasonably extrapolate if required.

3.1. Variogram analysis and kriging

Variogram analysis and kriging (VAK) is a weighted-average interpolation method that adjusts the averaging weights according to the spatial variability of the sample data. (The text books by Deutsch and Journel (1992) and Wackernagel (2003) give an excellent introduction into the subject.) VAK is performed in two steps. First, the spatial variability of the sample data set is quantified by mapping the variance of paired sample data against the distance between the data points. For this purpose, the total distance mapped is subdivided into *lags* of equal width. The resulting curve, the so-called *empirical variogram*, typically shows low variances at short distances and eventually reaches some saturation value
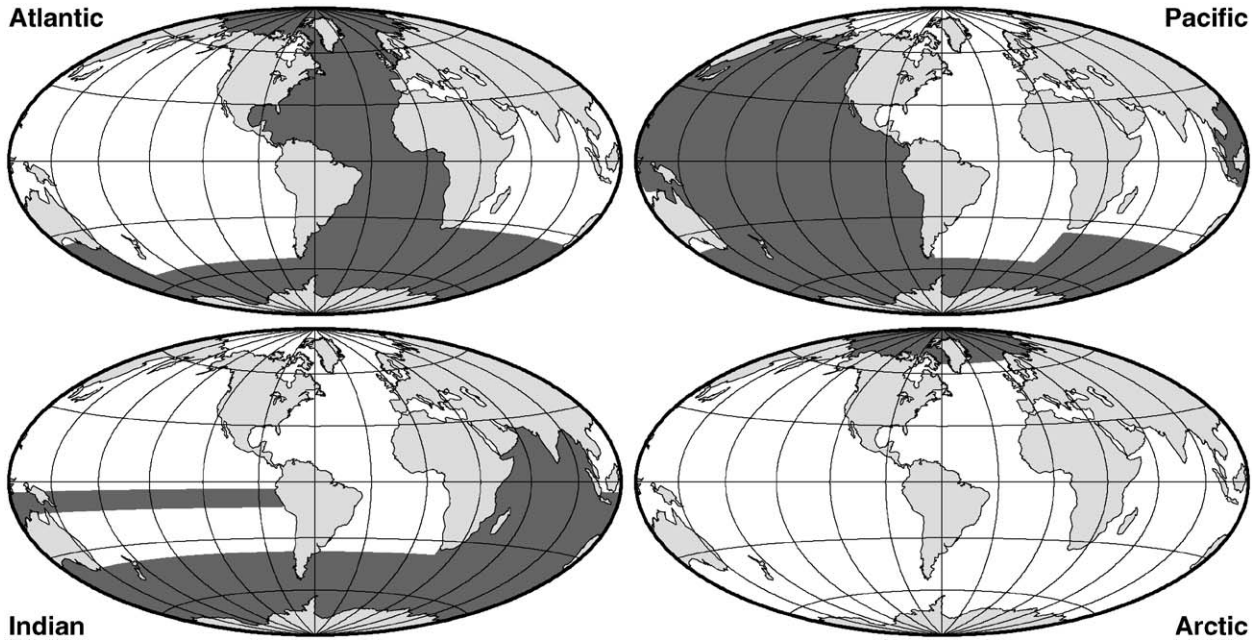
Fig. 2. Communication masks (grey shade) for the four major oceans (masks for the marginal seas were employed as well). Pairs of data points were allowed to contribute to the empirical variograms only if both points belong to a given mask. Likewise, the kriging and the Levitus estimate for a gridded point was computed only from those samples that share at least one ocean mask with the grid point.

for large distances (Fig. 1, bottom panels, solid lines). To feed this information into the kriging interpolation, an analytical *variogram model* is fitted to the empirical curve (dashed lines). Most of the commonly used models, such as the modified Gaussian that is employed here, are determined by two parameters: a *sill* that represents the highest variance for a large data point separation, and a *range* that denotes the distance at which the variogram model attains its sill. Therefore, the range is a measure of the distance over which the data are correlated. Because an empirical variogram does not necessarily start with zero variance at small distances, the variogram model may have to be adjusted by a constant offset, termed the *nugget effect*. In the second step, the actual kriging interpolation, the data values for a given set of locations are estimated, in our case, on the regular $2° \times 2°$ grid. For this purpose, kriging calculates the averaging weights according to the variogram model from the distances between the sample points and the location of a desired estimate. This weighting method minimizes the *kriging variance*, that is, the error variance of the estimates.

In this way, VAK accounts for the spatial variability present in the sampled temperature data. Given the fact that these variations are the result of ocean currents, mixing, and sea surface heat exchange, information on these processes is implicitly introduced into the interpolated field. Conceivably, if the global ocean is considered, there are regions that are closely related to each other because of the physical processes, whereas the linkages between others are much weaker. Therefore

it is not appropriate to use all possible data pairs for estimating the empirical variograms. We implemented a simple representation of these restrictions of inter-basin exchanges into the GSLIB variogram estimation routine "gamv2" (Deutsch and Journel, 1992, available at www.gslib.com) by employing communication masks (Fig. 2) that inhibit a pairing of data points that are unlikely to influence each other in the real ocean. These masks are used identically in the LOA method (see below). Another feature of the global temperature data set are the great differences of the spatial variability between different regions of the ocean. To account for this, "gamv2" was further revised to compute separate empirical variograms for three overlapping zonal belts that extend from the south pole to the equator, from $45°S$ to $45°N$, and from the equator to the north pole, respectively. A finer subdivision of the data set into overlapping $90° \times 90°$ boxes of the globe yielded only minor differences between the variograms for a given latitude (not shown here). Thus, we judge the belt-wise split as being sufficient, albeit necessary, as borne out by the great differences between the empirical variograms (Fig. 1, bottom, solid lines). A third modification was applied to "gamv2" to enable the use of spherical coordinates. This is required by the zonal cyclicity of the data set and the convergence of the meridians in high latitudes.

Parallel changes of the code were included in the GSLIB kriging routine "okb2d" that we employed for this study. The kriging was carried out separately for the three belts, and we combined the resulting belt-wise

fields by weighted averaging in the overlapping regions. The weights increased linearly from zero at a belt's southern and northern limits to one at its central latitude. The modified codes of "gamv2" and "okb2d", the communication masks, and the test data set may be obtained from www.palmod.uni-bremen.de/~csn.

There are other geostatistical packages that could be modified in a similar manner: *GSTAT* (Pebesma and Wesseling, 1998, available at www.gstat.org) provides a graphical interface and additional variogram models that could be more appropriate under some circumstances. For example, if a data set with cyclic boundaries like ours should be analyzed over zonal distances much larger than 90°, the periodic model might be a sensible choice. GSTAT also allows a masking of given areas by to user-defined polygons, comparable to the communication masks we use here. However, GSTAT is programmed for cartesian coordinates only and would thus require additional work to make it applicable to a sphere. Its code structure is much more complex than that of GSLIB. *Spherekit* (available at www.ncgia.ucsb.edu/pubs/spherekit) is already designed for spherical coordinates and uses a graphical front-end, but provides no masking or clipping of specified regions. Spherekit relies on specific versions of GMT,[1] Tcl/Tk,[2] and NetCDF[3] that have to be installed as well. The great advantage of GSLIB is that it comes in plain FORTRAN without any requirements except a compiler, and that code modifications can be easily implemented.

One might consider whether a realistic geodetic coordinate system could be more appropriate than the spherical one. Given the World Geodetic System 84, the earth's oblateness amounts to 1:298.257, and using spherical coordinates for kriging introduces a 0.3% anisotropy between the meridional and zonal directions. This is only a minor error and can be safely neglected in the context of the present application.

3.2. Levitus objective analysis

The interpolation by Levitus (1982, LOA) comprises three stages and starts with an initial guess by placing all 2° input data into their respective boxes on the 2° × 2° grid and filling the other boxes with the zonal mean of the input data for a given latitude. In the second stage, the first guess gets iteratively improved by adding the distance-weighted differences between input data and first guess to the grid points, using a hierarchy (Levitus, 1982) of predefined search radii (in our case 14°, 11°, 8°, and 5°). As already noted in the previous section, we

---

[1]GMT: The Generic Mapping Tools, http://gmt.soest.hawaii.edu.

[2]Tcl/Tk: Tool Command Language and graphical interface toolkit, http://www.tcl.tk/software/tcltk.

[3]NetCDF: Network Common Data Form, http://www.unidata.ucar.edu/packages/netcdf.

modified the original Levitus method by applying communication masks (Fig. 2) to prevent a spreading of information across the continents during the second stage, and all distances were computed within spherical coordinates. In a final step, the interpolated field is smoothed by the subsequent application of a median and a 3 × 3 point Shapiro (1971) filter. If there is no input data available for a grid point within the search radius, LOA preserves the first guess like the method of Reynolds and Smith (1999).

4. Results

4.1. Variogram analysis and kriging

We computed the empirical variograms for the three zonal belts omni-directionally with 50 lags of 1.5° width, that is, over a total distance of 75°. In the southern and the equatorial belt, the variograms show a well-defined Gaussian shape that reaches sills exceeding values of $100\,°C^2$ and $50\,°C^2$, respectively, at a range of roughly 50°–60° (Fig. 1, bottom panels, heavy solid lines). In the northern belt, no sill is visible. For a given latitude belt, both sills and ranges do not differ very much between the annual and seasonal data sets (thin lines) and we restrict our discussion to the annual-mean data.

Because of the general shape of the empirical variograms, it is advisable to use a Gaussian variogram model for kriging. However, Gaussian models are almost horizontal for small distances, which leads to many almost-zero entries in the coefficient matrix of the kriging equations. Even if the matrix remains invertible, there is a tendency to generate small-scale noise in the kriged field that can be reduced by adding a tiny artificial nugget effect (Englund and Sparks, 1991; Ababou et al., 1994; Pannatier, 1996). According to Wackernagel (2003, Chapter 16), a modified version of the Gaussian model,

$$\gamma(h) = c \cdot \left[1 - \exp\left(-\frac{h^p}{a^p} \right) \right], \qquad (1)$$

provides a much cleaner solution to this problem, where c and a denote the model's sill and range, and h represents the spatial distance. Instead of $p = 2$, the exponent of the standard Gaussian model, the modified model employs a reduced value, $p < 2$, that must be adjusted together with a and c when fitting the model to the empirical variogram. The fits were performed as follows: Direct least-square fits of model and empirical variograms resulted in values of $p > 2$, which violates the assumed variogram model. Therefore we varied p stepwise over the range $1.50 \leqslant p \leqslant 1.95$ and calculated a and c by least-square fitting the model for a given p. To narrow down the possible range of p, we visually inspected the fit of modeled and empirical variogram,

and the smoothness of the resulting kriged SST field. For all exponents $p \leqslant 1.7$, the resulting model slopes emerged distinctly smaller than the slope of the empirical data. If p was chosen $\geqslant 1.9$, the interpolated SST data were prone to the undesired small-scale disturbances. These findings leave $1.7 < p < 1.9$ as a suitable range, and we performed our analyses with $p = 1.8$.

The *practical range* of this modified Gaussian model, that is, the distance at which the model attains 95% of its sill, is indicated in Fig. 1 and Table 1 for all three belts, and given by

$$a_{95} = 3^{1/p} \cdot a \approx 1.84 \cdot a. \qquad (2)$$

As stated above, the empirical variogram for the northern belt shows no visible sill, caused by the large temperature difference between samples from the equatorial and Arctic Oceans. Given its curvature, a power variogram model given by

$$\gamma(h) = c \cdot h^a \qquad (3)$$

seems to be more appropriate in this case, where the exponent a is constrained to $0 < a < 2$. For values very close to 2, the power model is horizontal in its beginning, just as the standard Gaussian model, and generates the same small-scale noise in the gridded field. Unfortunately, in our case the fitted power model has an exponent of $a = 1.994$, and the power model is no suitable choice. Therefore we used the modified Gaussian model with an assumed sill of $600\,°C^2$. This is five times the sill found for the southern belt ($118\,°C^2$), corresponding to the ratio of the values that both empirical variograms assume at their inflection points, namely $\approx 60\,°C^2$ (southern belt) and $\approx 300\,°C^2$ (northern, Fig. 1), and consistent with the $\approx 25-30\,°C$ SST contrast between the equatorial and arctic regions. The range of the model was determined by least-square fitting as for the other latitude belts.

Kriging was then carried out with an omnidirectional search radius of $50°$ and a maximum of 40 data points to be included in the estimation of a given grid point. The small-scale features are well represented in the gridded result (Fig. 3, top), especially, if the areas of strong gradients (see above) are considered. There is a general agreement between dense spatial sampling

Table 1
Parameters of the variograms fitted to the annual mean SST data (see Fig. 1, bottom)

Latitude belt	Practical range (deg)	Sill (°C²)	Data points
North	193.6	600.0	267
Equatorial	69.1	51.5	354
South	52.2	118.1	277

The sill of the northern belt's variogram model represents no fit result but was directly set (see text).

(Fig. 1) and low kriging variance (Fig. 4). As a measure of the skill of the interpolation method, we chose the absolute difference between gridded and original data as displayed in Fig. 5; its per-ocean averages are listed in Table 2. The global mean difference is $1.22\,°C$ for the annual data and 1.24 and $1.28\,°C$ for northern Winter and Summer, respectively. By and large, low kriging variances coincide with small differences between the gridded and the original data (Fig. 5, top). However, there are discrepancies. The frontal systems associated with the Gulf Stream, the Kuroshio, the Brazil Current and the Agulhas retroflection, that are clearly visible in the original data, are hardly present in the gridded fields. These regions with strong oceanic fronts are the most difficult ones to interpolate, as can be seen from the differences between gridded and original field that exceed $5\,°C$ (Fig. 5, top). In case of the three currents, this comes as no surprise because none of these regions is adequately sampled. In the Gulf Stream and the Brazil Current, there are only two or three points along the respective current's axis, and there is literally nothing in the Kuroshio. This poor sampling leads to an under-estimation of spatial variability, especially across the currents, and the interpolation scheme does not know anything about the existence of the front. The only remedy to this problem would be a better sampling. For the region of the Agulhas retroflection, the sampling is quite dense, but the original data show high variability on a very small spatial scale that cannot be preserved by the interpolation, which is evident from the undulating pattern in the difference field south of Africa. If this were to be changed, the original data set had to be subdivided further into smaller regions (Schäfer-Neth and Paul, 2003) and the VAK be carried out on a smaller scale. In the Arctic Ocean, there are two regions where the kriged field differs considerably from the WOA data set: Along the Siberian coast the gridded SSTs fall well below the freezing point, whereas they are too high in the Barents Sea (Fig. 5). The cooling is due to a tendency of Gaussian (and almost-Gaussian) variogram models to extrapolate beyond the sample data range near the boundaries of the area (Wackernagel, 2003). Although there are no boundaries on a sphere, the coast of Siberia is farthest away from the sample data points, simply because of the presence of the large Eurasian land mass. This problem can easily be overcome by including some additional points in the data-void Arctic Ocean with a prescribed freezing-point temperature. Indeed, if the gridded field was cut off at the freezing point, the mean absolute temperature error would drop from $3.5\,°C$ to about $1.8\,°C$ (Table 2) in the Arctic Ocean. The unrealistic warming occurs because the steep and high variogram model fitted for the northern latitude belt, that is, the entire northern hemisphere, is not fully appropriate for the highest northern latitudes where there are only small
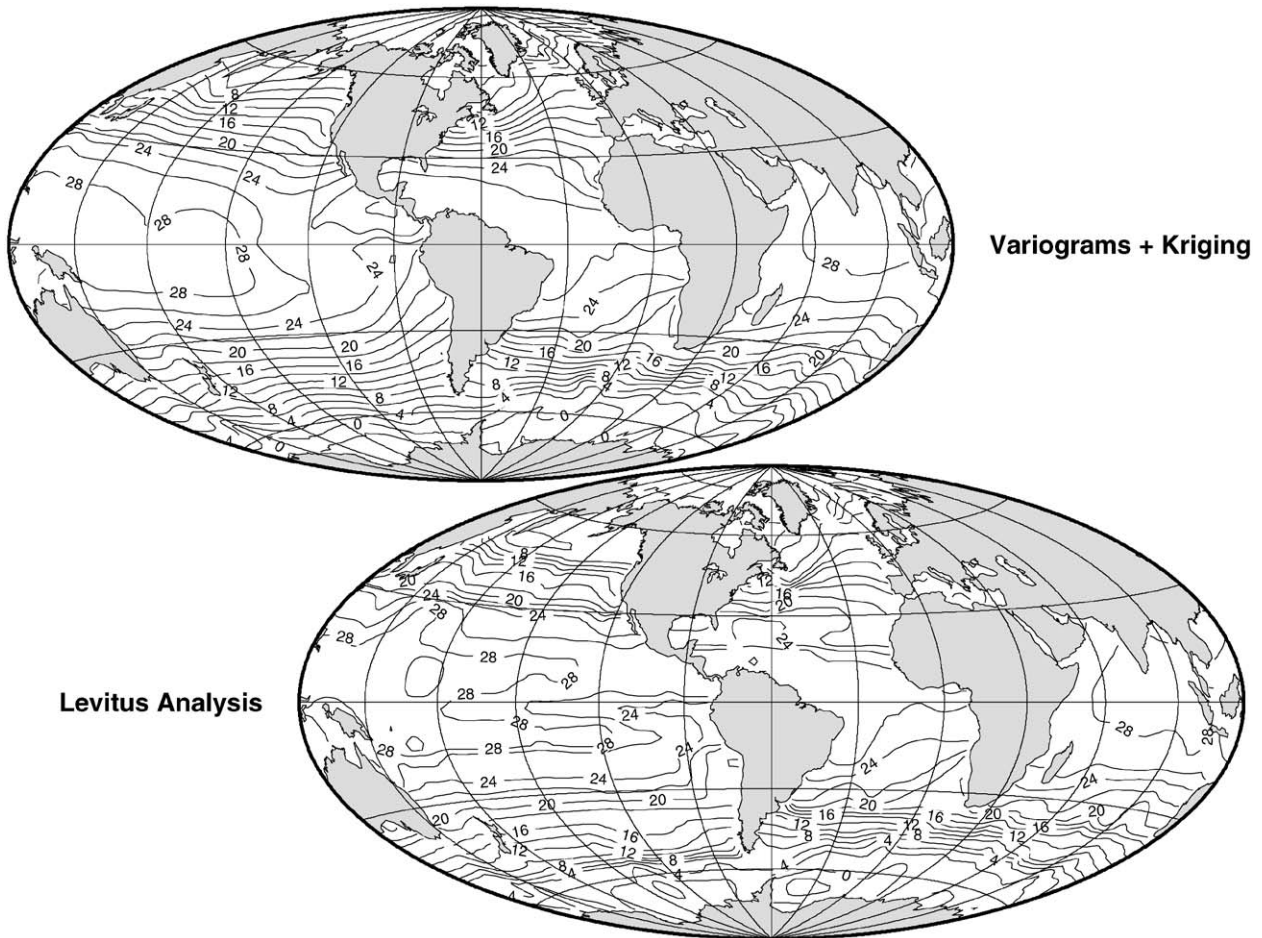
Fig. 3. Gridded annual mean temperatures (°C). The kriging result was obtained with the variogram models shown in the lower panels of Fig. 1 (dashed curves) and used up to 40 sample data within a maximum search radius of 50° around each grid point. The Levitus interpolation employed all available samples in a staggered hierarchy of four search radii of 14°, 11°, 8°, and 5°.
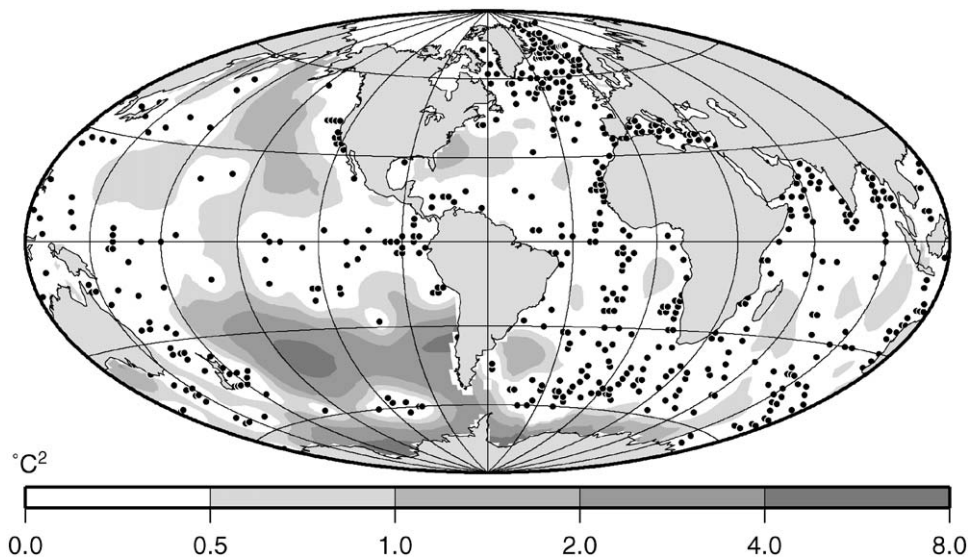

Fig. 4. Kriging variance for the interpolation of the annual WOA (1998) temperatures (°C^2) and locations of the sample points. The low values of the variance in areas of denser spatial sampling indicate that the variogram models were appropriately chosen.
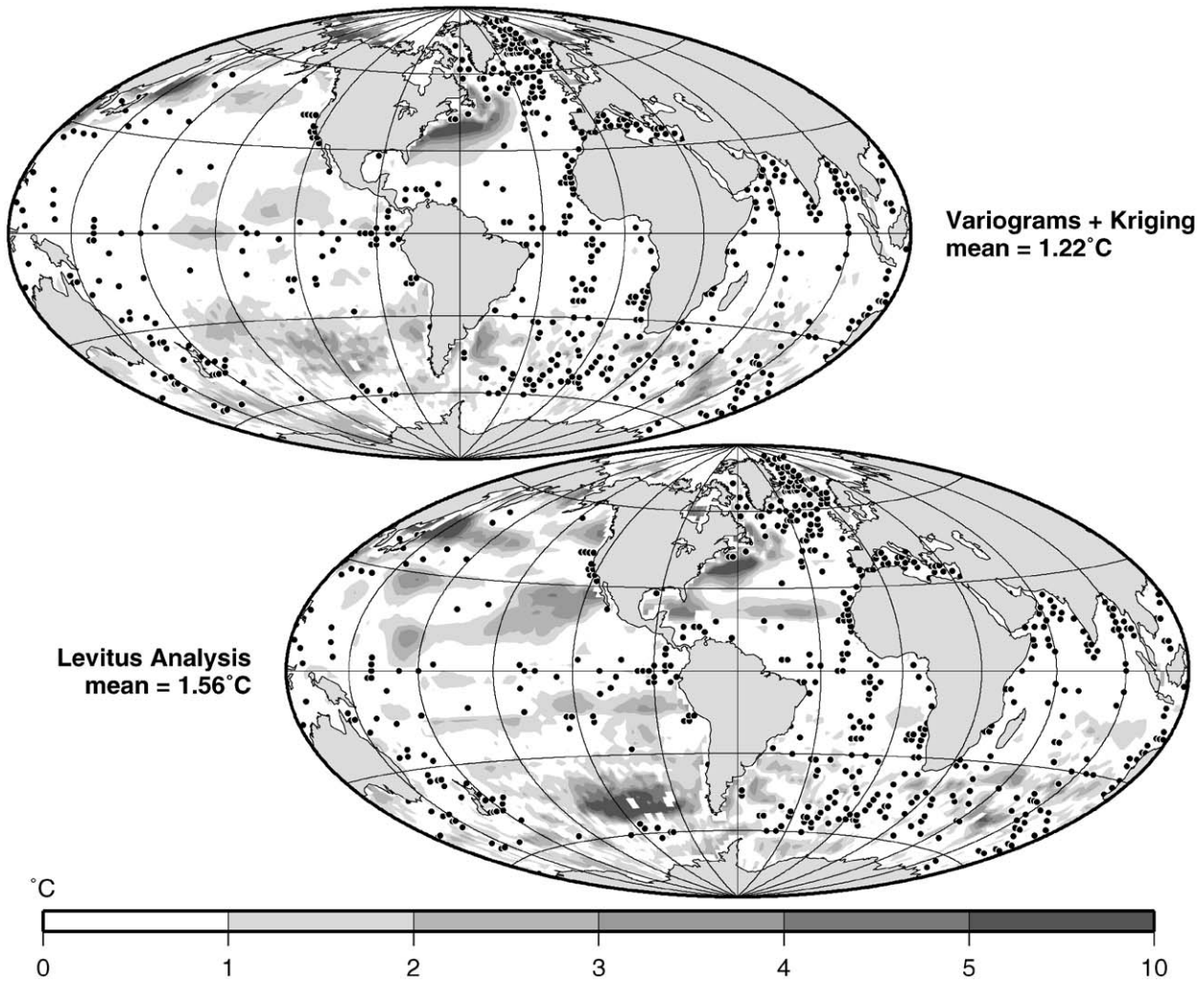
Fig. 5. Absolute difference between the interpolated and the unanalyzed annual temperatures (°C) and sample locations. In contrast to the kriging variance, the differences are not necessarily small in regions of dense spatial sampling and become large especially near oceanic fronts that emerge less sharp with both methods.

Table 2
Annual mean absolute differences (°C) between the original and the interpolated fields (see Fig. 5)

Ocean	Kriging	Levitus
Atlantic	1.29	1.40
Pacific	1.19 (1.15)	1.75
Indian	0.93	1.04
Arctic	3.52 (1.69)	1.84
Global	1.22 (1.15)	1.56

Values in parentheses result if all SSTs below −1.8 °C are set to this value in the kriged field.

temperature variations. In the present configuration, the Barents Sea is too strongly influenced by the higher mid- and low-latitude temperatures. The solution would be to further subdivide the northern belt and consider variograms of a shorter total range, and consequently

a lower sill, in the Arctic Ocean (Schäfer-Neth and Paul, 2003).

4.2. Levitus objective analysis

As can be seen from Fig. 3 (bottom), the LOA is characterized by a tendency to generate zonally oriented structures. Since many regions of the ocean are dominated by zonal current systems, the LOA yields fairly realistic results in these areas. For example, the 26 °C isoline in the southern Indian Ocean, the 8–10 °C belt in the southern Atlantic Ocean, and the 24–26 °C tongue in the equatorial Pacific Ocean show only small deviations from the original data (Fig. 5, bottom). Compared to the 28 °C line of the original (Fig. 1, top) field, the latter structure is somewhat over-emphasized. However, in areas with more meridional flows and low spatial sampling density, the zonal bias of the LOA may introduce unrealistic features such as the meanders of

the 22 and 24 °C isolines in the northern subtropical Atlantic Ocean and the squeezing of the 20–24 °C area off Baja California. Naturally, these pattern shifts result in large differences of 3–4 °C between the WOA data and the LOA interpolation (Fig. 5, bottom). The meridional temperature gradient, which is present in the original data from the North Pacific Ocean at about 40°N, becomes split into two weaker gradients around 30°N and 45°N. There, the LOA yields the highest absolute differences, beyond 5 °C, between the original and the interpolated data, except from the poorly sampled South Pacific Ocean and the Gulf Stream region for which the LOA fails to reproduce the location of the frontal system. This is reflected by the absolute difference between the original and the Levitus-interpolated data, which is highest for the Pacific Ocean (Table 2). It should be noted that the strength of the temperature gradient across the Gulf Stream is quite close to the original one, despite its displacement. In some cases of dense sampling at the coasts, the LOA preserves small-scale features very well, as exemplified by the 14° line west of North America and the almost straight 26° isoline west of Australia. The global mean differences relative to WOA amount to 1.56 °C in the annual mean and to 1.49 and 1.71 °C for Winter and Summer.

5. Discussion

According to the mean errors for the Atlantic Ocean (Table 2), the two methods are of comparable quality if the sampling density is high. However, if there are void regions wider than 20–30°, such as the subtropical North Atlantic Ocean, considerable differences may occur, as borne out by the meanders and loops of the 24° and 26° isolines in that area (Fig. 3), and by the considerably different absolute errors for the Pacific Ocean (Table 2).

5.1. Variogram analysis and kriging

To some extent, VAK resembles the optimal interpolation technique devised by Reynolds and Smith (1999). Their approach employs horizontal correlation lengths derived from the input data set to estimate gridding weights, quite similar to the variograms that can be viewed as an inverse measure of spatial correlation. Applied to satellite data, this method can rely on a much higher sampling density which enables an automatic fitting of the correlation length scales for any point of the globe. This is not generally possible for the presently available paleo-temperatures. The method of Reynolds and Smith accounts for the errors of the individual samples by adjusting the interpolation weights. This would be beneficial for the gridding of

the paleo-data and could in principle be accomplished by VAK. However, the statistical error of the individual temperature estimates is about ±1 °C for all MARGO reconstructions and thus provides no additional information. Hence, the kriging variances obtained by VAK for every gridded point do not reflect the quality of the data, but solely the spatial sampling density and the appropriateness of the variogram models fitted to the empirical variograms (Fig. 4). Nevertheless, the kriging variances provide an excellent indicator as to the quality of the interpolation, which will increase if the variance is minimized by a careful adjustment of the model variogram parameters. It should be noted that the temperatures reconstructed for a given core site by different methods or from different proxies may vary by much more than ±1 °C. If it can be ruled out in the future that this reflects different seasons of plankton growth or varying depth habitats, these deviations can be accounted for in the gridding procedure. With optimally tuned variogram parameters, VAK yields a very good interpolation over the data-void areas in-between the well-sampled ones, and reasonable extrapolations for the polar regions.

5.2. Levitus objective analysis

The LOA provides a fast method that can easily be repeated upon the arrival of new or revised samples, which is a clear advantage during the compilation of the data set and for the comparison of the individual temperature reconstruction methods. The characteristic zonal features produced by this method are due to the initialization with the zonal mean values of the sample data set. Once initialized, the field is updated at all grid points that are nearer to any of the sample points than the longest of the search radii, which is set to 14° in the present case. In the subtropical North Atlantic Ocean, the loops of the 22–24° isolines extend over a meridional distance corresponding to this radius. For the northern Pacific Ocean, which is void of samples over distances of 40–50°, this search radius implies that the LOA does not update the first guess at all. A clear indication for this is the split of the front associated with the North Pacific Current (Fig. 3, bottom), caused by the uneven meridional distribution of the samples that enter the initial guess for the North Pacific Ocean. Constructing the initial guess from the per-basin zonal means would improve this situation. These findings might depend on the search radius lengths and we repeated the Levitus interpolation with doubled and halved influence radii. Longer radii broadened the gradients and diminished the preference for zonally oriented structures, but did not remedy the split of the front in the north Pacific Ocean. The general widening of the gradients yielded even greater differences between the interpolated and the WOA data at the frontal areas. In the

under-sampled parts of the southern Pacific Ocean, the longer influence radii caused higher interpolated SSTs and as well a larger difference to WOA. The global mean difference was in this case 1.91 °C. With the shorter radii, the global mean error turned out to be only slightly lower, namely 1.87 °C. This similar difference was caused by too sharp gradients and an enhancement of the zonal structures. Thus, the original choice of 14°, 11°, 8°, and 5° constitutes the optimum, at least for a uniform application to the entire global grid. A local adaptation of the search radii to the changing sampling density should considerably improve the performance of the LOA. Regardless of the search radius, the requirement of a suitable first guess poses a problem outside the sampled area, that is, in the polar regions. Since there are no samples, the interpolation starts with the values of the northern- and southernmost samples, hence with temperatures that are much too high for the polar ocean. The subsequent iterative process partly adjusts this, but the final temperature distribution around Antarctica resembles the original one only very coarsely. To resolve this problem, additional sample points are necessary. If there is independent evidence for an ice cover, artificial "samples" with prescribed freezing point temperature are sufficient.

The subsequent Shapiro (1971) and median filters that employ a stencil of 3×3 grid points imply a reduction of the method's effective resolution by a factor of two, but compared to the VAK result, this does not flatten out the gradients in the interpolated fields. In fact, the VAK gradients are weaker and less realistic than those of LOA.

6. Summary

(1) The results of VAK are fairly reasonable even in under-sampled areas and when extrapolating poleward of the sampled regions. VAK fails only near strong oceanic fronts. Given the similar accuracy of the input data, the kriging variances reflect only the sampling density and provide no quality assessment for individual samples. However, minimizing the variance by tuning the variogram parameters generally reduces the difference between original and interpolated data set. A more sophisticated regional grouping of the sample data could avoid empirical variograms that show no sill, thereby opening a chance of automating the parameter adjustment. Then, VAK could be used as a convenient data-in/grid-out 'black-box' without the elaborate fitting of the variogram models.

(2) LOA is easily applied and yields good results in areas of dense sampling, especially in the well-sampled intermediate and high latitudes of the Atlantic Ocean, and in regions where zonal ocean currents match the method's preferential zonal spreading of information. This preference may cause artifacts where ocean currents are not predominantly zonal. For extrapolation beyond the sampled region, additional tie points are necessary, this would already be the case during the determination of the first guess. An adaptation of the influence radius to the local sampling density could lead to a better match of the interpolated and the original field.

(3) For gridding a global data set, all computations of distances and distance-related parameters must be carried out using spherical coordinates. The input data set should be subdivided into regions of similar spatial data dependencies. For VAK, this avoids mismatches between actual and modeled variability. For LOA, this allows search radii and an initial guess that are both adjusted to the spatial sampling density. Ocean communication masks are highly advisable, because they suppress unrealistic influences of the samples across the continental barriers.

(4) Given the different performances of both methods, we regard LOA as appropriate for work that is 'under way', for example, for the discussion of the (in part) large variations between the SST samples from different proxies and methods. If this discussion has eventually led to a consistent compilation of samples, VAK should be used for the gridding of the final MARGO SST data. Since both methods are problematic poleward of the sampled area, additional samples will greatly improve the results. These samples could be derived from proxy-based estimates of the seasonal ice covers.

References

Ababou, R., Bagtzoglou, A.C., Wood, E.F., 1994. On the condition number of covariance matrices arising in kriging, estimation, and simulation of random fields. Mathematical Geology 26 (1), 99–133.

CLIMAP Project Members, 1981. Seasonal reconstructions of the Earth's surface at the Last Glacial Maximum. Geological Society of America, Map and Chart Series MC-36, pp. 1–18.

Deutsch, C.V., Journel, A.G., 1992. GSLIB, Geostatistical Software Library and User's Guide. Oxford University Press, New York, Oxford.

Englund, E., Sparks, A., 1991. GEO-EAS 1.2.1 User's Guide. Technical report, United States Environmental Protection Agency.

Levitus, S., 1982. Climatological atlas of the World Ocean. NOAA Prof. Paper No. 13, 173pp.

Pannatier, Y., 1996. VARIOWIN: Software for Spatial Data Analysis in 2D. Springer, Berlin, Heidelberg.

Paul, A., Schäfer-Neth, C., 2004. How to combine sparse proxy data and coupled climate models. Quaternary Science Reviews, this issue, doi:10.1016/j.quascirev.2004.05.010.

Pebesma, E.J., Wesseling, C.G., 1998. GSTAT: a program for geostatistical modelling, prediction and simulation. Computers and Geosciences 24 (1), 17–31.

Reynolds, R.W., Smith, T.M., 1999. Improved global sea surface temperature analyses using optimum interpolation. Journal of Climate 7, 929–948.

Sarnthein, M., Gersonde, R., Niebler, S., Pflaumann, U., Spielhagen, R., Thiede, J., Wefer, G., Weinelt, M., 2003. Preface: Glacial atlantic ocean mapping (GLAMAP-2000). Paleoceanography 18, doi:10.1029/2002PA00769.

Schäfer-Neth, C., Paul, A., 2003. The Atlantic Ocean at the Last Glacial Maximum: 1. Objective mapping of the GLAMAP sea-surface conditions. In: Wefer, G., Mulitza, S., Ratmeyer, V. (Eds.), The South Atlantic in the Late Quaternary: Reconstruction of Mass Budget and Current System. Springer, Berlin, Heidelberg, pp. 531–548.

Shapiro, R., 1971. The use of linear filtering as a parameterization of atmospheric diffusion. Journal of Atmospheric Sciences 28, 523–531.

Taylor, S.P., Haywood, A.M., Valdes, P.J., Sellwood, B.W., 2004. An evaluation of two spatial interpolation techniques in global sea-surface temperature reconstructions: last Glacial Maximum and Pliocene case studies. Quaternary Science Reviews 23, 1041–1151.

Wackernagel, H., 2003. Multivariate Geostatistics, 3rd ed. Springer, Berlin, Heidelberg.

WOA, 1998. World ocean atlas 1998, http://www.nodc.noaa.gov/oc5/woa98.html. Technical report, National Oceanographic Data Center, Silver Spring, MD.

Wunsch, C., 1996. The Ocean Circulation Inverse Problem. Cambridge University Press, New York.