Aquarius Level 3 processing algorithm theoretical basis document. Version 0.9.

Jonathan M. Lilly and Gary S. E. Lagerloef

January 24, 2008

Abstract

A variety of methods for smoothing scattered data are reviewed in consideration for the Level 3 processing of the Aquarius data. It is argued that local polynomial smoothing has the best combination of features. The equations to be implemented are presented, and possible refinements are discussed.

1 Introduction

The Level 3 processing of Aquarius satellite data takes measurements at the boresight locations of the three radiometer beams (see Fig. 1), which have been already converted into physical units of salinity, and maps these onto a 1° grid. The Level 3 monthly product must satisfy the following requirements:

- The global RMS error is no greater than 0.2 psu.
- The noise decorrelation scale of the mapped field is $\sim$ 150 km.

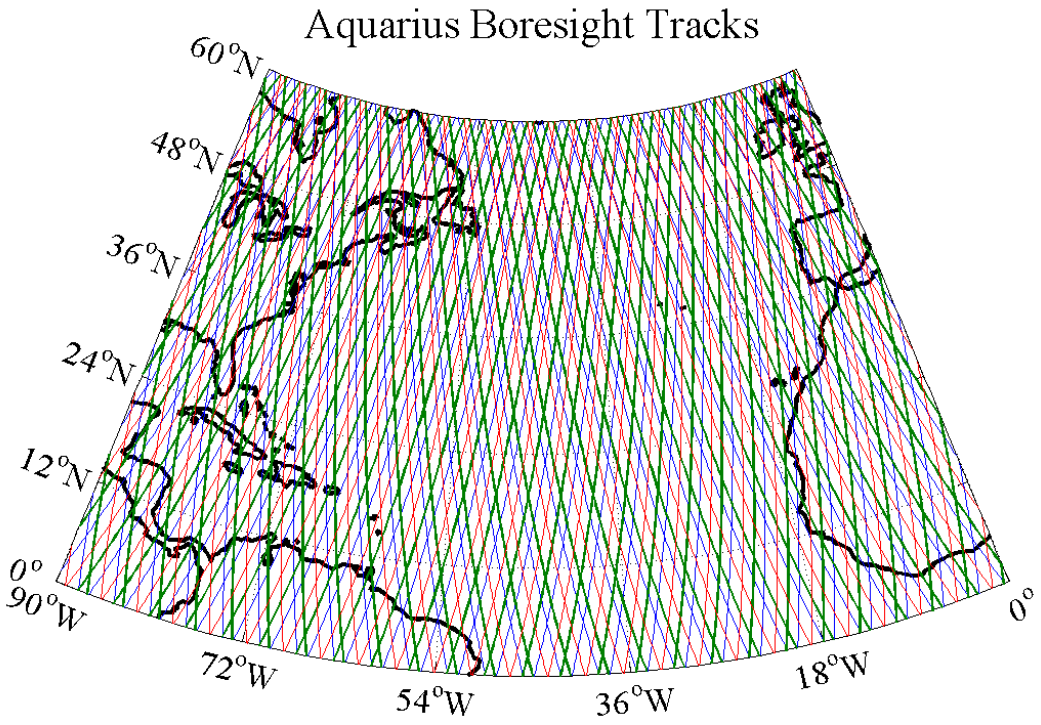

Figure 1: The Aquarius boresight tracks over the North Atlantic Ocean.

Of course, the simplest solution is simply to bin-average the data, leaving behind some "holes" at certain latitude/longitude locations. It is agreed that such a bin-averaged product should be produced. Obtaining a smoothed, spatially mapped field with no gaps is also desired. This problem is a subset of the problem of smoothing scattered data, which occurs frequently in oceanography and in particular in the treatment of satellite data. There are many solutions to this problem which differ both in philosophy as well as in appropriateness for particular settings. Here various approaches are reviewed in the light of the Aquarius mission requirements and sampling characteristics. Locally polynomial regression is recommended, as this is method offers explicit control over smoothness and locality, encompasses a range of possibilities from very simple to more refined, and is straightforward to both implement and to interpret.

2 Design considerations

In addition to the above requirements, we may note the following rule of thumb:

- The gridded data should be free from spurious structure.

the importance of which was suggested to us by M. Freilich. To the extent this goal may be realized, it will prevent misinterpretation of grid or mapping effects as representing physical phenomena. We may also demand that

- The mapped product should be accompanied by error estimates.

Earlier analysis suggested that, with realistic estimates of noise sources, the 0.2 psu can be met with even the simplest smoothing algorithm. The problem then becomes one of obtaining a mapped field, free from spurious structure, with a relatively small noise decorrelation scale. Significantly, we note later that the decorrelation scale of the sea-surface salinity (SSS) field itself is on the order of 500 km.

The Aquarius boresight pattern over the North Atlantic ocean is shown in Fig. 1. The relationship of the sampling pattern to a 1° grid is shown in Fig. 2, with each point representing a XX-second average. The sampling pattern is actually very tight, with most grid boxes having between one and four sample points. However, occasional grid boxes have zero sample points in them. The distribution of "empty" grid boxes exhibits latitudinal dependence owing to the peculiarities of the Aquarius sampling pattern; note that this pattern is not symmetric about the equator. Locations where two or even three beams coincide are seen. One of the challenges of the Aquarius Level 3 processing is to minimize the imprint of this spatially inhomogeneous sampling on the mapped product.

3 General problem

Let's say that we have observations g_n at data points x_n, $n = 1 \dots N$ and we wish to obtain an estimate $\widehat{g}(x)$ of the field $g(x)$. (Here, and in most of what follows, we consider a univariate field for simplicity.) There are several goals implicit in this problem: i) We wish to obtain estimates of the field $g(x)$ at locations x at which there are no data points; ii) we wish to smooth the observations in order to reduce the presumed noise in the observations, which we model as

$$g_n = g(x_n) + \epsilon_n(x_n) \tag{1}$$

where $\epsilon_n(x)$ is a discrete noise process of perhaps unknown statistics, which may depend upon the spatial location x; and finally iii) we wish to reconcile conflicting data points which reflect actual variability of $g(x)$ at smaller scales than those of interest. The problem is complicated by the fact that measurement locations x_n may be irregularly spaced. In a nutshell, we wish to smooth scattered data.

For the Aquarius satellite, each orbit is a sequence of latitude/longitude locations (ϕ_{nk}, θ_{nk}) $n = 1 \dots N$ observed at times t_n, where $k = 1, 2, 3$ is an index into the three different beams. We may take the observed quantity

$$g_{nk} = g(\phi_{nk}, \theta_{nk}) + \epsilon_{nk}(\phi_{nk}, \theta_{nk}) \tag{2}$$

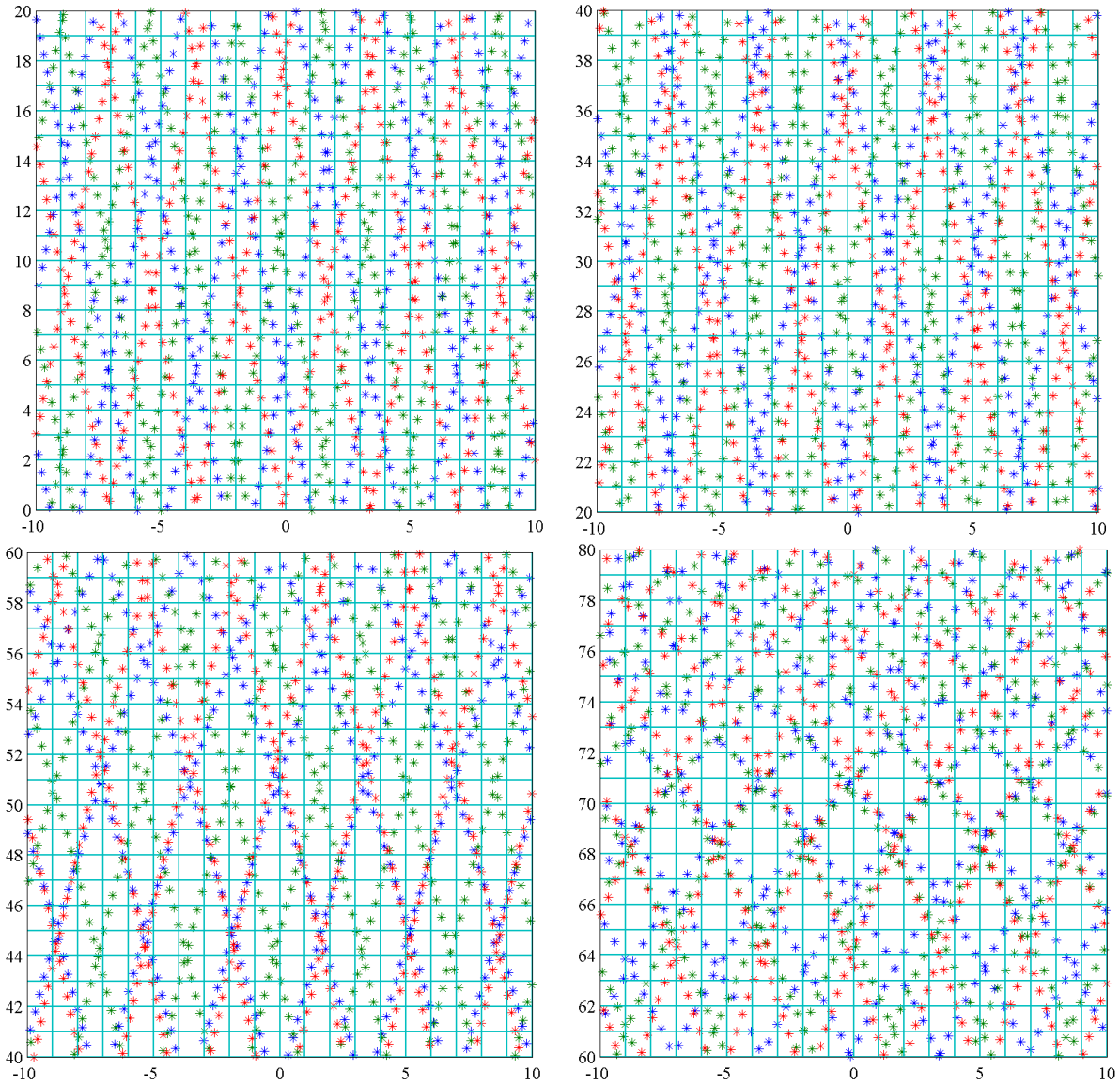

Figure 2: Northern hemisphere Aquarius boresight distribution with respect to an underlying 1°grid in various latitude bands.

to be the skin salinity $g(\phi, \theta)$ plus measurement noise, although of course this interpretation reflects a great deal of internal processing to generate a salinity signal from the radiometric measurements. Note that we let the noise process $\epsilon_{nk}(\phi_{nk}, \theta_{nk})$ be an explicit function of the beam number. Time dependence is neglected at the moment. Again we wish to smooth scattered data, but with additional complications of two-dimensionality and of multiple sensors.

4 Linear methods

There is a vast body of literature dealing with smoothing scatter data, since it is one of the key problems of statistical analysis. The majority of popular methods are linear in the data, which means the estimate $\widehat{g}(x)$ can be written as

$$\widehat{g}(x) = \frac{\sum_{n=1}^{N} g_n w_n(x)}{\sum_{n=1}^{N} w_n(x)} \tag{3}$$

for some weighting function $w_n(x)$ which perhaps depends upon the spatial location x. The simplest example of course is to choose

$$w_n(x) = w\left(|x - x_n|\right) \tag{4}$$

with $w(x)$ being, say, a Gaussian. Then (3) simply describes smoothing all available data with a Gaussian weight.

All the methods to be discussed here are linear in the data. This means that they can all be written as in (3). The different methods essentially amount to different means of choosing the weighting function $w_n(x)$. Thus these different methods are justified by different lines of argument for what constitutes the best weighting function.

In general, one may hope for two degrees of adjustability: *smoothness* and *locality*. One would like to control both how smooth or rough the resulting estimate $\widehat{g}(x)$ is, and also the radius $|x - x_n|$ over which observation points x_n influence the estimate at x. Not all methods permit adjustability of both of these degrees independently, and they generally differ in the ways in which these aspects are specified.

It is useful to mention some methods which, though not useful in practice, illustrate the continuum of possibilities. We may simply interpolate the data between observation points. This method leads to a highly localized estimate $\widehat{g}(x)$ at each point x, but involves no smoothing whatsoever and hence will be unsatisfactorily noisy. On the other hand, we may fit the entire dataset to a polynomial. This method is entirely global, and one controls the degree of smoothness by the order of the polynomial, with higher-order fits permitting rougher estimated fields $\widehat{g}(x)$. Global methods are not usually appropriate when one has spatially inhomogeneous structure in the underlying field $g(x)$.

5 Kernel estimators

The most straightforward method is simply to directly specify a fixed weighting function as in (4). Such methods are called linear smoothers, convolution estimators, or kernel estimators, and are reviewed by Fan and Gijbels (1996). The function $w(x)$ is generally written as

$$w(x) = K\left(x/h\right)/h \tag{5}$$

where $K(x)$ is a symmetric probability density function and is called the *kernel*, and the parameter h is called the *bandwidth*. As the bandwidth increases, the radius of influence increases and hence also does the number of observation points included in the estimate.

Kernel methods have strengths and weaknesses. Their primary strength is their simplicity. Since the smoothing kernel is fixed, one may obtain simple expressions for expected value and variance (Fan and Gijbels, 1996) in terms of the underlying field $g(x)$, the density of observation points $f(x)$, and the variance of the noise. Furthermore, with kernel methods we know precisely what has been done to the data, a very appealing property. The main drawback of such methods is that they do not allow smoothness and locality to be independently adjusted. Increasing the bandwidth simultaneously increases the smoothness and reduces the locality of the resulting estimate. This limitation arises because kernel methods are essentially the lowest-order member of local polynomial fitting, and locally fit the data to a constant (Fan and Gijbels, 1996). The missing degree of freedom would come from increasing the order of the polynomial fit, which decreases the smoothness with fixed locality.

Nevertheless, since they are the simplest method, kernel estimators provide a natural starting place. In earlier work, we applied the kernel estimator (using a 75 km standard deviation Gaussian) to synthetic Aquarius data, and found the agreement to be quite favorable. Our approach should only increase in complexity if it proves to be valuable in practice. Kernel smoothing also provides an extremely useful reference point for other methods. In order to understand other methods, we may cast them in the form (3), and demand to know to what smoothing kernel they correspond. In this way we can keep track of exactly what is being done to the data.

6 Optimal Interpolation

A very popular method in the oceanographic community is Optimal Interpolation (OI), known also as Gauss-Markov estimation and sometimes as Objective Analysis, which was introduced to oceanography by Bretherton et al. (1976). A straightforward treatment is given by McIntosh (1990). The idea of OI is to choose the weight function $w(x)$ such that the mean-square error

$$\mathcal{E}^2 = \sum_{n=1}^{N} \mathrm{E}\left\{|g(x_n) - \widehat{g}_n(x_n)|^2\right\} \tag{6}$$

is minimized; here "E" is the expectation operator. If the covariance of the field $g(x)$ between pairs of observation points x_n, x_m—written $\mathrm{Cov}\{g(x_n), g(x_m)\}$—is known, the covariance of the noise $\mathrm{Cov}\{\epsilon_n, \epsilon_m\}$ is known, and the cross covariance of the observed field and the noise $\mathrm{Cov}\{g(x_n), \epsilon_m\}$ is also known, then the form of weighting function $w(x)$ which minimizes (6) can be found. The method is also frequently used in a moving-window sense in which (6) is minimized over some local region.

In many situations it is the case that one has a sufficient history of observations, or else sufficient theoretical motivation, to confidently specify the data covariance function in a given region. It is also frequently the case that the noise is uncorrelated, and hence $\mathrm{Cov}\{\epsilon_n, \epsilon_m\} = \sigma^2\delta_{nm}$ where δ_{nm} is the Kronecker delta-function. Finally one often has the data and noise being independent such that $\mathrm{Cov}\{g(x_n), \epsilon_m\} = 0$. With these three conditions, the OI solution simplifies dramatically (McIntosh, 1990). The original application by Bretherton et al. (1976) was in an open-ocean case in which the data covariance could reasonably be prescribed on theoretical grounds.

The most desirable feature of OI is that, given the statistics of the data, the error, and their cross statistics, it yields the best linear estimator in the sense of having the smallest expected error. Its main difficulty is that the statistics upon which this interpretation hinges are often quite unknown. Additionally, strictly speaking, OI offers zero degrees of adjustability. It is intended to offer in a sense the correct answer.

Independent of the interpretation of the OI solution as the minimum-error solution for a given set of covariances, it is possible to adjust the solution by choosing different covariance functions. Specifying the (unknown) covariance function of the data also specifies the weighting function $w_n(x)$ of the linear estimate. Because of this, it is not necessary that the specified covariance function be a realistic representation of the true covariance function of the data in order for OI to yield pleasing results. However, if the covariance function is not known, it is preferable to leave aside OI and return to direct specification of the smoothing kernel.

There is a large literature both using OI and also analyzing its properties. McIntosh (1990) presented the theoretical relationship between OI and smoothing spline solutions, discussed subsequently. The sensitivity of the solution to changes in the specified covariance function was examined numerically by Franke (1991). Brankart and Brasseur (1996) used the "Generalized Cross-Validation" method of Wahba and Wendelberger (1980) to infer the unknown statistical information in an application to a Mediterranean Sea data set. A fast multiscale OI algorithm was presented by Menemenlis et al. (1997). Franke (1985) and Sokolov and Rintoul (1999) both compare OI with other methods.

The most relevant application of OI for the Aquarius satellite is that of Reynolds and Smith (1994). That paper describes using OI to generate global sea surface temperature maps on a 1° grid from a blend of satellite radiometer and in situ data. Those authors fit the covariance function of sea surface temperature observations to the anisotropic Gaussian model

$$C_{nm} = A \exp\left\{-\frac{(x_n - x_m)^2}{\lambda_x^2} - \frac{(y_n - y_m)^2}{\lambda_y^2}\right\} \tag{7}$$

where A is a normalizing constant. It was found that the zonal length scale λ_x is 500–1000 km, while the meridional length scale λ_y is 500–700 km; see Fig. 3. These large length scales arise because the sea-surface temperature is dominated by the gyre-scale patterns. One reason OI is used is that permits inclusion of a first-guess field. The radiometric sea surface temperature measurements suffer from substantial data dropout. The first-guess field is taken to be a complete map from the previous week, and this field is returned over locations for which there is no data during the current week.

OI is unlikely to be valuable to Aquarius for several reasons. Firstly, in contrast with Reynolds and Smith (1994), the large data dropout problem which OI is used to solve for sea surface temperature is not expected to be a major problem for Aquarius. Instead, the main problem will be locally fill in small "holes" in the sampling pattern. The Aquarius sampling pattern is actually quite dense, and therefore the gridding problem involves a sensible choice of small-scale smoothing. Secondly, the decorrelation lengths for sea-surface salinity—which will be comparable to those found by Reynolds and Smith (1994) for sea-surface temperature—are considerably larger than the decorrelation scale specified in the mission requirement. Finally, the OI analysis of Reynolds and Smith (1994) was implemented after a decade's worth of sea-surface temperature measurements from satellites. In contrast, sea-surface salinity measurements are considerably more sparse, and satellite measurements have never been carried out before. This means that there is a great deal of uncertainty in the correlation functions which are needed to complete the OI.

7 Smoothing splines

The idea of the smoothing spline solution (Wahba and Wendelberger, 1980; Gu, 2002) is to minimize the combination of the mean square error plus a roughness penalty. The quantity to be minimized is written as

$$\frac{1}{N}\sum_{n=1}^{N}\left|g_n-\widehat{g}_n(x_n)\right|^2+\lambda J\left(\widehat{g}\right) \tag{8}$$

where J is an operator which corresponds to a particular measure of smoothness, and λ is a smoothing parameter which controls the trade-off between solution error and smoothness. A typical choice of J is

$$J\left(\widehat{g}\right)=\int\left[\frac{d^2}{dx^2}\widehat{g}(x)\right]^2\,dx \tag{9}$$

in which case (8) expresses the trade-off between solution error and global-average curvature magnitude; in this case the solution is a cubic spline. The choice $\lambda=0$ then corresponds to exact interpolation, because the penalty on solution curvature has vanished, whereas $\lambda=\infty$ results in a linear regression model because the solution can have no curvature (Fan and Gijbels, 1996).

An important question involves what smoothing spline interpolation is actually doing to the data. It has been shown (Silverman, 1984, 1985) that the cubic spline corresponds approximately to a kernel smoothing with variable bandwidth. The expression will depend upon the probability density $f(x)$ of the sample points, which are taken to be randomly distributed. Specifically, for the sample size N large and the smoothing parameter λ relatively small, one may cast the solution in the form of the linear smoother

$$\widehat{g}(x)=\frac{1}{N}\sum_{n=1}^{N}g_n w_n(x) \tag{10}$$

with weighting function

$$w_n(x)\approx\frac{1}{f(x_n)h(x_n)}K_s\left(\frac{x_n-x}{h(x_n)}\right) \tag{11}$$

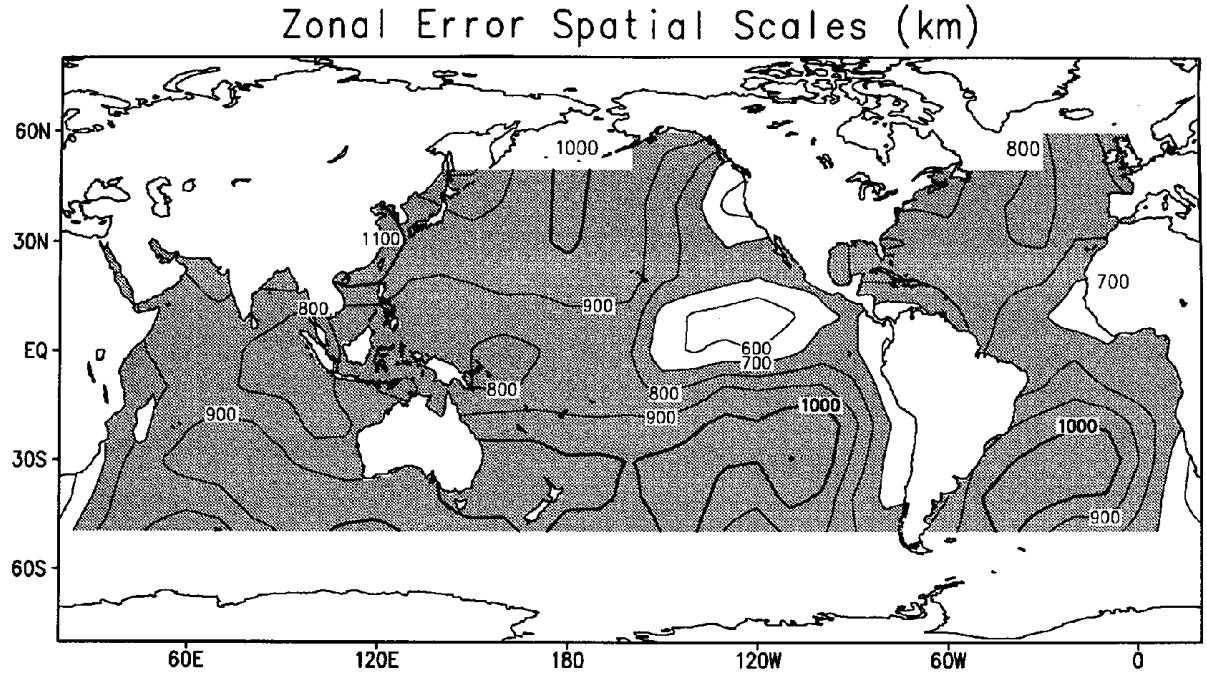

FIG. 14. The zonal error e-folding correlation scales, λ_x, derived from (6). The contour interval is 100 km. The 1000-km contour is a heavy line; values greater than 700 km are shaded.

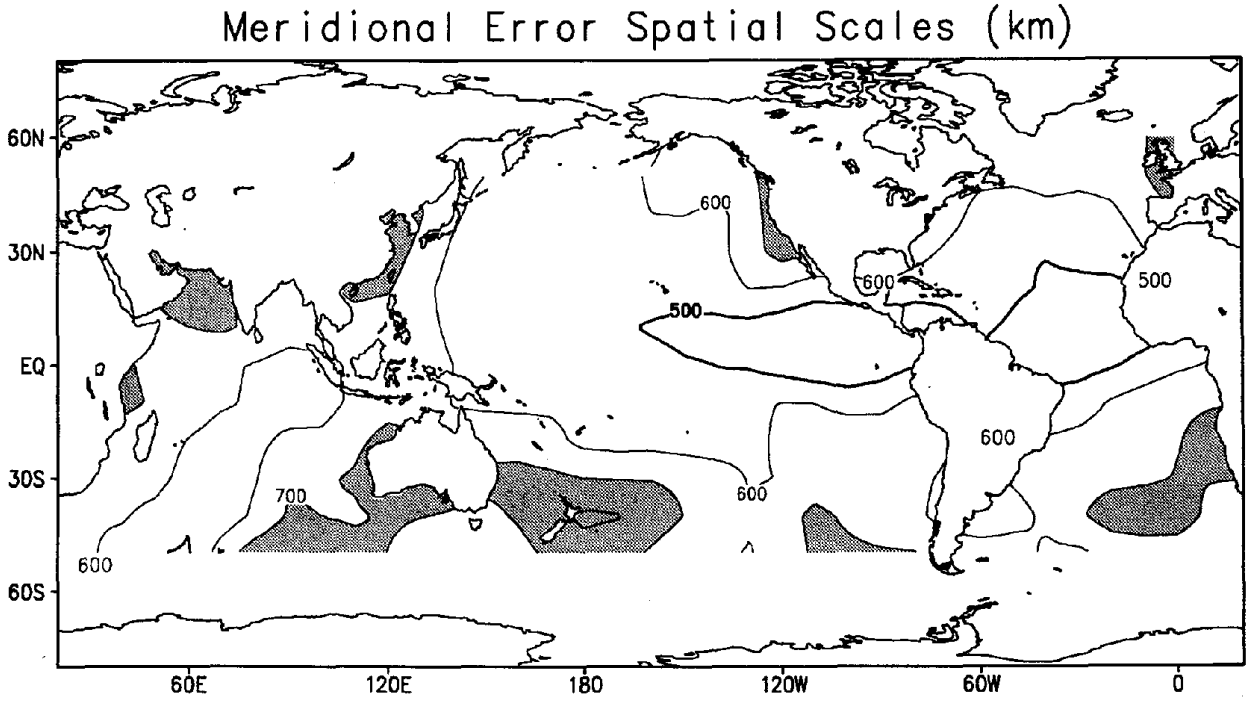

FIG. 15. The meridional e-folding error correlation scales, λ_y, derived from (7). The 500-km contour is a heavy line. Otherwise as in Fig. 14.

Figure 3: The zonal (upper) and meridional (lower) length scale in a Gaussian fit to the SST covariance function, from Reynolds and Smith (1994).

where the bandwidth and the smoothing kernel are given by

$$h(x_n) = \left[\frac{\lambda f(x_n)}{N}\right]^{1/4} \tag{12}$$

$$K_s(x) = \frac{1}{2} e^{-|x|/\sqrt{2}} \sin\left(|x|/\sqrt{2} + \pi/4\right) \tag{13}$$

as discussed in Silverman (1985) and Fan and Gijbels (1996). The most interesting feature is that the bandwidth varies spatially, and shrinks as the density of sample points increases.

Smoothing splines have a number of attractive properties, and have been frequently investigated and applied since the pioneering work of Wahba and Wendelberger (1980). The minimization of (8) can be given a statistical interpretation either in terms of a penalized likelihood method or a Bayesian framework (Gu, 2002). Varying the choice of operator J corresponds to different notions of "smoothness", with the practical consequence for the spectral response of the implied smoothing kernel, as discussed in McIntosh (1990) and Sokolov and Rintoul (1999). Much work has gone into methods for choosing the smoothing parameter λ, the most well-known of which is the method of generalized cross-validation (Wahba and Wendelberger, 1980) which corresponds to choosing λ in order to minimize an estimate of the mean square error. The smoothing spline method has also been generalized to spherical geometry by Wahba (1981). A long-standing issue has been computational cost; however, recent work on fast, efficient approximation (Kim and Gu, 2004) permits application to larger datasets. Smoothing splines and optimal interpolation, both being linear methods, are formally equivalent, and their connection was investigated in detail by McIntosh (1990).

Nevertheless, the smoothing spline method does not appear most well-suited for the Aquarius application. The degrees of freedom which are explicitly controlled by the smoothing spline method—namely, the choice of penalty operator J, and the smoothing parameter λ—do not appear to be the most useful parameters for this problem. The mission requirement of a 150 km decorrelation length means it is desirable to maintain explicit control over the bandwidth, which is only indirectly related to J and λ. Locality and smoothness are not controlled in a way which is convenient for the application to Aquarius.

Consider first penalty operator J. In some cases there is a natural choice for J; for example, if we seek to map the streamfunction $\psi(x, y)$, then minimizing the curvature has a physical interpretation in terms of the vorticity $\nabla^2\psi(x, y)$. Minimizing the curvature of the global salinity field has no such physical motivation, apart from being a convenient way to generate a smooth estimate. More importantly, the 150 km decorrelation mission requirement implies that λ is not free because increasing λ also increases the bandwidth. We can therefore not take advantage of methods for finding the optimal choice of λ; based on the discussion in the previous section, the optimal choice will likely imply a much larger bandwidth than the mission requirement.

There is another, more subtle issue. For a fixed choice of λ, the implicit bandwidth in the smoothing spline method automatically adjusts to account for differences in data density, as seen in (12). Thus, given the inhomogeneity of the Aquarius sampling grid, the effective radius over which the data will be smoothed will also vary spatially. Even if λ is chosen such that the mission requirement is satisfied, it is not immediately obvious that such spatially variable smoothing would be desirable. Rather, our initial impression is that the bandwidth should be kept spatially uniform in order to avoid the possibility of creating artifacts in the mapped field. The possibility of spatially nonuniform smoothing could be returned to later if more straightforward options prove unsatisfactory.

8 Local polynomial fitting

This brings us to local polynomial fitting (Fan and Gijbels, 1996), our method of choice. The idea here is to fit a Pth-order polynomial in the vicinity of each grid point x_m, $m = 1, 2, \ldots M$.

For data values g_n observed at locations x_n, $n = 1, 2, \ldots N$, this corresponds to minimizing

$$\sum_{n=1}^{N} \left| g_n - \sum_{p=0}^{P} \widehat{\beta}_p(x) \left[x_n - x\right]^p \right|^2 K_h(x_n - x) \tag{14}$$

at every grid point $x = x_m$, where $K_h(x) = K(x/h)/h$ is a decaying weighting function which depends upon the bandwidth h, with $K(x)$ being a probability distribution function. The regression coefficients $\widehat{\beta}_p(x)$, $p = 1, 2, \ldots P$ vary with spatial location, and are estimated at all grid point locations. The function $g(x)$ is estimated by the lowest-order coefficient, $\widehat{g}(x) = \widehat{\beta}_0(x)$, while higher-order regression coefficients estimate the derivatives of the field through $\widehat{g}^{(p)}(x) = p!\widehat{\beta}_p(x)$.

The adjustability of this method is primarily through the choice of bandwidth h, which may be either a constant or spatially varying, as well as through the order polynomial fit P. Increasing the order of the polynomial increases the degrees of freedom of the solution, reducing model bias at the expense of potentially increased variance due to under-smoothing the measurement noise. The $P = 0$ case corresponds to a constant fit, and the solution is equivalent to the kernel smoothing method presented earlier. This is immediately attractive for Aquarius because the degree of smoothness, specified by the polynomial order, is controlled independently from the bandwidth. The bandwidth may be specified, or alternatively, methods are available for choosing an optimal fixed or spatially varying bandwidth (Fan and Gijbels, 1996); however, it is not obvious that these will be consistent with the Aquarius mission requirements.

Various choices of weighting function $K(x)$ are also possible. It can be shown that an optimal choice is the parabolic function (Fan and Gijbels, 1996)

$$K_e(x) \;=\; \frac{3}{4} U(1 - |x|) \left[1 - x^2\right] \tag{15}$$

where $U(x)$ is the unit step function. The function $K_e(x)$, called the Epanechnikov kernel, is simply a parabola which descends from a maximum at the origin to a value of zero at $|x| = 1$. The Epanechnikov kernel is optimal in the sense that it minimizes the asymptotic mean square error of the resulting estimate. Another popular choice is the Gaussian weighting function, while some authors (e.g. Cleveland and Devlin, 1988) use a tri-cubic kernel proportional to $\left(1 - |x|^3\right)^3$. However, the exact choice of the weighting function does not appear to be particularly important for solution performance (Fan and Gijbels, 1996).

The simplicity of the method means that a broad range of theoretical results are readily available. For general order of fit P, it is straightforward to cast the solution in terms a linear smoother (3) with an equivalent kernel, in order that we may keep track of exactly what is being done to the data. The equivalent kernel depends upon the choice of $K_h(x)$ in (14), the fit order P, and the order of the coefficient $\widehat{\beta}_p$. The method also admits simple expressions for asymptotic bias, variance, and also confidence intervals of estimated field (Fan and Gijbels, 1996), a property that will be useful in eventually producing error maps associated with the salinity fields.

The "LOESS" method (Cleveland and Devlin, 1988) preferred by Chelton and Schlax (1994) is a variant of this class of smoothing method. Its main differences are a somewhat different notion of bandwidth (involving an nth nearest-neighbor consideration) as well as an iterative procedure intended to minimize the impact of statistical outliers. This iterative algorithm has advantages and disadvantages. It introduces complexity, and can lead to unpredictable behavior which cannot be captured by the simple theoretical expressions for error; but it can give practical advantages if outliers are indeed a problem.

The above discussion focuses on a univariate application, while of course for Aquarius we are mapping a surface. The generalization to this bivariate case is straightforward up to the

local linear fit (Fan and Gijbels, 1996), and can be extended to higher orders if necessary. For convenience, we use the Cartesian coordinates $\mathbf{x} = [x \;\; y]^T$ and ignore the effects of sphericity. We now replace the kernel function with

$$K_B(\mathbf{x}) \equiv K\left(\left|\mathbf{B}^{-1}\,\mathbf{x}\right|\right) \tag{16}$$

where $|\mathbf{x}| \equiv \sqrt{\mathbf{x}^T\mathbf{x}}$ denotes the Cartesian length of the vector $\mathbf{x}$, and $\mathbf{B}$ is a 2×2 matrix of unit determinant which permits anisotropic smoothing. For example, $\mathbf{B}$ may consist of a rotation of plus a differential stretching in order to preferentially smooth along one direction rather than another; $\mathbf{B} = \mathbf{I}$, with $\mathbf{I}$ being the identity matrix, is of course the isotropic case. The bivariate version of the Epanechnikov kernel is (Fan and Gijbels, 1996)

$$K_e(\mathbf{x}) \;\; = \;\; \frac{2}{\pi} U(1 - |\mathbf{x}|)\left[1 - |\mathbf{x}|^2\right] \tag{17}$$

which, as in the univariate case, is optimal.

With data values g_n observed at locations $\mathbf{x}_n \equiv [x_n \;\; y_n]^T$, $n = 1, 2, \ldots N$, the local linear fit is obtained by minimizing

$$\sum_{n=1}^{N} \left| g_n - \widehat{\beta}_0(\mathbf{x}) - \widehat{\beta}_x(\mathbf{x})\,[x_n - x] - \widehat{\beta}_y(\mathbf{x})\,[y_n - y] \right|^2 K_{\mathbf{B}}(\mathbf{x}_n - \mathbf{x}) \tag{18}$$

at every grid point $\mathbf{x} \equiv [x \;\; y]^T = \mathbf{x}_m$, $m = 1, 2, \ldots M$. The estimated two-dimensional surface $\widehat{g}(\mathbf{x})$ is then given by $\widehat{g}(\mathbf{x}) = \widehat{\beta}_0(\mathbf{x})$, and the gradient vector $\nabla g(\mathbf{x})$ is directly estimated as $\widehat{\nabla g}(\mathbf{x}) = [\widehat{\beta}_x(\mathbf{x}) \;\; \widehat{\beta}_y(\mathbf{x})]^T$. The solution to the locally linear two-dimensional smoothing problem is known immediately from least squares theory and can be cast conveniently in matrix form. Define the $N \times 3$ matrix

$$\mathbf{X}(\mathbf{x}) \equiv \begin{bmatrix} 1 & x_1 - x & y_1 - y \\ 1 & x_2 - x & y_2 - y \\ \vdots & \vdots & \vdots \\ 1 & x_n - x & y_n - y \end{bmatrix} \tag{19}$$

together with the $N \times N$ weight matrix $\mathbf{W}(\mathbf{x}) \equiv \mathrm{diag}\,\{K_{\mathbf{B}}(\mathbf{x}_n - \mathbf{x})\}$, the data vector $\mathbf{g} \equiv [g_1 \;\; g_2 \;\; \ldots g_N]^T$, and the solution vector $\boldsymbol{\beta}(\mathbf{x}) \equiv [\widehat{\beta}_0(\mathbf{x}) \;\; \widehat{\beta}_x(\mathbf{x}) \;\; \widehat{\beta}_y(\mathbf{x})]^T$. The solution is then given by (**??**)

$$\boldsymbol{\beta}(\mathbf{x}) = \left(\mathbf{X}^T\mathbf{W}\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{W}\mathbf{g} \tag{20}$$

where the superscript "-1" denotes the usual matrix inverse; the functional dependence on location $\mathbf{x}$ is suppressed on the right-hand side for clarity. Implementing this at every grid point $\mathbf{x} = \mathbf{x}_m$, $m = 1, 2, \ldots M$, we obtain the desired estimate of the mapped field and its gradient.

Because of a variety of factors—its simplicity, the direct adjustability of order and bandwidth, the availability of theoretical results for bias and variance, and the existence of a variety of more advanced extensions—the local polynomial smoothing seems the ideal method for Aquarius. We will begin with the locally linear model (18) with isotropic, spatially fixed variance and with options for both the Epanechnikov and Gaussian kernels. Initially, we will investigate the advantage gained by the local linear fit with respect the locally constant fit implied by the kernel smoothing method (3). The role of varying bandwidth within the constraints of the mission requirement, and the differences between the two kernels, will be explored. The following modifications are available for the future, should they be needed: extension to a local quadratic fit; anisotropic or spatially varying smoothing; and robustification along the lines of the LOESS method.

9 Other methods

We mention in passing some other methods.

9.1 Radial basis functions

Another type of solution involves the use of "radial basis functions", such as the "multiquadric-biharmonic" (MQ-B) method advocated by Nuss and Titley (1994). Despite its complicated name, this method is conceptually easy to understand. At each observation point, one places an azimuthally symmetric decaying function. These radial basis functions are weighted by coefficients which have yet to be determined. The global field is simply a superposition of contributions from the different data points, spatially extended by the radial basis functions. The coefficients are found by forcing agreement between the global field at the observation points and the observed values. If the data are noisy, then a certain amount of mismatch may be incorporated into the model, by introducing a "smoothing parameter" and estimating the noise variance (Nuss and Titley, 1994).

The family of radial basis function methods consists of different choices for what decaying function to choose as a basis, together with means for deciding on the value of the smoothing parameter. In a comparison of different methods, Sokolov and Rintoul (1999) find good results using MQ-B for the analysis of hydrographic data, though those results are likely highly dependent upon the particular application.

The main advantage of this method is that from any set of sparse measurements, one obtains a smoothed field which is known everywhere. That is, once the coefficients are obtained—which depend only on the observation points—the solution everywhere is also known. A disadvantage as far as Aquarius is concerned is that obtaining a suitably smooth map becomes tied to the problem of estimating the noise. Given the Aquarius sampling, it would seem that this method would lead to results which are rather "bumpy", as resolution is concentrated where the data density is high. In any case, the method of radial basis functions do not appear to have as broad applicability to the noisy data as do smoothing splines and local polynomial modeling.

9.2 Spatio-spectral localization

An interesting global method is the spatio-spectral localization solution of Simons et al. (2006). Those authors revisit the time/frequency optimization problem of Slepian (1978) for the case of spherical geometry. One may construct a set of eigenfunctions for the sphere which are simultaneously localized in a given spatial region and a given spectral (i.e. spherical harmonic) band. Increasing the number of included eigenfunctions leads to a more uniform coverage over the target region, but with the trade-off of additional leakage. An example is shown in Fig. 4 for simultaneous localization in a spectral band and in the land regions of the earth's surface. The overshoot away from the continents is a manifestation of Gibbs' effect.

Inverting the land solution leads to a set of eigenfunctions localized in the ocean; one could also create eigenfunctions for the ocean fraction covered by the Aquarius satellite. These eigenfunctions can become a smoothing method by simply projecting scattered data onto them. This method is however inappropriate for mapping Aquarius at Level 3 because it is completely global – a given data point could affect the solution everywhere on the globe. A fine control over the degree of smoothness can be obtained through the choice of spectral band together with the number of included eigenfunctions, but the eigenfunctions are by nature global functions. Spatio-spectral localization could nevertheless be an interesting way to analyze Aquarius data during earlier levels of processing.

9.3 Spherical wavelet basis

Another interesting method (Holschneider et al., 2003) decomposes the data set onto a local wavelet basis appropriate for two-dimensional data on the sphere. The idea of this method is to concentrate resolution where data density is high, which can be of great value for data which

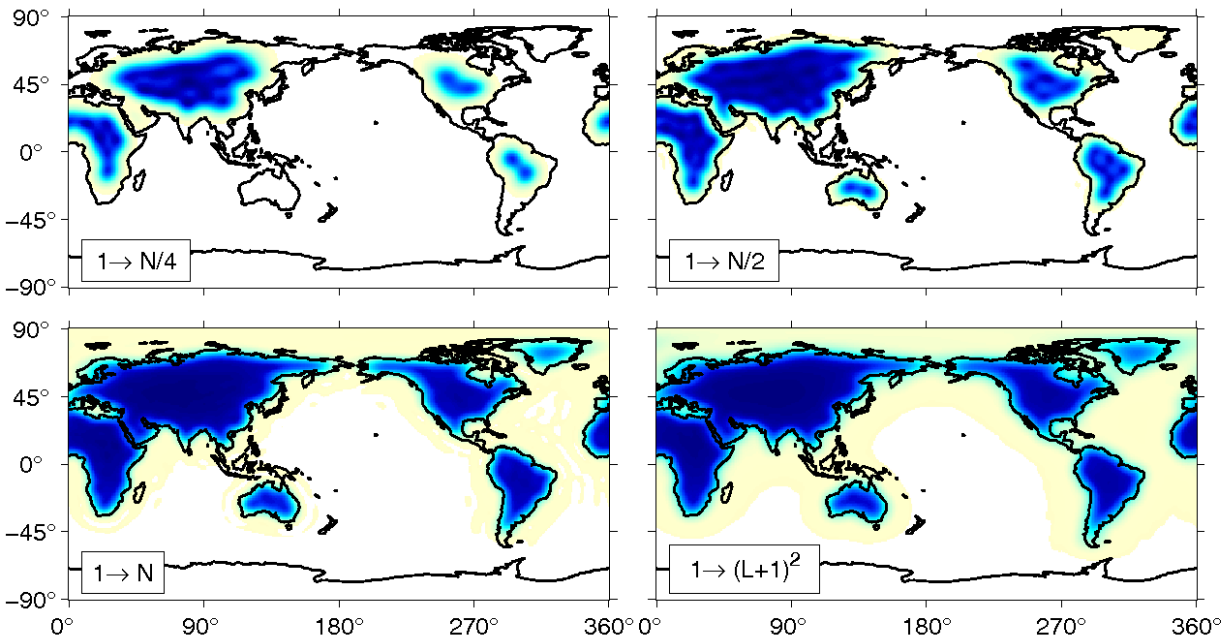

Figure 4: Total energy for a set of spectrally and spatially localized basis functions on the sphere, as the number of included basis functions increases. Coverage of the target region, the land surfaces, increases as more eigenfunctions are included, as does leakage into the ocean region. Taken from Simons et al. (2006).

exhibits inhomogeneous sampling patterns. By contrast, global projection methods have the same resolution everywhere – and hence lead to artificial structure in data-poor regions. The wavelet themselves also have an interesting physical interpretation as representing the solution of the Laplace equation to multipole sources within the sphere (Holschneider et al., 2003).

Philosophically, this is the opposite of what we seek for the Aquarius mapping problem, which is to obtain uniform spatial resolution despite spatially inhomogeneous sampling. While it may be possible to set some coefficients to zero in order to control the smoothness of the resulting field, this appears to be a not particularly explicit means of achieving our goals.

10 Summary

Of the various methods for the smoothing scattered data that are in common use, the most appropriate for the Aquarius gritting problem seems to be the method of local polynomial modeling. This method permits explicit control the bandwidth, which is directly to the decorrelation scale, and also permits smoothness to be independently controlled by the order of the fit. The "LOESS" method (Cleveland and Devlin, 1988) preferred by Chelton and Schlax (1994) is a member of this class. An attractive feature here is the ability to increment complexity as necessary. We may proceed from a locally constant fit, to a locally linear fit, to a higher-order fit; we may introduce an iteration for robustness in the presence of outliers if required; and we may move from a spatially constant isotropic smoothing window to an automatically adjusting or anisotropic smoothing window if this is seen to lead to an advantage. This method permits us to keep track of exactly what has been done to the data through simple expressions for equivalent smoothing kernels, a feature which should help to guard against the misinterpretation of smoothing effects as being physical structure. Finally, theoretical expressions for bias and variance are also available, which will be useful and eventually producing error maps along with the salinity field.

References

Brankart, J.-M. and Brasseur, P.: Optimal analysis of in situ data in the western Mediterranean using statistics and cross-validation, J. Atmos. Ocean Tech., 13, 477–491, 1996.

Bretherton, F. P., Davis, R. E., and Fandry, C. B.: A technique for objective analysis and design of oceanographic experiments applied to MODE-73, Deep-Sea Res., 23, 559–582, 1976.

Chelton, D. B. and Schlax, M. G.: The resolution capability of an irregularly sampled dataset: with application to Geosat altimeter data, J. Atmos. Ocean Tech., 11, 534–550, 1994.

Cleveland, W. S. and Devlin, S. J.: Locally weighted regression: an approach to regression analysis by local fitting, J. Acoust. Soc. Am., 83, 596–610, 1988.

Fan, J. and Gijbels, I.: Local polynomial modelling and its applications, Chapman and Hall, 1996.

Franke, R.: Sources of error in objective analysis, Mon. Wea. Rev., 113, 260–270, 1985.

Franke, R.: Sensitivity of the error in multivariate statistical interpolation to parameter values, Mon. Wea. Rev., 119, 815–832, 1991.

Gu, C.: Smoothing spline ANOVA models, Springer, 2002.

Holschneider, M., Chambodut, A., and Mandea, M.: From global to regional analysis of the magnetic on the sphere using wavelet frames, itit, 135, 107–124, 2003.

Kim, Y.-J. and Gu, C.: Smoothing spline Gaussian regression: more scalable computation via efficient approximation, J. Roy. Stat. Soc. B Met., 66, 337–356, 2004.

McIntosh, P. C.: Oceanographic data interpolation: objective analysis and splines, J. Geophys. Res., 95, 13 529–13 541, 1990.

Menemenlis, D., Fieguth, P., Wunsch, C., and Willsky, A.: Adaptation of a fast optimal interpolation algorithm to the mapping of oceanographic data, J. Geophys. Res., 102, 10 573–10 584, 1997.

Nuss, W. A. and Titley, D. W.: Use of multiquadric interpolation for meteorological objective analysis, Mon. Wea. Rev., 118, 1611–1631, 1994.

Reynolds, R. W. and Smith, T. M.: Improved global sea surface temperature analyses using optimum interpolation, J. Climate, 7, 929–948, 1994.

Silverman, B. W.: Spline smoothing: the equivalent variable kernel method, Ann. Stat., 12, 898–916, 1984.

Silverman, B. W.: Some aspects of the spline smoothing approach to non-parametric regression curve fitting, J. Roy. Stat. Soc. B Met., 47, 1–52, 1985.

Simons, F. J., Dahlen, F. A., and Wieczorek, M. A.: Spatiospectral localization on a sphere, siam, 48, 504–536, 2006.

Slepian, D.: Prolate spheriodal wave functions, Fourier analysis, and uncertainty– V: The discrete case, Bell Syst. Tech. J., 57, 1371–1430, 1978.

Sokolov, S. and Rintoul, S. R.: Some remarks on interpolation of nonstationary oceanographic fields, J. Atmos. Ocean Tech., 16, 1434–1449, 1999.

Wahba, G.: Spline interpolation and smoothing on the sphere, 2, 5–16, 1981.

Wahba, G. and Wendelberger, J.: Some new mathematical methods for variational objective analysis using splines and cross validation, Mon. Wea. Rev., 108, 1122–1143, 1980.