A METHOD FOR RECONSTRUCTING BOTTOM TOPOGRAPHY FOR AN ENCLOSED BASIN FROM SCATTERED, SPARSE MEASUREMENT DATA

Alexander F. Shchepetkin[1,2], Olga S. Volodko[3,4]

[1]*P. P. Shirshov Institute of Oceanology, Russian Academy of Sciences,*
36, Nahimovskiy prospekt, Moscow, Russia, 117997
[2]*Moscow Institute of Physics and Technology, 9 Institutskiy per., Dolgoprudny,*
Moscow Region, 141700, Russia, e-mail: `shepetkin.af@ocean.ru`
[3]*Institute of Computational Modeling SB RAS, Krasnoyarsk, Akademgorodok, 660036, Russia*
[4]*Siberian Federal University, Krasnoyarsk, 79 Svobodny pr., 660041, Russia*
e-mail: `osv@icm.krasn.ru`

Numerical simulation of circulation and internal waves in a basin requires the knowledge of bottom topography, defined as a continuous and continuously differentiable field (unless there are known features of the relief to justify the opposite), which is, unfortunately, not always available with sufficient resolution and coverage. In this article we review existing techniques for producing regularly gridded field from scattered bathymetry data – in our case raw field data measured by a boat equipped with an echo sounder and GPS – and propose a new one, which we believe is the most optimal for this situation. The technique essentially goes along the line of approach of Sandwell (1987) using Green functions to construct biharmonic spline interpolation, which we augment by adding coastline and introduce special preprocessing of measured data to identify and eliminate (by averaging out) potentially contradictory and unreliable measurements which may cause spurious oscillations of biharmonic spline.

Introduction

The subject of this study is lake Shira, located at $54°30'38''$N, $90°12'09''$E in Republic of Khakassia, Russia. The lake is approximately 9.5*km* long and 5*km* wide, it is salty, and, despite the fact that it is relatively shallow, $\sim 25m$ deep, it is stratified in both temperature and salinity. During the last several years a significant, multidisciplinary research program was undertaken, including observations and measurements of physical and biological/ecosystem processes. Historically the lake was classified as meromictic, however in fall 2014 there were observations of complete mixing of water mass throughout the entire water column, which was not observed in all several decades. So one the primary concerns is the rate of exchange of its hypolimnion water with the upper layers, and ultimately, surface. Velocity measurements using ADCP show the presence of detectable internal waves (Kompaniets et al., 2017), which is expected to provide a more efficient vertical mixing mechanism than diffusion (Lemckert and Imberger, 1998; Wüest

and Lorke, 2003). A more recent analysis *in situ* data established pattern of circulation (Volodko et al., 2017). In summer 2018 GPS-tracked drifters were used to further study circulation. In its turn, these developments stimulate the need for numerical modeling of lake Shira, echoing some past experience of monitoring and modeling ecological impact on lake Baikal decades earlier (Penenko , 1989), and more advanced in sea of Azov (Shabas , 2014).

One of the first task on the path to hydrodynamic numerical simulation is to set up suitable bathymetry, which implies having is as a continuous 2D field, realistically accurate, but should respecting the necessary precautions associated with terrain-following coordinates (Beckmann and Haidvogel, 1993; Shchepetkin and McWilliams, 2003). The bathymetry of lake Shira was measured in summer 2010 from a boat equipped with GPS and an echo sounder, traversing the lake by making 9 nearly parallel passes in NE-SW direction, approximately $1km$ apart from each other, as well as some randomly placed measurements along the way, Fig. 1. The coastline for lake Shira is available from ESRI (a shape file `.shp` in its proprietary format), it contains 578 data points, and it provides resolution in tens of meters, which is more than sufficient for our goal. The area around the lake, and presumably, its bottom relief are characterized by rather gentle topography.
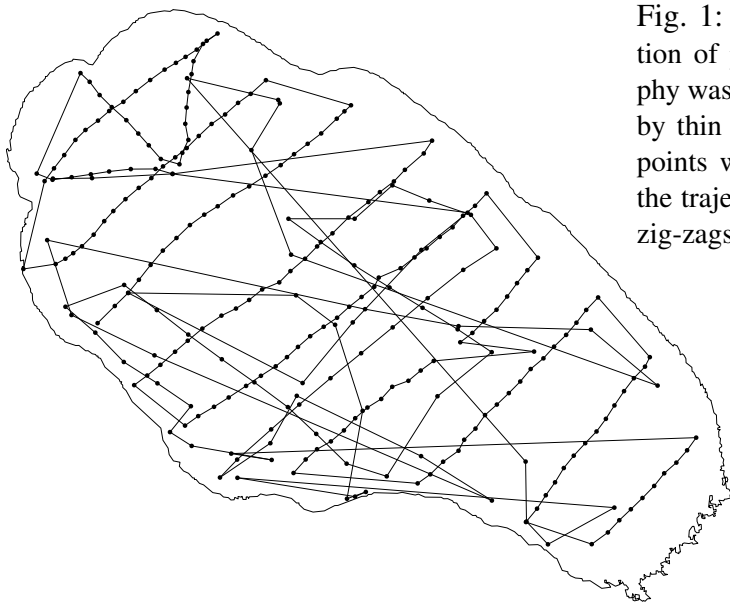

Fig. 1: Coastline of lake Shira and location of points where the bottom topography was measured. The dots are connected by thin line which is the sequence of the points within the file - presumably along the trajectory of the boat, but some erratic zig-zags and crossings are hard to explain.

The task of reconstructing continuous field from irregularly scattered point data is very common in numerical modeling and engineering, resulting in abundance of numerous methods. Yet, there is no tendency to converge toward one or few universal approaches. This is mostly explained by the speculative nature of the problem – essentially, in addition to the available data one should utilize some *a priori* knowledge about the properties of the object being modeled, and use it as a guideline to select or construct an algorithm. From this moment the problem can no longer be viewed as purely mathematical interpolation, but rather as making a physical model. In mathematical terms this is often formulated through a variational principle: postulation the desired properties of the field to be reconstructed, and expressing it in terms of an integral functional which needs to be minimized. It is equivalent to formulation a *goal* or *penalty* function, and, because

this can be done in a non-unique way, this leads to variety of methods.

Overview of existing techniques

Perhaps by far the most common technique to handle interpolation of scattered data is Delaunay triangulation followed by some interpolation algorithm – in the simplest case linear within each triangle, but there are other options as well. They are readily available in Matlab as `griddata` function, so it would be natural to explore them first.

An immediate restriction of triangulation-based methods is that they cannot provide any values outside the convex area connecting the points, and because there are too few measurements near the coast, so it is natural to augment the available bathymetry data with points on the shoreline where the depth is presumed to be zero (or set to a user-specified minimum value). At this stage we made an algorithm to scan the entire combined data set to identify and exclude potentially contradictory data: either because of GPS errors or echo sounder errors, when the depth is measured again in nearly the same location (perhaps a later time, or on different/return pass by the boat), the measured value does not repeat itself exactly. Such data entries were averaged (including their coordinates and measured values) and merged into single point. In addition, one measured point ended up on land and was discarded. The placement of combined data is shown in Fig. 2. Once the data is extended, triangulation can cover the entire lake.

A note of caution related to the selection of an appropriate coordinate system in which interpolation is performed: because `griddata` interprets its input arguments as Cartesian coordinates, it computes distances simply as

$$d = \sqrt{(x_j - x_k)^2 + (y_j - y_k)^2},$$

and uses them for deciding which points are connected into triangles. As the result, an innocent-looking

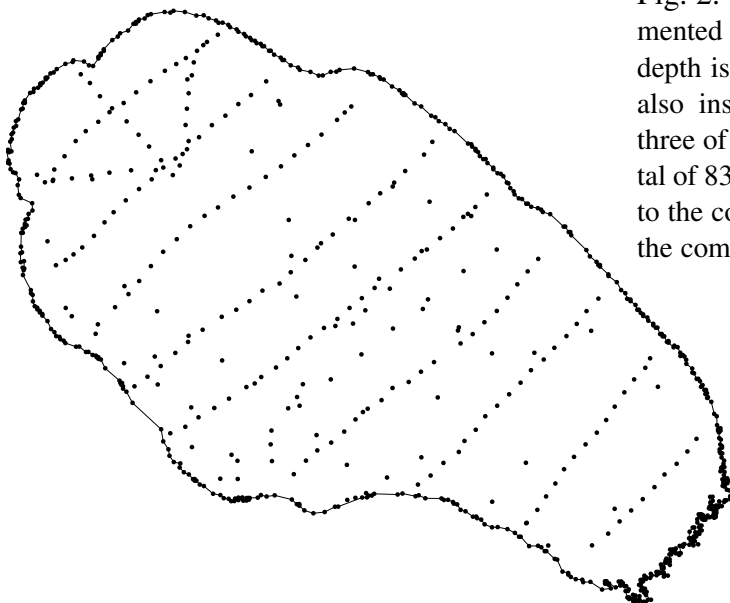$$\text{zout} = \text{griddata}(\text{zlon}, \text{zlat}, \text{z}, \text{rlon}, \text{rlat});$$

Fig. 2: Topographic data from Fig. 1 augmented by the coastline points, where the depth is expected to vanish. The data was also inspected for coincident points and three of them were removed. There are total of 839 points here, 578 of which belong to the coastline. This data was used for all the comparisons performed in this section.
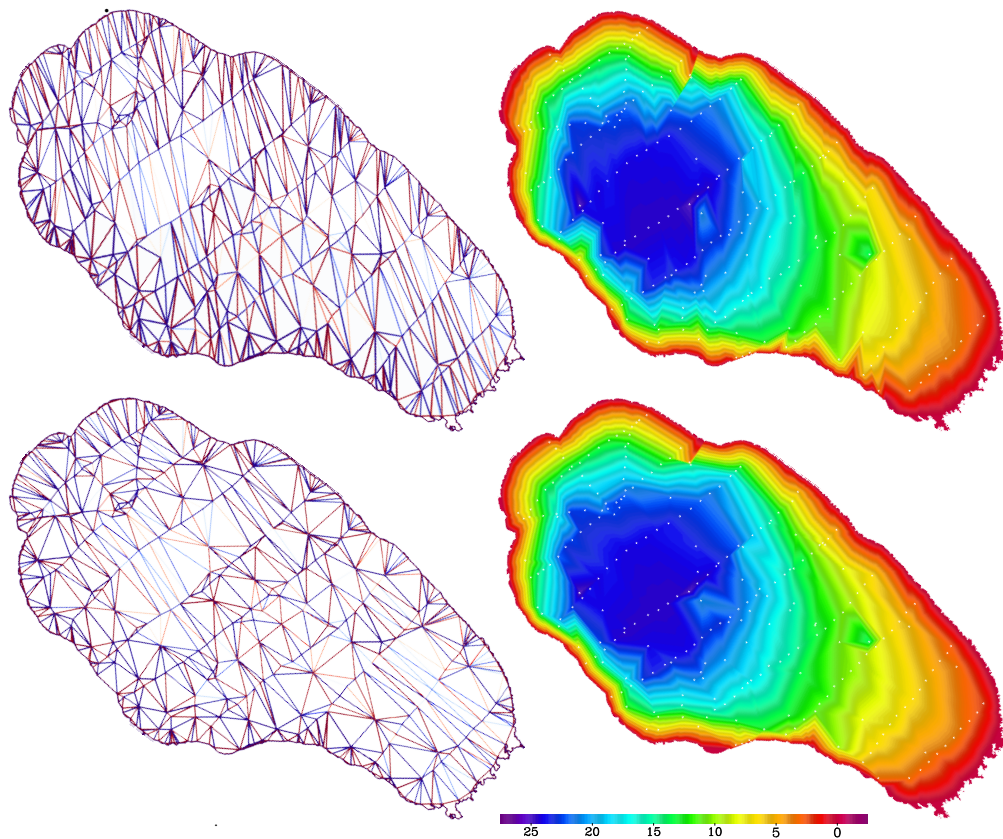
Fig. 3: Triangulation (*left column*) and bottom topography (*right column*) of lake Shira reconstructed from scattered data by linear interpolation within each triangle shown on the left side in the corresponding row. *Upper row* – applying Matlab function `griddata(...,'linear')` directly in geographical lat-lon coordinates; *lower row* – after conformally transforming coordinates to flat-plane Cartesian system. White dots in each panel on the right indicate the location of the actual data points.

results in a very erroneous triangulation, if `zlon,zlat` are geographical longitude and latitude coordinates of scattered data points expressed in degrees; and so do `rlon, rlat` – longitude and latitude of the nodes of regular grid. This is illustrated in Fig. 3. Although this seems to be obvious – e.g., at $60°N$ the distance in north-south direction corresponding to $1°$ increment in latitude is twice as much to the distance in east-west direction corresponding to $1°$ increment in longitude – there are still numerous Matlab scripts circulating within the ROMS modeling community where `griddata` has is used directly in *lon-lat* coordinates in such manner. The remedies are readily available: An intuitive way to correct the situation is to conformally transform the horizontal coordinates (both ROMS grid and data points) into flat Cartesian coordinates: in the simplest case of very small domain just to multiply longitude by cosine of the median latitude; a more diligent way would be to use Lambert conformal conical projection with its two standard latitudes optimally chosen for the particular modeling domain (minimum distortion). For large grids (the size of the domain becomes comparable with the radius of the Earth) the distances can no longer be accurately compared on a conformal projection map, however

it is possible to create ROMS grid with perfectly isotropic resolution: local metric factors $pm = pn$ everywhere, despite the fact both pm and pn change significantly along the entire grid. In this case a natural approach is to use grid-index coordinates of ROMS grid, and, correspondingly, express the locations of measured data points in terms of indices on ROMS grid and fractional parts of indices for the precise location within each cell. In essence, this relies on the property of conformal grids which locally are almost Cartesian.

While the version of interpolated topography shown the bottom row of Fig. 3 is technically correct, it is clearly not acceptable for the purpose of numerical modeling: it is non-smooth, has jagged contour lines, and some of its features – notably the kink next to coastline close to the upper-left corner are questionable and potentially spurious. Besides `"linear"`, `griddata` supports three other potentially viable options: `"cubic"`, `"natural"`, and `"v4"`. Their outcomes are shown in Fig. 4.

In each row shown on the right is point-wise Laplacian computed from the field on the left. The absolute scale does not matter, but the scale in exactly the same on all plots presented here. Computing point-wise Laplacian is useful to expose discontinuities of first and second derivatives, and this way evaluate the quality of interpolation. Also it is possible to track down both resolution/placement of the initial source data, as well as the algorithm itself (if unknown). Thus, in the case of `"linear"`, Laplacian vanishes is exactly inside each triangle, but because of the discontinuity of the first derivatives at the edges, computing discrete Laplacian yields large values there, exposing the triangulation itself, with color intensity of line proportional to the discontinuity in the first derivatives (essentially angle of bent between to flat facets).

In the case of `"cubic"`, the facets are no longer flat, but the triangulation lines are still clearly visible. This indicates discontinuity of the first derivatives. Overall there is no noticeable improvement in comparison with `"linear"`: the contour lines are still jagged sharply at the lines connecting discrete points.

The natural neighbor (`"natural"`) is another triangulation-based algorithm, but it produces no triangulation lines in its Laplacian, and the resultant field is continuously differentiable, except at the data points themselves. One obvious artifact is that that the data points act like tension points resulting in reverse curvature of the contour lines between the points, and sharp kinks at the points themselves (a very similar pattern would occur if one solves Laplacian equation $\nabla^2 f = 0$ everywhere, except a set of discrete points $f_k = f(x_k, y_k)$, where the function is forced to have the prescribed value). All nine ship tracks in NE-SW direction are clearly identifiable in Laplacian, and in the bathymetry itself.

The last option, `"v4"`, is biharmonic spline interpolation. It produces continuous and continuously differentiable field, which is smooth everywhere including at the data point locations. Its drawback is that, in comparison with other methods, and to a much stronger degree than they, it has tendency to produce artificial bumps and holes, which we believe are spurious. Thus, on the bottom-left panel in Fig. 4, the bump next to the northern coast, and the hole in deep area next to it, are more pronounced than in all other plots, and close examination of placement of the data points (indicated by white dots in all panels in this figure) reveals that neither bump, nor hole have any measurements inside to confirm that the extrema are real. Similar, although not as dramatic, observations can be made about some other extrema found in this panel.
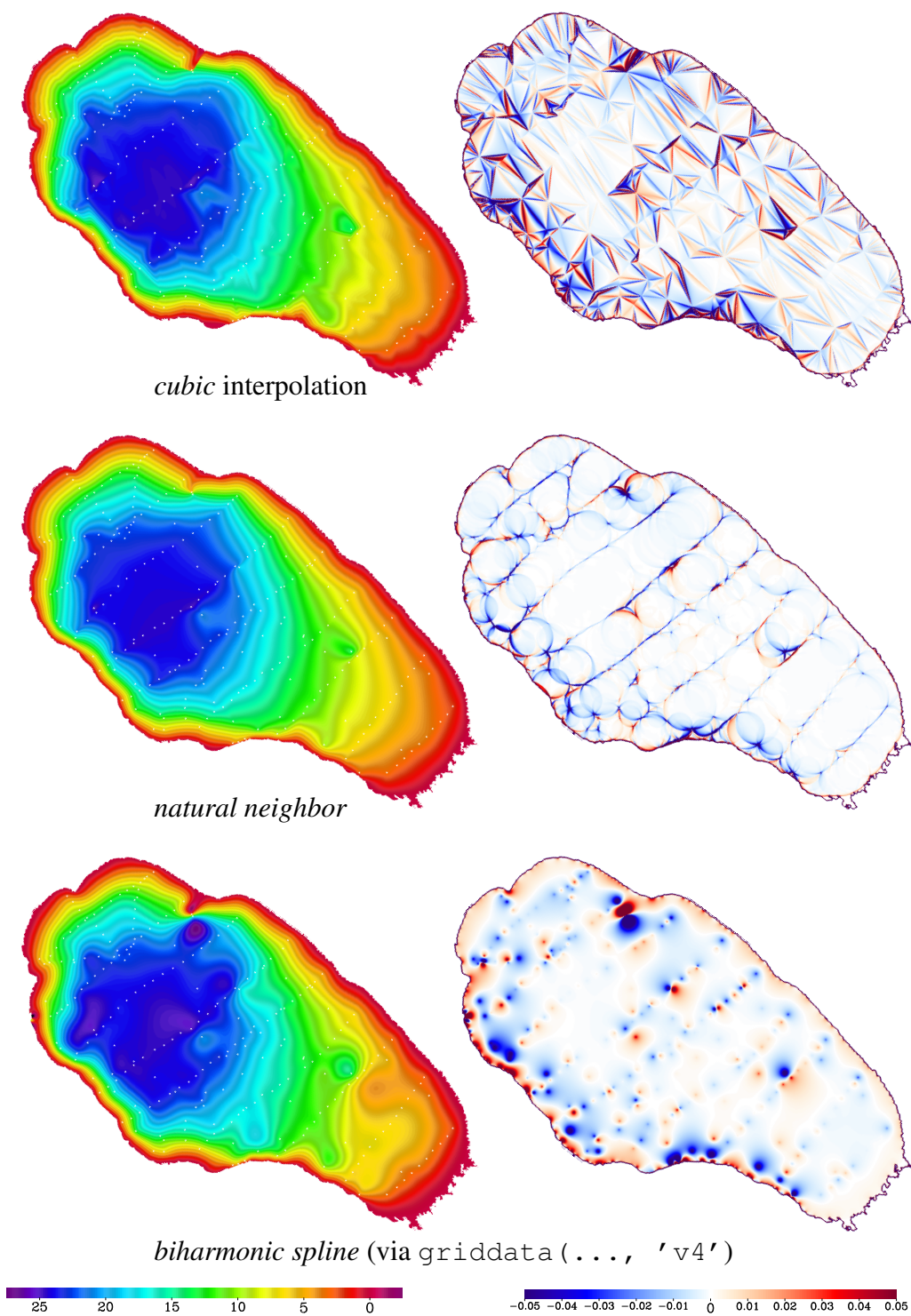
71

cubic interpolation

natural neighbor

biharmonic spline (via `griddata(..., 'v4')`)

Fig. 4: Comparison of commonly available techniques to interpolate scattered data onto regular grid: *left column*: resultant bathymetry; *right:* point-wise Laplacian of the field on the left shown here as a measure of its smoothness. White dots in the left panels indicate the location of the actual data points.

So far, none of the techniques considered by us produced an acceptable version of bottom topography suitable for hydrodynamic modeling of lake Shira. In the next section we examine this algorithm in more detail with the intent to understand the causes of this behavior, and ultimately to eliminate them.

Biharmonic spline interpolation

The last technique shown on Fig. 4, biharmonic spline interpolation, was successfully applied by Sandwell (1987) to create gridded fields from altimetry data from Geos-3 and Seasat altimeter satellites. Because satellites produce measurements with higher resolution along the track, while having sparse coverage in perpendicular direction, and because in addition to that ascending and descending track yield approximately 60-degree criss-cross patterns, the data placement is similar to that of irregular triangular grids. With some modifications to control the undesired oscillating, this technique was applied by Smith and Wessel (1990) to produce gridded bathymetry for from ship measured data.

Sandwell (1987) procedure is outlined as follows:

Similarly to one-dimensional cubic spline, which satisfies the variational principle of being a continuous function having the minimum possible integral of the square of its second derivative, while at the same time, constrained by the condition of be equal to the prescribed value at every data point, the two-dimensional interpolation is constructed as a 2D function $f = f(\xi, \eta)$ such that it has minimum possible integral of the square of its Laplacian,

$$\Phi = \iint\limits_{\xi,\eta \in \mathscr{D}} \left(\nabla^2 f\right)^2 \mathrm{d}\xi \mathrm{d}\eta\,, \qquad \nabla^2 = \frac{\partial^2}{\partial \xi^2} + \frac{\partial^2}{\partial \eta^2} \tag{1}$$

while at the same time it assumes all the given values at data points,

$$f(x_k, y_k) = f_k, \qquad \forall k = 1, ..., K\,. \tag{2}$$

The 2D integral above over the entire computational domain. In its turn, variational derivative of such integral with respect to the function is bilaplacian (Laplacian of Laplacian) of the function itself,

$$\frac{\delta \Phi}{\delta f} = \nabla^4 f\,, \tag{3}$$

e.g., Briggs (1974). Then, we demand that the bilaplacian is equal to zero everywhere, except the data points, where it is delta function.

$$\nabla^4 f = 0\,, \qquad (\xi, \eta) \neq (\xi_k, \eta_k), \quad k = 1, ..., K\,. \tag{4}$$

This leads to Green function

$$G(\xi, \eta) = \left(\xi^2 + \eta^2\right) \cdot \left[\ln \sqrt{\xi^2 + \eta^2} - 1\right] \tag{5}$$

which is continuous, and has its first derivatives continuous at $\xi = 0, \eta = 0$ as well, despite having $\xi^2 + \eta^2$ under logarithm,

$$G\big|_{(\xi,\eta)\to(0,0)} \to 0\,, \qquad \frac{\partial G}{\partial \xi}\bigg|_{(\xi,\eta)\to(0,0)} \to 0\,, \qquad \frac{\partial G}{\partial \eta}\bigg|_{(\xi,\eta)\to(0,0)} \to 0\,, \tag{6}$$

73

and its bilaplacian is zero everywhere, except in $\xi = 0, \eta = 0$, where it is delta function. The interpolation f is therefore constructed as

$$f(\xi, \eta) = \sum_{j=1}^{K} w_j \cdot G(\xi - \xi_j, \eta - \eta_j) \tag{7}$$

where (ξ_j, η_j) are the data points. Weights w_j are determined from the set of conditions that $f(\xi, \eta)$ assumes prescribed values at every data point, $f_k = f(\xi_k, \eta_k)$, or

$$f_k = \sum_{j=1}^{K} w_j \cdot G(\xi_k - \xi_j, \eta_k - \eta_j), \qquad \forall k = 1, ..., K \tag{8}$$

where within the summation one should skip $j = k$ to avoid singularity of logarithm, mathematically it does not matter because $G \to 0$ as $(\xi, \eta) \to (0, 0)$. Therefore the problem reduces to solving a full-matrix linear system of equations matrix size $K \times K$. The matrix of the system is symmetric. It is expected to have some of the coefficients very large, others be very small resulting in the need for special care to avoid accumulation of errors when solving this system. Direct solution of 8 requires $\mathcal{O}(K^3)$ operations. The interpolation itself, requires computing and adding up K Green functions at every point of the modeling grid (e.g., ROMS grid), hence needing $\mathcal{O}(N_\xi \cdot N_\eta \cdot K)$ operations more.

This clearly indicates the limits of practical applicability of Sandwell's method. Specifically for this problem we designed a version of fully-pivoted Gauss-Jordan linear solver, where the criterion for selecting pivot (both in raw and column) not just the largest by absolute value element, but the one which yields the maximum ratio of the largest by absolute value element within the raw to the next largest within the same raw, and then maximum over all raws. Other than that, the solver generally follows guideline of Press et al. (1992, see Sec. 2.1 there), and in addition to this it is parallelized for multi-threaded shared-memory environment via `Open MP`. In practice this solver can handle systems resulting from several thousand data points – well beyond the needs to the data set for lake Shira we have (~ 1000 points, topography + coastline).

The other noticed drawback of Sandwell's method is the tendency to produce spurious oscillations when applied to rough data. In fact, Sandwell himself discusses the approach to relax the requirement that interpolation function exactly reproduces values at all the data points, proposing to allow some tolerances and use a least-square fit. Of course, an immediate dilemma on this path is how to assign weighs to each difference – effectively decide which measured values to be trusted more, and which less.

To address the spurious oscillations Smith and Wessel (1990) proposed to replace the bi-harmonic equation (4) with

$$A\nabla^4 f + B\nabla^2 f = 0, \qquad (\xi, \eta) \neq (\xi_k, \eta_k), \quad k = 1, ..., K. \tag{9}$$

which is motivated my the concept of *splines under tension*, and is equivalent to another form of functional to be minimized

$$\Phi = \iint_{\xi, \eta \in \mathscr{D}} \left\{ A\left(\nabla^2 f\right)^2 + B\left(\nabla f\right)^2 \right\} \mathrm{d}\xi \mathrm{d}\eta, \tag{10}$$

where the addition of pure Laplacian in the first- and $(\nabla f)^2$ term in the second equation indicates that the interpolation function should not me just a surface with minimum possible curvature, but should also go from one data point to the next with as uniform ascend/descend rate as possible. In the absence of $\nabla^4 f$-term pure Laplacian produces surface which satisfies the monotonicity principle, however this also results in sharp kink in slope at every data point. Including $\nabla^2 f$-term smoothes the kink to a degree, depending on the selected values of weights A, B. In the general case $A, B \neq 0$ no analytical forms of Green function are known, and furthermore, Smith and Wessel (1990) not even interested in such, because they pursued approach of solving (9) numerically with the expectation that the data set they intent to deal with is too big for the Green function approach be useful due to the inherent $\mathcal{O}(K^3)$-operation barrier.

It should be noted that although Smith and Wessel (1990) state that in the case of $B = 0$ their approach reverts back to Sandwell's, strictly speaking it does not: as a mater of fact, Green function for the relevant differential operators are defined only for a specified sets of boundary conditions. Green function (5) corresponds to the infinite domain. This also makes it translationally invariant, and therefore universal: it does not depend on the point (ξ_k, η_k) around which it is centered. If, in addition to equation (4), $G(\xi, \eta)$ must also satisfy boundary conditions at the outer edge of the domain, then each data point (ξ_k, η_k) should have its own unique form of Green function.

Because of the highly scale-selective nature of biharmonic operator, common relaxation techniques are too inefficient to solve (9). Smith and Wessel (1990) pursue multigrid approach, resulting in complicated, but powerful software, which is also openly available,

```
http://gmt.soest.hawaii.edu/doc/latest/surface.html.
```

The optimal A, B-weighting is still to be determined empirically as a compromise between smoothness of the resultant interpolated field and robustness of the method, with no adaptativity to local conditions in mind.

Biharmonic spline with pre-processed scattered data

In this section we revisit Sandwell (1987) approach to identify the causes of spurious oscillations with the intent to prevent them by preprocessing scattered data.

By the construction biharmonic spline assumes all the prescribed values at the data points. The interpolated values at a given location are essentially weighted sums of data points, and because of Green function spatial structure, the weights decreasing with as inverse second power of the distance. As the result, if there is a pair of data points close to each other, while all other surrounding data points are further away in comparison with the distance between the two, then the local gradient of the interpolant is set by the ratio of the difference between the values at the two points and the distance between them. This may yield very erroneous gradient because the distance is affected by the errors in GPS coordinates, and the smaller the distance, the less reliable are the measurements. Linear methods do not produce oscillations, but in response to such data entries the tend to produce discontinuities – almost vertical walls in the interpolated field. Indeed, visual inspection of Fig. 3 reveals the presence of such pairs of "bad" points.

To address the issue, we implemented a procedure to inspect to identify the pairs which are closer to each other than a user-specified threshold, and reconcile them: the two
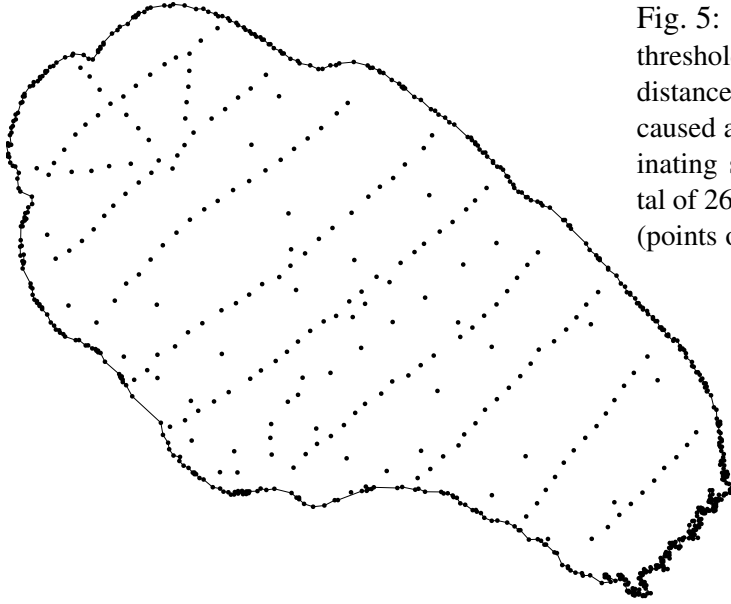
75

Fig. 5: Preprocessed data. In this case the threshold value for the minimum allowed distance was set to $dL_{\min} = 150m$ which caused averaging out 39 pairs, hence eliminating same amount of points out of total of 264 where topography was measured (points on the coastline were not affected).

points are replaced with one having the half-and-half averaged coordinates and measured value. In practice it is a bit delicate because of the inherent data dependency: if three or more points come too close to each other, the location of the averaged point may change the status of logical condition of whether or not some of the source points should be averaged – in may end up at some distance further away from its neighbor. To resolve this, the procedure begins with a very small threshold – only a fraction of user-specified value, and gradually increasing it until reaching the user-specified value. This way, only two-point averaging takes place at any stage. An example of data preprocessed this way is shown in Fig. 5. Perhaps visually indistinguishable from Fig. 2, close comparison shows only few instances where close points from Fig. 2, were replaced with single ones, but obviously, there is no overall decimation of data.

The effect of averaging with progressively increasing threshold of minimum allowed distance $dL_{\min}$ between the data points is illustrated in Fig. 6. It is significant: bumps-and-deeps which we believe spurious disappear. Fig. 7 shows this in detail: by comparing locations of the data points (indicated by white dots) from one panel to the next, it is possible to identify which pair of close points causes appearance of each spurious feature.

However, after exceeding $dL_{\min} = 100m$ there is also noticeable removal of useful data points, which is, obviously, undesirable. This motivates us to design a more selective criterion for identifying unreliable data: not only by the distance, but by the measured values as well. After some experimentation we came up with

$$\left(\frac{z_j - z_k}{dZ_{\max}}\right)^2 + dL_{\min}^2 > dL^2 \qquad \text{where} \quad dL^2 = (x_j - x_k)^2 + (y_j - y_k)^2 \qquad (11)$$

where $dL_{\min}$ is minimum allowed distance, and $dZ_{\max}$ is maximum allowed topographic slope. The rationale is that if two data points indicate that point-wise slope

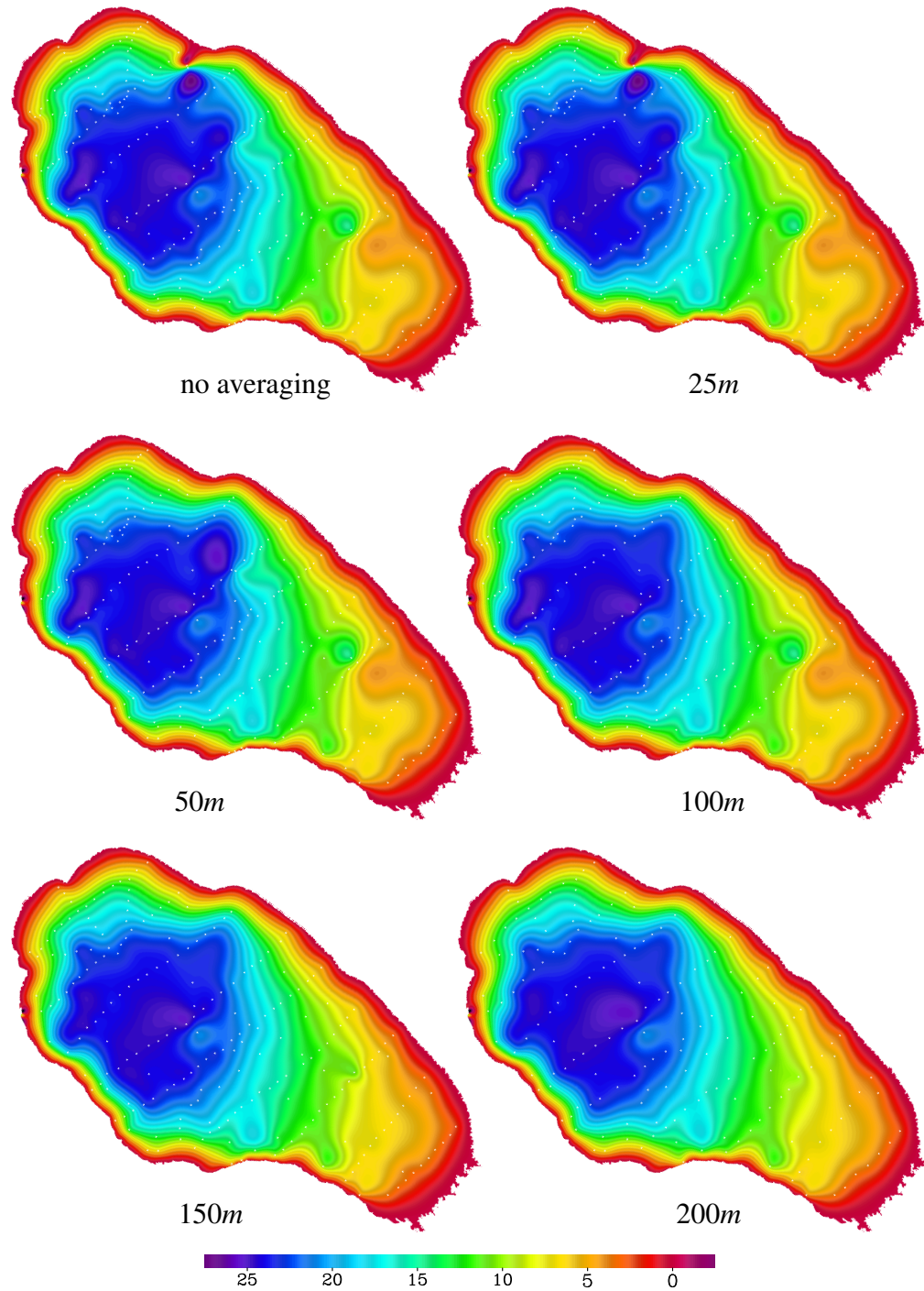$$\frac{|z_j - z_k|}{dL} > dZ_{\max}, \qquad (12)$$

Fig. 6: Sensitivity of reconstructed bottom topography to the threshold of minimum allowed distance below which the data points are merged, *left-to-right*, *top-to-bottom*: no averaging, 25*m*, 50*m*, 100*m*, 150*m*, and 200*m*. Note progressive disappearance of pairs of adjacent deep and shallow spots as the threshold increases.
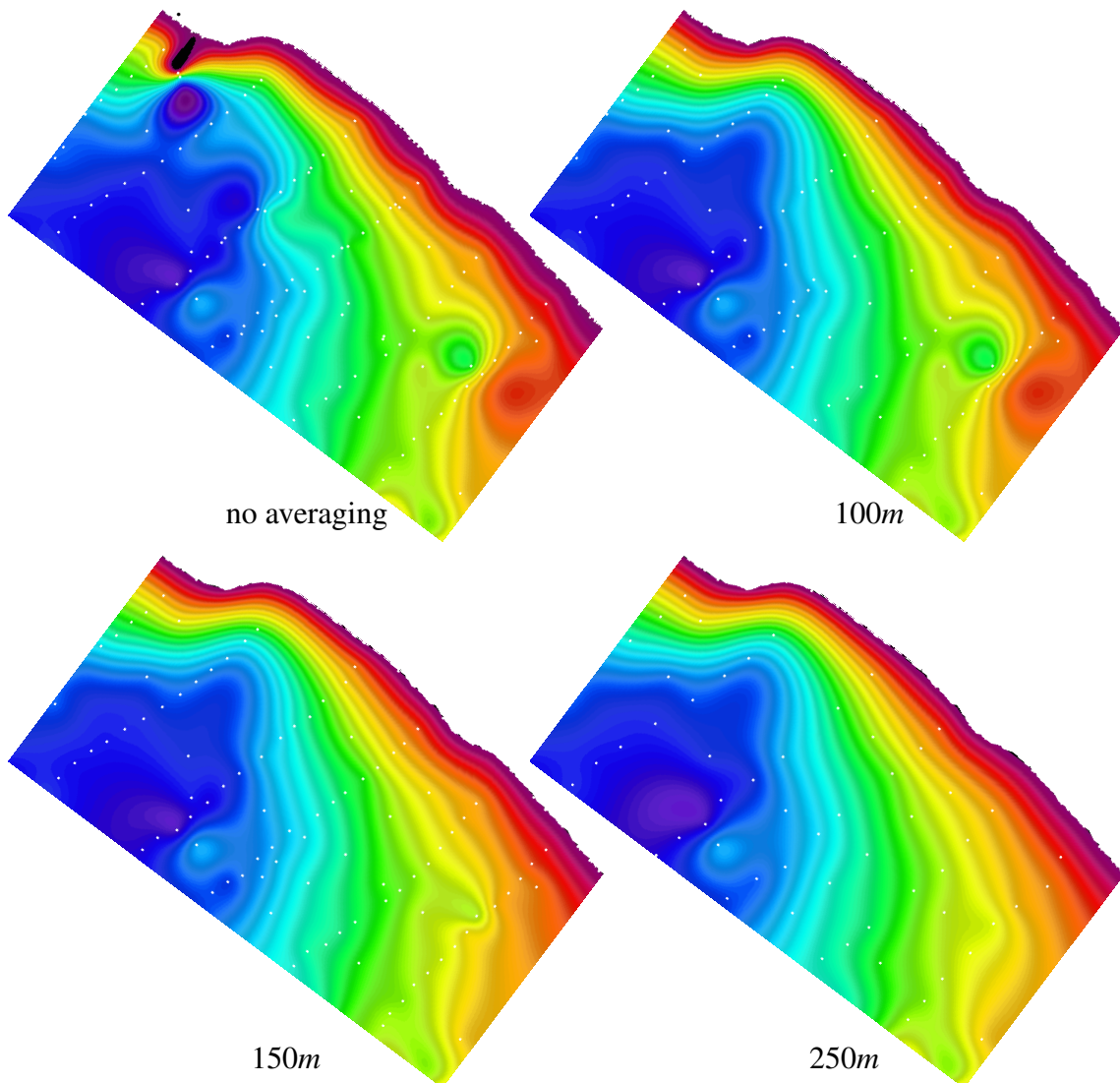
Fig. 7: Detailed view explaining the appearance of spurious topographic features due to closely-placed data points in the case of bi-harmonic spline interpolation: on all panels the white dots indicate location of the actual data points which were used for reconstruction of 2D field in each particular case - these are the remaining points *after* imposing the constraint of minimum distance. In the case of *raw* data (top left) the presence of the bump and hole in the upper-left corner is solely due to the presence of two data points between them at a very small distance from each other. A small discrepancy in measured values in these two points, and the fact that the interpolated field assumes exactly the two values, causes assume steep gradient between the points, and the smoothness of the spline causes it to produce bump and hole. Because neither bump, nor hole contain any data points recording their extremal values, these features should be considered as spurious (at least, their existence are not established by the data). Once the two points are merged into one (right panel on top, and both panels below), the bump and hole disappear. A similar bump-and-hole pair pattern on the shallow area close to the right edge in top two panels, reduces to just an indent once the minimum allowed distance is set to 150*m*), and can be considered as spurious as well, however pattern in the deep part contains some data points close to the extrema, and can be considered as realistic.

exceeds some limit $dZ_{\max}$ beyond which it is considered unrealistic, then the pair $\mathbf{x}_j$ and $\mathbf{x}_k$ should replaced with a single one,

$$\mathbf{x}_{j'} = (\mathbf{x}_j + \mathbf{x}_k)/2, \qquad z_{j'} = (z_j + z_k)/2, \qquad (13)$$

i.e., averaged. This does not mean that the data is discarded completely – the averaged point still contributes to the calculation of interpolated bathymetry, but it no longer affects the local gradient of the interpolant. The other threshold, minimum allowed distance $dL_{\min}$, is related to the accuracy of GPS, and should be chosen to be at least an order of magnitude larger in order to make sure that relative distances between the data points are computed reliably. Again, if two points are found too close to each other, computing gradient is considered unreliable, but their averaged value should be taken into account. The combined criterion (11) essentially acts like a penalty function, taking into account both the distance and the slope, with both factors are contributing toward decision to treat data is not reliable.

Fig. 8, *top panel* shows a version of reconstructed topography obtained by using combined criterion. In comparison with Fig. 6 we note that all spurious features have been removed, but this is achieved at much lesser price of of removing potentially useful data: setting $dL_{\min} = 50m$ and $dZ_{\max} = 0.04$ leads to removal of only 13 points, however, using purely minimum distance criterion requires $dL_{\min} = 100 \rightarrow 150$ to remove hole-and-bump feature on the upper-right corner on the corresponding panels in Fig. 7, and at this point decimation of useful data points becomes already noticeable.

The remaining artifact on Fig. 8, *top* is the noticeable wavy patterns in the isolines which can be visually correlated to the location of data points: e.g., tracing the contour lines going along the north-eastern coast, where the bathymetry is shallow and smooth, an can get an impression that the lines are forced go closer to the coast in the vicinity of a each data point, causing them to make a turn with a small radius, then relax back away from the coast in between, only to make tight turn again near the next point. This is related to the non-analytical behavior of Green functions (5) at the data points: while the $G = G(\xi, \eta)$ and its first derivatives are continuous, the Laplacian is not. This artifact can be counteracted by applying biharmonic smoothing,

$$f^{(m+1)} = f^{(m)} - \tau^{(m)} \cdot \Delta\xi^4 \cdot \nabla^4 f^{(m)} \qquad (14)$$

where $\Delta\xi$ is horizontal grid spacing (in our case the horizontal grids are designed in such a way that resolution is locally the same in both directions, $\Delta\xi = \Delta\eta$ at every location, despite the fact that grids are curvilinear, and resolution is nonuniform), and $\nabla^4 f = \nabla^2 (\nabla^2 f)$ where ∇^2 is discrete 9-point Mehrstellen Laplacian operator,

$$\Delta\xi^2 \cdot \nabla^2 = \frac{2}{3} \cdot \nabla^2_+ + \frac{1}{3} \cdot \nabla^2_\times = \begin{bmatrix} 1/6 & 2/3 & 1/6 \\ 2/3 & -10/3 & 2/3 \\ 1/6 & 2/3 & 1/6 \end{bmatrix} \qquad (15)$$

which has the property that its leading-order truncation error can be expressed as Laplacian of Laplacian. As the result, if applied to a discrete field $f_{i,j}$ obtained by sampling an analytical function at grid points, and Laplacian of this function is zero, then the discrete Mehrstellen operator vanishes with fourth-order accuracy, while the conventional
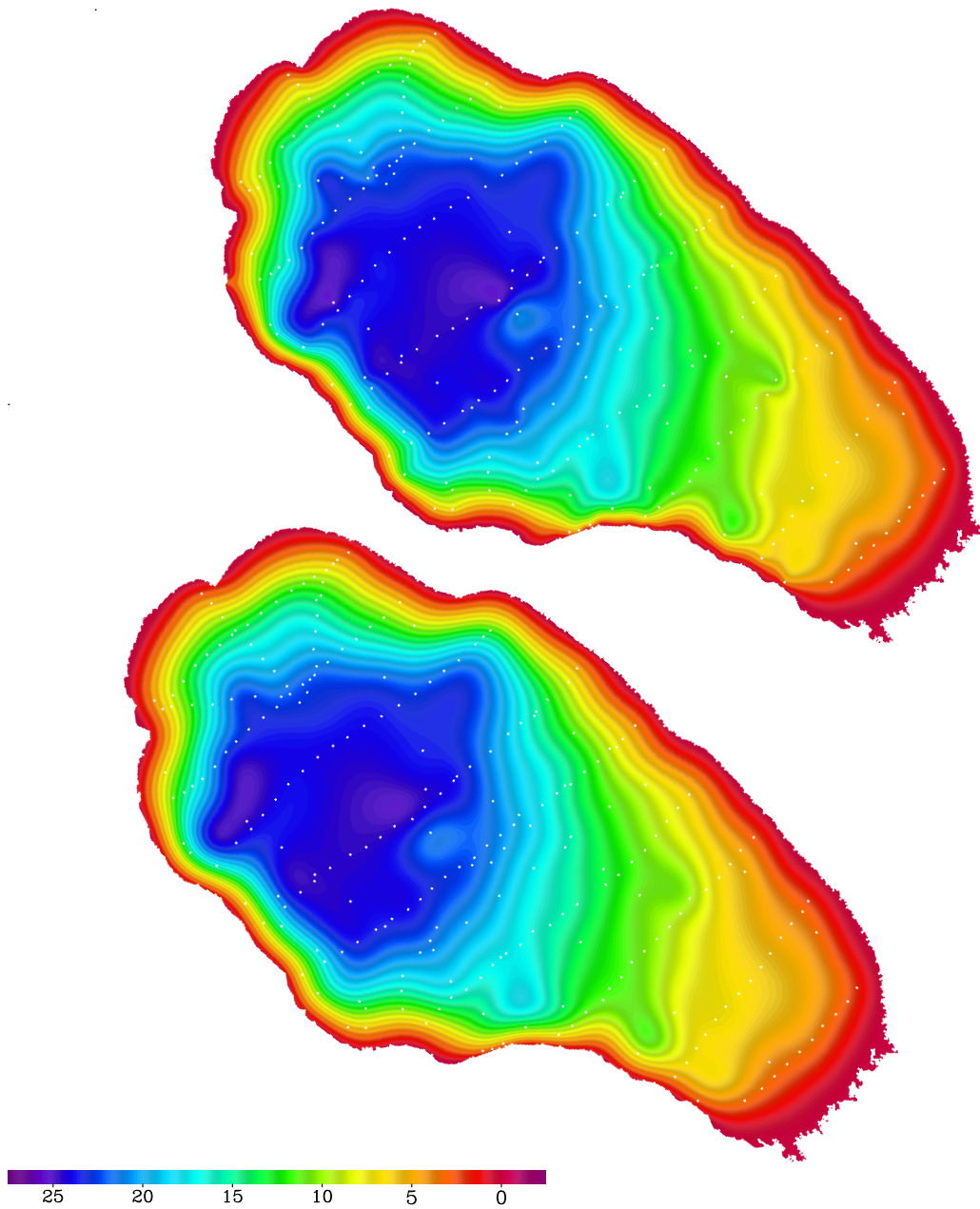
Fig. 8: *Top panel*: Bathymerty of lake Shira reconstructed from scattered data using combined criterion (11) with minimum allowed distance $dL_{\min} = 50m$ and maximum topographic slope $dZ_{\max} = 0.04$. In this case only 13 points from the original data set of 264 were removed by averaging. *Bottom*: Same as above, but after applying 120,000 biharmonic smoothing iterations. The most noticeable difference is lesser wavy pattern in the isolines along the north-eastern coast. In the upper version this clearly corresponds to the locations of the data points visually "pushing" the isolines toward the coast.

5-point cross $+$ or $\times$ Laplacians only with the second. Another notable property of (15) is that it is the most isotropic combination of $+$ or $\times$ Laplacians, hence, if used as a smoothing operator on a discrete grid, it yields in the best possible isotropy of the spread of the field being smoothed. The bilaplacian is computed by repeated Laplacian. Using

$\tau^{(m)} = \tau_0 = (3/16)^2$ corresponds to the exact cancellation checkerboard mode as every iteration. Twice as much, $\tau^{(m)} = 2\tau_0 = 2(3/16)^2$, is the maximum beyond which the iteration becomes unstable, if it is applied at every step m. To accelerate the iterations, we use spread spectrum approach, by setting different τ for consecutive m in a repeated sequence, $\tau^{(m)} = \tau_0, 2\tau_0, \tau_0, 4\tau_0, \tau_0, 2\tau_0, ...$ The result is shown in Fig. 8 *bottom panel*, where we have performed 120,000 iterations. The reason why so many iterations are needed before any change becomes noticeable is because by the construction, bilaplacian of a field reconstructed as sum of Green functions (5) has its bilaplacian equal to zero everywhere, except at the data points. As the result, the primary effect of smoothing (14) occurs in the vicinity of data points, where it restores analyticity of the field at the expense of some departure from the measured values. Bathymetry field in Fig. 8 *panel* is continuously differentiable and its Laplacian (not shown) no longer exposes the location of data points.

Conclusion

Realistic data always contains a certain amount of erroneous entries. Properly identifying and eliminating them is a nontrivial algorithmic dilemma, very specific to the nature of the data. However, it is also the most efficient way to address the cause of errors at the earliest possible stage, at "source", rather than deal with filtering out the consequences later. In this article we develop and tested a technique for reconstructing bottom topography of an enclosed basin from sparse scattered measurements.

The approach we found successful consist on augmenting scattered bathymetry data with the coastline, where the depth is presumed to vanish or to go to some user-specified minimum value. Doing so eliminates the possibility of admitting potentially inaccurate extrapolated values near the coast, where measured data is typically not available, and provide robust side boundary conditions for the slops when interpolation interior data. This augmentation is followed by an automatic inspection of the combined data in order to identify and eliminate potentially contradictory values which may trigger spurious oscillations of the subsequent biharmonic spline interpolation onto regular grid. In essence, this implies averaging of selected data points which are too close to each other, so minor errors in their coordinates cause significant relative errors in distance between them, resulting in potentially erroneous values of bathymetric slope. At the same time, we try to keep preprocessing of the raw data at minimum in order to avoid excessive departure from the measurement.

Overall, this approach is a drastic departure from the usual ROMS modeling practices which mostly rely on gridded data from major sources (Etopo, GEBCO, and SRTM30 PLUS, Becker et al., 2009), and use some degree of smoothing of topography: in the case of lake Shira the data is simply too sparse to make any smoothing undesirable. The technique developed in this article was implemented as software completely free of using any proprietary or license-restricted software, and is made publicly available to ROMS modeling community.

Shchepetkin A.F., Volodko O.S.

Tolomeyev from the Institute of Biophysics Siberian Branch of Russian Academy of Science for their help in organzing and conducting *in situ* measurements.

References

Becker, J. J., Sandwell, D. T., Smith, W. H. F., Braud, J., Binder, B., Depner, J., Fabre, D., Factor, J., Ingalls, S., Kim, S.-H., Ladner, R., Marks, K., Nelson, S., Pharaoh, A., Sharman, G., Trimmer, R., VonRosenburg, J., Wallace, G. and Weatherall, P. Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30_PLUS *Marine Geodesy*, 2009, Vol. 32, No. 4, pp. 355–371, `doi:10.1080/01490410903297766`.

Beckmann, A. and Haidvogel, D. B. Numerical simulation of flow around a tall isolated seamount Part I. Problem formulation and model accuracy *J. Phys. Oceanogr.*, 1993, Vol. 23, pp. 1736–1753, `doi:10.1175/1520-0485(1993)023<1736:NSOFAA>2.0.CO;2`.

Briggs, I. C. Machine contouring using minimum curvature. *Geophysics*, 1974, Vol. 29, No. 1, pp. 39–48, `doi:10.1190/1.1440410`.

Kompaniets, L. A., Yakubaylik, T. V., Volodko, O. S. Analysis of characteristics of lake Shira based on in situ observations. *Vestnik Buryat. Gos. Uvivers.: matematika i informatika*, Ulan-Ude, Izd. BGU, 2012, issue 9, pp. 167-176. Original text in Russian: *Kompaniets, L. A., Yakubaylik, T. V., Volodko, O. S.* Analiz kharakteristic ozera Shira na osnove naturnykh dannykh. *Vestnik Buryatskogo gosudarstvennogo universiteta: matematika i informatika* Ulan-Ude: Izd. BGU, 2012, vypusk 9, s. 167-176.

Lemckert, C. and Imberger, J. Turbulent benthic boundary layer mixing events in fresh water lakes. *Physical Processes in Lakes and Oceans*, J. Imberger, editor, AGU, Washington, D. C., 1998, Vol. 54, pp. 503–516, `doi:10.1029/CE054p0503`.

Penenko, V. V. Estimate of anthropogenic impact onto lake Baikal region using numerical simulation. *Meteorologiya i gidrologiya*, 1989, No. 7, pp. 76-84. Original text in Russian: *Penenko, V. V.* Otsenka antropogennogo vliyaniya na region ozera Baikal s pomoshchyu chislennogo modelirovaniya. *Meteorologiya i gidrologiya*, 1989, No.7, s. 76-84.

Press, W. H., Teukolsky S. A., Vetterling, W. T. and Flannery, B. P. Numerical Recipes in Fortran 77. The Art of Scientific Computing, 2nd Edition *Cambridge University Press*, 1992, ISBN 0-521-43064-X, 818pp.

Sandwell, D. T. Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data *Geophys. Res. Lett.*, 1987, Vol. 14, No. 2, pp. 139–142, `doi:10.1029/gl014i002p00139`.

Shabas I. N. Mathematical modeling of multicomponent material transport in Sea of Azov using multiprocessor computers. *Izvestia YuFU, Tech. Sci.* 2014, No. 12, pp. 200-210. Original text in Russian: *Shabas I. N.* Matematicheskoe modelirovanie zadach perenosa mnogokomponentnykh primesey v Azovskom more na mnogoprotsessornykh vychislitelnykh sistemakh. *Izvestia YuFU, Tekhnicheskie nauki*, 2014, No.12, s. 200-210.

Shchepetkin, A. F. and McWilliams, J. C. A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate // *J. Geophys. Res. Oceans*, 2003, Vol. 108(C3), Art. no. 3090, `doi:10.1029/2001JC001047`.

Smith, W. H. F. and Wessel, P. Gridding with continuous curvature splines in tension *Geophysics*, 1990, Vol. 5, No. 3, pp. 293–305, `doi:10.1190/1.1442837`.

Volodko, O., Kompaniets, L., and Gavrilova, L., Analysis of the velocity profile in Lake Shira in summer using principal component analysis. *International Multidisciplinary Scientific GeoConference SGEM*, 2017, Vol. 17, pp. 831–838.

Wüest, A. and Lorke, A. Small-scale hydrodynamics in lakes *Annual Rev. Fluid Mech.*, 2003, Vol. 35, pp. 373–412, `doi:10.1146/annurev.fluid.35.101101.161220`.

МЕТОД ПОСТРОЕНИЯ РЕЛЬЕФА ДНА ДЛЯ ЗАКРЫТОГО ВОДОЁМА ИЗ ПРОИЗВОЛЬНО РАСПРЕДЕЛЕННЫХ ТОЧЕЧНЫХ ДАННЫХ ИЗМЕРЕНИЙ

А. Ф. Щепёткин[1,2], О. С. Володько[3,4]

[1]*Институт океанологии им. П. П. Ширшова, Российской академии наук 117997, Москва, Нахимовский проспект, д. 36*
[2]*Московский физико-технический институт, 141700, г. Долгопрудный, Московской обл., Институтский пер. 9*
e-mail: `shepetkin.af@ocean.ru`
[3]*Институт вычислительного моделирования СО РАН 660036, Красноярск, Академгородок, 50, стр. 44*
[4] *Сибирский Федеральный Университет, 660041, Красноярск, пр. Свободный, 79*
e-mail: `osv@icm.krasn.ru`

Чтобы моделировать течения и внутренние волны в водоёме необходимо задать его рельеф дна в виде непрерывной, желательно непрерывно дифференцируемой функции (если, конечно, не не существует физических особенностей определяющих обратное). К сожалению, данные рельефа с достаточным пространственным разрешением существуют далеко не всегда. В этой статье мы рассмотрим ранее известные методы построения полей на регулярных сетках из пространственно разбросанных данных – в нашем случае это топография дна измеренная эхолотом с катера оборудованного GPS – и представим метод который мы считаем оптимальным для нашей ситуации. Наш подход в целом следует методике работы Sandwell (1987), который предложил использовать бигармонические сплайны выраженные через функции Грина для интерполяции данных спутниковых измерений геофизических полей. В нашем методе мы дополнили измеренные топографические данные береговой линией, после чего подвергли получившийся массив специальной обработке чтобы выявить и исключить противоречивые и/или ненадежные данные, так чтобы впоследствии предотвратить нежелательные численные эффекты (осцилляции) бигармонических сплайнов.

Ключевые слова: Построение сеточной функции из разбросанных данных; Бигармоническая сплайн интерполяция; Выявление и отбраковка недостоверных и ненадежных данных; Региональное моделирование океана

Литература

Компаниец Л. А., Якубайлик Т. В., Володько О. С. Анализ характеристик озера Шира на основе натурных наблюдений // Вестник Бурятского государственного

университета: математика, информатика. Улан-Удэ: Изд-во БГУ. – 2012. – выпуск 9. – С. 167-176.

Пененко В.В. Оценка антропогенного влияния на регион озера Байкал с помощью численного моделирования // Метерология и гидрология. 1989. №7. С. 76-84.

Шабас И.Н. Математическое моделирование задач переноса многокомпонентных примесей в Азовском море на многопроцессорных вычислительных системах. Известия ЮФУ. Технические науки. 2014. №12. С. 200-210.

Becker J. J., Sandwell D. T., Smith W. H. F., Braud J., Binder B., Depner J., Fabre D., Factor J., Ingalls S., Kim S.-H., Ladner R., Marks K., Nelson S., Pharaoh A., Sharman G., Trimmer R., VonRosenburg J., Wallace G., Weatherall P. Global bathymetry and elevation data at 30 arc seconds resolution: SRTM30_PLUS // Marine Geodesy. 2009. Vol. 32. No.4, P. 355–371, `DOI:10.1080/01490410903297766`.

Beckmann A., Haidvogel D. B. Numerical simulation of flow around a tall isolated seamount. Part I. Problem formulation and model accuracy // J. Phys. Oceanogr. 1993. Vol. 23. P. 1736–1753.
`DOI:10.1175/1520-0485(1993)023<1736:NSOFAA>2.0.CO;2`.

Briggs I. C. Machine contouring using minimum curvature // Geophysics. 1974. Vol. 29. No.1, P. 39–48. `DOI:10.1190/1.1440410`.

Lemckert C., Imberger J. Turbulent benthic boundary layer mixing events in fresh water lakes // Physical Processes in Lakes and Oceans, J. Imberger, editor, AGU, Washington, D. C., 1998. Vol. 54. P. 503–516. `DOI:10.1029/CE054p0503`.

Press W. H., Teukolsky S. A., Vetterling W. T., Flannery B. P. Numerical Recipes in Fortran 77. The Art of Scientific Computing, 2nd Edition // Cambridge University Press. 1992. ISBN 0-521-43064-X. 818pp.

Sandwell D. T. Biharmonic spline interpolation of GEOS-3 and SEASAT altimeter data // Geophys. Res. Lett. 1987. Vol. 14. No.2. P. 139–142. `DOI:10.1029/gl014i002p00139`.

Shchepetkin A. F., McWilliams J. C. A method for computing horizontal pressure-gradient force in an oceanic model with a nonaligned vertical coordinate // J. Geophys. Res. Oceans. 2003, Vol. 108(C3). Art. no. 3090, `DOI:10.1029/2001JC001047`.

Smith W. H. F., Wessel P. Gridding with continuous curvature splines in tension // Geophysics. 1990. Vol. 5, No.3. P. 293–305. `DOI:10.1190/1.1442837`.

Volodko O., Kompaniets L., Gavrilova L., Analysis of the velocity profile in Lake Shira in summer using principal component analysis // International Multidisciplinary Scientific GeoConference SGEM 2017. Vol. 17. P. 831–838.

Wüest A., Lorke A. Small-scale hydrodynamics in lakes // Annual Rev. Fluid Mech. 2003. Vol. 35, P. 373–412. `DOI:10.1146/annurev.fluid.35.101101.161220`.