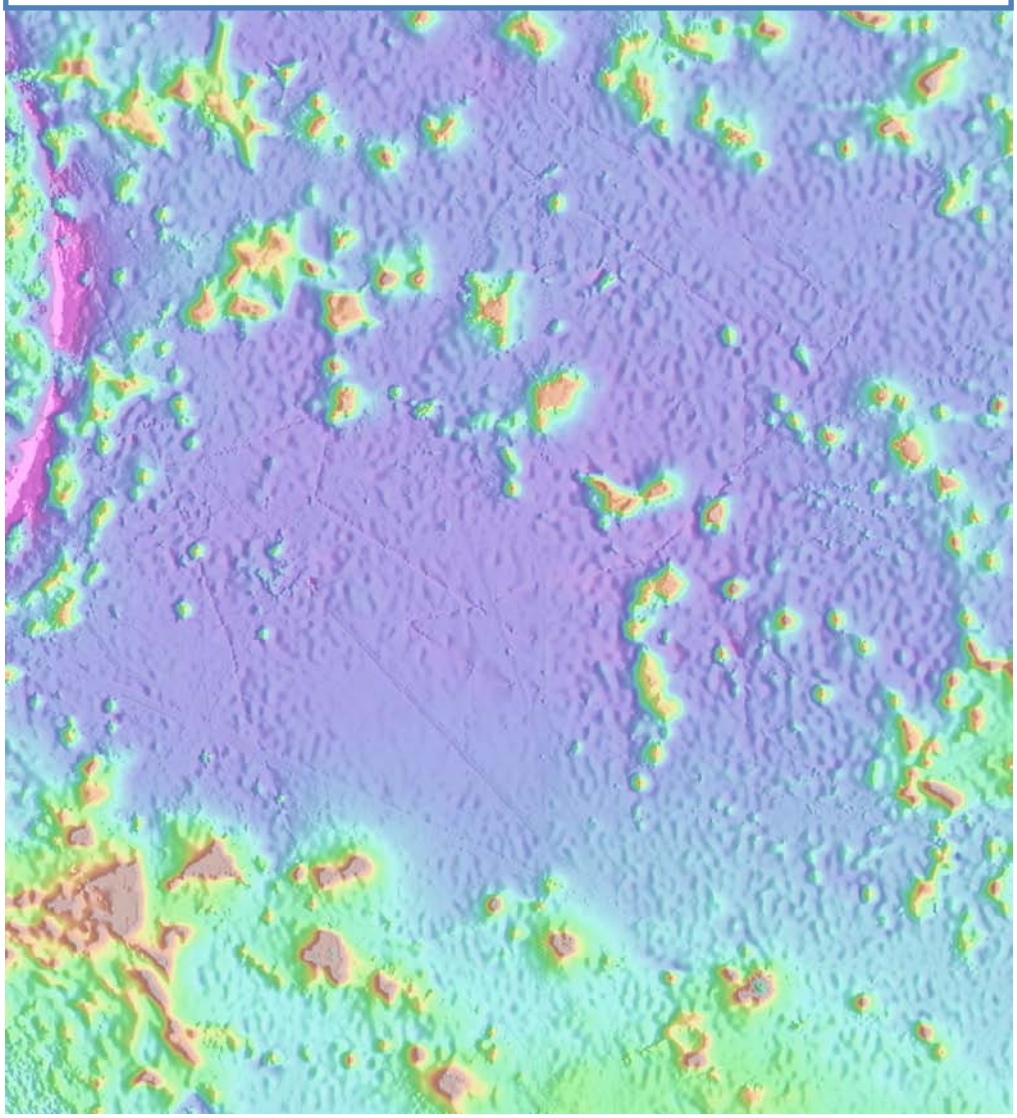
The IHO-IOC GEBCO Cookbook



November 2011

Contributors

Directorate of Hydrography and Navigation (DHN)
Center of Hydrography and Navigation (CHN)
Brazilian Navy, Brazil
http://www.mar.mil.br/dhn/

British Oceanographic Data Centre (BODC), UK
http://www.bodc.ac.uk

Canadian Hydrographic Service, Canada
http://www.charts.gc.ca

CARIS Geospatial Software Solutions, Canada
http://www.caris.com

Center for Coastal & Ocean Mapping
Joint Hydrographic Center
University of New Hampshire, USA
http://ccom.unh.edu

Centro de Investigáciones Oceanograficas e Hidrográficas (CIOH), Columbia
http://www.cioh.org.co/

Cooperative Institute for Research in Environmental Sciences (CIRES)
University of Colorado at Boulder
NOAA National Geophysical Data Center, USA
http://www.ngdc.noaa.gov

DRAFT

International Hydrographic Bureau, Monaco
http://www.iho.int

Geological Survey of Israel, Israel
http://www.gsi.gov.il

National Oceanography Centre, UK
http://noc.ac.uk/

Naval Research Laboratory, USA
http://www.nrl.navy.mil

NOAA Laboratory for Satellite Altimetry, USA
http://ibis.grdl.noaa.gov

NOAA National Geophysical Data Center, USA
http://www.ngdc.noaa.gov

Geological Institute of Russian Academy of Sciences
Laboratory of Geomorphology and Ocean Floor Tectonics, Russia
http://atlantic.ginras.ru/index.html

Department of Geology and Geochemistry
Stockhom University, Sweden
http://www.su.se

DRAFT

© Copyright International Hydrographic *Organization year of first publication*

This work is copyright. Apart from any use permitted in accordance with the Berne Convention for the Protection of Literary and Artistic Works (1886), and except in the circumstances described below, no part may be translated, reproduced by any process, adapted, communicated or commercially exploited without prior written permission from the International Hydrographic Bureau (IHB). Copyright in some of the material in this publication may be owned by another party and permission for the translation and/or reproduction of that material must be obtained from the owner.

This document or partial material from this document may be translated, reproduced or distributed for general information, on no more than a cost recovery basis. Copies may not be sold or distributed for profit or gain without prior written agreement of the IHB and any other copyright holders.

In the event that this document or partial material from this document is reproduced, translated or distributed under the terms described above, the following statements are to be included:

"Material from IHO publication [reference to extract: Title, Edition] is reproduced with the permission of the International Hydrographic Bureau (IHB) (Permission No ……./…) acting for the International Hydrographic Organization (IHO), which does not accept responsibility for the correctness of the material as reproduced: in case of doubt, the IHO's authentic text shall prevail. The incorporation of material sourced from IHO shall not be construed as constituting an endorsement by IHO of this product."

"This [document/publication] is a translation of IHO [document/publication] [name]. The IHO has not checked this translation and therefore takes no responsibility for its accuracy. In case of doubt the source version of [name] in [language] should be consulted."

The IHO Logo or other identifiers shall not be used in any derived product without prior written permission from the IHB.

DRAFT

Table of Contents

Contributors **2**
Table of Contents **5**
List of Figures **9**
List of Tables **10**
Preface **11**
Chapter 1.0 Introduction **12**
Approach 12
Chapter 2.0 Creating a Grid from XYZ Data **15**
2.1 Gridding XYZ Data with Generic Mapping Tools (GMT) 15
2.1.1 Downloading Example Data Set 15
2.1.2 Generic Mapping Tools (GMT) 18
2.1.3 Plotting Data with GMT 19
2.1.4 Gridding XYZ Data with GMT 20
2.1.5 What does a Bad Track in a Grid look like? 24
Chapter 3.0 Gathering Software **26**
3.1 Generic Mapping Tools 26
Chapter 4.0 Software Overview **28**
4.1 ArcGIS 28
4.1.1. What is ArcGIS Desktop Software? 28
Overview of ArcGIS Desktop products 28
4.1.2. Is the software free, what licenses are available and how can you access the software? 29
4.1.3. What type of computer platform does it run on? 29
4.1.4. What does the software do? 29
ArcGIS Desktop application packages - what are they and what do they do? 29
In what formats and structures does ESRI ArcGIS Desktop software store data? 30
How do you work with your data in ArcGIS Desktop packages – do you have to reformat it? 32
Managing your data sets 33
Can you create a 3D surface from your source data sets? 34
Working with 3D surfaces 35
Viewing and editing your data in ArcMap 36
How can I get help with using the software? 39
4.2 CARIS HIPS (Hydrographic Information Processing System) 39
4.2.1 Introduction 39
4.2.2 Acknowledgements 40
4.2.3 HIPS Workflow 40
4.2.4 Project Data Tree and General Information 41
4.2.5 HIPS Workflow- Detailed 43
4.3 Generic Mapping Tools (GMT) 64
4.4 r2v, Surfer, Global Mapper 64
4.4.1 Able Software Corporation's r2v (raster to vector) 64
4.4.2 Global Mapper Software LLC - Global Mapper 12 64
4.4.3 Golden Software Inc. Surfer® Version 9 65

DRAFT

Grid Construction 65
Chapter 5.0 Gathering Data 68
5.1 NGDC Bathymetry Data 68
5.2 JAMSTEC Data 68
Chapter 6.0 Data Cleaning 75
6.1 Quality Assessment 75
6.1.1 Sources of uncertainty 75
Elevation value 75
Morphologic change since data collection 76
Data not measuring ground surface 76
6.1.2 How to determine data uncertainty 77
6.2 Data Cleaning Techniques 77
6.2.1 Automatic (non-interactive) data cleaning 77
6.2.2 Manual (interactive) data cleaning 77
6.2.3 Clipping to polygon or elevation value (e.g., zero) 78
6.2.4 Common multibeam swath sonar errors 78
Spikes under nadir 78
Tracking water-column noise 78
Roll-over on inward-facing steep slopes 79
Anomalies on outward-facing slopes 79
Navigation missing on first ping 79
Inaccurate sound speed profile used 79
6.2.5 Common lidar errors 79
Bomb craters 79
Vegetation and buildings 80
Water-surface values 80
6.3 Data Validation 80
References 80
6.4 Use of uncertainty surfaces 81
6.5 Validation Procedures 81
Chapter 7.0 Metadata Documentation 82
7.1 Introduction 82
7.2 Metadata 82
7.3 Point Data Attribution 82
7.4 Bathymetric Model Attribution 83
Chapter 8.0 Gridding Data 84
8.1 Introduction to gridding 84
8.1.1 Polynomials 84
8.1.2 Near-neighbor searches 84
8.1.3 Triangulation 84
8.1.4 Splines 84
8.1.5 Kriging 84
8.2 Examples of gridding methods 84
8.2.1 Gridding JAMSTEC multibeam data 84
8.2.2 GMT Surface and Nearneighbor 85
8.2.3 GMT Triangulate 86

DRAFT

8.2.4 Spline pyramid 86
8.2.5 Kriging 86
8.2.6 Spline Interpolation 86
8.2.7 Global Mapper 88
8.2.8 r2v 88
8.2.9 Surfer 88
8.2.10 Gridding in ArcMap 89
Importing .XYZ file in ArcMap 89
Reprojecting the data 91
Defining cell resolution for the grid 92
Creating a gridded surface 96
Comparison with original data values 99
8.3 Testing of gridding methods 101
8.3.1 Distance-to-control 101
8.3.2 Histograms 101
8.3.3 Statistical tests 101
Chapter 9.0 Uncertainty 102
9.1 Sources of Uncertainty 102
9.2 Measuring uncertainty 103
9.2.1 Methods of quality assessment, artifacts 103
9.2.2 Comparisons with multibeam 103
Local 103
Regional 103
Global 103
9.2.3 Monte Carlo Technique 103
9.2.4 Bayesian Network Technique 104
Bayesian Network Design and Access 105
Bayesian Network Training Procedures 107
Operational Flow 111
9.2.5 Windowed Regression with Kriging Corrections 112
Linear Smoothing by Locally Weighted Regression (Loess) 112
Combining windowing with ordinary kriging corrections 115
9.2.6 Error Growth Model 121
9.2.7 Split-Sample Approach 122
SS_UNC 125
SS_UNC.SCM 134
RSPLIT 144
9.2.8 Spectral analysis 150
9.2.9 Slope 150
9.2.10 Kriging discussion and cautionary note 150
9.3 Sharing uncertainty results 150
9.3.1 ASCII files 150
9.3.2 netCDF files 150
9.3.3 “BAG” files 150
Chapter 10.0 Additional Resources 151
10.1 Multibeam Data 151

DRAFT

10.2 Global Bathymetry Models 151
Chapter 11.0 Glossary 152
11.1 S-44 Terms 152
Chapter 12.0 Acronyms and Abbreviations 154

DRAFT

List of Figures

Figure 2.1.1.1 NOAA National Geophysical Data Center (NGDC) GEODAS website. 15
Figure 2.1.1.2 Criteria Selection web page. 16
Figure 2.1.1.3 Portion of Search Results web page. 16
Figure 2.1.1.4 Download Format web page. 17
Figure 2.1.1.5 Compress and Retrieve data web page. 17
Figure 2.1.1.6 Retrieve download. 18
Figure 3.1.1 University of Hawaii website for GMT. 26
Figure 3.1.2 GMT download and installation web page. 27
Figure 5.2.1 JAMSTEC Data Site for Research Cruises.....69
Figure 5.2.2 KAIREI Research Data web page. 70
Figure 5.2.3 KR05-01 Cruise web page. 71
Figure 5.2.4 KR05-01 Bathymetry Data web page. 72
Figure 5.2.5 KR05-01 Bathymetry Data files available for download. 73
Figure 5.2.6 JAMSTEC Data Search Portal. 74
Figure 5.2.7 Data Search Portal Results. 74
Figure 8.2.1.1 Color shaded-relief image of gridded KR05-01 multibeam points. 85
Figure 8.2.2.1 Results of gridding depths from V12.1* at constrained grid cells (black dots): GMT "surface" gridding routine with tension set to 0 (left) and set to 1 (middle), and "nearneighbor" gridding routine (right). 86
Figure 9.2.3.1 Process flow for the Monte Carlo procedure of Jakobsson et al (2003).....104
Figure 9.2.4.1 Conceptual flow of the Bayesian Network approach.....105
Figure 9.2.4.2 Bayesian network topology as displayed by the Netica software GUI. The "Horizontal_Error" node variables follow the navigation errors provided in Table 1, with the numbers for surface navigation error in meters. The "Bottom_Gradient" and "Uncertainty_ Estimate" nodes follow a logarithmic binning scheme. 106
Figure 9.2.4.3 High level view of the Bayesian network training process. The figure illustrates computation of the standard deviations as modified from Jakobsson (2003), and use of these results to construct the conditional probability tables (CPT) for the Bayesian network. 108
Figure 9.2.4.4 Low level flow chart of the Monte Carlo simulations block for the Bayesian network training process. Here, *X* and *Y* are the east and north positions of the output grid points in meters after conversion to Universal Transverse Mercator (UTM) projection and specification of a UTM origin in the grid, *Z* is interpolated depth, S_o is an unperturbed interpolation, S_{mn} is the perturbed grid for the nth Monte Carlo loop and m^{th} Horizontal Error Category, E' is error from smoothing. The bold face means that the quantity is a vector. .109
Figure 9.2.4.5 Low level flow chart of the Standard Deviations block for the Bayes net training process. 110
Figure 9.2.4.6 Low level flow chart of the Bayes network programming block for the Bayesian network training process. The network is stored in a library file that is accessed by API. .111
Figure 9.2.4.7 Integrated flow chart of the DBDB-V Uncertainty Expert System (DUES) after programming of the Bayes network component is complete. 112
Figure 9.2.5.1 Equivalent kernel weights for linear and quadratic loess windows. 115
Figure 9.2.7.1 Example histogram output from Split-Sample program SS_UNC. 123

DRAFT

List of Tables

Table 9.2.4.1 Navigation uncertainty categories used in Elmore et al. (2009) for the Bayesian Network approach to uncertainty estimation. The abbreviations are as follows: GPS – Global Positioning System; SINS – Ship's Inertial Navigation System; DR – Dead Reckoning; NAVSAT – Navy Navigation Satellite System; LORAN – Long Range Navigation. The "/" means that two or more techniques are combined. 107

DRAFT

Preface

TEXT IN PREPARATION

DRAFT

Chapter 1.0 Introduction

GEBCO, the General Bathymetric Chart of the Oceans, is a non-profit organization bringing together an international group of experts whose aim is to provide the most authoritative publicly available bathymetry of the world's oceans, for scientific and educational use. GEBCO welcomes and encourages new contributions of data and expertise from all over the world.

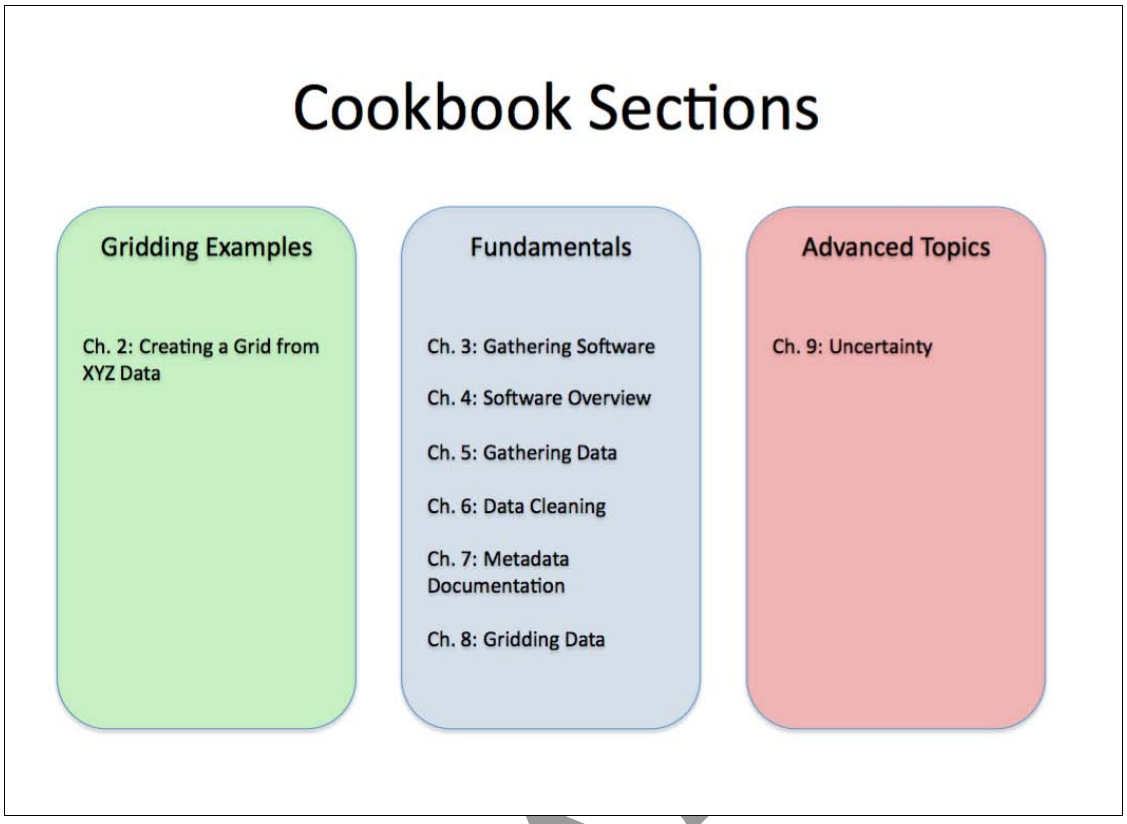
One of GEBCO's products is a global gridded bathymetry data set. This grid contains depth (and, on land, elevation) estimates on a 30 arc-second mesh, which near the equator is about a one-kilometer spacing. Over the oceans, seafloor depths are from ship soundings, and where there are no soundings depths are interpolated, using estimates from satellite altimetry as a guide. The global bathymetry grid is periodically updated by incorporating additional bathymetric survey data, new compilations, and improved data. With this document, the "GEBCO Cookbook," we hope to enable more people to contribute data and gridded compilations to GEBCO.

The Cookbook contains chapters that span basic to advanced topics. These have been written by expert GEBCO contributors from more than a dozen international research organizations, universities, governments, and companies. The Cookbook is a "living document" that will be continually updated and expanded as new contributions are received, and as the GEBCO community's sense of best practices evolves. The Cookbook is available for free download from the GEBCO website (http://www.gebco.net).

DRAFT

Approach

The Cookbook is composed of three main sections- "Gridding Examples," "Fundamentals," and "Advanced Topics."



We start with gridding examples so that beginning users desiring to straightaway produce a grid from xyz data can do so by delving into the first section. We make available an example xyz data set through Internet download and also show an alternate method to obtain it through the National Geophysical Data Center (NGDC) website. A brief overview of gridding software packages, some of which are freely available for download from the Internet, is provided. Illustrated step-by-step instructions that include command lines to process and grid xyz data using the various software packages are listed. The command lines can be cut-and-paste onto the user's computer and, along with the example xyz data set, the user can reproduce the results shown in the cookbook and then apply the same technique to their own data. The results are displayed in maps and plots that can be reproduced with the given command lines as well. This step-by-step approach also allows us to encourage users, once they are familiar with running the example command lines and producing grids and maps, to try changing the command options so that they can see for themselves how the changes affect the final result.

The second section, "Fundamentals," gives a more in-depth look at topics related to preparing, processing, and gridding xyz data. An overview of various gridding software packages, including ArcGIS, CARIS HIPS, Generic Mapping Tools (GMT), r2v, Surfer, and Global Mapper, is included. Also covered are the topics of data cleaning, assessing data quality, validation, metadata documentation, and discussion on the various gridding techniques.

The third section, "Advanced Topics," contains discussions that more experienced users may find useful. Uncertainties are covered in detail- its sources, and methods of measurement

including Monte Carlo and Baysian Network techniques, windowed regression, error growth models, and more. Methods of sharing uncertainty results in the form of ASCII, netCDF, and "BAG" files is included.

Lastly, additional resources, references, a glossary, and appendixes are provided to enable the user to find more information. Included is a list of websites from institutions and organizations that make multibeam data available for public Internet download.

DRAFT

Chapter 2.0 Creating a Grid from XYZ Data

This chapter provides step-by-step instructions that will show a user how to process and grid xyz data. An example data set is made available so the user can follow along on their own computer.

2.1 Gridding XYZ Data with Generic Mapping Tools (GMT)

Contributed by K. M. Marks, NOAA Laboratory for Satellite Altimetry, USA

2.1.1 Downloading Example Data Set

In this section we demonstrate how to create a grid from xyz bathymetry data. The data set used in this demonstration is available for download from ftp://ibis.grdl.noaa.gov/pub/karen/outgoing/ngdc.topo.xyz.

These data may also be downloaded directly from the NOAA National Geophysical Data Center (NGDC) website as follows. First, access the GEODAS (GEOphysical DAta System) webpage at http://www.ngdc.noaa.gov/mgg/gdas/, which is shown in Fig. 2.1.1.1.

Figure 2.1.1.1 NOAA National Geophysical Data Center (NGDC) GEODAS website.

Select "Marine Trackline Search Criteria Entry (bypass Interactive Map) to reach the "Criteria Selection" page, which is shown below.

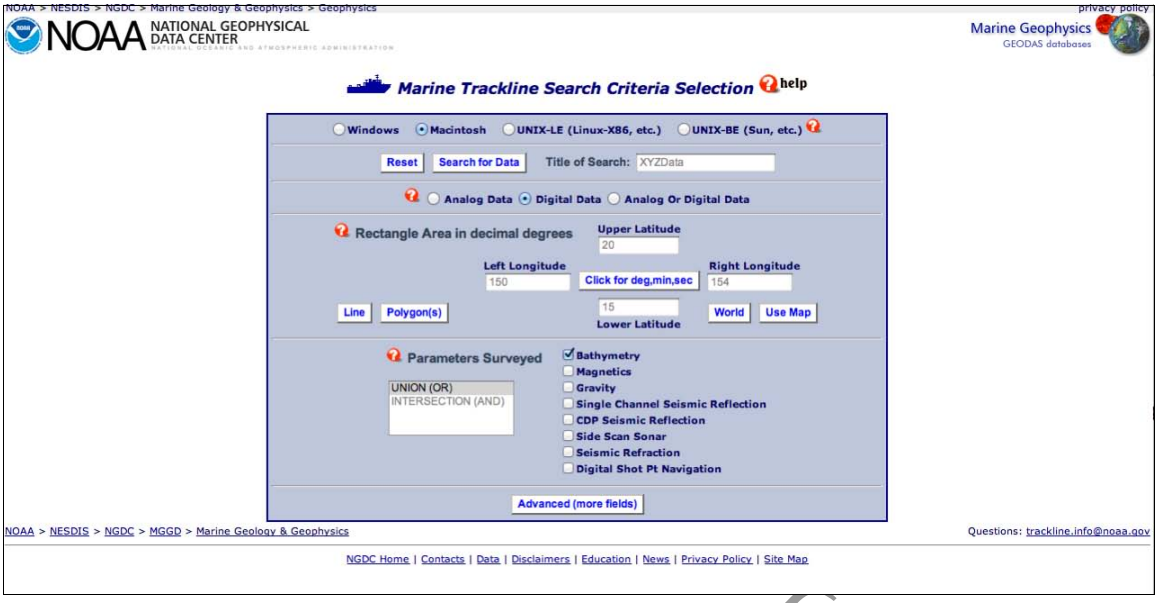


Figure 2.1.1.2 Criteria Selection web page.

Our data set ranges from 15 to 20 N latitude, and from 150 to 154 E longitude. The user should enter these geographic coordinates into the corresponding boxes, select "Digital Data" and "Bathymetry," and then press the "Search for Data" button on the web page. There is no need to enter more "Advanced" fields for this example.

The results of the search are displayed (see Fig. 2.1.1.3 below). Select "Download Search Results" by pushing the web page button.

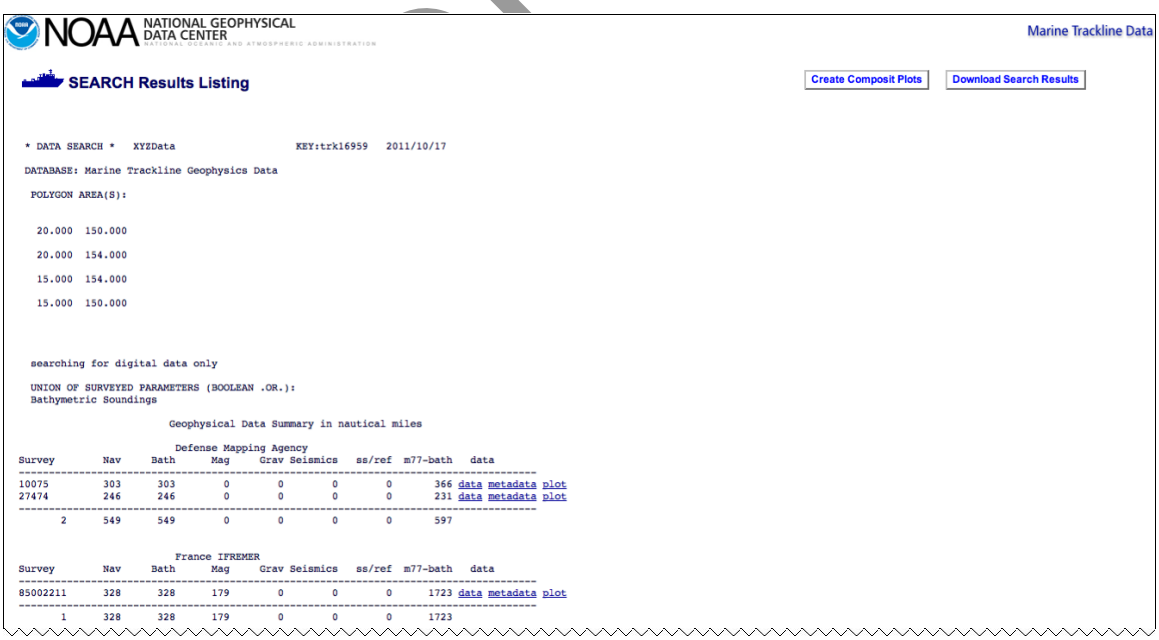


Figure 2.1.1.3 Portion of Search Results web page.

DRAFT

The desired format of the downloaded data can be designated next (see Figure 2.1.1.4 below). For our data set, select "Single File of All Survey Data in Area," "Space Delimited XYZ Format," check "Convert uncorrected depths to corrected meters using Carter's Tables," and select Longitude, Latitude, and Corrected Depths (meters) in the box. Then press the "Process Digital Data" button.

Figure 2.1.1.4 Download Format web page.

A web page notifying the user that the data have been processed and are available for retrieval comes up next (see Fig. 2.1.1.5 below). Press the "Compress and Retrieve Your Data" button. The GEODAS system has assigned the arbitary name "trk16959.tgz" to the data set.

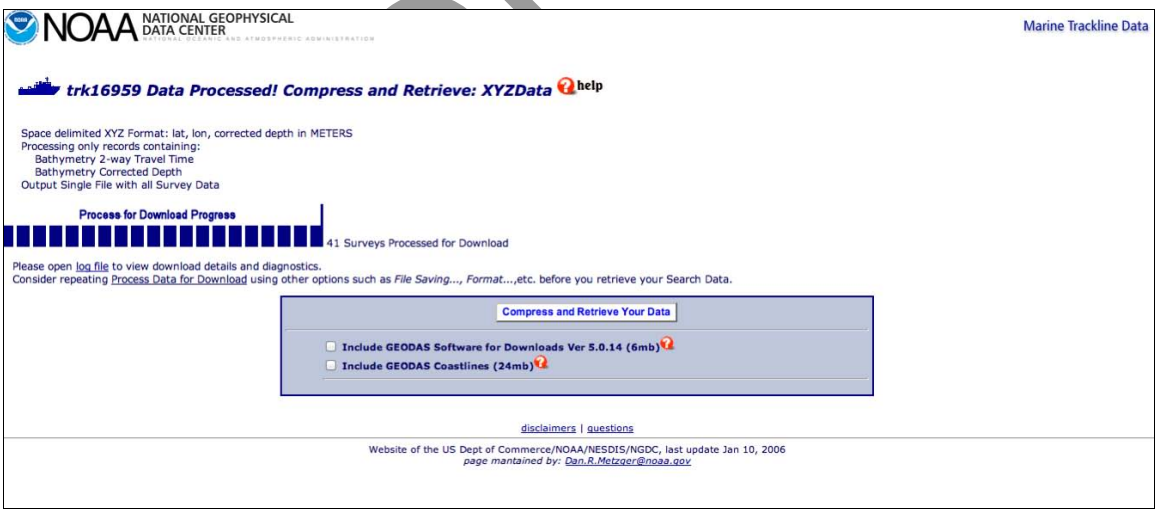


Figure 2.1.1.5 Compress and Retrieve data web page.

The next webpage provides information about the size of the download and the expected download times (Fig. 2.1.1.6 below). Push the "Retrieve" button to download the data set. If the download file size is particularly large, select "Retrieve Compressed File" next to the retrieve

button to reduce the size of the download file. The data file is then downloaded onto the user's computer.

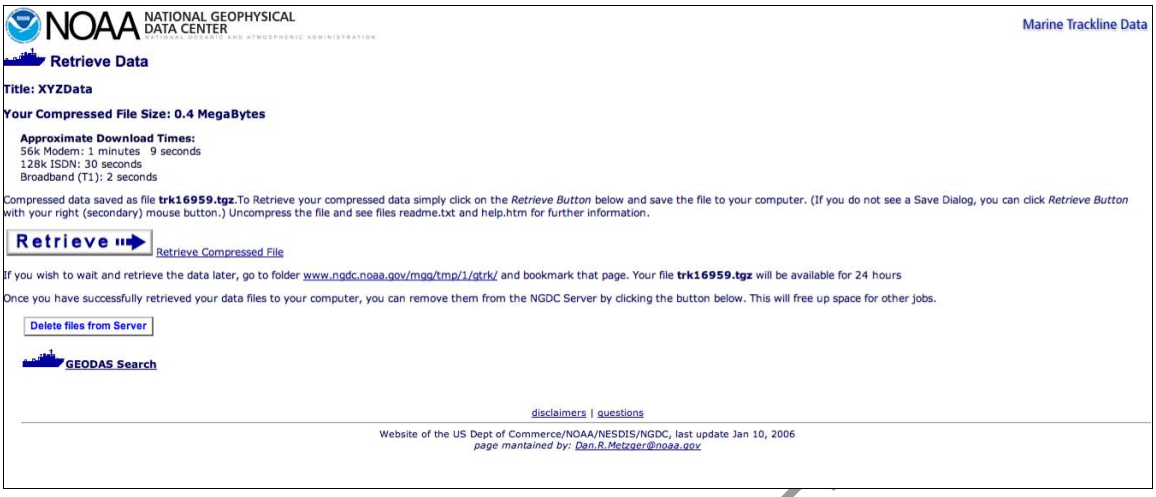


Figure 2.1.1.6 Retrieve download.

The downloaded data file is provided in a tar (trk16959.tar) format. The command

tar –xvf trk16959.tar

forms a directory containing the xyz data set (/trk87076_data/trk87076/trk87076.xyz). We chose to rename the xyz data set "ngdc.topo.xyz."

2.1.2 Generic Mapping Tools (GMT)

GMT (Generic Mapping Tools) (Wessel and Smith, 1998) is a collection of open source mathematical and mapping routines for use on gridded data sets, data series, and arbitrarily located data. The GMT package is available for download from the University of Hawaii website (http://gmt.soest.hawaii.edu/) (see Fig. 2.1.2.1). We utilized GMT routines for our gridding and mapping examples shown in this section. Actual GMT command lines are provided in the discussions that follow.

DRAFT

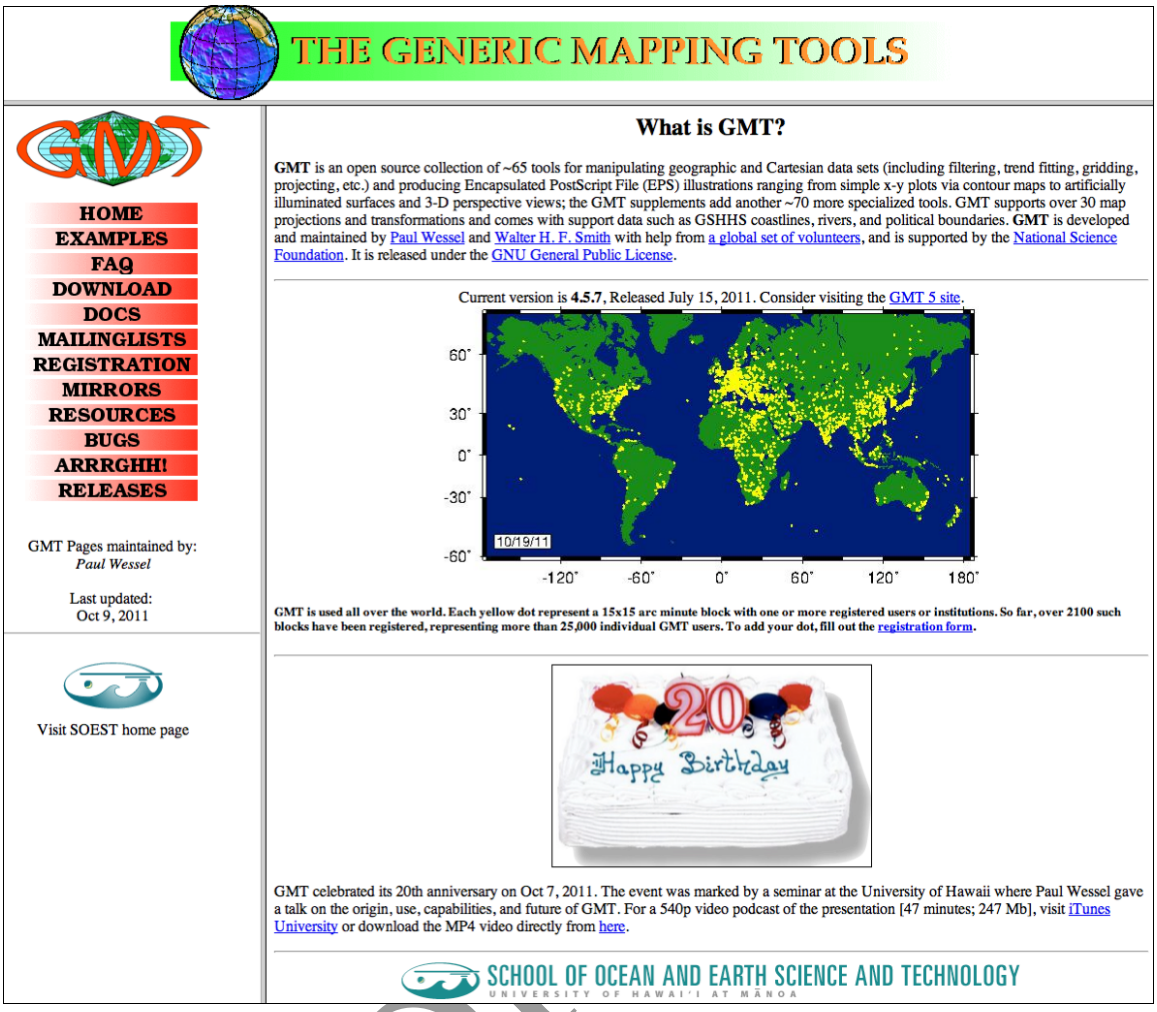


Figure 2.1.2.1 University of Hawaii website for GMT.

2.1.3 Plotting Data with GMT

It is desirable to plot the xyz data on a map to see the locations of the points. The xyz data may be plotted on a map using Generic Mapping Tools (GMT).

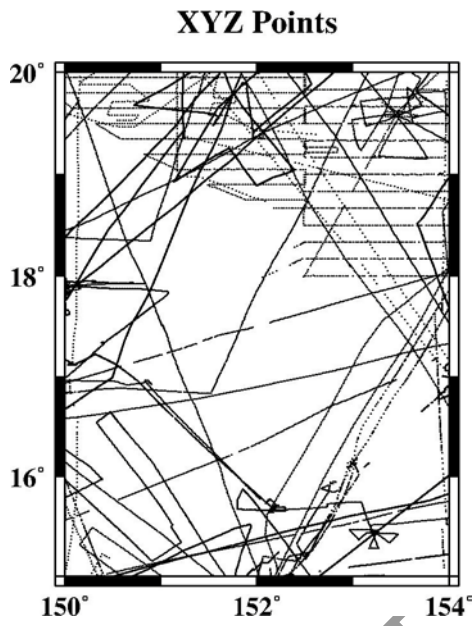


Figure 2.1.3.1 Map of XYZ point locations.

Figure 2.1.3.1 (above) was produced with the following GMT commands:

```
psbasemap -R150/154/15/20 –Jm0.5 -Ba2f1:."XYZ Points":WeSn -K > Fig.2.1.3.1.ps
psxy ngdc.topo.xyz -R -Jm -Sp -O >> Fig.2.1.3.1.ps
```

It is useful to note the distribution of the data points. In the center of the map there are few ship tracks, leaving large areas with no bathymetry measurements. In the northeastern corner, there is systematic survey coverage.

2.1.4 Gridding XYZ Data with GMT

It is possible to create a grid from the randomly located xyz data points. GMT routine "surface" is an adjustable tension continuous curvature gridding algorithm that can be used to grid the xyz data. We input the xyz data into "surface," setting the tension factor to 0.25 (surface tension set to "0" gives the minimum curvature solution, and set to "1" gives a harmonic surface where maxima and minima are only possible at control points). Figure 2.1.4.1 (below) shows the resulting grid.

DRAFT

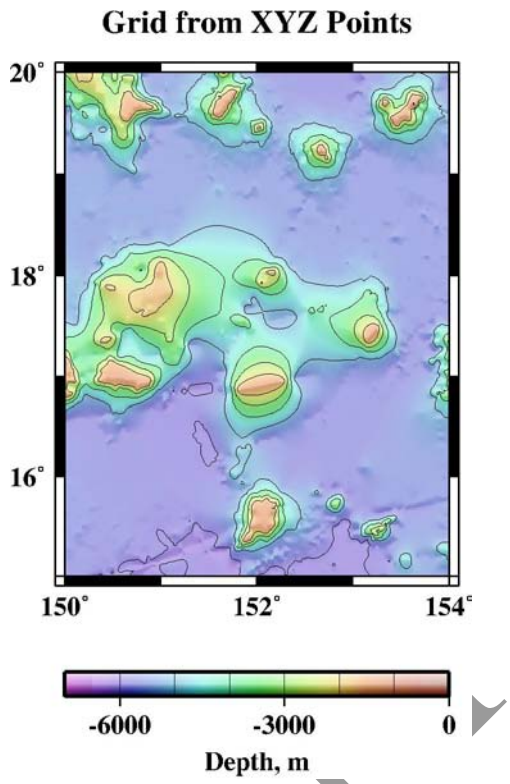


Figure 2.1.4.1 Grid of xyz data points, using GMT "surface" with tension set to 0.25.

The GMT commands used to produce the gridding solution and plot it in Fig. 2.1.4.1 follow.

```
surface ngdc.topo.xyz -R150/154/15/20 -I1m -T0.25 -Gngdc.topo_grd
grdgradient ngdc.topo_grd -A0 -Ne0.2 -Ggrad_grd
grdimage ngdc.topo_grd -Igrad_grd -Ctopo.cpt -Jm0.5 -K>Fig.2.1.4.1.ps
pscoast -G175 -R150/154/15/20 -Jm0.5 -W2 -Df -Ba2f1:."Grid from XYZ Points":WeSn
        -K -O>>Fig.2.1.4.1.ps
grdcontour ngdc.topo_grd -R -Jm -C1000 -W1 -K -O>>Fig.2.1.4.1.ps
psscale -D1/-.5/2/.125h -Ctopo.cpt -Ba3000g1000:"Depth, m": -I -N300 -O >>Fig.2.1.4.1.ps
```

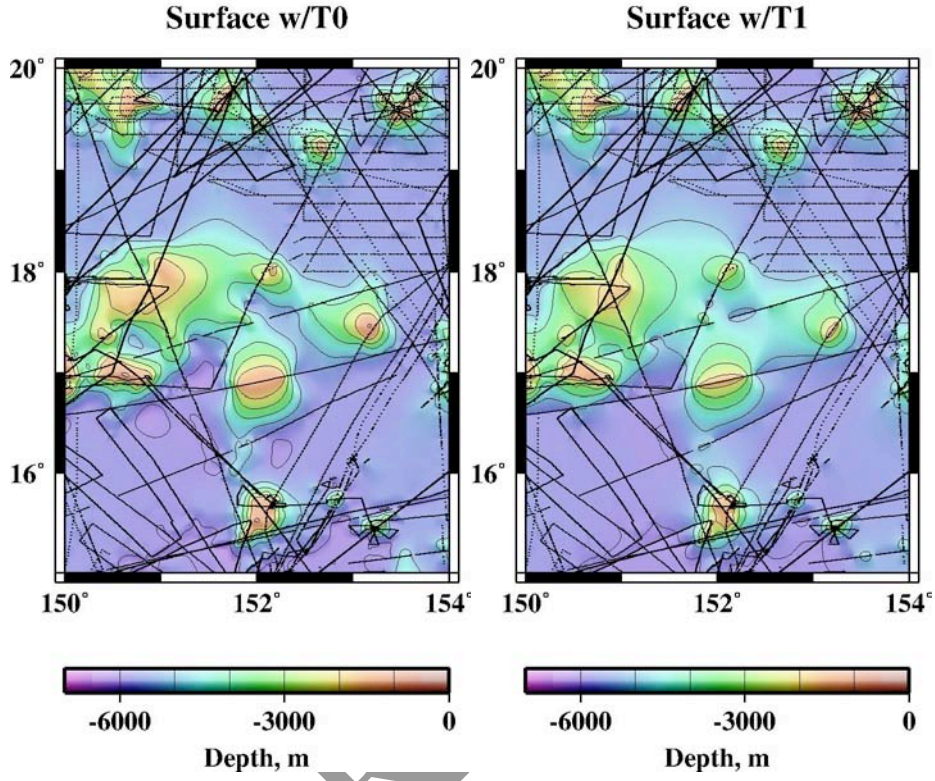
The color palette table (topo.cpt) is:

```
-7000  300        .350000 .85   -6500  275.625      .350000 .85
-6500  275.625    .300000 .90   -6000  250.3125     .300000 .90
-6000  250.3125   .300000 .90   -5500  225          .300000 .90
-5500  225        .300000 .90   -5000  199.6875     .300000 .90
-5000  199.6875   .300000 .90   -4500  175.3125     .300000 .95
-4500  175.3125   .300000 .95   -4000  150          .350000 .95
-4000  150        .350000 .95   -3500  125.625      .350000 .95
-3500  125.625    .350000 .95   -3000  99.375       .350000 .95
-3000  99.375     .350000 .95   -2500  75           .350000 .95
-2500  75         .350000 .95   -2000  50.625       .350000 .95
-2000  50.625     .350000 .95   -1500  25.3125      .350000 .95
-1500  25.3125    .350000 .95   -500   10           .250000 .85
-500   10         .200000 .85   0      10           .150000 .80
B      300        .350000 1
F      0          0.00000 1.0
```

There are other algorithms that can be used to grid the xyz data points, and each will produce a different solution; the differences are most evident in areas where there are large gaps between

DRAFT

measurements. Changing the tension parameter in GMT "surface" will also produce a different solution. "Nearneighbor" uses a nearest neighbor algorithm to assign an average value within a radius centered on a node. "Triangulate" performs optimal Delaunay triangulation. We show the results of GMT "surface," "nearneighbor," and "triangulate" gridding algorithms below.



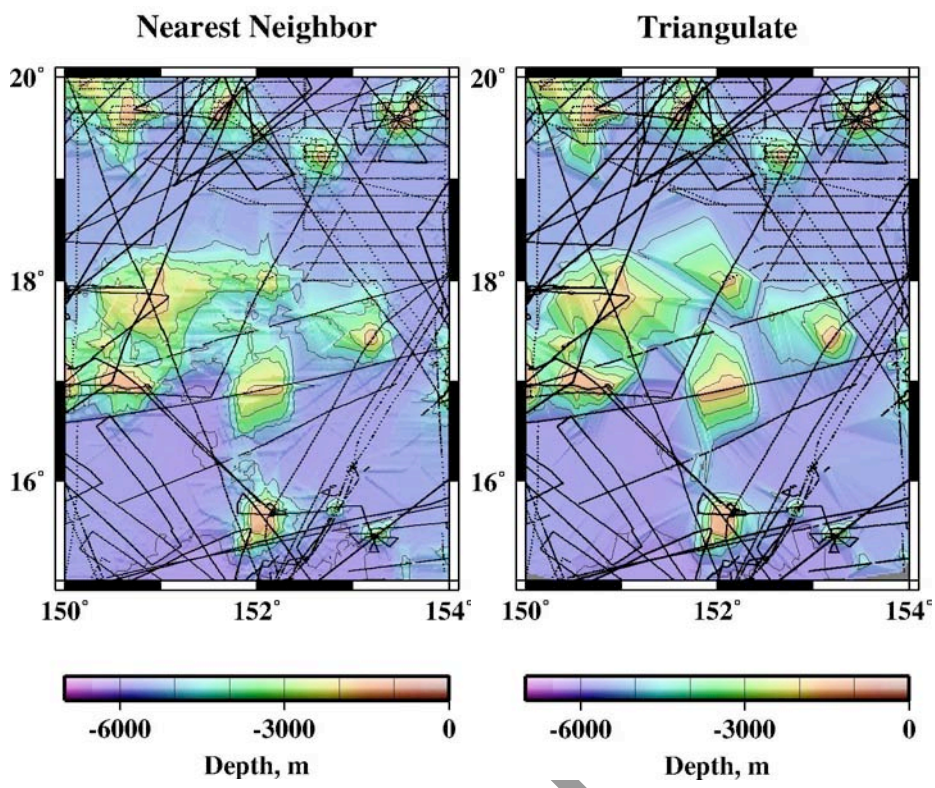


Figure 2.1.4.2 Results of gridding xyz data (black dots) with GMT gridding routines "surface" with tension set to 0 (top left) and set to 1 (top right), "nearneighbor" (bottom left), and "triangulate" (bottom right).

The command lines used for the corresponding GMT gridding routines follow. The commands to plot the figures are the same as listed for Fig. 2.1.4.1, and the color palette file is the same.

```
surface ngdc.topo.xyz -R150/154/15/20 -I1m -T0 -Gngdc.topo_grd
surface ngdc.topo.xyz -R150/154/15/20 -I1m -T1 -Gngdc.topo_grd
nearneighbor ngdc.topo.xyz -R150/154/15/20 -N4/1 -S100k -I1m -Gngdc.topo_grd
triangulate ngdc.topo.xyz -R150/154/15/20 -I1m -Gngdc.topo_grd > junk
```

Each gridding algorithm produces a very different result because each one uses a different method of interpolating depth estimates in the gaps between data points. Where the gaps are large, areas of shallower seafloor (orange-yellow colored anomalies) appear to follow the ship tracks. This is because ships collect depth measurements along one-dimensional tracks without detecting seafloor away from the track. Seafloor can be mapped in two dimensions when the survey coverage is adequate (for example, in the northeast corner of the study area).

Bathymetry models that estimate depths from a combination of ship measurements and depths derived from satellite gravity can also map the seafloor in two dimensions. Fig. 2.1.4.3 below shows the Smith and Sandwell (1997) bathymetry model (version 13.1) over our study area.

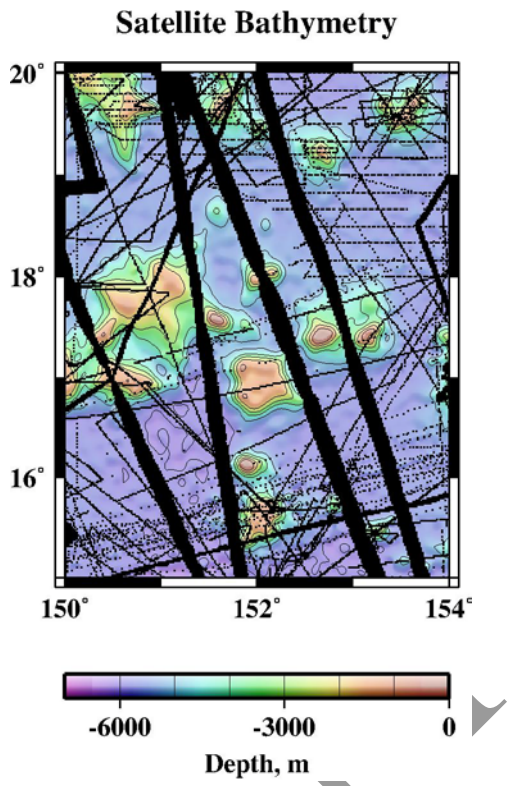


Figure 2.1.4.3 Grid of estimated depths from altimetric bathymetry model, black dots are ship controls used in model

It is important to note that some bathymetry anomalies lie in gaps between ship controls (black dots), and are also mapped away from the ship tracks. Because bathymetry models incorporate depths estimated from satellite gravity and the satellite tracks have ~4 km uniform spacing over the oceans, bathymetry anomalies are accurately mapped in two dimensions. Grids created only from xyz data containing large gaps may not accurately map the seafloor located between the ship tracks.

In our experience, GMT "surface" with a tension set to 0.25 produces a nice-looking bathymetry grid.

2.1.5 What does a Bad Track in a Grid look like?

Gridding algorithms simply ingest the xyz data supplied, and produce a grid. If the xyz data contain errors, they may be visible in the resulting grid. A common problem is a "bad track," where the ship data along a track are poor, which can result from a variety of reasons. To demonstrate what a "bad track" in a grid looks like, we added 2000 m to the depths of track 01010060 (NGDC ID number) and combined it with the example xyz data set and gridded it using routine GMT "surface" (see Fig. 2.1.5.1).

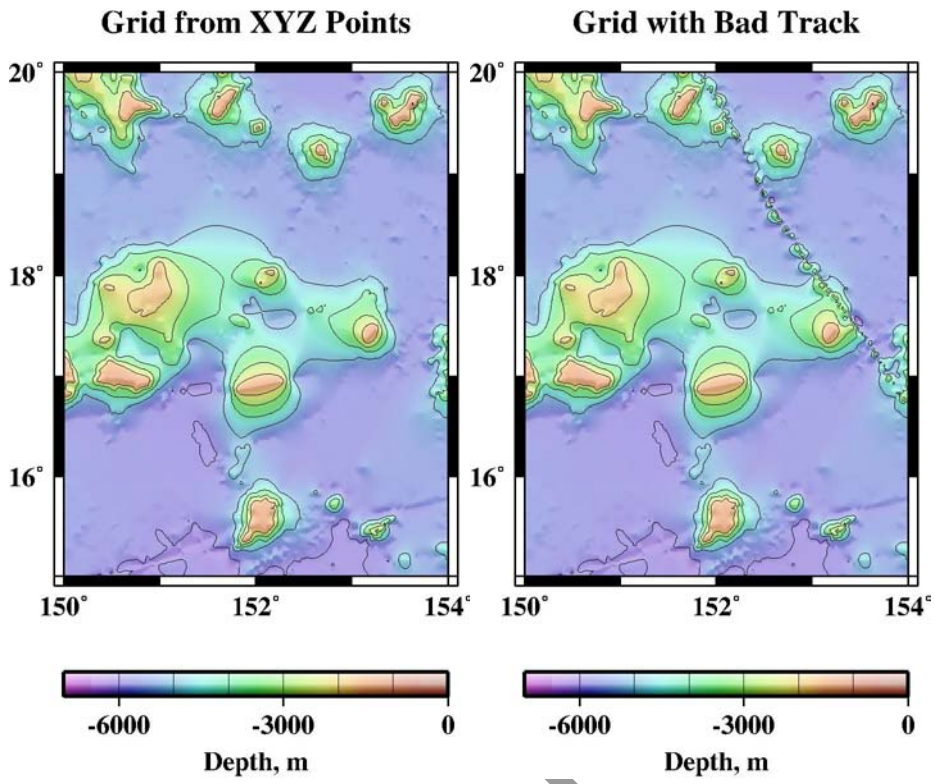


Figure 2.1.5.1 Grid of xyz data points (left), and grid of same xyz data points with "bad track" included (right).

In the right panel of Fig. 2.1.5.1, the bad track stands out as a line of local shallow seafloor anomalies. Contours adjacent to the line are contorted and jagged. Sometimes a bad track or other data with errors can be easy to find as in this demonstration, but they can also be subtle. It is important to use data that are as clean and free of errors as possible.

DRAFT

Chapter 3.0 Gathering Software

Some software packages are freely available for download from the internet. The purpose of this chapter is to show where to find such software and how to obtain it.

3.1 Generic Mapping Tools

Contributed by K. M. Marks, NOAA Laboratory for Satellite Altimetry, USA

GMT (Generic Mapping Tools) (Wessel and Smith, 1998) is a collection of open source mathematical and mapping routines for use on gridded data sets, data series, and arbitrarily located data. The GMT package is available for download from the University of Hawaii website (http://gmt.soest.hawaii.edu/) (see Figure 3.1.1). We utilized GMT routines for much of our data analyses, gridding, and mapping discussed in selected chapters that follow. Software packages such as MATLAB, IMSL, ArcGIS and others may also provide similar mathematical and mapping capabilities, but these are not free.

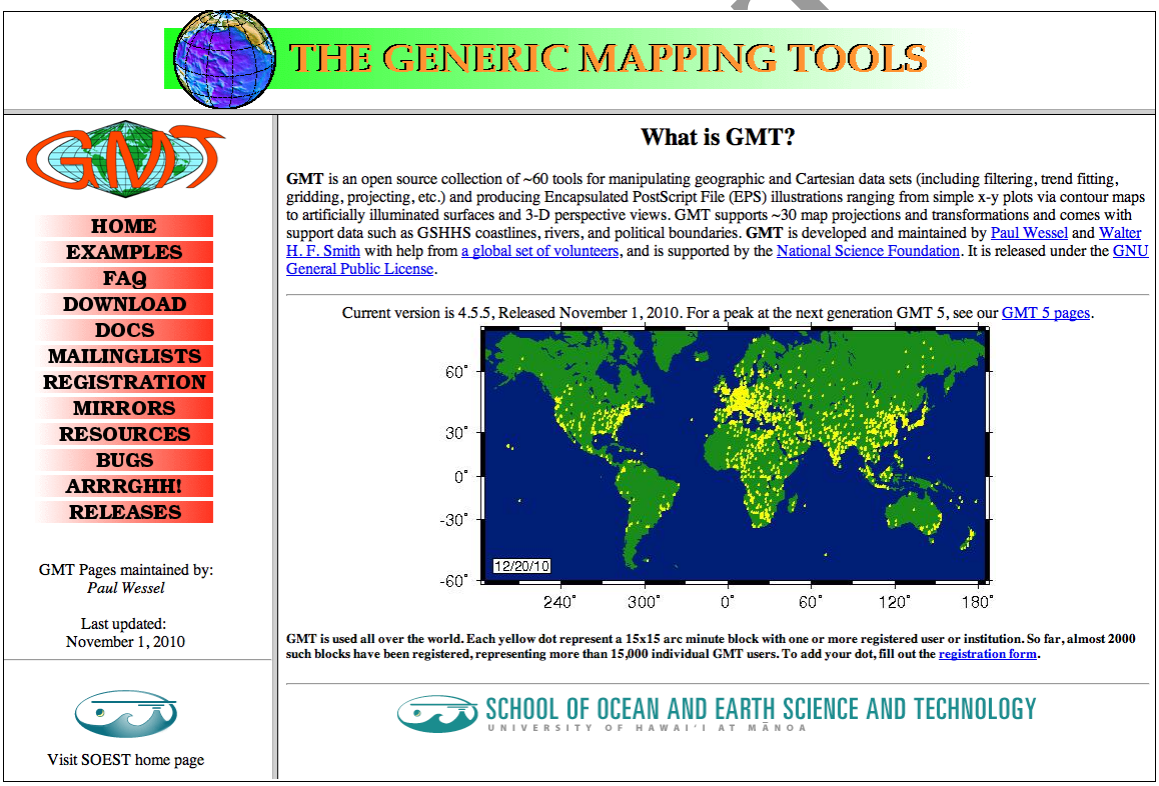
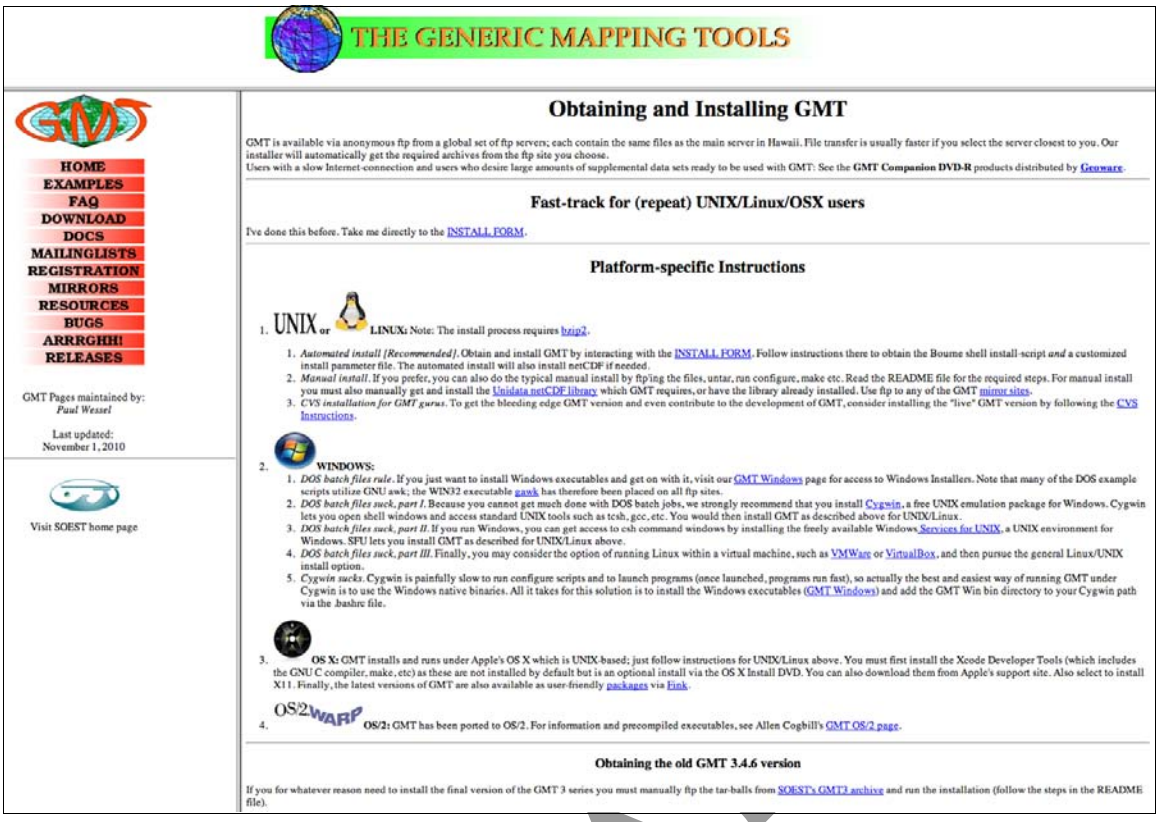


Figure 3.1.1 University of Hawaii website for GMT.

To download the GMT software package, click on "download" shown in the left column of the figure above. Instructions for how to obtain and install the GMT software are given on the download webpage (see Figure 3.1.2).



THE GENERIC MAPPING TOOLS

HOME
EXAMPLES
FAQ
DOWNLOAD
DOCS
MAILINGLISTS
REGISTRATION
MIRRORS
RESOURCES
BUGS
ARRRGHH!
RELEASES

GMT Pages maintained by:
Paul Wessel

Last updated:
November 1, 2010

Visit SOEST home page

Obtaining and Installing GMT

GMT is available via anonymous ftp from a global set of ftp servers; each contain the same files as the main server in Hawaii. File transfer is usually faster if you select the server closest to you. Our installer will automatically get the required archives from the ftp site you choose.
Users with a slow Internet-connection and users who desire large amounts of supplemental data sets ready to be used with GMT: See the **GMT Companion DVD-R** products distributed by Geoware.

Fast-track for (repeat) UNIX/Linux/OSX users

I've done this before. Take me directly to the INSTALL FORM.

Platform-specific Instructions

1. UNIX or LINUX: Note: The install process requires bzip2.
   1. *Automated install [Recommended]*. Obtain and install GMT by interacting with the INSTALL FORM. Follow instructions there to obtain the Bourne shell install-script *and* a customized install parameter file. The automated install will also install netCDF if needed.
   2. *Manual install*. If you prefer, you can also do the typical manual install by ftp'ing the files, untar, run configure, make etc. Read the README file for the required steps. For manual install you must also manually get and install the Unidata netCDF library which GMT requires, or have the library already installed. Use ftp to any of the GMT mirror sites.
   3. *CVS installation for GMT gurus*. To get the bleeding edge GMT version and even contribute to the development of GMT, consider installing the "live" GMT version by following the CVS Instructions.
2. WINDOWS:
   1. *DOS batch files rule*. If you just want to install Windows executables and get on with it, visit our GMT Windows page for access to Windows Installers. Note that many of the DOS example scripts utilize GNU awk; the WIN32 executable gawk has therefore been placed on all ftp sites.
   2. *DOS batch files suck, part I*. Because you cannot get much done with DOS batch jobs, we strongly recommend that you install Cygwin, a free UNIX emulation package for Windows. Cygwin lets you open shell windows and access standard UNIX tools such as tcsh, gcc, etc. You would then install GMT as described above for UNIX/Linux.
   3. *DOS batch files suck, part II*. If you run Windows, you can get access to csh command windows by installing the freely available Windows Services for UNIX, a UNIX environment for Windows. SFU lets you install GMT as described for UNIX/Linux above.
   4. *DOS batch files suck, part III*. Finally, you may consider the option of running Linux within a virtual machine, such as VMWare or VirtualBox, and then pursue the general Linux/UNIX install option.
   5. *Cygwin sucks*. Cygwin is painfully slow to run configure scripts and to launch programs (once launched, programs run fast), so actually the best and easiest way of running GMT under Cygwin is to use the Windows native binaries. All it takes for this solution is to install the Windows executables (GMT Windows) and add the GMT Win bin directory to your Cygwin path via the .bashrc file.
3. OS X: GMT installs and runs under Apple's OS X which is UNIX-based; just follow instructions for UNIX/Linux above. You must first install the Xcode Developer Tools (which includes the GNU C compiler, make, etc) as these are not installed by default but is an optional install via the OS X Install DVD. You can also download them from Apple's support site. Also select to install X11. Finally, the latest versions of GMT are also available as user-friendly packages via Fink.
4. OS/2 WARP: GMT has been ported to OS/2. For information and precompiled executables, see Allen Cogbill's GMT OS/2 page.

Obtaining the old GMT 3.4.6 version

If you for whatever reason need to install the final version of the GMT 3 series you must manually ftp the tar-balls from SOEST's GMT3 archive and run the installation (follow the steps in the README file).

Figure 3.1.2 GMT download and installation web page.

DRAFT

Chapter 4.0 Software Overview

The purpose of this chapter is to give an overview of various software packages that can be used to clean, process, manipulate, analyze, and plot data used in scientific research and production.

4.1 ArcGIS

Contributed by Pauline Weatherall, British Oceanographic Data Centre (BODC), UK

4.1.1. What is ArcGIS Desktop Software?

The software company, ESRI (www.esri.com), produce a number of Geographic Information System (GIS) products suitable for use on PCs and for serving and accessing GIS data over the internet and via mobile devices. Further information about available products can be found on ESRI's web site: http://www.esri.com/products/index.html

The information below concentrates on the use of the ESRI **ArcGIS Desktop product suite**, designed for PC-based GIS work. As part of its functionality, the software can be used to

- View, overlay, edit and interrogate/analyze and capture geo-referenced data
- Produce maps and imagery
- Grid data
- Carry out datum and projection transformations

There are a number of **licensing levels** for the ArcGIS Desktop products. The functionality of the software depends on the licensing level and is also reflected in the price of the software. Some of the products are available **free of charge** but have much reduced functionality compared with the commercially-available products.

DRAFT

Overview of ArcGIS Desktop products

Within the ArcGIS Desktop product suite there are three main commercially-available packages:

- **ArcView** – the basic level product, allowing data visualisation, simple query and analysis functions, and also the ability to create and edit simple geographic features.
- **ArcEditor** – extends the functionality of ArcView to include additional tools to create, edit, and carry out quality checks on your data.
- **ArcInfo** - includes all the functionality of ArcEditor and adds more advanced spatial analysis, data manipulation, and cartography tools.

ArcInfo is at the highest licensing level and therefore has the most functionality. In addition, there are 'extension' packages (such as: **3D Analyst** and **Spatial Analyst**) which give access to more tools and functionality in addition to the basic ArcGIS Desktop packages.

Free software packages are also available

- **ArcReader** - is a free, desktop mapping application for viewing maps and images authored in other ArcGIS Desktop products. You cannot import data from other sources. The software can be downloaded from ESRI's web site: http://www.esri.com/software/arcgis/arcreader/index.html
- **ArcGIS Explorer** - is a free GIS viewer that lets you explore and visualise GIS information. In addition to ESRI format files, you can import imagery in the form of GeoTIFFs, JPEG and KML and also WMS. The software can be downloaded from ESRI's web site: http://www.esri.com/software/arcgis/explorer/index.html

4.1.2. Is the software free, what licenses are available and how can you access the software?

As described above, there is a charge made for the majority of the ArcGIS Desktop packages (**ArcInfo**, **ArcEditor** and **ArcView**) and the amount may depend on the licensing level of the software and the number of licenses purchased. For the most up-to-date information, consult the 'Pricing' section for each product on ESRI's web site: (http://www.esri.com/products/index.html#desktop_gis_panel) and/or contact ESRI directly (http://www.esri.com/about-esri/contact.html).

ArcReader and **ArcGIS Explorer** software are supplied free of charge but do not have the full functionality of the commercially available ArcGIS Desktop packages. More information about the levels of functionality is given in the overview of ArcGIS Desktop products: http://www.esri.com/software/arcgis/about/gis-for-me.html

Information about accessing the various software packages can be found on ESRI's web site: http://www.esri.com/products/index.html#desktop_gis_panel

DRAFT

4.1.3. What type of computer platform does it run on?

ArcGIS Desktop products are designed for use in a MS Windows environment. Details about supported operating systems and hardware and software requirements can be found on ESRI's web site: http://resources.arcgis.com/content/arcgisdesktop/10.0/arcgis-desktop-system-requirements

4.1.4. What does the software do?

As outlined above, the software can be used to visualise, edit, capture and analyse geo-referenced data.

Firstly, we need to look at what the various component packages of the ArcGIS Desktop products do and what types of data and data formats they can work with. The following is a broad overview of their functionality, further information can be found in the product's documentation.

ArcGIS Desktop application packages - what are they and what do they do?

The main packages are: **ArcMap** and **ArcCatalog** which are included with the ArcView, ArcEditor and ArcInfo ArcGIS Desktop products.

ArcMap is used for creating, displaying, exploring and editing your data; assigning styling to the display and for building images and map layouts from individual data layers or data sets.

ArcCatalog lets you organise, preview and manage your data sets. It acts like MS Windows Explorer in providing an overview of the contents of the directories on your PC. However, it displays the information in a form that is useful for working with ESRI software, i.e. although a shapefile is made up of at least three separate files, within ArcCatalog it is just represented as one file. ArcCatalog also lets you view and enter metadata for your data sets in a number of standard formats.

ArcToolbox is a collection of data processing tools collected into related 'toolboxes' and 'toolsets'. They have a wide range of functionality, including data analysis and conversion tools. The availability of the various toolboxes and toolsets depends on the level of your ArcGIS Desktop license.

The packages, **ArcScene** and **ArcGlobe** are part of the 3D Analyst extension package and can be used to produce animations (ArcScene) and display of data on a globe (ArcGlobe).

Further information can be found in the product documentation on ESRI's web site: http://resources.arcgis.com/content/product-documentation?fa=listDocs&PID=43

DRAFT

In what formats and structures does ESRI ArcGIS Desktop software store data?

ESRI ArcGIS Desktop software works with and stores data in a number of ways, as individual data files or as collections of related data files. The data sets may be stored individually on disc or be part of a database. The data sets can be in the form of 3D surfaces (e.g. rasters), vectors (e.g. points, lines or polygons) or images.

The following outlines the main data storage types used by ESRI's ArcGIS Desktop software products.

Geodatabases

This is the main data model for ArcGIS and is the primary data format used for editing and data management. It is made up of a collection of files (feature classes, raster data sets and tables). **Feature classes** are collections of features of the same type, such as points, lines or polygons. They also share a common set of attribute fields. Within a Geodatabase they can be grouped into **feature data sets** which are collections of feature classes with the same geographic reference information (i.e. datum and projection).

Geodatabases work across a range of database management system (DBMS) architectures and file systems, can be of various sizes, and be single or multi-user.

Within some ArcDesktop GIS packages, storing your data in a geodatabase allows you to carry out advance editing and quality control procedures on vector data sets, i.e. looking at how the vector data sets which make up the geodatabase 'interact' or share geometry – this is called topology. For example making sure the outlines of individual polygons do not overlap. This functionality is not available for shapefiles. For information about geodatabase topology can be found on ESRI's web site:
http://webhelp.esri.com/arcgisserver/9.3/dotnet/index.htm#geodatabases/topology_basics.htm

Further information about geodatabases can be found on ESRI's web site:
http://resources.arcgis.com/content/geodatabases/9.3/about

Shapefiles

The Shapefile is a format that ESRI software uses to store the shape, geometric location and attribute information about a data set. Individual shapefiles can contain only one feature 'type', i.e. points, lines or polygons. Shapefiles can be edited but they cannot be used as part of a topolygolgy for more advanced quality control procedures.

Further information about shapefiles can be found on ESRI's web site:
http://downloads2.esri.com/support/whitepapers/mo_/shapefile.pdf

3D Surfaces

ESRI products work with the concept of three types of surface models:

- **Rasters** – rectangular array of grid cells (holding a data value) arranged in rows and columns

- **Triangular Irregular Networks (TINs)** - a surface consisting of nodes (or points) joined by edges into various sizes of triangles. Because the nodes can be irregularly spaced, a TIN can show higher resolution information where there is more variation in terrain and be lower resolution where there is less variation.

- **Terrain surfaces** – are a set of 'pyramided' TINs, i.e. a set of TIN's at multiple levels of resolution. This helps when working with high volume data sets.

Find out more about how ArcGIS software works with 3D surfaces:
http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=3483&pid=3482&topicname=About_3D_surfaces

DRAFT

How do you work with your data in ArcGIS Desktop packages – do you have to reformat it?

As described above, ArcGIS Desktop packages work with data in a number of forms, as individual 'stand-alone' files or as part of a geodatabase.

Data in some formats can be imported or used directly with ArcGIS Desktop applications, others need to be converted to ESRI formats.

Information about supported data formats for import and export can be found on ESRI's web site:
http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Data_formats_supported_in_ArcGIS

Through the **ArcGIS Data Interoperability Extension** tool set, data can be imported, exported and converted to a number of formats.

Further information can be found on ESRI's web site:
http://www.esri.com/software/arcgis/extensions/datainteroperability/index.html

Further information about importing data is given below:

Importing an existing raster

You can convert existing raster data sets into an ESRI raster, the data can be in ASCII format, binary (float) or as a US Geological Survey DEM. The conversion can be done through the ArcToolbox 'Conversion tools – to raster' options. Further information about this and how to use this tool can be found on ESRI's web site:
http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=An_overview_of_the_To_Raster_toolset

In addition to creating rasters from input point data sets, you can also create rasters from polyline and polygon data sets. Further information is available at the above link.

Connecting to Open Geospatial Consortium (OGC) Web Services

You can connect directly to Web Map Services (WMS), Web Coverage Services (WCS) and Web Feature Services (WFS). Further information can be found on ESRI's web site:
http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?topicname=overview_of_ogc_and_iso_support

Working with KML files

Information on working with KML files can be found on ESRI's web site at:

DRAFT

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?topicname=an%20overview%20of%20kml%20support%20in%20arcgis

Adding X Y data

Point data in ASCII format can be imported into ArcMap as a layer using the **Add XY** option, attribute information can also be imported along with the data. The data can then be exported as a shapefile or feature class.

Additional tools for data import

There are 'add on' tools such as ETGeowizards (http://www.ian-ko.com/ET_GeoWizards/gw_main.htm) which can be used to help with the import and export of vector and raster data sets for use with ArcMap.

Managing your data sets

ArcCatalog is the ArcGIS Dektop package for managing and organising your data sets; looking at the contents of directories and database; creating new data files; previewing data sets and managing your data sets' metadata. Further information can be found on ESRI's web site:

http://webhelp.esri.com/arcgisdesktop/9.3/pdf/arccatalog_tutorial.pdf

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Workspace_and_dataset_management_with_ArcCatalog

The following are examples of some simple tasks that you can carry out, for the full range of functionality see the above documentation.

Previewing your data

The Preview tab lets you explore your selected data set without first having to create a map in ArcMap – allowing a quick overview or 'snap shot' of your data set. You can select 'Geography View' to look at the geographic coverage of your data set. For data sets that contain both geographic data and attribute information, you can also use 'Table' view to look at the attribute data.

How can I check the metadata associated with my data set?

Navigate to the location of your data file in ArcCatalog, highlight the file by clicking on it, then in the right-hand display window click on the 'Metadata' tab. You can select the style sheet used to display the information by clicking on the stylesheet drop down list on the main toolbar.

Updating or creating metadata

DRAFT

You can update existing metadata, create new metadata records and/or import and export metadata in a number of recognised standards. More information about working with your data sets metadata in ArcGIS can be found on ESRI's web site:

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=About_metadata

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Exploring_an_item's_metadata

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Importing_and_exporting_metadata_with_ArcCatalog

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Editing_metadata

Can you create a 3D surface from your source data sets?

DRAFT

The following outlines how you can use ESRI's software to create a raster or Triangular Irregular Networks (TIN) from your input data.

Creating a raster from point data

As part of the ArcGIS 3D Analyst extension software package you can create a raster from a data set of points using a number of interpolation methods:

- Inverse distance weighted
- Spline
- Kriging
- Natural neighbours

Firstly, you will need to convert your point data set to an ESRI format such as a shapefile or feature class. Then using the 3D Analyst package you can generate a raster, defining the parameters to be used during the gridding process is done through a dialog box. The example below shows how this works.

Further information about creating rasters can be found in the ArcGIS 3D Analyst product documentation, available from ESRI's web site:

http://downloads2.esri.com/support/documentation/ao_/775Using_3D_Analyst.pdf

How to generate a TIN

Through a dialog box, you can create a TIN from your input data using ArcGIS 3D Analyst extension routine create/modify TIN.

Further information can be found in the product documentation:

http://downloads2.esri.com/support/documentation/ao_/775Using_3D_Analyst.pdf

Converting between rasters and TINs

Through the 3D Analyst extension toolset there are tools available to convert between rasters and TINs. Further information can be found in the 3D Analyst documentation.

Working with 3D surfaces

The following are examples of some simple tasks that you may want to carry out, for the full range of functionality see the ArcGIS Desktop documentation.

Analyzing surfaces and generating contours

You can analyse the surface of your 3D model, looking at slope and aspect and hill shade and also **generate contours** at specified intervals. This can be done through the 3D Analyst extension. Further information is given in the documentation.

http://downloads2.esri.com/support/documentation/ao_/775Using_3D_Analyst.pdf

Calculations on rasters

You can carry out mathematical operations on cells in a raster through the 3D Analyst raster math tools. This includes carrying out calculations on grid cells such as add, subtract or divide. This can be done by either adding or subtracting one grid from another or by adding or subtracting single values for each grid cell.

The following table lists the available toolsets in 3D Analyst and gives a brief description of each:

http://resources.esri.com/help/9.3/arcgisengine/com_cpp/gp_toolref/geoprocessing_with_3d_analyst/an_overview_of_the_3d_analyst_toolbox.htm

Some tools are also available through the Spatial Analyst Extension raster calculator:

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?topicname=the%20raster%20calculator

Extracting data from rasters

DRAFT

You can extract data from rasters either based on geographic area 'clipping' or by cell values (attributes). For example, you may to select a geographic sub-area from a grid or select cells from a grid with a depth greater than say 100m.

The majority of the extraction tools are available with the **Spatial Analyst Extension**; however, the 'Clip' tool is available with any ArcGIS license.

Further information about the 'Clip' tool can be found at:

http://resources.esri.com/help/9.3/arcgisengine/com_cpp/gp_toolref/data_management_tools/clip_data_management_.htm

For information about data extraction as part of the spatial analyst extension see:

http://resources.esri.com/help/9.3/arcgisengine/com_cpp/gp_toolref/spatial_analyst_toolset_overview/an_overview_of_the_spatial_analyst_toolbox.htm

http://resources.esri.com/help/9.3/arcgisengine/com_cpp/gp_toolref/spatial_analyst_toolset_overview/an_overview_of_the_extraction_toolset.htm

Resampling, generalising and filtering rasters

A set of resampling, generalisation and filtering tools is provided as part of the ArcGIS Desktop products.

Resampling a raster, i.e. changing the size of the grid cells can be done through the 'resample' tool, part of the Data Management Toolbox:

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=1928&pid=1922&topicname=Resample_(Data_Management)

The ArcGIS Spatial Analyst extension contains a number of tools for generalising and filtering rasters:

http://resources.esri.com/help/9.3/arcgisengine/com_cpp/gp_toolref/spatial_analyst_toolset_overview/an_overview_of_the_spatial_analyst_generalization_toolset.htm

Viewing and editing your data in ArcMap

ArcMap is the ArcGIS Desktop package which can be used to view and edit your data sets.

Further information can be found on ESRI's web site:

http://webhelp.esri.com/arcgisdesktop/9.3/pdf/arcmap_tutorial.pdf

DRAFT

An overview of displaying and visualising data in ArcMap is given at the following links:

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?topicname=mapping_and_visualization_in_arcmap

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Layers,_data_frames,_and_the_table_of_contents

The following is a brief overview of some of simple functions that you might want to do.

Viewing data sets

To view a data set, it needs to be in a format which can be recognised by ESRI ArcGIS Desktop software (see above for more information). Click on the 'Add Data' button from the ArcMap toolbar, then search your data directories to select the file that you want to load.

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Adding_layers_to_a_map

Once a data set is loaded you can look at changing how the displayed layer looks – altering the colours or symbols used to display the data and adding label and text information. This can help to highlight aspects of the data set that you want to enhance, e.g. displaying data in a particular depth range all one colour.

The layer will appear in the layers list in the left hand panel of the display window. Right click on the file name and then click on 'properties'. For raster data sets, you can choose to display the grid using stretched values along a colour ramp, you define the start and end colour values, using values and classified intervals, e.g. equal intervals, natural breaks, defined intervals etc. or unique values for each separate cell attribute. You can define a colour scheme, choosing from pre-existing colour schemes or defining and saving your own colour scheme.

You can change the properties of individual colour intervals – perhaps changing one interval to be transparent – this helps with overlaying data sets. You can also make a selected range of values one colour – say for land areas.

Further information about choosing display colours can be found in 'help' system which accompanies ArcMap under 'Displaying rasters'.

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=3157&pid=3144&topicname=Improving_the_display_of_raster_data

For vector data, you can change the symbology used to display the data set to get the most out of your data set. Symbols can be defined on size of attributes and/or colour and attribute information can be displayed on the map.

DRAFT

For example, when plotting sounding points – you can colour code them for attribute information such as depth or for contributing organisation. Overlaying source sounding point data on grids can help to identify any artifacts in the gridded data set created from bad data points.

Overlaying data sets

ArcMap displays your data sets as layers. You can change the drawing order of the layers, for example to plot points on top of a map.

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=Managing_group_layers

Adding images

You can images in a number formats into ArcMap. Further information about recognised formats can be found in the documentation:

http://www.esri.com/library/brochures/pdfs/arcgis10-desktop-functionality-matrix.pdf

If needed, you can **georeference** the image, adding geographic location information, using the routines which are part of the Georeferencing toolbar.

http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?id=3153&pid=3144&topicname=Georeferencing_a_raster_dataset

Capturing and editing vector data

You can create and edit vector data, changing their attribute information and also the shape and extent of features. There are a number of tools which you can use to carry out these procedures. Further information can be found on ESRI's web site:

http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=An_overview_of_editing_and_data_compilation

Creating a profile plot

Using the 3D Analyst Extension tools in ArcMap you can create a profile plot across your raster data set. Further information can be found on ESRI's web site:

http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?id=2950&pid=2943&topicname=Creating_profile_graphs

Can you create a plot or map from your data?

As discussed above, in ArcMap you can build up layers and define the colours that you want to use to display your data.

DRAFT

The layout view enables you to add in additional information such as a map graticule, legend and title. See the 'Working with map elements' section of the ArcMap tutorial:

http://webhelp.esri.com/arcgisdesktop/9.3/pdf/arcmap_tutorial.pdf

How can I get help with using the software?

There are documents and links on ESRI's web site under 'support' http://support.esri.com/en/

ArcDesktop tutorials: http://resources.arcgis.com/content/product-documentation?fa=viewDoc&PID=43&MetaID=1403

Product documentation: http://resources.arcgis.com/content/product-documentation

Techincal arcticles: http://resources.arcgis.com/content/kbase

Tip and tricks: http://downloads2.esri.com/support/documentation/ao_/arcgis-desktop92-tips_1006.pdf

About ArcGIS: http://webhelp.esri.com/arcgisdesktop/9.3/pdf/what_is_arcgis.pdf

ArcGIS DeskTop help:
http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=welcome

Getting started tutorials:
http://webhelp.esri.com/arcgisdesktop/9.3/tutorials/getstart/getstart_tutorial_index.htm

Commonly used tools -
http://webhelp.esri.com/arcgisdesktop/9.3/index.cfm?TopicName=An_overview_of_commonly_used_tools

Geoprocessing tools - quick guide
http://webhelp.esri.com/arcgisdesktop/9.3/pdf/Geoprocessing_Quick_Guide.pdf

DRAFT

4.2 CARIS HIPS (Hydrographic Information Processing System)

Contributed by Cmdr. Izabel Jeck, Directorate of Hydrography and Navigation (DHN), Center of Hydrography (CHM), Brazilian Navy, Brazil

4.2.1 Introduction

The purpose of this document is to present to the user the basic workflow of the program CARIS HIPS (*Hydrographic Information Processing System*), version 7.0, to process multibeam data.

HIPS is a suite of comprehensive hydrographic data processing tools that can be used to process simultaneously multibeam, backscatter, side scan sonar, LiDAR and single beam data. CARIS HIPS is a program developed by a Canadian company named CARIS (www.caris.com).

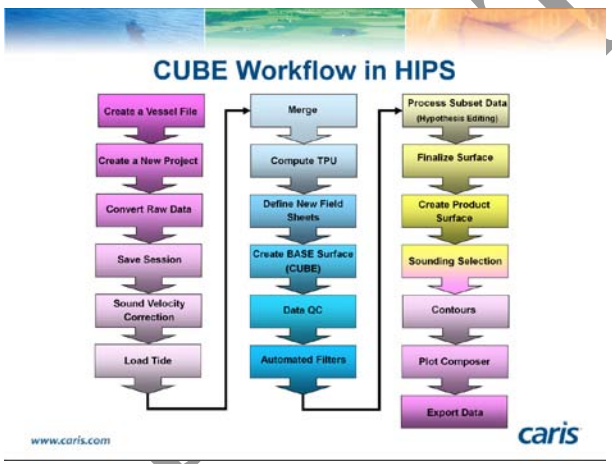
To use this software it should be note the computer minimum configuration, this information and any extra information can be obtained in the software documentation or in CARIS web page.

4.2.2 Acknowledgements

This document was made with the contribution and the authorization of CARIS - Universal System Ltd. Company, and further information can be obtained in www.caris.com.

4.2.3 HIPS Workflow

The figure below presents the software workflow.



Create a Vessel File (HVF): Setup the sensor locations and uncertainties in the vessel reference frame.
Create a New Project: Setup the *Project - Vessel - Day* data structure.
Convert Raw Data: Raw data is converted into HIPS data format.
Save Session: Save the current workspace (data and current view).
Sound Velocity Correction: Load and edit sound velocity profiles and apply the correction.
Load Tide: Load tide data from one or more tide stations.
Merge: Combine vertical and horizontal information to produce geo-referenced data.
Compute TPU: Utilize uncertainty values entered in the HVF in an effort to compute the total propagated uncertainty of each individual sounding.
Define New Field Sheets: Define the map projection and location of the data.
Generate BASE Surface (CUBE): Merged data is used to produce CUBE surface.
Data QC: Sensors, such as navigation, gyro, heave, etc. are examined if problems have been identified in the BASE surface.

Automated Filters: Filter soundings using swath geometry and/or according to IHO survey order accuracies.
Process Subset Data: Validate CUBE surface and edit the geo-referenced soundings directly, on many lines simultaneously, where the CUBE surface has been adversely affected by erroneous soundings.
Recompute: Update the BASE Surface after the data has been edited and the surface has been validated.
Finalize: Finalize BASE Surface to ensure designated soundings are carried through to bathymetric products.
Create Product Surface: Produce a generalized product surface from the BASE Surface.
Contours: Use either the BASE Surface, product surface or tile set to output contours.
Sounding Selection: Use a height source for selection of a representative sounding set.
Export Data: HIPS soundings and surfaces can be exported to various formats for data transfer.

Before start the processing the user must know:

- dX,dY,dZ sensor offsets
- ecosounder used and number of beams
- sensor timing corrections
- sensor bias corrections
- positioning format (geographic or ground coordinates recorded?)
- have calibration offsets been applied?
- has vessel motion been removed?
- has remote heave been removed?
- has draft been applied?
- has sound velocity corrections been applied?

4.2.4 Project Data Tree and General Information

The Control Window displays the data in an expandable tree. This display is organized to reflect the Project, Vessel, Day, Line directory structure. Data deleted from the Project tree is sent to the Recycler folder as defined in an Environment variable in the Options dialog.
Project information can be viewed in the **Project Properties** dialog box. General project information and information about the coordinate system is stored here. The project geodetics can be transformed using this dialog box if required.

Default Options
- Presentation and Performance
The default options can be accessed using the **Tools > Options** menu. The **Options** dialog box is organized into six tabs.
The **General** tab controls zoom and pan factors, and settings for displaying sounding size, the SVP location symbol and the navigation point symbol. The pick aperture setting determines how close the cursor has to be to an item in order to select it. A number of other options can be enabled or disabled in the General tab.
The **Display** tab controls the colour of lines and backgrounds in different parts of HIPS and related editors.

DRAFT

Under the **Display** tab the user can configure the presentation of Geographic or Projected Grids & Scale Bars for the HIPS Display window. Grids, which can be displayed as lines, ticks or crosses, will be displayed, with a scale bar, appropriately for the current zoom factor. Under the **Units** section users can configure the display units for their data.
All data in HIPS is stored in metres and so the conversion will be applied to the display of the data.

- Directories and Environment
The **Directories** tab enables the user to modify the location of project dependent data files, either one at a time or by using the **Root Path,** to set all the HIPS and SIPS data paths at once, if they all have the same base path. This information is saved to the registry.
The **Environment** tab enables the user to modify the locations of the HIPS and SIPS support information, such as the master file, symbol file, datum file, scratch location, etc.
By default, this information is located in the ..\CAIRS\HIPS\System directory.

- S-52 and S-57
The **S-52** tab controls the display of S-57 ENC data, there are options to select colours and transparency for depth areas, symbols, contours etc.
The **S-57 Environment** tab controls the location of S-57 dependent configuration files.

– 3D View
The **3D** tab controls the behavior of the keyboard and mouse controllers when navigating the 3D View:
• **Controller Type**
- Terrain Flyer: navigate from the perspective of the height source.
- First Person: navigate from a camera view of the current 3D scene.
• **Flight Speed:** Controls the speed at which you move through the display during a fly-through.
• **Zoom Speed:** Controls how fast the view is changed when using the zoom controls.
• **Smooth-Fly-To Speed:** Controls how fast the view is changed when a user double-clicks an area of the 3D View.

DRAFT

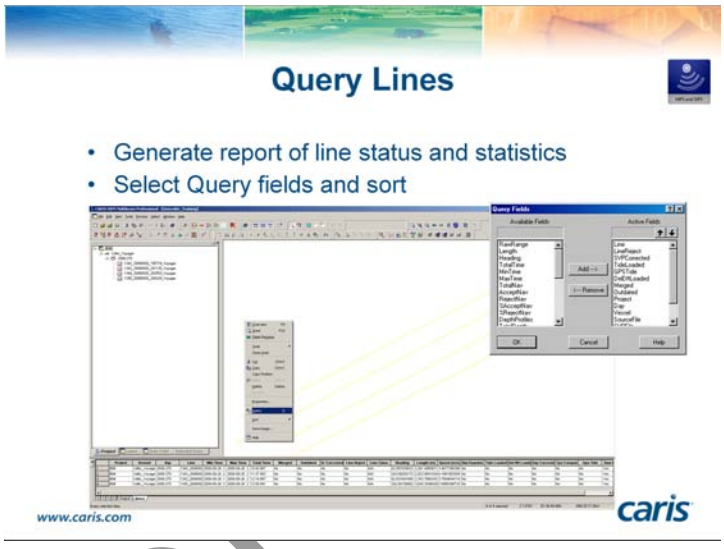
<u>Open Background Information</u>
Both vector and raster information can be loaded into HIPS as background information. This option can be accessed using the **File > Open Background Data.**

<u>Layers Data Tree</u>
The **Control Panel** displays the data in an expandable tree. This display is organized to reflect the layers of data displayed in the HIPS Display Window. It is possible to toggle these layers on and off using the check boxes in the **Layers** tab, and the draw order of the layers can be changed by dragging and dropping the layers in the **Draw Order** tab.
Entire Fieldsheets can be turned off from the right-click context menu.
In order to select data such as track lines in the Display Window**,** the layer must first be highlighted in the **Layers** tab of the Control Window.

<u>Selecting Data in HIPS and SIPS</u>

Survey lines can be selected by highlighting them in the Project tab of the Control Window, or by clicking on them in the Display window. Multiple lines can be selected using the shift and control keys. Lines can also be selected by dragging a box over the lines in the Display Window.
Selection in the Display Window can be done in two ways: **Select by Range** or **Select by Lasso.**
Lines cannot be selected when the cursor is in constant zoom mode. When in constant zoom mode. To exit from zoom mode, click on the zoom icon or use the ESC key.
Other selection options are **All in Display, All and Clear Selection.**
Note: The Select by Range tool is depressed by default. If you are unable to select lines in the Display Window check to see that this tool is selected.

Query Lines



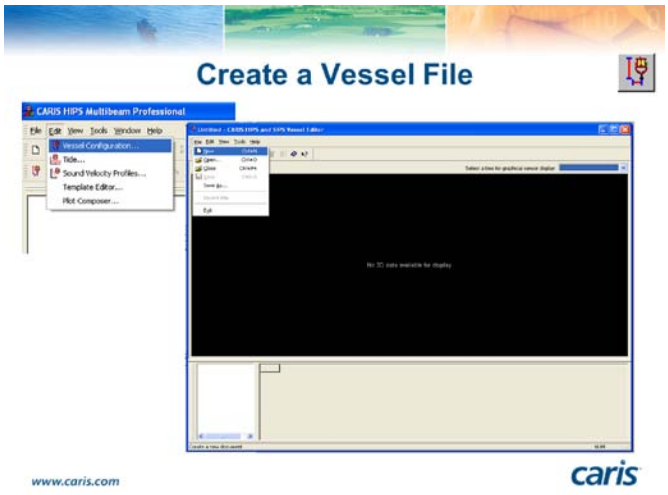
All the information listed when a line is queried is calculated on-the-fly so it may take a few minutes to list. The data is displayed in the Worksheet Window and the columns can be sorted by clicking on any of the headings. By right-clicking on the headings and selecting **Query Fields…** from the pop-up menu you can choose which fields are displayed. The options **Select All**, **Copy** and **Paste** and **Save As…**allow the contents of the Worksheet Window to be taken into software like Excel.
Right-click in the Display Window brings up a pop-up menu you can use to copy the geographic coordinates of the cursor position to the clipboard.

4.2.5 HIPS Workflow- Detailed

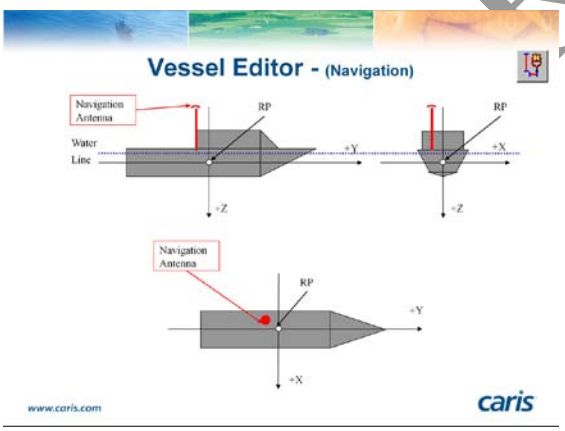
Vessel editor
The Vessel Editor is used to create and edit HIPS Vessel Files (HVF). The HVF contains information necessary for combining all sensor data to create a final position/depth record. Sensor information is used during the merge process. All sensor entries are time stamped, therefore one HIPS vessel file can be used for the life of a vessel. If a sensor is moved or added, the new offset information is given a new timestamp. The merge process compares the date and time of the observed data with the date and time of the HVF sensor information to ensure that the appropriate offsets for that time are used.

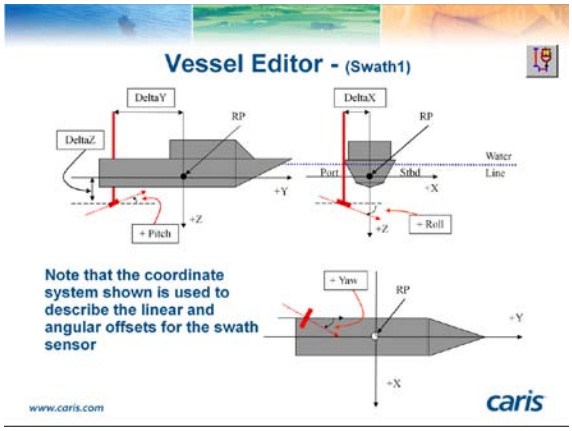
Note:
It is not possible to process sensor data time stamped prior to the earliest HVF timestamp, as the program applies the most recent information.
The positions of the sensors in the HVF may not be the same as the actual configuration of the vessel. For example, if sensors offsets are applied in the acquisition software, the HVF should describe the 'corrected' position, not the actual position of the sensor.

The vessel editor can be accessed using the edit > vessel configuration…from the main menu.

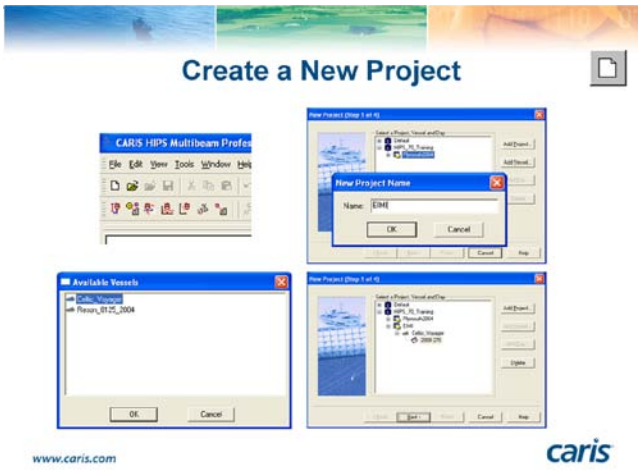


The figures below present HIPS coordinate system to be considered in the VESSEL EDITOR - SENSOR SECTIONS)





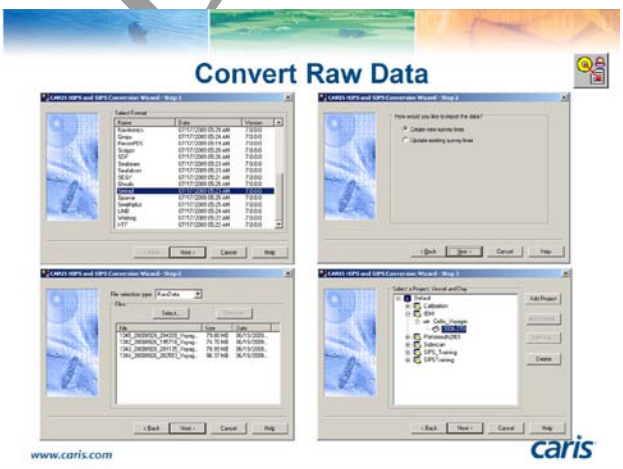
Create a New Project



A list of existing projects will be displayed in the option **file > open project**, you can also create a new project using **file > new.**
Note: Do not use spaces in the name of the project.
In the **file > new,** highlight the project created, click the **Add Vessel…** button and select the hvf vessel file; click the **Add Day…** button. Select the geographic parameters for the project.
Note: Regions of the world appear in the **Group** list. When a group is selected, the user can select from a pre-built set of map **Zones**. These parameters are referenced by the Map Definition File (Hips\System\mapdef.dat). This file can be appended to if additional definitions are required or another file can be selected by accessing **Tool > Options** from the main HIPS menu and then selecting the **Environment** tab.
The user may also opt to have the program automatically select the appropriate UTM zone.
Define a geographic window for the project. The default region is the entire globe. If an existing project in the area is open, the **Current View** button can be used to get the extents.

DRAFT

<u>Convert Raw Data</u>



- Select the **Conversion Wizard** icon or **File > Import > Conversion Wizard...** from the menu to start the conversion wizard.
- Select **data format** from the list of formats.

- Select the **File selection type** from the list, in this case **RawData**
- Select all the lines.

Raw data files are usually stored in PreProcess so this is the default location.

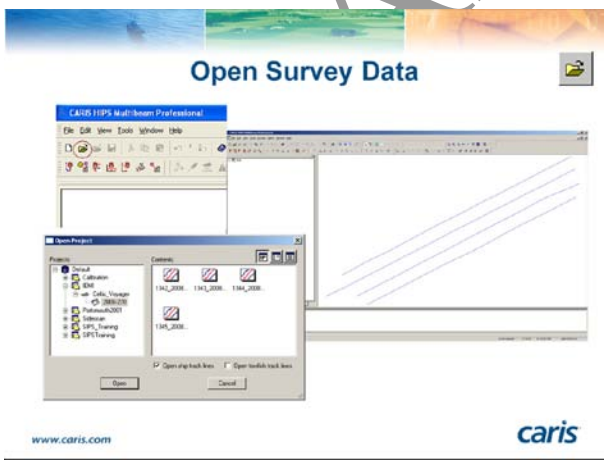
- Select the destination Project, Vessel, and Day folder for the converted
- Select the geographic reference system used for the acquisition of the raw data. If Ground Coordinates are used, the appropriate zone definition needs selecting from the list. If Geographic Coordinates are used nothing more needs to be set.
- Do not set **Navigation** or **Depth** filters. The **Navigation** filter is useful to restrict data to a specific geographic region. The geographic region can be defined manually or by Project File extents. The **Depth** filter is used to reject large bathymetric spikes that fall outside of the min and max depths of the survey area.

The first five steps of the Conversion Wizard are the same for all data formats. From Step 6 the wizard differs to accommodate individual requirements for each format type.

The **<Back** button can be used to check that no mistakes have been made in the settings. Once the line data has finished converting the **<Back** button changes to **<< Restart**.

DRAFT

Open Survey Data

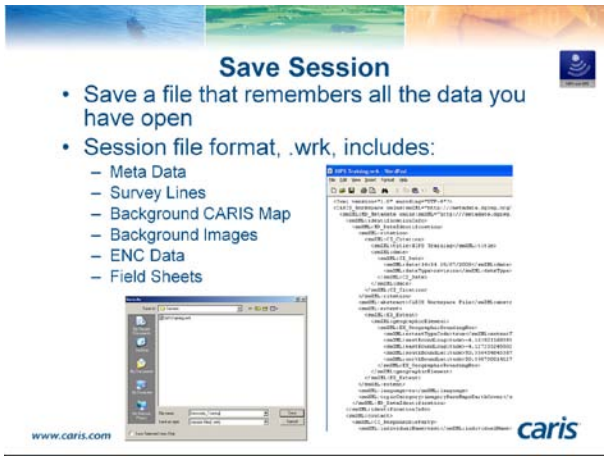


Once converted the data can be opened in the main interface.

The lines will be listed in the Contents window.

If side scan data was converted and resides in the same files as the bathymetry data then the tracklines showing the vessel and towfish positions should both be selected.

Save Session



The session file is a text file located in the HIPS\Session directory. It consists of a series of data objects in XML format describing:

• **Workspace** – Creator of the session, HIPS project name, creation and last modification times, and location of the HIPS source data.

• **View Extent** – The last used window size and location.

• **Project Data** – Survey lines, CARIS files, background images, and field sheets.

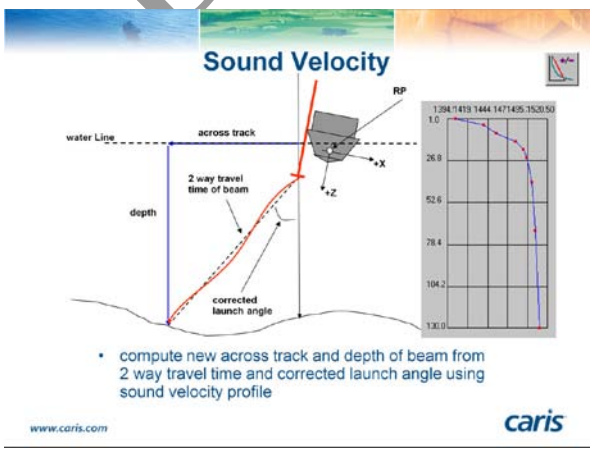
Double clicking a session file in Windows will automatically open the session in HIPS.

Sessions should be used to efficiently organize the data that you are working with.

Having a series of small work areas as separate sessions instead of one large project will allow for faster, more effective data access.

DRAFT

<u>Sound Velocity</u>



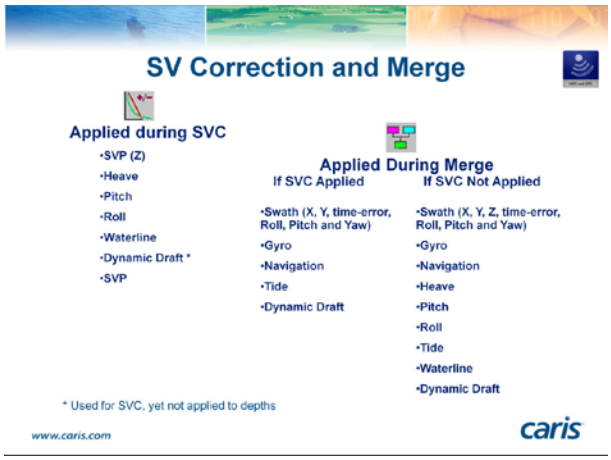
Raw data formats such as Simrad contain two-way travel time and beam launch angles.

Producing a geographically referenced sounding position and depth from this data is a two-stage process.

• The procedure for calculating the length and path of the sound wave through the water column for each beam is called the Sound Velocity Correction (SVC).
The result is an along-track/across-track/depth for each beam.
• The Merge process (which is explained later) converts the along-track/across-track/depths into latitude, longitude, depth by combining the ship navigation with horizontal and vertical offsets in the HIPS vessel file.
The SVC algorithm calculates the ray path of the sound wave through the water column for each beam. In order to do this the program needs to know several things:
• Any rotations that have to be applied to the sensor head – both static (i.e., setup and calibration parameters) and dynamic (measured by the attitude sensors).
• The acoustic velocity of the water column. This information is loaded into the program as a sound velocity profile.
There are different options for applying velocity information to a swath:
• **Previous in time** – This is the method traditionally used by HIPS, which applies the profile taken immediately before the collected swath.
• **Nearest in time** – In this situation HIPS will apply the SVP with the time stamp closest to that of the collected swath.
• **Nearest in distance** – Uses the position attributes of the profile to determine the nearest profile to a given swath.
• **Nearest in distance within time** – Uses the position attributes of the profile to determine the nearest profile to a given swath within a time window which the operator has to enter.

Note:
The profile selection methods listed above are made on a swath-by-swath basis and so it is possible to have several SVPs used in the same survey line.
For Simrad data, the sound velocity and attitude corrections have been applied during acquisition, and do not need to be reapplied.
The process checks the HVF to see if the heave, pitch, and roll are set to Apply "Yes". Since Simrad has already corrected for Sound Velocity, it does not have to be re applied. Doing so would skew the data.
Note that in this case there is no sensor called SVC to the Vessel File (hvf). This sensor would need to be configured in order to apply (or reapply) sound velocity.

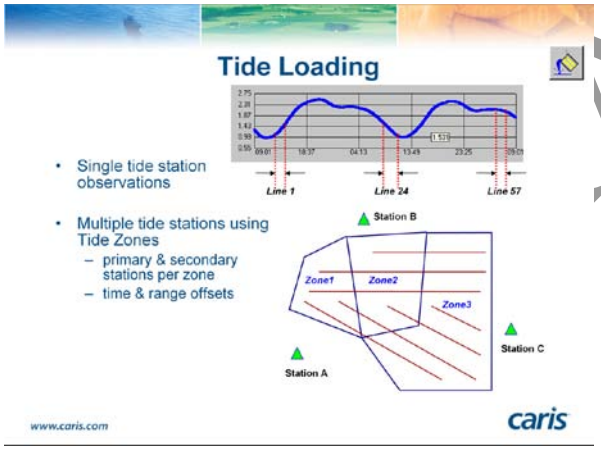
<u>Sound Velocity Correction</u>

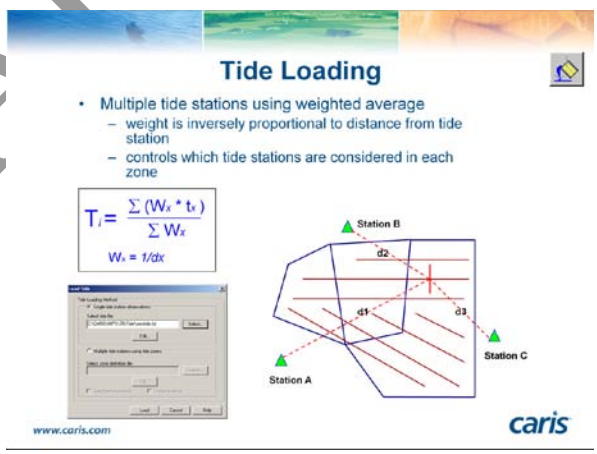


DRAFT

As stated earlier, some of the parameters that are usually applied in the merge process can be applied during the SVC process.
If you sound velocity correct your raw data to produce a new observed depth, the merge process knows what has been applied and will not reapply it as SVC is optional.
The following are taken into account during SV correction:
• Travel Time - angle observations
• Sound Velocity Profile
• All entries in the SVP section.
• Water line
• Dynamic heave, pitch and roll, along with any errors. Including pitch and roll induced heave calculated from MRU lever arm offsets (X, Y and Z of motion sensor)
• Dynamic Draft - uses a table of speed versus draft values in the HVF.
• The Dynamic Draft values are only used to find the proper sound velocity layer to start SVC. The draft values will be applied to the observed depths in the Merge Process.
Note: The dynamic heave, pitch and roll will only be applied if **Apply** is set to **Yes** in the HVF for each individual parameter. However, the offset values for Roll, Pitch and Yaw entered into the Swath/SVP will always be applied.

Tide Loading





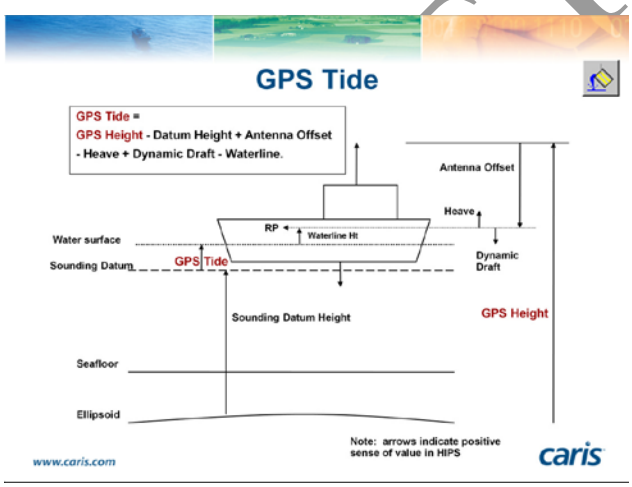
Tide data can be applied in two ways:
• The tide data from a single tide station can be applied to each survey line as a whole.
• Tide data from multiple tide stations can be compiled. The tide data is loaded for each part of a line as determined by the zone the part falls into.
To define a tide zone several parameters are required:
• Zone boundary in geographic coordinates.
• The location of the primary tide station.
• Locations of up to 3 optional secondary tide stations.
• Time offset in minutes for each station.
• Range offset / tide scalar for each station.
• Outage limit in minutes controls when data will be extracted from secondary stations.

• Interpolation interval in seconds for final interpolated tide loaded into each survey line data structure.
Note: When using data from multiple stations, a weighted average of tidal observations along the survey line is generated. This only works if the information about the tide coefficients is given in the tide zone file. The weight given to the tide data is inversely proportional to the distance between the station and the swath data, which is why the zone definition file should include latitude and longitude positions for tide stations.

To apply tide you have to select all of the lines and click on the **Load Tide** tool button or select **Process > Load Tide**.

The tide weighting formula can be expressed as the following;
Ti = The tide value for the ith profile along the line
tx = The tide value from the xth tide station for the time and location of the ith profile
Wx = The weight of the tide value coming from the xth tide station
dx = The distance between the ith profile along the line and the xth tide station.

GPS Tide

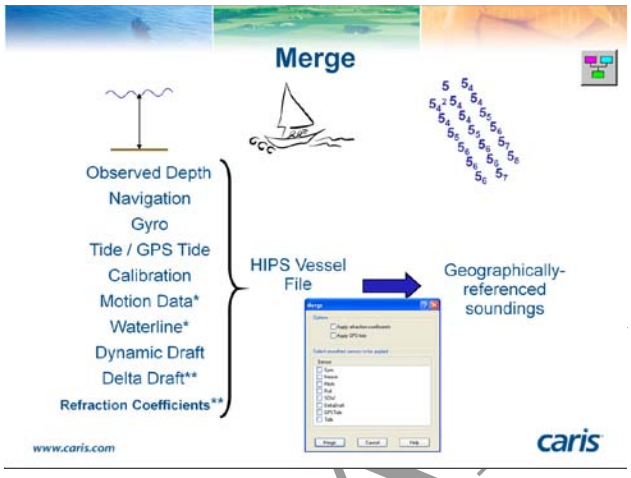


HIPS has the ability to use ellipsoid heights from GPS RTK instead of traditional tide data.
GPS Tide data is applied optionally as part of the Merge process. It is used as a direct replacement for regular tide data. The vessel configuration is used as normal with respect to transducer offsets, calibration values, application of heave, pitch, and roll, etc.
In order to apply GPS tide you first need to calculate it. This calculation is performed with the **Process > Compute GPS Tide…** command.
To be able to use the GPS height information for tidal corrections, the separation between the ellipsoid and the vertical survey datum needs to be known. When computing the GPS tide this separation can be entered as a single value or a model file can be used. Two binary sounding datum models can be used for computation of GPS tide:
• GEOID99
• EGG97 (European Gravimetric Geoid 1997)
Alternatively, an ASCII sounding datum model can be used. The data must be in the format Lat,Long,Z, for example: -33.848326,151.192435,20.4

Not in all cases raw GPS height will be recorded in the raw data files. The used acquisition system may already have made corrections on this data. To allow for this it, is possible to select which terms of the GPS tide computation should be applied and which not. If a sensor has already been applied on the GPS heights by the acquisition system, it should not be applied by HIPS. Only corrections that are selected in the Compute GPS Tide dialog will be applied in the computation.

Merge

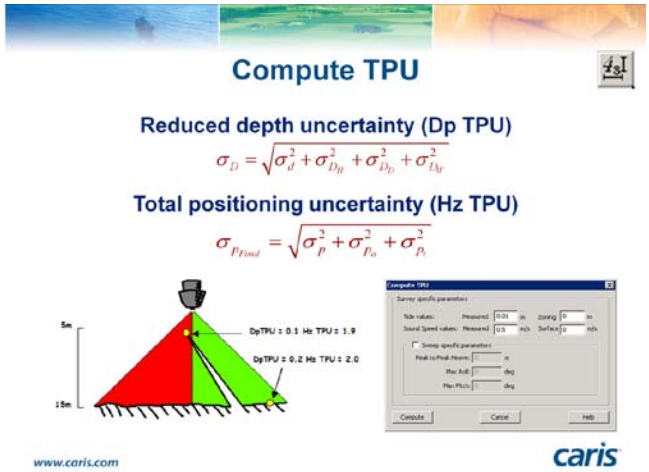
Merge takes into consideration all of the horizontal and vertical offsets in the HVF to produce a processed sounding.



Parameters marked with an * have been applied in the SVC process. Some additional parameters need to be selected as part of the Merge process. These include:

- Apply Refraction coefficients
- Apply GPS tide
- Smooth Sensor data

Compute TPU



In order to generate TPU values for each sounding, the definition of uncertainty estimates for

DRAFT

each of the contributing sensor measurements has to be combined using a propagation algorithm. The result is a separate uncertainty estimate for the depth, **DpTPU**, and horizontal position of the sounding, **HzTPU**. The results are presented at the 95% confidence interval which is equivalent to 1.96 x the standard deviation.
The compute TPU function must be performed after the Sound Velocity correction. Whether it is applied before or after Merge will depend on your adopted workflow. When you compute TPU it will always be done on the Observed depths, therefore tide and draught offsets will not affect the values.
This calculation is performed with the **Compute TPU** tool button or **Process > Compute TPU**.
The TPU values will be calculated for the soundings based on the errors defined in the HVF, the Total Propagation of Uncertainty algorithms and the survey specific uncertainty estimates defined.

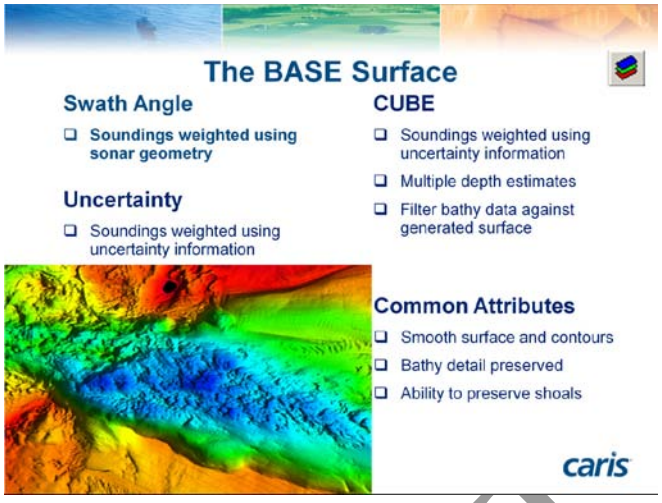
Defining New Field Sheets

Field sheets are essential for the creation of Weighted Grids and BASE Surfaces.
Field Sheets can be created in locations other than the directory directly under HIPS.
To create a New Field Sheet the user can use the **New Field Sheet** tool button or select **Process > New Field Sheet.**
- Enter a Name for the Field Sheet.
Note: Field Sheet names should be made up of alphanumeric characters only and should not contain spaces.
- Enter a Scale.
- Leave the default values for the **Horizontal** and **Depth Resolutions**. This will define the measurable accuracy of the coordinate and depth resolutions.
- Confirm the **UTM Zone / Mercator**
- Define Field Sheet extents by three ways:
1) Select the Current Display button to use the current display as the perimeter.
2) Define the extents with the mouse by clicking on the User Defined Area button.
3) Type in the coordinates for the bottom left and top right corners of the field sheet.
- Create a user defined Field Sheet large enough to accept all of the data in the survey area.

Once a new Field Sheet has been created it will appear as a new layer in the Layer tab of the Control Window. The display of the Field Sheet name and outline can be controlled by the Properties dialog box.
The Field Sheet icon in the Control Window is green, meaning that this is the active Field Sheet. If you have more than one Field Sheet open then it is possible to change which is set as active. This is achieved by right-clicking on the Field Sheet name in the Control Window and selecting the **Set as Active Field Sheet** option from the pop-up menu.

DRAFT

The BASE Surface

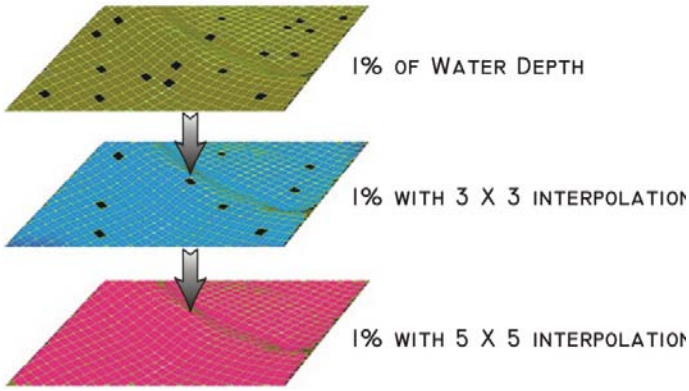


HIPS supports the ability to create three different surface types:
• the Swath angle surface,
• the Uncertainty surface, and
• the CUBE surface.
Through the BASE Surface Wizard the user can select which type will be created. The surface types use different weighting algorithms to produce gridded images of the sonar data. All three surface types will produce a smooth surface that retains the sonar resolution. The user also has the ability to preserve shoal or critical values.
The CUBE surface is also a powerful, semi-automated cleaning tool that can be used to increase processing efficiency. Both Swath and Uncertainty surfaces can also be used for cleaning to some degree.

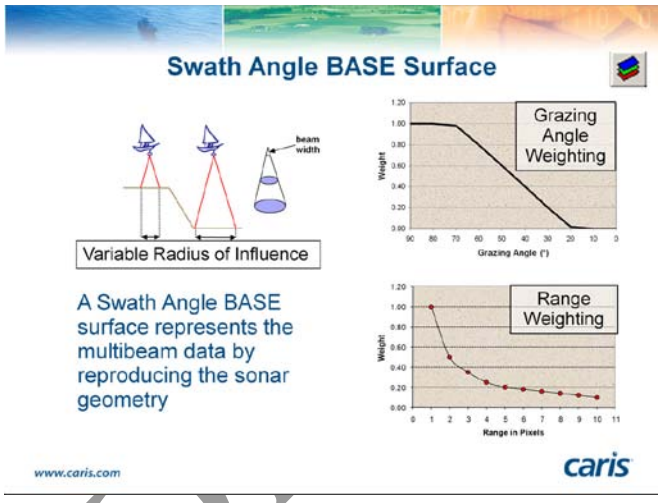
DRAFT

BASE Surface

When the cursor is moved over the BASE Surface, tool tips are displayed showing the value for the nearest Base Surface node. This option can be toggled on/off in the **Tools > Options** menu option under the **General** tab.
BASE Surface nodes can also be selected and queried. To query surface node values, enable the display of the surface in the Control Window and highlight that layer. Leftclick and drag to select an area of nodes. Select the **Edit > Query** menu option to have the XYZ values of the selected nodes displayed in the Query window.
Note: To obtain the maximum BASE Surface resolution it is often best to create a small BASE Surface and work upwards. If working with a single resolution the first BASE Surface you create may be 1% of the average water depth. The other option is to create a surface with a resolution that varies with the water depths of the survey area.
The initial BASE Surface may contain holes; the holes represent areas where the data was not dense enough for the resolution selected. In such areas you can interpolate to fill in the holes.



Swath Angle Surface – Theory



The Swath Angle BASE surface is a rasterization method specifically designed for multibeam data. It is an accurate representation of multibeam data because it considers the actual geometry of the sonar system. It does this in three ways:

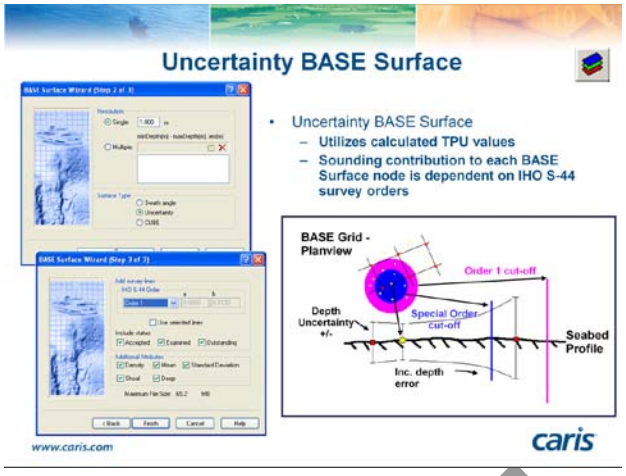
Variable radius of influence - The radius of influence of each sounding is calculated from the beam width of the selected sonar, and increases with depth and grazing angle. This addresses the increase in footprint size with distance from the sonar head.

Range weighting - The range weighting for each sounding decreases with distance from a node. This lessens the effect that soundings further away from the node will have on the BASE surface.

Grazing angle weighting - Errors in multibeam data tend to increase in magnitude in the outer beams due to the longer ray path distances. This tends to magnify refraction problems and other errors. The grazing angle weight function is applied to each sounding to reduce the effects of the outer beam soundings on the BASE Surface.

This weight function is controlled in an ASCII file called GrazingAngleWeights.txt in the HIPS\System directory. This file can be modified by the user

Uncertainty Surface – Theory

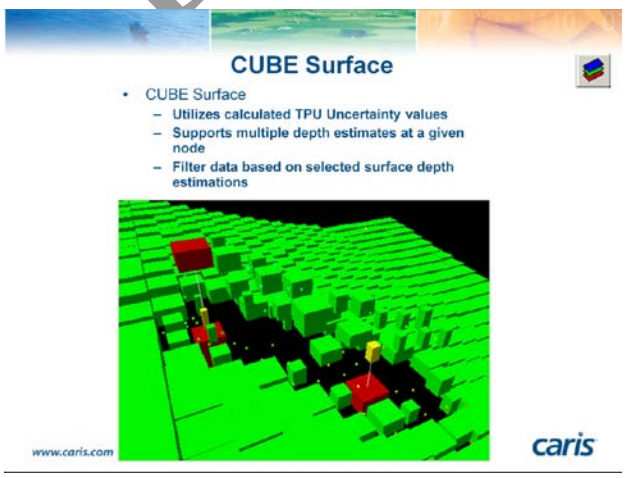


As mentioned previously, the horizontal TPU (HzTPU) and depth TPU (DpTPU) values are used in the creation of an Uncertainty BASE Surface. When creating an Uncertainty BASE Surface each sounding value will have an associated depth uncertainty and horizontal uncertainty. The depth uncertainty that each sounding has increases with the horizontal distance from the surface nodes. Therefore the greater the distance, the greater the uncertainty, and the less the sounding will contribute to the node.

The point at which a sounding will no longer contribute to a node is dependent upon the IHO S-44 Order selected. A sounding will only contribute to a node if the propagated vertical uncertainty of the sounding does not exceed the vertical accuracy requirement specified by the order.

DRAFT

CUBE Surface – Theory



The CUBE (Combined Uncertainty and Bathymetry Estimator) Surface can be used as a cleaning tool in the HIPS workflow, besides being an end product or for product creation.

During the CUBE Surface creation, soundings are weighted and contribute to surface grid nodes based on TPU values and distance from the nodes. The weighting method is similar to the Uncertainty BASE Surface creation.
The CUBE Surface however, allows for multiple depth estimates or hypotheses to exist at a single grid node, depending on the variation of the sounding data. CUBE then uses "Disambiguation" to determine which hypothesis at each node is the most "correct".
Users can verify, and if necessary override, CUBE decisions in Subset Editor. Once any necessary edits have been made to the CUBE surface, the CUBE filter can be applied to the data. Any sounding data that is not in agreement with the selected hypotheses will be flagged as rejected.

Before generating a CUBE Surface, all lines to be included must be merged and TPU must be computed.
- Highlight the field sheet name and then select the **New BASE Surface** tool or select **Process > BASE Surface > New.**
- Enter a name for the BASE Surface.
Note: It is a good idea to give the grid a name that relates to the resolution. Do not use spaces or non-alphanumeric characters.
- Enter a **Resolution.**
- Select **CUBE** for the **Surface Type.**
The Depth Filter can be used to limit the depth range of the generated surface to a specific interval.
- Set **IHO S-44 Order.**
- Select **Density & Locale** as the **Disambiguation** method.
- Click the **Advanced** button and ensure that the **Default Configuration** is selected on the Advanced Options dialog.
- Ensure that the **Initialization Surface** option is deselected.
The Disambiguation method will determine the method CUBE will use to determine the most likely bathymetry estimates. In this case it will use Density but if it is not confident in a selection it will utilize the Locale method.
The Advanced Options tab allows users to select or define the parameters to be used for the CUBE Surface creation. There are several pre-configured parameter sets available for selection. The configuration selection should be made based on the water depths and seafloor complexity of the survey area.
Note: Only experienced users should adjust or define the CUBE parameters. It is also possible for CUBE can use an Initialization Surface to ensure that grossly erroneous soundings will not be utilized in the CUBE surface creation. The CUBE surface will be visible in the Display window and the surface child layers, which are created by default, will be added to the Layers tab.

The CUBE surface contains the following layers:
• **Density:** The number of soundings that contributed to the selected hypothesis.
• **Depth:** The depth estimation of the selected hypothesis.
• **Hypothesis Count:** The number of different depth estimations at the surface node.
• **Hypothesis Strength:** A value to describe how mathematically confident CUBE is that the correct hypothesis was selected during disambiguation. Like Uncertainty, lower values are better.
• **Mean:** The mean of the soundings that contributed to the selected hypothesis.

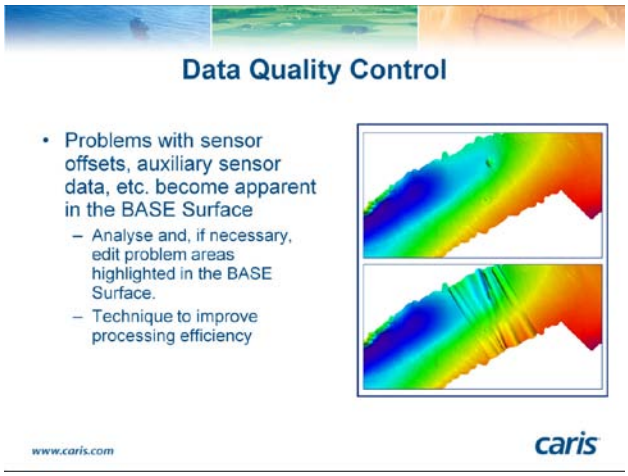
DRAFT

• **Node Standard Deviation:** The combined standard deviation of the soundings that contributed to the selected CUBE hypothesis.
• **Standard Deviation:** The average standard deviation of the soundings that contributed to the selected CUBE hypothesis.
• **Uncertainty:** The vertical uncertainty associated with the selected hypothesis. Lower uncertainties are better.
• **User Nominated:** This layer indicates if the hydrographer has nominated a hypothesis during hypothesis editing. This process will be explained further in a later exercise.
• **Guide_Depth:** This layer represents the depth values of the initialization surface but resampled and displayed at the same resolution as the CUBE surface. This layer is only displayed if an initialization surface was selected during the surface creation.
• **Guide_Uncertainty:** This layer represents the uncertainty values of the initialization surface but resampled and displayed at the same resolution as the CUBE surface. This layer is only displayed if an initialization surface was selected during the surface creation.
• **Bounding Polygon:** This layer is a coverage polygon representing the boundaries of the data. The polygon includes the outer boundary of the data as well as an holes in the data.

The CUBE Surface can be shaded to assist in feature recognition. This and other image manipulation tools can be accessed via the Properties tab. This window can be used to manipulate the display of the various layers displayed in the Layers tab well as change the Disambiguation method used in the CUBE surface Creation.
Furthermore, the layer selected in the Layers tab will be displayed in the **Properties** tab. The following can be accessed when a BASE surface layer is selected:
• **Colour Map** and **Colour Range –** Control the display of images using **Colour Maps** (colours defined by pixel values) or **Colour Ranges** (colours defined by surface values). Colour Maps and Colour Ranges can be edited by click the Browse button to launch their respective editors**.**
• **Transparency (%)** – Set a transparency percentage to be applied to the display of the image. This is useful when displaying multiple images in the HIPS Display.
• **Depth Filter –** Filter the display of surface nodes by specifying a minimum and maximum depth value.
• **Raster Legend** – Enable the display of a Raster Legend to show the correlation between the surface values and selected colours in the HIPS Display window.
• **Shading and Vertical Exaggeration –** Shade the image to bring out surface texture. The shading can be changed by moving the sun symbol or entering values for Azimuth and Angle. Use the **Vertical Exaggeration** option to control the apparent vertical height of the surface, as this can also help to distinguish features.
Whether a surface is updated when changes are made in Subset Editor is defined using the "Keep Up To Date" flag for a surface.

<u>Data Quality Control</u>

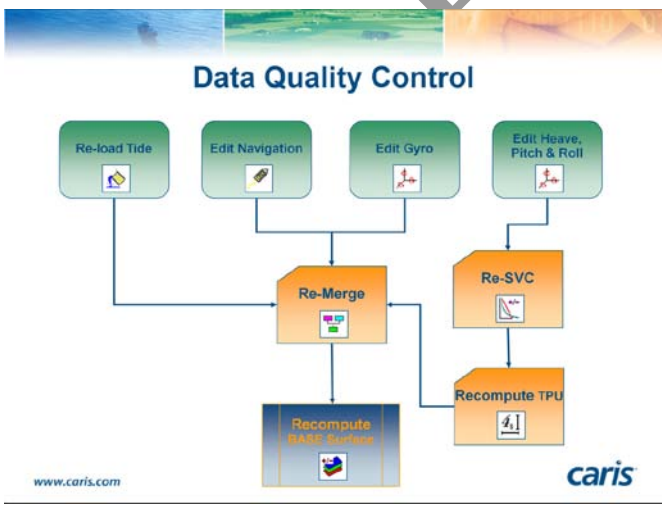
DRAFT



Prior to surface creation the auxiliary sensor data has not been investigated. The BASE surface can be used to highlight problems that may exist in the attitude and/or navigation data, problems which can be addressed using the Navigation and Attitude Editors.

By relying on the gridded data to highlight problems in the auxiliary sensors, processing efficiency can be improved, as the user is no longer required to investigate the motion and navigation of each survey line.

Note: A survey line must be selected before any of the line editors can be opened.



If changes are made to the navigation and/or motion data through the QC process it will be necessary to repeat previous steps in the HIPS workflow. The workflow diagram displayed here outlines the operations that need to be recalculated if changes are made.

Navigation Editor

The Navigation Editor lets you examine and clean individual position fixes as recorded by the vessel's positioning system.

- Select the first survey line in either the Control or Display Window.
- Select **Tools > Navigation Editor** or click on the Navigation Editor button in the Tools toolbar.

The **Navigation Editor** tab will be added to the Control window. This tab has four areas containing navigation editor controls:

• The **General** section allows the size of the navigation fixes to be changed. You can also specify if you want the points to be connected. You can select to display the graphs vertically or horizontally, and choose whether the minimum and maximum values displayed are for the entire line or the current view.
• The **Interpolation method** section determines the interpolation method used. This can be either linear or a tight, medium, or loose bezier curve. These are explained on the next slide.
• The **Spike detection** sections contain controls for finding speed jumps or time jumps.
• The final section deals with the **Number of points that are displayed** in the Display Window, this can either be based on a zoom factor or a fixed value. Users may also enable the display of Gyro Indicators and set the length and colour to provide an visual indication for the direction that the survey line was run.
When checking navigation, keyboard shortcuts can be used to speed up the procedure.

Navigation Interpolation:
Position observations do not usually occur at exactly the same instant in time as a depth observation (ping). It is also unlikely that there is a position for every ping.
In most cases it will be necessary to interpolate positions to match the times for each ping. The interpolation method is set in the Interpolation method section of the Navigation Editor tab.
The default method is **Linear**, where interpolation between each successive position is obtained by simply connecting the positions with a straight line. Linear interpolation is suitable when the original navigation positions are clean and do no significantly deviate from the neighboring positions.
The other three options (Loose, Medium, Tight) are various degrees of **Bezier Curve**. In these cases the line of interpolation will not follow the navigation positions exactly.

Edit Navigation:
There are four main edit functions that are used when editing data in the navigation or attitude editors.
• **Accept** - This will change the status of all selected sounding positions to Accepted. This can also be activated using the **A** key as a shortcut.
• **Reject - Break Interpolation** - This will reject selected sounding fixes without trying to interpolate their position. (Navigation for the pings is interpolated from the navigation data.) This function causes all soundings between the navigation fixes immediately before and after the gap to be rejected. Use **B** as a the shortcut key.
• **Reject -With Interpolation** - This rejects selected navigation fixes but interpolates a position over the gap. This can also be activated using the **W** key.
• **Query** - This gives the details of the fixes in the Worksheet window. The data can then be sorted in ascending or descending order by double-clicking the column heading. This can be useful for finding jumps and spikes. The **Q** key can also be used.
The whole line can be queried by right-clicking and selecting **Query Line** from the popup menu. This information is in memory, so it lists quickly, and the data can be pasted or saved as a text file for use in other software.

Attitude Editor
The attitude editor is used to display and edit sensor information.

DRAFT

You have to select a line, and then select the **Attitude Editor** tool button or **Tools > Attitude Editor.**
Gyro, Heave, Pitch and Roll sensor information is displayed by default. More sensors can be added by selecting them from the available sensors list, which can be accessed from the **Tools > Sensor Layout** menu.
Other sensor data could be GPS Tide, Side Scan Gyro, ROV Altitude etc.
The four default windows in the Attitude Editor display time series graphs for Gyro, Heave, Pitch and Roll. The tools to manipulate this data are available in the Control Window under the Attitude Editor tab.
When the cursor is placed on one of the time series graphs, the value is displayed as a tool tip.
Gyro data in CARIS HIPS is displayed as a positive value when a clockwise rotation is experienced.
Heave data in CARIS HIPS is displayed as a positive value when the vessel is heaved upwards.
Pitch data in CARIS HIPS is displayed as positive when the vessels bow is down. Roll data in CARIS HIPS is displayed as positive when the vessels starboard side is up.

Filtering:
The **Moving Average** option works by averaging data within a window. The window size can be based on seconds or number of data points. Increasing the size of the window increases the level of smoothing.
The **Fast Fourier** option takes out all frequencies which have a shorter wavelength than the box size set in points or seconds.
The **Threshold** option sets the cut-off value as a multiple of the standard deviation. The cut-off value is applied to the difference data. The **Difference** between the filter line and the sensor data is then calculated. This difference will be greater in areas in which local variability is high. Any difference values outside of the threshold limit will be rejected with interpolation.
Filtering can be applied in three ways: either the present screen only, filter to the end of the line, or filter the whole (selected) line. Filtering can be applied with or without interpolation. As in the Navigation Editor, sensor values may be **Queried**, **Accepted**, **Rejected** (with interpolation), or **Rejected** (break interpolation).
Motion data that contains high-variability noise can be addressed by using the smoothing routine. Smoothing parameters will be stored separately and applied during Sound Velocity Correction.

Swath Editor
The Swath editor is used to view and edit the multibeam swath data. This can be either a manual process or a semi-automated one using filters based on knowledge of the data.
After select the first survey line in the Project tab of the Control Window click on the **Swath Editor** button or select **Tools > Swath Editor.**
The Swath Editor has several windows:
• **Swath Editor** tab - A new tab appears in the Control Window, containing a number of sub-panels that are used to manipulate the data in the Swath Editor.
• **Plan View -** Displays the beams and swaths in plan view. The bottom of the display is the start of the line. The scroll bar can be used to page through the data.
• **Side View -** This is an along track profile of all the swaths in the Plan View viewed from the starboard side, so that forward is to the right of the window.

DRAFT

• **Profile View -** This view represents a single swath at a time viewed from the rear. The scroll bar can be used scroll through the individual profiles.
• **Rear View -** This is the data displayed in the Plan View window viewed from the rear of the ship.
• **Amplitude -** If the sonar data contains intensity information then it can be displayed in this window. This view can assist in feature recognition.
• **3D View** - In addition to image manipulation in the 3D View, the data can be selected, rejected, queried and accepted.

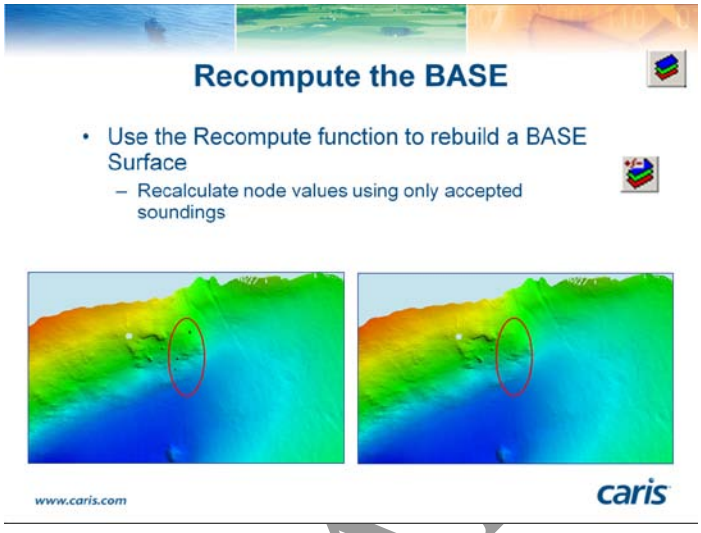
Subset Editor
The Subset Editor is activated by either clicking on the **Subset Editor** icon or by selecting **Tools > Subset Editor > Open** from the main menu.
A new Subset tab will appear in the Control window, and an additional toolbar will be displayed. The cursor in the Display Window will change to crosshairs to define the outline of the subset to be edited.
To define a subset drag a box around the data to be included using the left mouse button.
The subset outline can be rotated by holding down the **Ctrl** key while drawing. The subset can also be resized by dragging the corner anchor points.
The yellow 2D profile window can be resized or moved by dragging the anchor points.
To move the 2D slice along the subset use the keyboard arrows.
The toolbar has the following options:
• 2D and 3D windows can be toggled on and off.
• The status of subset tiles can be set to incomplete, partially complete or complete.
• The subset can be locked, so that the no accidental resizing, moving, and rotation can take place causing a redraw. The subset can also be locked by pressing the <**L**> key.
The Subset Editor tab has the various Subset Editor items organized into a **Data** tree. By selecting the items in the **Data** tree, the corresponding properties will be displayed in the **Properties** section of the Subset Editor tab.
Subsets should not be too large. If they contain too many soundings they will take toolong to open.

Editing windows:
The user can pan (middle mouse button) and zoom in/out (Crtl+right mouse button) on the **2D View**. Other functionality includes:
• The **View Direction** of the data in the window is relative to the subset orientation. The yellow arrow on the subset outline in the Display points towards the top. The view direction should be selected with this in mind. Select **Automatic** to always have the 2D View display along the longest edge of the slice.
• Set the display by selecting **Auto** Exaggeration, **min/max** values or by adjusting it manually using the Vertical Exaggeration in the 2D View. You can also choose to **Include Rejected** soundings in the scaling. Soundings can be displayed as digits and graticule labels can be turned on and off.
The **3D View** is controlled in the same way as the 3D View for the Swath Editor. There are however functions that are specific to the Subset Editor.
• The **Data Display** options allow for the 3D view display **Type** to be set as Points, Spheres, Cylinders, Surface or Surface Wireframe. The Subset **Outline** and **2D view slice** can also be

DRAFT

enabled for display in the 3D View. **3D Dynamic Skip** reduces the number of points displayed while the 3D View is being manipulated by the user.
• Under the **Controls** the user can **Allow rotation under data.** This allows the operator to rotate the data beyond the horizontal and view the underside of the sounding data. The display of the Compass, Lighting and Exaggeration can also be enabled/disabled.

Recompute the BASE Surface



After you have finished editing the data in the Subset Editor and exited out, or cleaned the bathymetry using the Surface filter, the BASE Surface may appear outdated. This is indicated by a red exclamation mark on the BASE Surface icon in the Session tab of the Control Window. The **Recompute** function enables quick recomputing of a BASE Surface to update any changes that have been made.
If all manual edits had been performed in the Subset Edit with the **Automatic BASE surface update** option selected under **Tools** > **Options** the surface would already be up-to-date and would not need to be recomputed. **Automatic BASE surface update** recomputes small sections of the surface as the user moves across the point cloud in Subset Editor.
If the surface is outdated select the BASE Surface from the Control Window, right mouse click and select **Recompute** from the pop-up menu.
All changes made during the editing process will be applied to the BASE Surface.

The data should now be geo-referenced and cleaned.
The HIPS data can now be used to create bathymetric products. These are generated in the Field Sheet Editor.
Bathymetric products include Selected Soundings, Contours and others.

Product Surface

The first step to generating smoother contours is to create a Product Surface. The Product Surface maintains the designated soundings from the finalized BASE Surface and can be used to create more cartographically correct contours. A generalized surface is created by rolling an

DRAFT

imaginary ball over the BASE Surface. The user can specify the scale and radius of the ball. The ball will be rolled over the surface at the interval defined by the resolution.
The Product Surface is activated highlighting the Field Sheet layer and selecting **Process > BASE Surface > New Product Surface**.

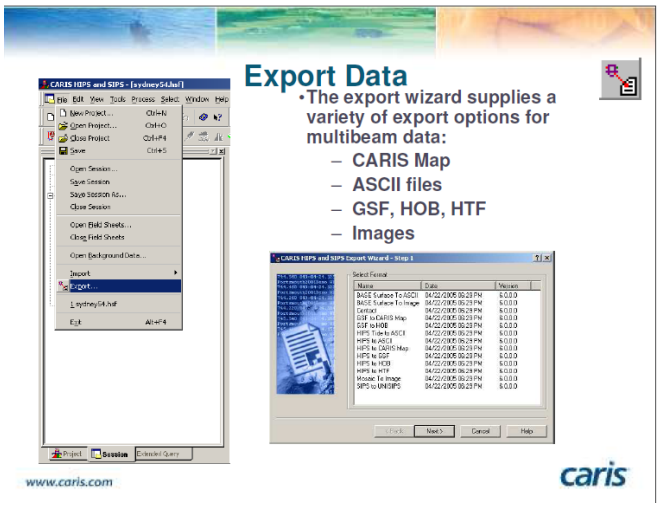
Sounding Selection

It is impossible to display all soundings produced by a multibeam survey on a map or chart and maintain legibility. The selection of soundings that appear should be dependent on chart scale and purpose. Soundings can be created from a Tile layer or BASE Surface.
Click on the **Create Sounding Layer** icon or select the **Process > Products > Soundings > New Sounding layer.**
Soundings can be displayed in two ways, the soundings can be engineering style, with decimals, or they can appear as slanted soundings with drop decimeters.
Note: In most situations it is wise to use the same data source for soundings and for contours. This removes the possibility of producing soundings on the wrong side of contours.

Contouring

Contours are created using a three step wizard.
Open the Contouring wizard by clicking on the New Contour Layer icon or by selecting the **Process > Products > Contours > New Contour Layer.**
Contouring from different sources produces different results. The BASE Surface (swath angle method) is a mean surface, whereas tiles or bins can be shoal-biased. If you are creating products that will be used for navigation then a bin or uncertainty BASE Surface maybe a more appropriate source for building contours.
- Select the Depth layer of the generalized Product Surface as the height source.
- From the **Standard Contour** tab select an interval.
- The **Theme Number** will determine how layers are named and organised, this information can be used by CARIS GIS to differentiate the data.

Export Data



DRAFT

This will launch the Export Wizard. There are various export options available. Data, grids, and images can all be exported using this wizard.
The BASE Surface generated in HIPS can be exported to image file or ASCII file. The user can specify the units for the exported depth values and additional sounding information can also be exported as optional attributes.
The depth can be exported as an ASCII file.

4.3 Generic Mapping Tools (GMT)

Generic Mapping Tools (GMT) (Wessel and Smith, 1998) is a collection of open source mathematical and mapping routines for use on gridded data sets, data series, and arbitrarily located data. The GMT package is available for download from the University of Hawaii website (http://gmt.soest.hawaii.edu/).

4.4 r2v, Surfer, Global Mapper

Contributed by John Hall, Geological Survey of Israel (Retired), Israel

Recent compilations of the Caspian Sea and the Red-Arabian Seas illustrate the preparation of high resolution grids using the r2v (www.ablesw.com), Global Mapper 12 (www.globalmapper.com), Surfer 9 (www.goldensoftware.com) software.

An overview of each of the packages follows:

4.4.1 Able Software Corporation's r2v (raster to vector)

A license for r2v costs $395 for an academic user, or $795 for non-academic. A USB dongle is now employed. As used for bathymetry, a scanned map is geo-referenced, either using Global Mapper (more convenient), or r2v. Once a .tfw (true world file) exists, r2v can work in map coordinates, which are usually Mercator. Very few of r2v's capabilities are used for this work. Mainly the automatic line following is used, following conversion to grayscale, and then contrast selection to accentuate contours while eliminating background. The program will rapidly follow a line until a gap or branch is encountered. One then toggles to manual mode, passes the obstruction, and continues in automatic mode. The program gives complete control over the nodes, adding, moving, smoothing, deleting, and joining. Use of the F2 and F3 keys allows almost unlimited magnification for very steep slopes. Navigation across the map is very easy. Contour values are easy to apply, either individually, or by identifying a beginning value and increment and then drawing a line across sequential contours. For complicated contours it is possible to color code the contours. Output of r2v is generally via a .dxf vector file, or as .xyz points, although other formats are available. The program is also able to digitize points, but each point requires a number of steps. Dr. Ted Wu is the owner of Able Software and readily answers questions. A trial version of r2v can be downloaded.

4.4.2 Global Mapper Software LLC - Global Mapper 12

DRAFT

A license for Global Mapper 12 is $349, and somewhat less expensive for multiple licenses. Global Mapper was started in 1991 by Mike Childs, who is very responsive to additions and modifications to the program, which undergoes regular improvements. In short, GM12 is a comparatively easy to use program which deals with a wide variety of raster images, grids and vectors, especially those available on the Internet. It deals with a continually increasing number of projections and datums, as well as formats, and allows speedy conversion between them. For our purposes, GM12 allows rapid geo-referencing of charts, rapid TIN display of .xyz files prepared by r2v, projection change of .dxf files, rapid display of multiple grids, superposition of various raster and vector layers, highspeed input of grids (ASTER-GDEM, SRTM, SRTM30-Plus, GEBCO_08, IBCAO etc.), and their manipulation (cropping, clipping, translation, scaling, offsetting etc.). Global Mapper has a tight non-lossy grid format (.gmg).

Global Mapper also offers display and logging of GPS data, draping of imagery or maps on grids with 3-D display, digitization tools, contouring, view-shed and water-shed analysis, screen capture, topographic profiles with slopes, palette construction, shading options etc. It has also become an important connectivity and display tool for sophisticated multi-platform users of programs such as GMT, ArcGIS, CARIS, and IVS Fledermaus. Two features lacking on Global Mapper are the ability to convert .dxf files to xyz (which can be done with the free program dxf2xyz from www.guthcad.com), and the use of degrees minutes and seconds for geographic coordinates output to .csv format. This is important for merging new digitized contours and shorelines with the spot soundings digitized years ago, to form a single file for interpolation by Surfer. The work-around of course is to use Cartesian meridional parts for Mercator maps, or else UTM coordinates.

DRAFT

4.4.3 Golden Software Inc. Surfer® Version 9

A license for Surfer's latest version 9 is $699. Global Mapper has the 3-D display, contouring and hypsometric shading and palette selection capabilities of Surfer. However Surfer is the program of choice for interpolation of grids from the tens of thousands of points generated by r2v and the historical digitization of soundings. The Surfer gridding interface allows rapid investigation of the input files, and the Filter Data function deals with duplicates as well as averaging points that are so close that steep gradients would be formed.

Grid Construction

As an example, over 55% of the Mediterranean has now been mapped with multibeam sonar. The MediMap Group (http://www.ciesm.org/marine/morphomap.htm) has pooled their data and it is currently available within the group at a node spacing of 500 m. Efforts to decrease this spacing has apparently been opposed by commercial interests. Basically this coverage is sufficiently good to make a first pass at a 0.1' grid for IBCM-II. Israel is now concluding a decade of pro-active mapping with a Kongsberg-Simrad EM1002 owned by John Hall, installed on the Israeli government vessel R/V Etziona. Over 10,000 km² of the Mediterranean offshore have been surveyed, from minimum depths of 15 m to about 700m with the EM1002, and then recently out to 1,600 m depth with a rented L-3 ELAC Gmbh SB3050 system. While the Israeli government regards the digital data as classified, grids of 100 m on the shelf and 50 m in deeper

waters are available. Hopefully this will encourage other countries to allow use of higher definition deep water data.

The seas that are of interest follow, in order of size etc.:

Water Body	Total Area	Max Depth	Shoreline	Countries
Mediterranean	2,966,000 km²	5267 m	46,000 km	22
Red Sea	438,000 km²	2211 m	11,254 km	8
Black Sea	474,000 km²	2206 m	4,340 km	7
Caspian Sea	371,000 km²	1025m	6,397 km	4
Persian Gulf	251,000 km²	90 m	3,334 km	8

Russian charting at about 1:200,000 covers the periphery of all these bodies. Thus grids can be obtained by first taking the land coverage at 30 m from the latest ASTER-GDEM. These come in one degree squares (called granules) and can be downloaded, up to 2000 at a time, from http://wist.echo.nasa.gov/api/. It is necessary to register, and after choosing and verifying the coverage wanted, one receives a verification and link to an ftp site. The granules can be quickly downloaded with an ftp client program like Filezilla (http://filezilla-project.org). The files consist of the height values (.dem format) and the number of scenes used (.num). Global Mapper 12 will rapidly ingest the height data while ignoring the .num component, and can then output single .gmg grid files with the data for the overall area. The next step is to remove the data over the sea. This can be done by using Tools>Control Center> Options>Alter Elevation Values to reject anything below 0.1 m. While this will generally work, sometimes the alternative is to single out specific granules and crop them using an area boundary in Global Mapper or a blanking file in Surfer. Such a file is usually the 0 m contour from the digitizing done in r2v.

The clipped ASTER-GDEM file gives a pretty good indication of where the land is, and for checking the marine data which often is pre-WGS84 or WGS72 datums. Older Russian charts sometimes referred to the Russian Pulkovo S42 datum of 1942, or in the case of early British surveys, relative to one of the 700 datums used in UKHO surveys. Global Mapper's ability to overlay and translate charts can correct some of these geodetic displacements.

The next step is to geo-reference the 300 dpi color scans of the charts, one at a time, using Global Mapper. The georeferencing can be checked by bringing up Tools>Configure>General>Grid Display>Lat/Long Grid and specifying the map graticule interval. Visual inspection will then indicate the quality of the georeferencing. General four corner control points will suffice, but for a long body like the Red Sea, perhaps as many as 20 points may be necessary. Using File>Export Raster Image Format>GeoTiff allows export of a GeoTiff with .tfw true world file that can be used in r2v. At present 92 different projections can be used. Generally it is best to use Cartesian coordinate like meridinial parts or UTM.

Within r2v the coastline and available contours are digitized and their depths noted. Where inset charts at larger scale are available, the spot soundings were separately digitized, and data from them can be digitized separately and then later merged by loading the larger scale .dxf contour file into the smaller scale chart. Where necessary, spot soundings can also be picked up. In order to seamlessly place the bathymetry under the land coverage, the high elevations shown on the

DRAFT

navigational chart may be digitized so that the kriging process will apply a landward slope commensurate with the actual topography. Inshore contours at 3, 6, 10, 20, and 50 m are often almost continuous, so that the shallows are reasonably well represented. For the southern Red Sea, where there are numerous reef areas indicated from space imagery, fictitious contours at 20 and 3 m depths help to define the reefs. In general, where ships travel, the hydrographers have soundings, but of course, regardless of what the final grid spacing is, the results must be clearly marked as "Not To Be Used For Navigation."

Coastal navigational charts generally overlap one another, so if each is digitized and then interpolated to a grid, their grids can later be cropped and stacked appropriately within Global Mapper for output as the final grid. In the case of the Red Sea, UTM grids were interpolated at 100 m for Zones 36 and 37. Global Mapper keeps the grids at the finest scale used, so with ASTER-GDEM 30 m data the default output would be 100 m.

REFERENCES

Carron, M. J., P. R. Vogt, and W.-Y. Jung. 2001. A proposed international long-term project to systematically map the world's ocean floors from beach to trench: GOMaP (Global Ocean Mapping Program). Int. Hydrographic Rev., 2, p. 49–55.

Doytsher, Y., and Hall, J. K., 1997. Gridded affine transformation and rubber-sheeting algorithm with FORTRAN program for calibrating scanned hydrographic survey maps. Computers & Geosciences, 23(7), p. 785-791.

Hardy, R. L., 1971. Multiquadric equations of topography and other irregular surfaces. J. Geophys. Res., 76(8), p. 1905-1915.

Hall, J. K., Schwartz, E., and Cleave, R. L. W., 1990. The Israeli DTM (Digital Terrain Map) Project. In: Microcomputer Applications in Geology, II (J. T. Hanley and D. F. Merriam, eds.). pp. 111-118.

Computers and Geology, Vol. 6, Pergamon Press, New York. Computer program on diskette as part of the COGS (Computer Oriented Geological Society) public-domain library.

Hall, J. K., Udintsev, G. B. and Odinokov, Yu. Yu., 1994. The bottom relief of the Levantine Sea. In: Geological structure of the North-eastern Mediterranean (Cruise 5 of the Research Vessel 'Akademik Nikolaj Strakhov'). (V. A. Krasheninnikov and J. K. Hall, eds.), p. 5-31. Historical Productions-Hall, Jerusalem, April 1994.

Vogt, P. R., W. -Y. Jung, and D. J. Nagel, 2000. GOMaP: A matchless resolution to start the new millennium. Eos Transactions, American Geophysical Union, 81(23), p. 254, 258. June 6, 2000.

DRAFT

Chapter 5.0 Gathering Data

There are many types of geophysical data available for free download from the internet. In this chapter we show where to obtain data useful for researchers desiring to create grids from publicly available bathymetric data.

5.1 NGDC Bathymetry Data

TEXT IN PREPARATION

5.2 JAMSTEC Data

Contributed by K. M. Marks, NOAA Laboratory for Satellite Altimetry, USA

Many types of geophysical data are available for download from the JAMSTEC website (http://www.godac.jamstec.go.jp/cruisedata/e/), which is shown in Figure 5.2.1. Here we demonstrate how to obtain multibeam data from the JAMSTEC website.

DRAFT

Figure 5.2.1 JAMSTEC Data Site for Research Cruises

Multibeam data may be downloaded from the JAMSTEC website as follows. The user first selects the "data archives" link for one of the five research vessels on the main JAMSTEC website (Fig. 5.2.1). The link opens a global map of cruises color-coded by collection year (Fig. 5.2.2). The user selects the desired year, and then the cruise ID. Selecting "Apply" opens the

individual cruise web page (Fig. 5.2.3). To reach the bathymetry web page, the user selects "Bathymetry" from the right side of the cruise web page (see Fig. 5.2.3).

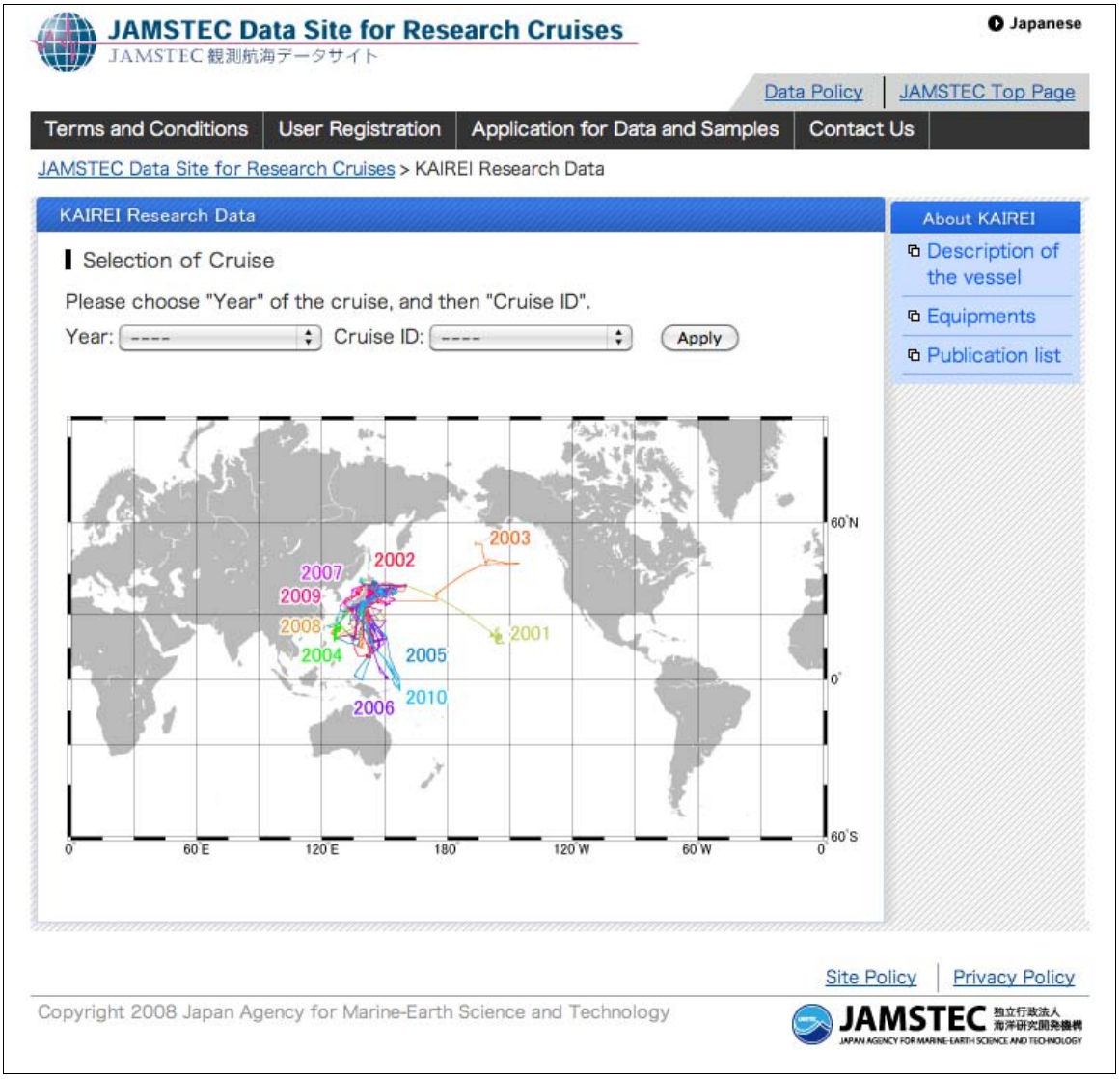


Figure 5.2.2 KAIREI Research Data web page.

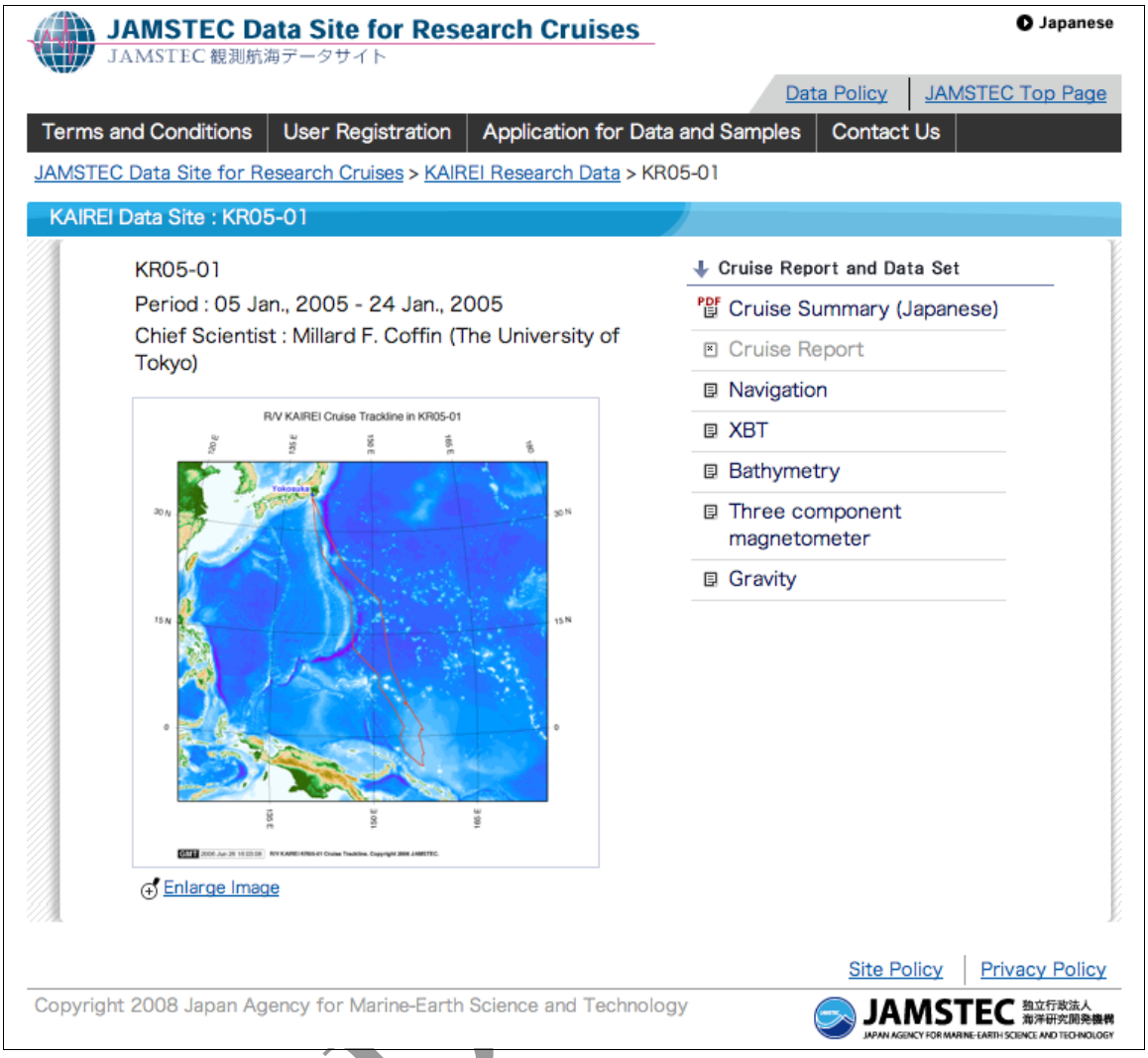


Figure 5.2.3 KR05-01 Cruise web page.

Bathymetry data may be downloaded from the bathymetry web page (Fig. 5.2.4) via the "Data" link, as well as information on the instruments, data format, and collection system details (i.e., the "Instruments," "Format description," and "Readme" links, respectively). The bathymetry data may be downloaded as a number of compressed files in "zip" format (see Fig. 5.2.5) that need to be uncompressed ("unzipped") and then assembled sequentially into a single file. The bathymetry data are ASCII *xyz* (longitude, latitude, depth) multibeam echo sounder ping records. There may not be data along the entire track plot shown on the bathymetry web page and making a plot of the coverage will reveal any gaps.

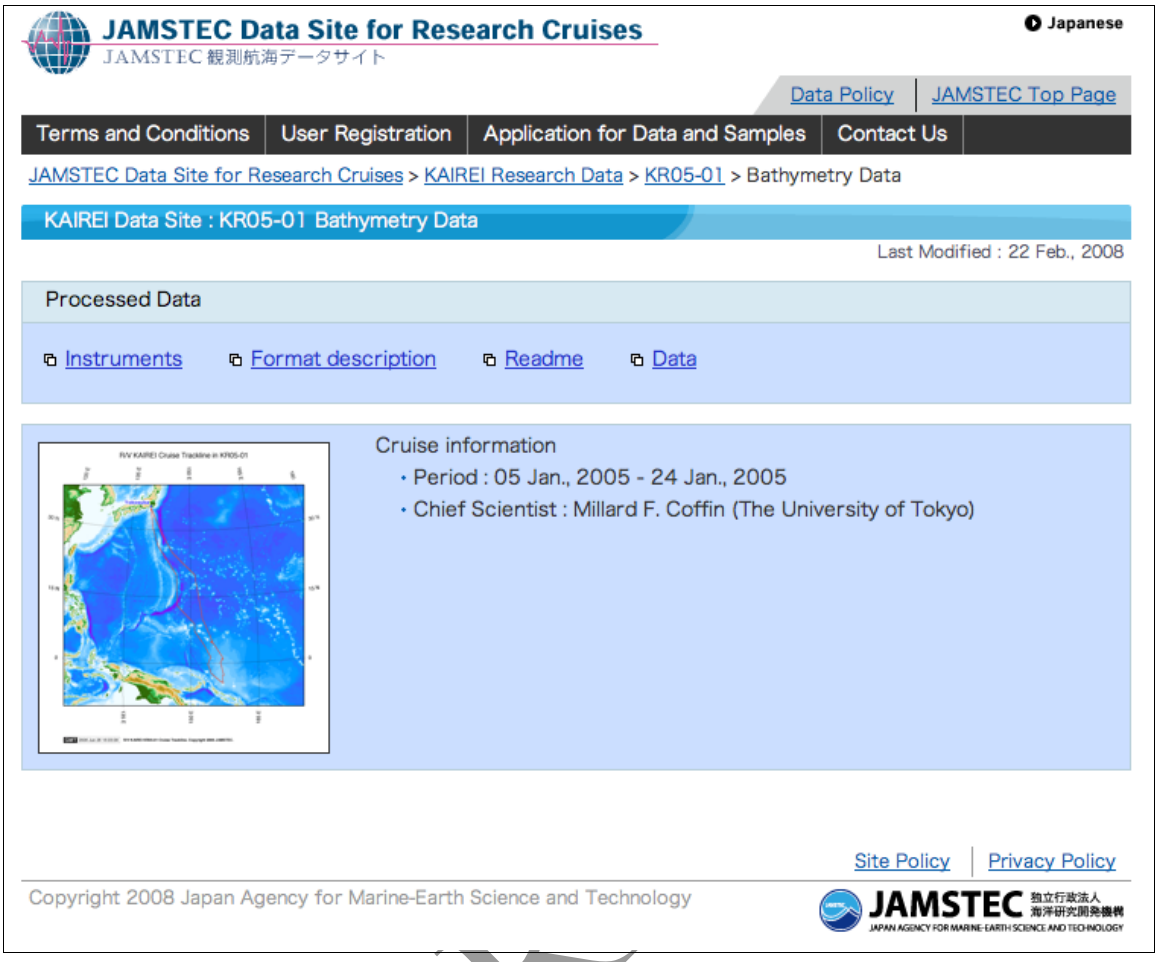
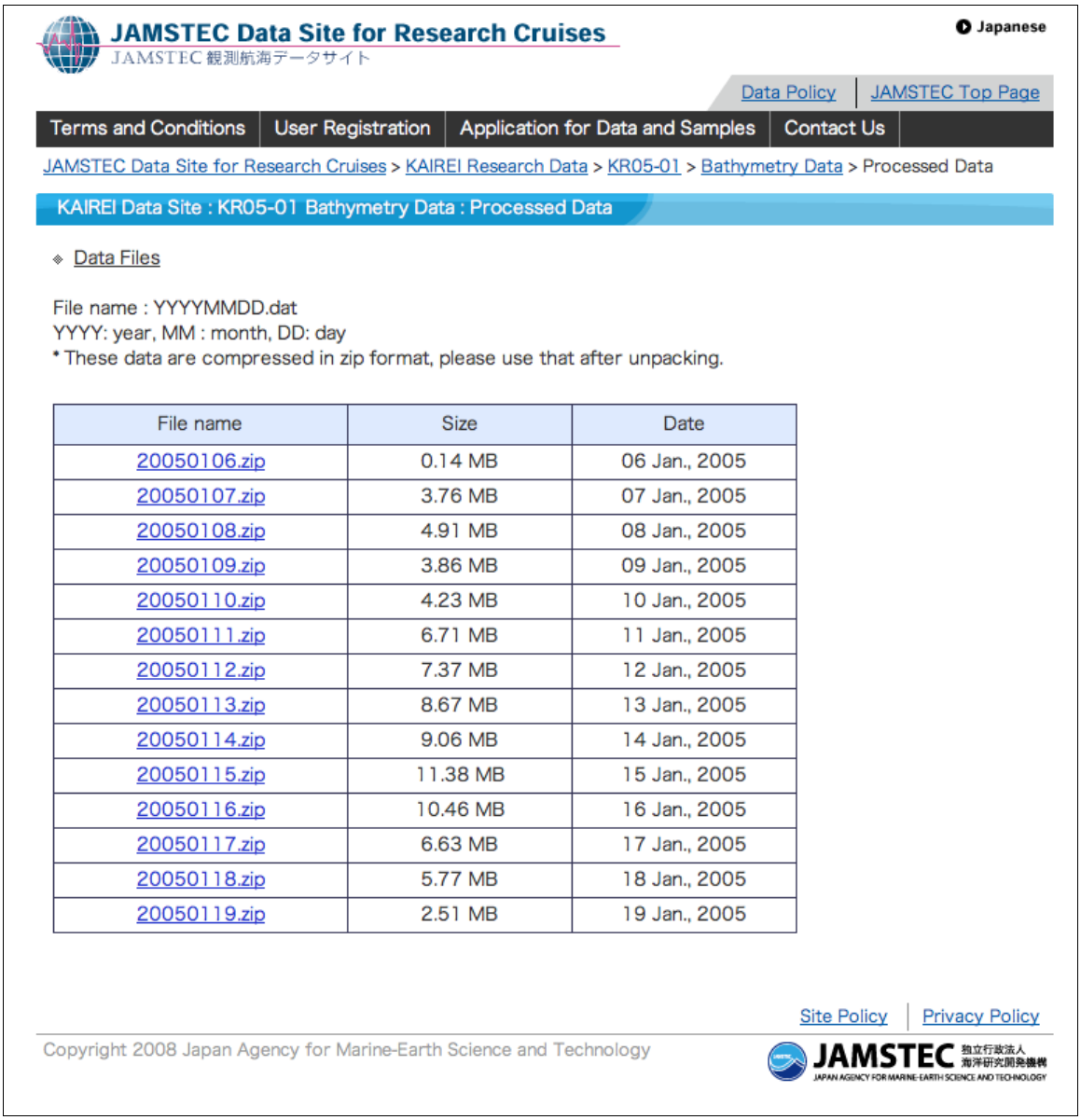


Figure 5.2.4 KR05-01 Bathymetry Data web page.

DR



JAMSTEC Data Site for Research Cruises
JAMSTEC 観測航海データサイト

Japanese

Data Policy | JAMSTEC Top Page

Terms and Conditions | User Registration | Application for Data and Samples | Contact Us

JAMSTEC Data Site for Research Cruises > KAIREI Research Data > KR05-01 > Bathymetry Data > Processed Data

KAIREI Data Site : KR05-01 Bathymetry Data : Processed Data

◈ Data Files

File name : YYYYMMDD.dat
YYYY: year, MM : month, DD: day
* These data are compressed in zip format, please use that after unpacking.

File name	Size	Date
20050106.zip	0.14 MB	06 Jan., 2005
20050107.zip	3.76 MB	07 Jan., 2005
20050108.zip	4.91 MB	08 Jan., 2005
20050109.zip	3.86 MB	09 Jan., 2005
20050110.zip	4.23 MB	10 Jan., 2005
20050111.zip	6.71 MB	11 Jan., 2005
20050112.zip	7.37 MB	12 Jan., 2005
20050113.zip	8.67 MB	13 Jan., 2005
20050114.zip	9.06 MB	14 Jan., 2005
20050115.zip	11.38 MB	15 Jan., 2005
20050116.zip	10.46 MB	16 Jan., 2005
20050117.zip	6.63 MB	17 Jan., 2005
20050118.zip	5.77 MB	18 Jan., 2005
20050119.zip	2.51 MB	19 Jan., 2005

Site Policy | Privacy Policy

Copyright 2008 Japan Agency for Marine-Earth Science and Technology

JAMSTEC 独立行政法人 海洋研究開発機構 JAPAN AGENCY FOR MARINE-EARTH SCIENCE AND TECHNOLOGY

Figure 5.2.5 KR05-01 Bathymetry Data files available for download.

An alternate way to reach bathymetry web pages is for the user to click on the "JAMSTEC Data Search Portal" tool (see Fig. 5.2.1). This accesses instructions on how to use the tool as well as provides an entryway to the tool's ArcIMS interface (see Fig. 5.2.6). On the ArcIMS interface, the user may drag the red box on the global map (upper left corner) to the desired study location, then select "Quick Search" to drag a box on the main map. This brings up a search result panel (see Fig. 5.2.7) that lists the cruises that traverse the study box. The bathymetry web page for the cruise of interest can then be reached by clicking on "Jump."

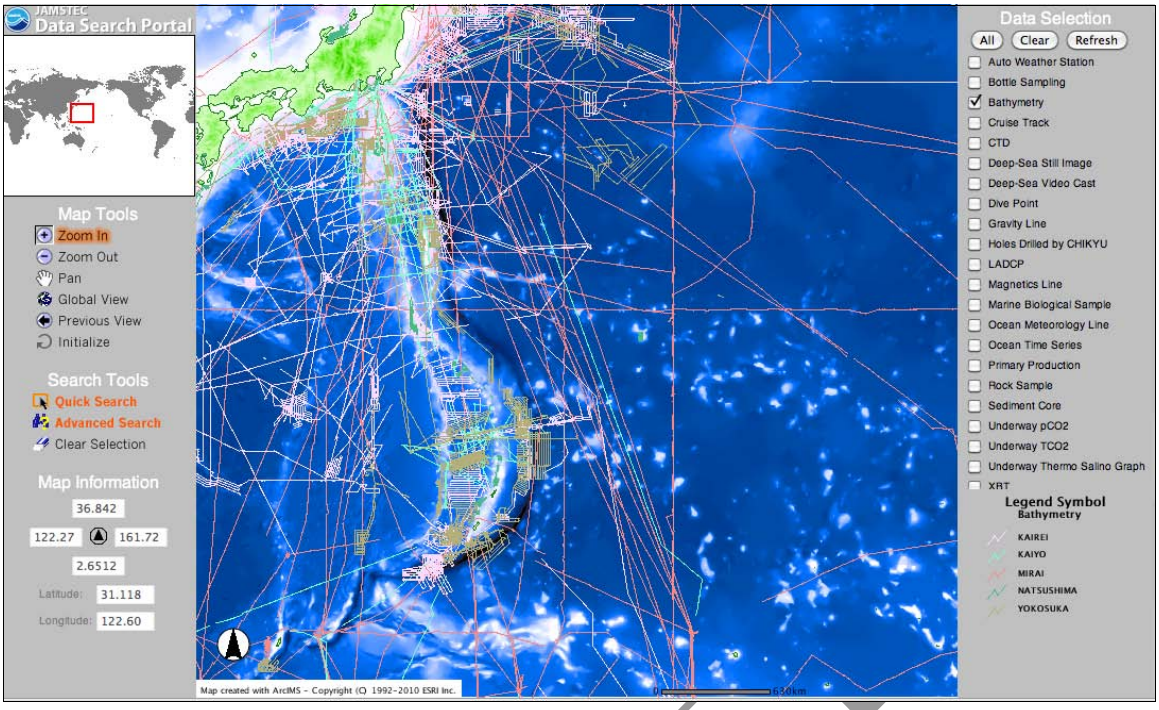


Figure 5.2.6 JAMSTEC Data Search Portal.

Search Result (Select Data Set : Bathymetry) 16 Selected

No.	Data ID	Latitude	Longitude	Period (start)	Period (end)	Ship Name	Cruise ID	DSRV Name	Dive No.	Web Site
1	0013-00000027			2002/02/20	2002/03/29	MIRAI	MR02-K02			Jump
2	0013-00000035			2003/02/19	2003/03/29	MIRAI	MR03-K01			Jump
3	0013-00000052			2004/11/16	2004/12/08	MIRAI	MR04-07			Jump
4	0013-00000054			2005/01/13	2005/02/18	MIRAI	MR04-08_leg2			Jump
5	0013-00000056			2005/05/24	2005/06/30	MIRAI	MR05-02			Jump
6	0013-00000064			2006/02/04	2006/03/17	MIRAI	MR06-01			Jump
7	0013-00000073			2007/02/15	2007/03/25	MIRAI	MR07-01			Jump
8	0013-00000075			2007/05/30	2007/07/13	MIRAI	MR07-03			Jump
9	0013-00000085			2008/07/01	2008/08/05	MIRAI	MR08-03			Jump
10	0013-00000211			2005/01/04	2005/01/23	KAIREI	KR05-01			Jump
11	0013-00000228			2005/12/09	2005/12/24	KAIREI	KR05-17			Jump
12	0013-00000242			2006/09/11	2006/09/22	KAIREI	KR06-12			
13	0013-00000264			2007/11/22	2007/12/01	KAIREI	KR07-16			Jump
14	0013-00000326			2002/09/29	2002/10/14	KAIYO	KY02-10_leg1			
15	0013-00000639			2001/01/03	2001/01/14	YOKOSUKA	YK01-01			
16	0013-00000915			2009/11/02	2009/12/11	MIRAI	MR09-04			

Figure 5.2.7 Data Search Portal Results.

Chapter 6.0 Data Cleaning

6.1 Quality Assessment

Contributed by Rob Hare and Paola Travaglini, Canadian Hydrographic Service, and Martin Jakobsson, Stockholm University, Sweden

6.1.1 Sources of uncertainty

Elevation value

Instruments such as multibeam swath sonars and airborne lidar do not actually measure elevation, but rather travel time (of sound or light, respectively). Many different sources therefore contribute to the uncertainty of a calculated depth/elevation value. Although the following text focuses on uncertainties in data acquired with multibeam swath sonar systems, it should be noted that it is in principle applicable to data acquired with any elevation measurement system; a single-beam echosounder is just a special case of a multibeam echosounder (i.e. nadir beam only). Agencies conducting surveys are encouraged to provide uncertainty budgets, or total propagated uncertainty (TPU) estimates for their own systems.

i. *Sound or light speed uncertainty – how accurately was sensor measurement or time converted into a depth/elevation value?*

   Multibeam swath sonar systems rely on frequent measurements of sound speed, typically taken an XBT (expendable bathythermograph) or CTD (conductivity-temperature-depth package), and sometimes measured directly using an SVP (sound speed profiler). These measurements should be taken at least daily, or with changing weather conditions, as longer time intervals decrease the accuracy of the soundings due to improper ray-tracing. Measurements should be taken more frequently in areas of known oceanographic fronts and off the mouths of major rivers.

   The sound speed profile is used to calculate depth for each beam, via ray tracing, and is embedded in the binary sonar files. The profiles can be extracted and replaced with different profiles using multibeam processing software such as MB-System, which can recalculate depths and positions for each beam footprint.

   Because sound travels farthest for the outermost, lowest grazing angle beams, especially in deep water, and because ray bending increases as the tangent of the beam angle, these beams have the largest depth uncertainty due to sound speed uncertainty caused by ray bending, and often contain anomalous values inconsistent with other, overlapping data. As such, they are frequently flagged for exclusion. If sound speed profiles cannot be taken at a rate which minimizes the uncertainty due to refraction, then line spacing should be decreased in order to increase the overlapping region, and provide additional confidence in the bathymetry with many of the outer beams eliminated from the gridding process.

   Uncertainty in the measurement of vessel roll with also contribute significantly to the TPU of depth in the outer beams.

DRAFT

ii. Position uncertainty – how well known was the position of the sensor?

With swath sonar systems the distance between the sounding on the seafloor and the positioning system antenna can be very large, especially in deep water. Because of this, sounding position uncertainty is a function of the errors in vessel heading, beam angle solution, refraction correction model and the water depth, in addition to the uncertainty of the positioning system itself.

Roll and pitch errors will also contribute to the uncertainty in the positions of soundings. Overall, it may be very difficult to determine the position uncertainty for each sounding as a function of depth. The uncertainties are a function not only of the swath system but also of the location of, offsets to and accuracies of the auxiliary sensors.

The use of non-vertical beams introduces additional uncertainties caused by incorrect knowledge of the ship's orientation at the time of transmission and at time of reception of sonar echoes.

Uncertainty associated with the calculated depth of an individual swath sonar beam will include the following components:

a) Positioning system uncertainty (e.g. of the GPS antenna)
b) Sensor location uncertainty (of the sonar's transducer and hydrophones relative to the GPS antenna)
c) Beam angle uncertainty (relative to the sensor), including beam steering uncertainty when sound speed at the transducer face in not directly measured
d) Uncertainty associated with beam detection of seabed reflection echo (angle or time)
e) Uncertainty in the orientation of the vertical motion sensor (relative to the sonar); pitch bias and roll bias
f) Vessel motion sensor uncertainty, i.e. roll, pitch and heave
g) Uncertainty in vessel heading (distinct from path over ground)
h) Uncertainty associated with the ray path model (including the sound speed profile)
i) Sensor offset uncertainty (relative to local water level)
j) Ellipsoidal / vertical datum transformation model uncertainty (if applicable)
k) Time synchronization / latency between all of the systems

All uncertainties should be combined statistically to obtain a *total vertical uncertainty* (TVU).

DRAFT

Morphologic change since data collection

Morphologic change is an ongoing process that will, over time, decrease the accuracy of survey data. In the coastal zone especially, the movement of sand alters seabed morphology. This may occur as the result of a major storm, or long-term sediment dispersal or accumulation. Regional tectonic uplift or subsidence may also play a role. The older the data is the greater its uncertainty, with the greatest rates of change occurring in shallow water.

Data not measuring ground surface

Not all instrument measurements detect Earth's solid surface. Many individual sonar depths may actually reflect water column noise. Spikes often occur directly along the ship track, where the sound energy returning to the transducer is strongest. Returns from the water surface or from

tree canopy are also common in topographic lidar. These values need to be indentified and deleted.

6.1.2 How to determine data uncertainty

a. Typical for modern data types
   i. Lidar [*Fugro Earthdata*, 2007]

Point spacing	Vertical accuracy (RMSE)
1 m	0.09 m
2 m	0.20 m
3 m	0.30 m
4 m	0.40 m
5 m	0.51 m

   ii. Multibeam swath sonar
   iii. Hydrographic surveys undertaken prior to 1998

Depth	Vertical accuracy
< 30 m	0.3 m
> 30 m	1% of depth

b. Estimating navigation uncertainty for older data
c. Comparison with other data

6.2 Data Cleaning Techniques

Contributed by Rob Hare and Paola Travaglini, Canadian Hydrographic Service, and Martin Jakobsson, Stockholm University, Sweden

DRAFT

6.2.1 Automatic (non-interactive) data cleaning

During this stage, the coordinates (i.e. positions and depths) obtained should be controlled automatically by a program using suitable statistical algorithms which have been documented, tested and demonstrated to produce repeatable and accurate results. When selecting an algorithm, robust estimation techniques should be taken into consideration as their adequacy has been confirmed. Many high-density bathymetry processing packages have built-in statistical processing tools for detecting and displaying outliers. Generally speaking, higher-density data sets with large amounts of overlap between lines provide an increased likelihood of detecting blunders. In addition to statistics, threshold values for survey data can be used to facilitate the detection of blunders. Each agency is responsible for the validation of the algorithm used and the procedures adopted.

All blunders and erroneous and doubtful data should be flagged for subsequent operator control. The type of flag used should indicate that it was set during the automatic stage.

6.2.2 Manual (interactive) data cleaning

Following automated processing procedures, there is a requirement for an experienced and responsible hydrographer to review the automated results and validate those results and/or resolve any remaining ambiguities.

For this stage the use of 3-D visualization tools is strongly recommended. Decision making about whether to accept or reject apparently spurious soundings can often be enhanced by viewing combined data sets in three dimensions. These tools should allow viewing the data using a zoom facility. The interactive processing system should also offer different display modes for visualization, e.g. depth plot, uncertainty plot, single profile, single beam, backscatter imagery etc. and should allow for the visualization of the survey data in conjunction with other useful information e.g. shoreline, wrecks, aids to navigation etc. Editing the data should be possible in all modes and include an audit trail. When editing sounding data, it can often be useful to understand the spatial context of the examined data points. What may appear to be bad soundings (blunders) out of context may be recognized as real seafloor artifacts (submerged piles, wrecks, etc.) when viewed in the context of a chart backdrop for example. If feasible, data displays should be geo-referenced. The ability to compare surfaces from newly collected data to ones generated from historical information can often be useful in validating the quality of the new information, or alternatively, for notifying the collecting agency of an unresolved systematic bias that requires immediate attention.

If feasible, these tools should include the reconciliation of normalized backscatter imagery with bathymetry and, provided that automated object detection tools were used, the display of flagged data for both data modes should be possible.

The rules to be observed by operators during this stage should be documented.

The flags set during the automatic stage, which correspond to depths shallower than the surrounding area, should require explicit operator action, at least, for Special Order and Order 1 a/b surveys. If the operator overrules flags set during the automatic stage, this should be documented. If a flag is set by the operator, the type of flag used should indicate this.

It can be helpful to create a map plotting ship track and date/time to assist with correcting anomalies discovered during visualization.

DRAFT

6.2.3 Clipping to polygon or elevation value (e.g., zero)

6.2.4 Common multibeam swath sonar errors

Spikes under nadir

Spikes may be present directly under nadir for many multibeam swath sonar systems, especially older systems. This appears to result from excessive sound energy reflected back toward the sonar and being received in the side lobes of non-nadir beams. Care should be taken when flagging these values as actual seafloor features may appear similar to these spikes.

Tracking water-column noise

Modern multibeam swath sonar systems generally include automated bottom tracking algorithms. Such algorithms search for seafloor echoes in a window centered on the previous (or adjacent) ping's depths. However, the system can begin to track water column noise, and may continue to do so if the system is not actively monitored. Depth variability between pings can be

large, the across-track depth profiles are not consistent with seafloor topography, and most of the beams/pings should be discarded.

Roll-over on inward-facing steep slopes

In areas of steep slopes, especially where the seafloor rises towards the sea surface, some sonar systems have a tendency to "roll over" the outermost beams to create a backside or lip. These values must be flagged. These errors can often be detected when editing the data in the context of a priori knowledge of the seabed shape, especially man-made vertical structures or known topography of canyons or fjord walls.

Anomalies on outward-facing slopes

Where steep hills and seamounts are located off ship track, little sound energy from the sonar pulse that strikes the outward-facing slope may be reflected toward the sensor, and the quality of the returning echo is severely degraded. In such cases, the along-track distance between beams is significantly increased, and the variability in depths between pings can be large. Most of these values should be flagged. These types of errors can often be easily discovered when comparing against the overlapping deeper water swaths.

Navigation missing on first ping

In some cases, the first ping of a sonar file may be missing the associated position information supplied by the ship's onboard navigation system. This is generally a time synchronization issue, resulting from the sonar system having not yet received the position from the GPS system. This can be corrected by editing the sonar file's navigation and extrapolating the latitude and longitude backward from subsequent pings.

Inaccurate sound speed profile used

If an inaccurate sound speed profile was used in ray-tracing, the resulting across-track depth profiles tend to tilt upwards on the outermost beams on both sides ("smiles") or downwards ("frowns"). These data require the application of a better sound speed profile, which might be obtained from other, nearby sonar data, or historical averages for the region. In any event, the uncertainty in the resulting depths is expected to be large, especially on the outermost beams. Some processing software packages have the facility to modify the observed sound speed profile in order to bend the outer beams of multiple overlapping swaths until a best fit is achieved in the overlap area.

6.2.5 Common lidar errors

Bomb craters

Most lidar data sets contain scattered too-low points, or negative blunders, perhaps produced when a specular reflection or too-close ground saturates the detector and produces an internal echo. If vegetation reflections are removed by a find-the-lowest-point-in-the-vicinity algorithm,

DRAFT

true ground points adjacent to the negative blunders may be misidentified as vegetation reflections and removed. The result can be a conical crater that is entirely an artifact. (from Puget Sound Lidar Consortium, 2011)

Vegetation and buildings

First returns in lidar surveys commonly come from the tops of tree canopies, buildings, etc., not ground surface. The lidar data need to be classified, allowing removal of all non-bare earth returns. Some lidar systems provide backscatter intensity values that can be used to validate assumptions about types of ground cover. If geo-registered colour imagery is available, this should also be used to validate assumptions about types of ground cover.

Water-surface values

Lidar surveys (topographic) at the coast are typically flown at low tide for surveying of the inter-tidal zone. As such, water-surface values over the ocean are common, and individual waves may be identifiable. A lidar classification schema may be used to remove these values (water surface returns are generally of low intensity). Alternatively, the data may be clipped to a coastline, though clipping to a high-water coastline may eliminate valid measurements in the exposed inter-tidal zone. As a last resort, edit the data manually by paying attention to where the slope of the inter-tidal zone transitions to wave caps and troughs.

6.3 Data Validation

Contributed by Rob Hare and Paola Travaglini, Canadian Hydrographic Service, and Martin Jakobsson, Stockholm University, Sweden

The final survey data should be subject to independent in-house validation employing documented quality control procedures.

6.3.1 Internal self-consistency –

comparison with other survey soundings (e.g., from overlapping swaths) on the same survey

6.3.2 External consistency –

comparison with other datasets (e.g., systematic offsets), checklines from independent systems (or vessels, or dates)

6.3.3 Comparison with reported survey specifications –

analysis of statistics at cross check points/lines, or in grid cells over flat seafloor, provided a sufficient number of points are available in each cell to compute reliable statistics.

References

DRAFT

Fugro Earthdata, 2007. Lidar Mapping Fact Sheet.
http://www.fugroearthdata.com/pdfs/FCT_Lidar-Educational_11-07.pdf

International Hydrographic Bureau, 2008. *IHO Standards for Hydrographic Surveys, 5th ed., Special Pub. No. 44*, Monaco. http://www.iho-ohi.net/iho_pubs/standard/S-44_5E.pdf

Puget Sound Lidar Consortium, 2011. *Lidar artifacts*, Seattle.
http://pugetsoundlidar.ess.washington.edu/lidardata/#LIDAR%20artifacts

The following has been extracted from 5th Edition, S-44 (2008) Annex B by Rob Hare and edited for context

6.4 Use of uncertainty surfaces

Many statistical bathymetry processing packages also have the ability to generate an *uncertainty surface* associated with the bathymetry using either input error estimates or by generating spatial statistics within grid cells. Displaying and codifying these *uncertainty surfaces* is one method of determining whether the entire survey area has met the required specifications. If some areas fall outside the specifications, these areas can be targeted for further data collection or use of alternative systems in order to reduce the uncertainty to within an acceptable tolerance. When performed in real-time, the sampling strategy can be adapted as the survey progresses, ensuring the collected data are of an acceptable quality for the intended use. Each agency is responsible for the validation of these processing capabilities prior to use.

6.5 Validation Procedures

The final data should be subject to independent in-house validation employing documented quality control procedures.

DRAFT

Chapter 7.0 Metadata Documentation

The following has been extracted from 5th Edition, S-44 (2008) Chapter 5 by Rob Hare and edited for context

7.1 Introduction

To allow a comprehensive assessment of the quality of survey data it is necessary to record or document certain information together with the survey data. Such information is important to allow exploitation of survey data by a variety of users with different requirements, especially as requirements may not be known when the survey data is collected.

7.2 Metadata

Metadata should be comprehensive but should comprise, as a minimum, information on:

- the survey in general e.g. purpose, date, area, equipment used, name of survey platform;
- the geodetic reference system used, i.e. horizontal and vertical datum including ties to a geodetic reference frame based on ITRF (e.g. WGS84) if a local datum is used;
- calibration procedures and results;
- sound speed *correction* method;
- tidal datum and reduction;
- uncertainties achieved and the respective *confidence levels*;
- any special or exceptional circumstances;
- rules and mechanisms employed for data thinning.

Metadata should preferably be an integral part of the digital survey record and conform to the "IHO S-100 Discovery Metadata Standard", when this is adopted. Prior to the adoption of S-100, ISO 19115 can be used as a model for the metadata. If this is not feasible similar information should be included in the documentation of a survey.

Agencies responsible for the survey quality should develop and document a list of *metadata* used for their survey data.

7.3 Point Data Attribution

All data should be attributed with its *uncertainty estimate* at the 95% *confidence level* for both position and, if relevant, depth. The computed or assumed scale factor applied to the standard deviation in order to determine the uncertainty at the 95% *confidence level*, and/or the assumed statistical distribution of errors should be recorded in the survey's metadata. (For example, assuming a Normal distribution for a 1D quantity, such as depth, the scale factor is 1.96 for 95% confidence. A statement such as "Uncertainties have been computed at 95% confidence assuming a standard deviation scale factor of 1.96 (1D) or 2.45 (2D), corresponding to the assumption of a Normal distribution of errors," would be adequate in the metadata.) For soundings this should preferably be done for each individual sounding; however a single uncertainty estimate may be recorded for a number of soundings or even for an area, provided the difference between the individual *uncertainty* estimates can be safely expected to be

DRAFT

negligible. The attribution should, as a minimum, be sufficient to demonstrate that the requirements of these Standards have been met.

7.4 Bathymetric Model Attribution

If a *Bathymetric Model* is required, *metadata* should include: the model resolution; the computation method; the underlying data density; uncertainty estimate/*uncertainty surface* for the model; and a description of the underlying data.

DRAFT

Chapter 8.0 Gridding Data

8.1 Introduction to gridding

TEXT IN PREPARATION

8.1.1 Polynomials

TEXT IN PREPARATION

8.1.2 Near-neighbor searches

TEXT IN PREPARATION

8.1.3 Triangulation

TEXT IN PREPARATION

8.1.4 Splines

TEXT IN PREPARATION

8.1.5 Kriging

TEXT IN PREPARATION

DRAFT

8.2 Examples of gridding methods

8.2.1 Gridding JAMSTEC multibeam data

Contributed by K. M. Marks, NOAA Laboratory for Satellite Altimetry, USA

A grid can be formed from the individual multibeam *xyz* points downloaded from the JAMSTEC website (see Figure 5.2.4). Because there are so many data points it is advantageous to first take their block average using GMT routine "blockmedian," calculating the median *z* (at the *x, y* location of the median *z*) for every non-empty grid cell on a 6 arc-second mesh. The next step is to use GMT routine "surface," an adjustable tension continuous curvature surface gridding algorithm, to form a grid at 6 arc-second spacing in longitude and latitude from the median depths. Routine "grdmask" is then used to create a mask that is applied to the grid using "grdmath" so that it holds values only in cells that contain one or more of the original *xyz* points. In Appendix B.2 we list the GMT routines used to create the grid from KR05-01 *xyz* multibeam points that is shown in Figure 8.2.1.1. In Appendix B.3 we list the GMT routines used to produce Figure 8.2.1.

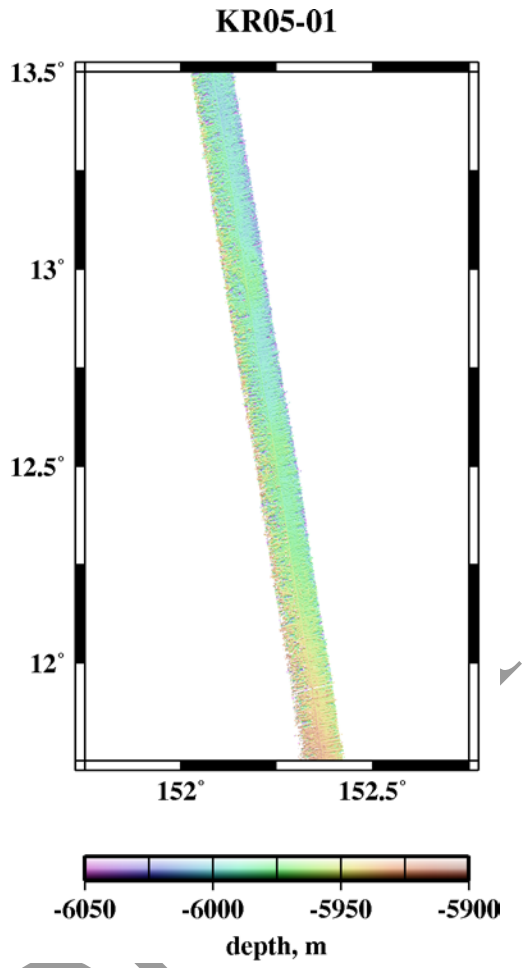


Figure 8.2.1.1 Color shaded-relief image of gridded KR05-01 multibeam points.

8.2.2 GMT Surface and Nearneighbor

We used depth values from the constrained grid cells from V12.1* in the study area as input into different gridding algorithms: GMT routine "surface," employed with the tension set to "0" and set to "1," and GMT routine "nearneighbor." Surface tension set to "0" gives the minimum curvature solution, and set to "1" gives a harmonic surface where maxima and minima are only possible at control points. "Nearneighbor" uses a nearest neighbor algorithm to assign an average value within a radius centered on a node. The GMT command lines used to produce the gridding solutions shown in Figure 8.2.2.1 are listed in Appendix B.14.

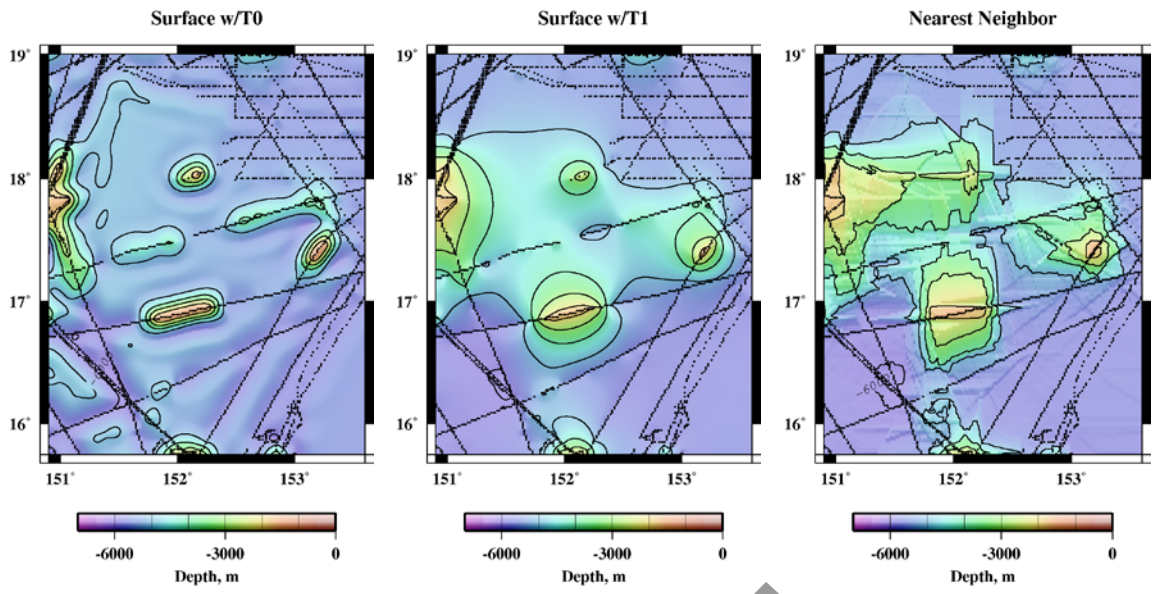


Figure 8.2.2.1 Results of gridding depths from V12.1* at constrained grid cells (black dots): GMT "surface" gridding routine with tension set to 0 (left) and set to 1 (middle), and "nearneighbor" gridding routine (right).

DRAFT

8.2.3 GMT Triangulate

TEXT IN PREPARATION

8.2.4 Spline pyramid

TEXT IN PREPARATION

8.2.5 Kriging

TEXT IN PREPARATION

8.2.6 Spline Interpolation

Contributed by Chris Amante and Matt Love, NOAA National Geophysical Data Center (NGDC)

Spline interpolation estimates elevation values using a mathematical function to create a continuous surface that passes exactly through all input elevations while minimizing the overall squared curvature. Oscillation effects of a minimum curvature surface can be reduced by using tension. The surface can be thought of as a flexible rubber plate that can be adjusted using

tension, with a tension of zero corresponding to the minimum curvature surface. Two open-source software programs, the Generic Mapping Tools (GMT; http://www.soest.hawaii.edu/gmt/), and MB-System (http://www.ldeo.columbia.edu/res/pi/MB-System/html/mbsystem_home.html), offer spline gridding algorithms, 'surface' and 'mbgrid,' respectively.

Prior to using GMT's 'surface,' it is recommended that the user pre-process the data with GMT 'blockmean', 'blockmedian', or 'blockmode' to avoid spatial aliasing and to eliminate redundant data. As an example, for two adjacent NOS surveys in xyz format, H10393, and H10394, pre-process the data using 'blockmedian'. Required user defined parameters include the input xyz file(s), the grid spacing (-I), the region of interest (-R), and the name of the output xyz file. Numerous other options can be specified (See GMT Manual Pages).

GMT 'blockmedian' example:

blockmedian H10393.xyz H10394.xyz -I1c/1c -R-88.13/-88.02/30.07/30.15 > blockmedian.xyz

Required user defined parameters for GMT 'surface' include the xyz input data, the name of the output binary 2-D .grd file (-G), the grid spacing (-I), and the region of interest (-R). Numerous other options can be specified (See GMT Manual Pages). Most notably, the user can define a tension value between 0 and 1 (-T). A tension value equal to zero indicates a minimum curvature solution.

GMT 'surface' example:

surface blockmedian.xyz -Gsurface_example.grd -I1c/1c -R-88.13/-88.02/30.07/30.15 -T0.25

DRAFT

Required user defined parameters for MB-System's 'mbgrid' include the name of the datalist containing a list of the input data files and their format (-I), and the character string to be used as the root of the output filenames (-O). There are numerous other user-defined parameters (See MB-System Manual Pages). Most notably, the user can define the datatype (-A), the distance from the data that the spline interpolation may be applied and the interpolation mode (-C), the grid spacing (-E), the gridding algorithm mode (-F), the format of the output grid file (-G), the region of interest (-R), the tension value between 0 and infinity (-T), and the extension of the internal grid so that the output grid is a subset from the center of a larger grid allowing for data outside of the output grid to guide the spline interpolation at the edge of the output grid (-X).

MB-System 'mbgrid' example:

mbgrid -Igebco.datalist -Ombgrid_example -A2 -C10/2 -E0.00027777778/0.00027777778/degrees! -F1 -G3 -R-88.13/-88.02/30.07/30.15 -T100 -X0.05

DRAFT

8.2.7 Global Mapper

TEXT IN PREPARATION

8.2.8 r2v

TEXT IN PREPARATION

8.2.9 Surfer

TEXT IN PREPARATION

8.2.10 Gridding in ArcMap

Contributed by Anastasia Abramova, Geological Institute Russian Academy of Sciences, Russia

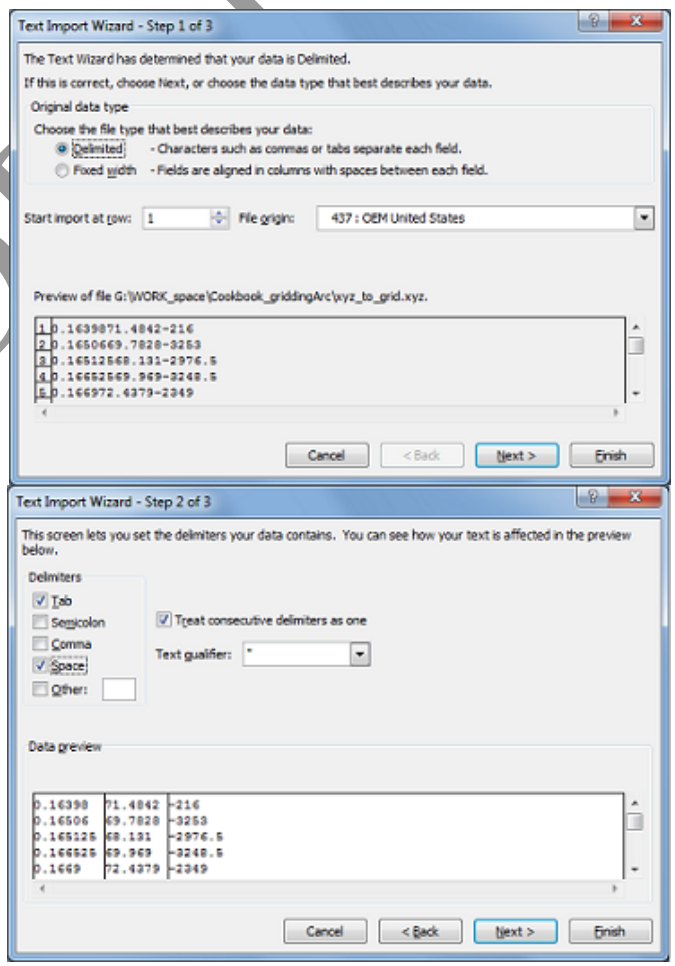
Here we present step by step example on how to open .xyz space delimited column data in ArcMap and to interpolate data points into a grid of appropriate resolution. As an example .xyz file with Latitude, Longitude and Depth in Geographic projection covering area between 2-11E and 72-74N is used. We would like to create a grid in UTM projection with appropriate cell resolution using Inverse Distance Weighting Interpolation algorithm. The procedure of reprojecting the data, assessing the data density to choose suitable grid size is described. Tools available in ArcMap in order to select optimal interpolation parameters and to compare resultant surface with original data points are discussed. In order to be able to carry out above noted you will need to have Spatial Analyst and Geostatistical Analyst extensions.

Importing .XYZ file in ArcMap

In order to open .xyz file in Arc, you need to modify it: text file needs to have header, data should be comma separated and file name should not have spaces in it.

In order to modify .xyz text file, import it into MS Excel:

- File>Open .xyz
-
- Select Delimiters that you have then press Finish

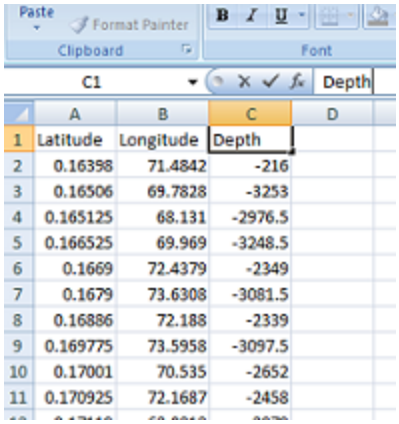


DRAFT

- Insert header row, type header name for each column (no spaces!)

- File> Save as .CSV (Comma delimited)

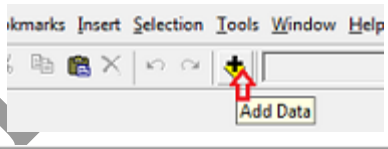
- Close file In MS Excel! (Arc doesn't want to open it if it is opened somewhere else)



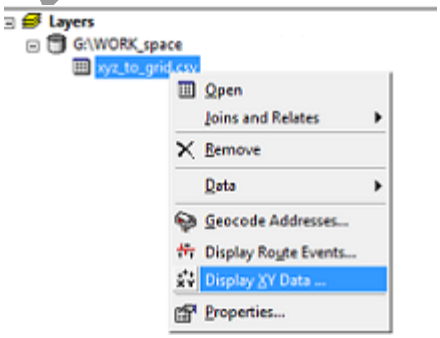
	A	B	C
1	Latitude	Longitude	Depth
2	0.16398	71.4842	-216
3	0.16506	69.7828	-3253
4	0.165125	68.131	-2976.5
5	0.166525	69.969	-3248.5
6	0.1669	72.4379	-2349
7	0.1679	73.6308	-3081.5
8	0.16886	72.188	-2339
9	0.169775	73.5958	-3097.5
10	0.17001	70.535	-2652
11	0.170925	72.1687	-2458

- Open ArcMap: File> New > Blank Document

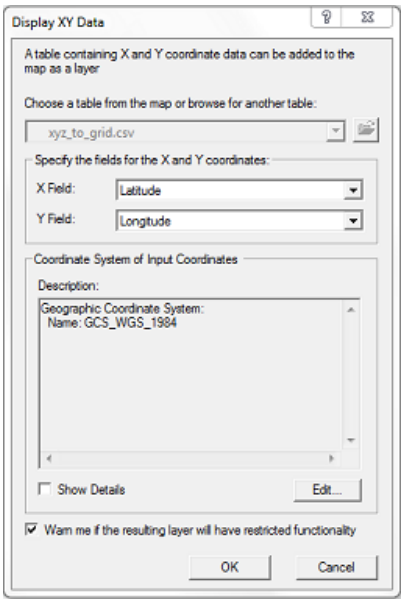
- Select Add Data from Standard Toolbar (Or go to File> Add Data) and add your .csv file



- To display your data right click on the layer> Display XY Data

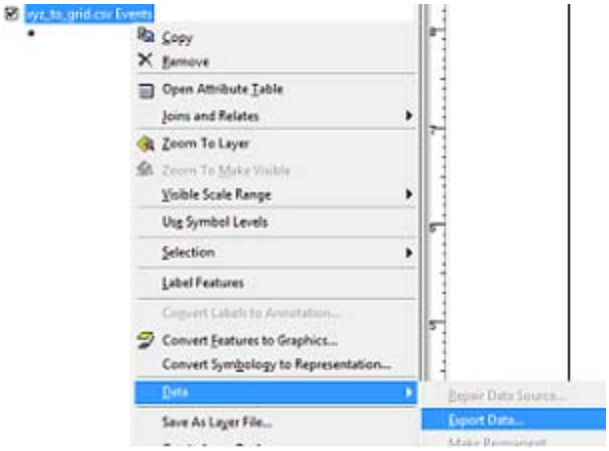


- Select the appropriate fields: X as Latitude and Y as Longitude. In order to define the input data projection: press Edit > Select> Geographic> World > WGS 1984. Press Ok



DRAFT

- In order to save file as ArcMap shape file, right click on the layer> Data> Export Data> Save as .shp file

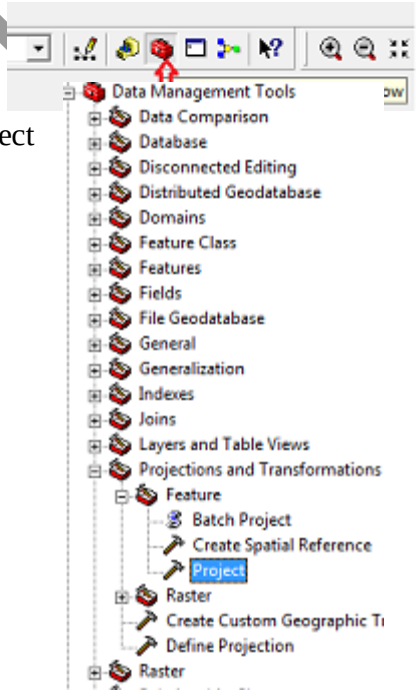


Reprojecting the data

- We would like to create a grid in UTM projection with units in meters, instead of degrees. For that we need to reproject our data.

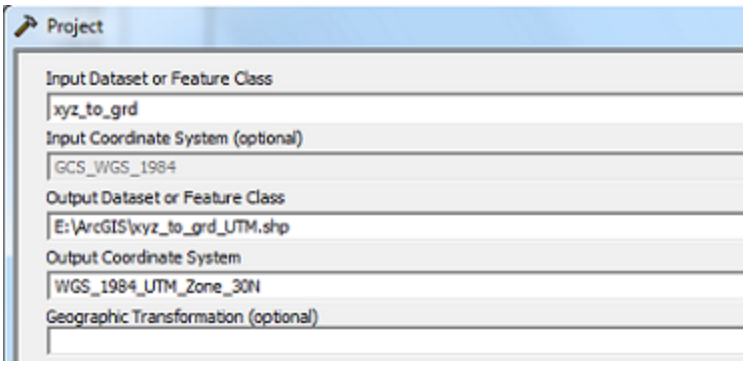
- Click on Show/Hide ArcToolbox:

- In opened ArcToolbox choose Data Management Tools> Projections and Transformations> Feature> Project



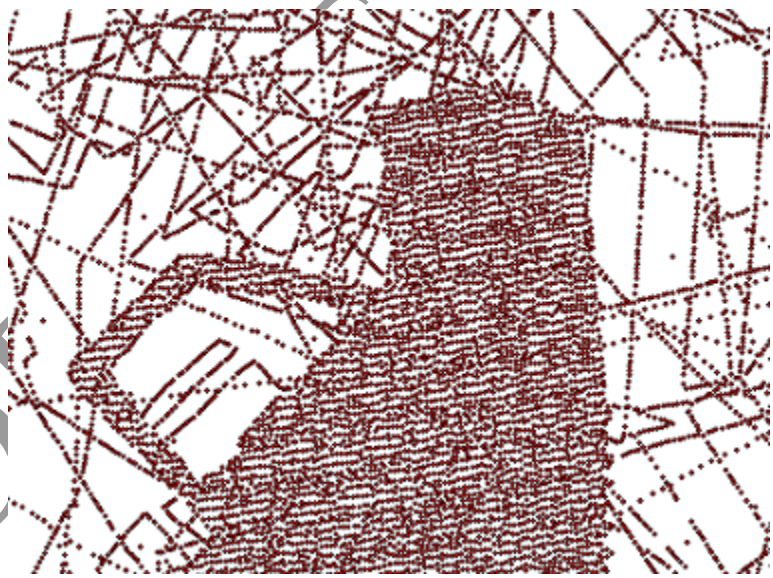
DRAFT

- Define the Input and Output files and Output Coordinate System (In our case it is WGS 1984, UTM Zone 30N). Press Ok.



- You might want to change the data frame projection view from Geographic to UTM_Zone_30N now (projection in the data frame view is defined by projection of the first opened file). Right click on Layers> Properties> Coordinate Systems. Select from Predefined list Projected> UTM> WGS 1984> UTM Zone 30N. Press Ok. Now data frame view has the same projection as reprojected file.

- Finally our added data looks like this. As you can notice, the data point density varies over the area



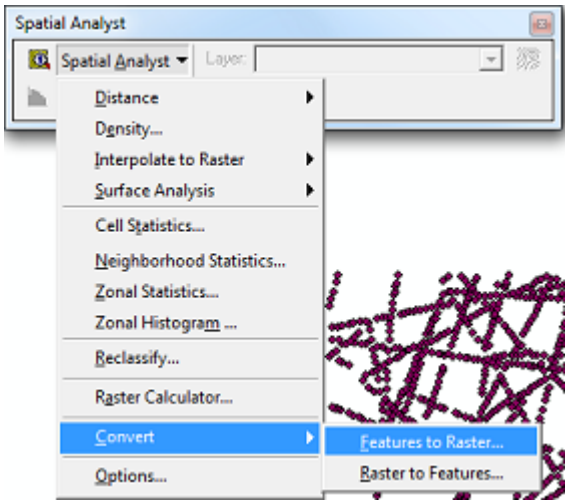
Defining cell resolution for the grid

- The choice of grid resolution depends on your purposes. Meanwhile, data distribution, density, complexity of modeled terrain and scale of the final map should be taken into account as well.

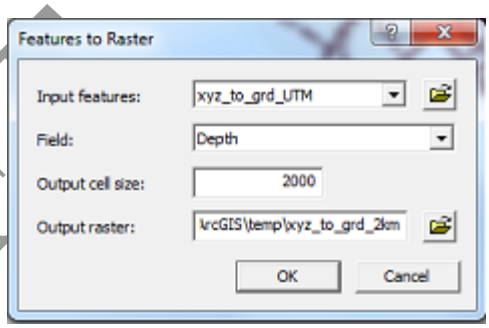
- We would like to create grid with 2 km cell resolution. Here we show how to identify the percentage of cells which values will be defined by interpolation and how to estimate the data density within each cell for the chosen grid size.

- Make sure that Spatial Analyst Extension is activated: go to Tools> Extensions> check Spatial Analyst. Also, add Spatial Analyst toolbar: go to View> Toolbars> check Spatial Analyst.

- From Spatial Analyst toolbar select Convert> Features to Raster

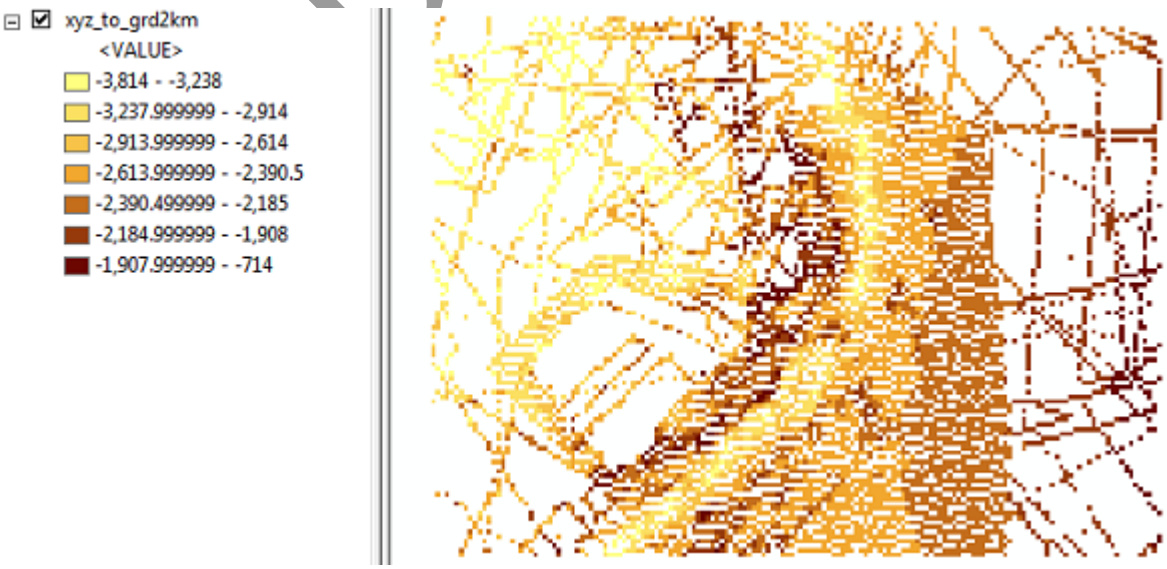


- Select appropriate field and cell resolution:

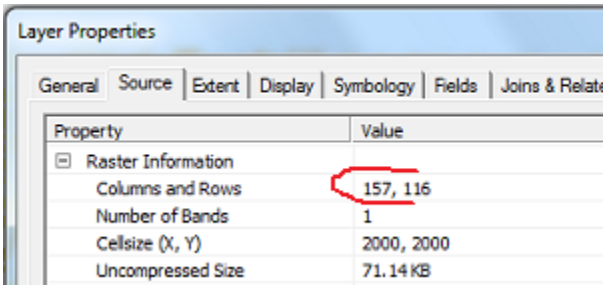


DRAFT

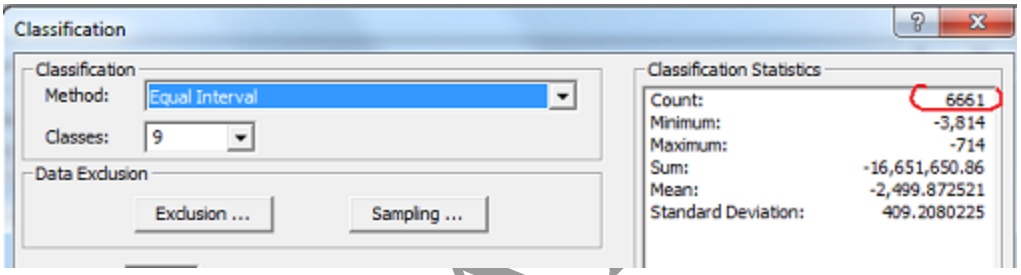
- As you can see, in the created raster, only cells with data points in it have non zero values:



- In order to find out what is the percentage of cells with “no data values”, go to Layer> Properties> Source. Here you can look up the total number of cells in the grid created by multiplying total number of rows by number of columns (157*116=18212)



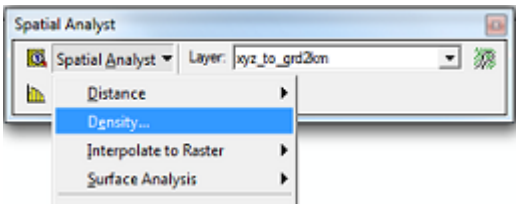
- In order to find out how many cells have non zero values, go to Symbology tab in Layer Properties window and click on Classified. Select Classify tab. The classification table will appear. In Classification Statistics Window, Count will tell the amount of non zero cells (6661). This is approximately 36 % of total amount of cells. Therefore, the values for 64% of cells will have to be estimated by interpolation.



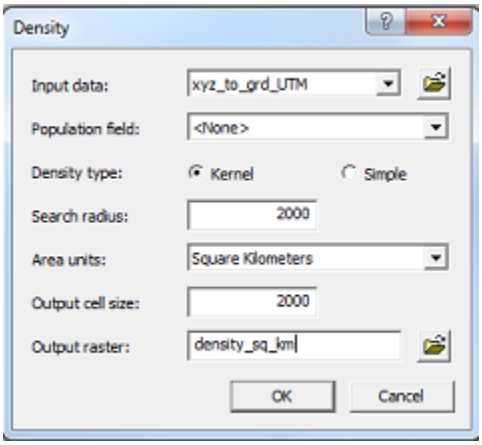
DRA

- Now we would like to find out what is the maximum density of data points per cell for the chosen grid size (we don’t want it to be too small or too big - although in the end in depends on your purpose).

- From Spatial Analyst toolbar select Density:



- Leave population field to “None”, define search radius and output cells size in projection units (meters), and desired output area units (square km)
  For more information on how the density tools work go to:
  http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/An_overview_of_the_Density_tools/009z0000000r000000/



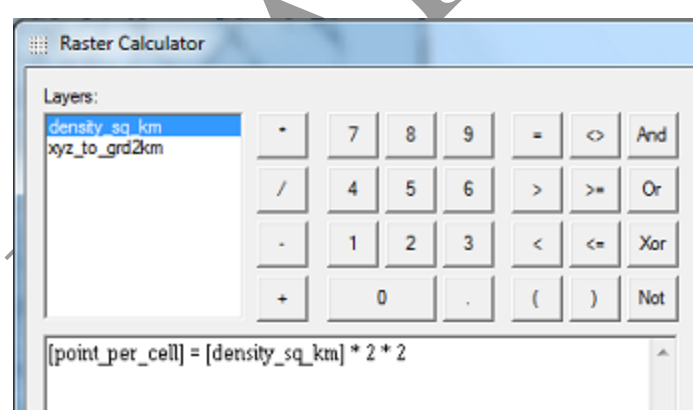
- Calculated result represents density of data points per unit area (squared km) within the neighborhood of defined search radius (2000 by 2000 square meters).

- In order to calculate amount of data points per grid cell of chosen resolution, go to Spatial Analyst toolbar and select Raster Calculator.

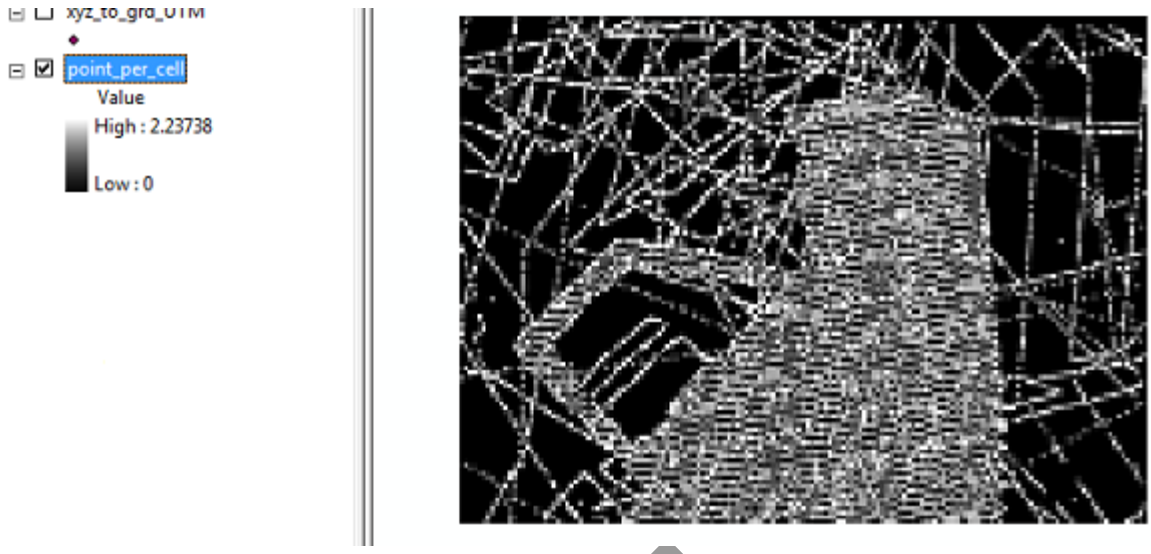
- In Raster Calculator syntax is very important. For more information on map algebra rules and on how to use raster calculator:

  http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#/An_overview_of_the_rules_for_Map_Algebra/00p600000006000000/
  http://help.arcgis.com/en/arcgisdesktop/10.0/help/index.html#//009z000000z7000000.htm

- We would like to create a new raster layer with the name [point_per_pixel] (put layer name into square brackets and better avoid spaces in the names). We would calculate number of points per cell from the density layer by multiplying it by the area of search radius (2000 by 2000 square meters). Note that units should be the same as units in the density layer: density map was in km squared, therefore we multiply density layer by 2*2 km sq:



- As a result we get the following raster:



- As you can see, density of points is maximum 2.2 points per grid cell. This can be considered not enough; therefore you might want to change grid size to coarser resolution. We will use this grid spacing in our example.

Creating a gridded surface

- There are several interpolation algorithms available in ArcMap, these include Global and Local Polynomials, Kriging, Spline, Natural Neighbor and Inverse Distance Weighting available within Spatial Analyst and 3D extensions . The description of these interpolation algorithms can be found in other sections of Chapter 8.2. More information on interpolation from point to surface methods can be found:

- http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=An_overview_of_the_Raster_Interpolation_toolset
http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=An_overview_of_the_Interpolation_toolset

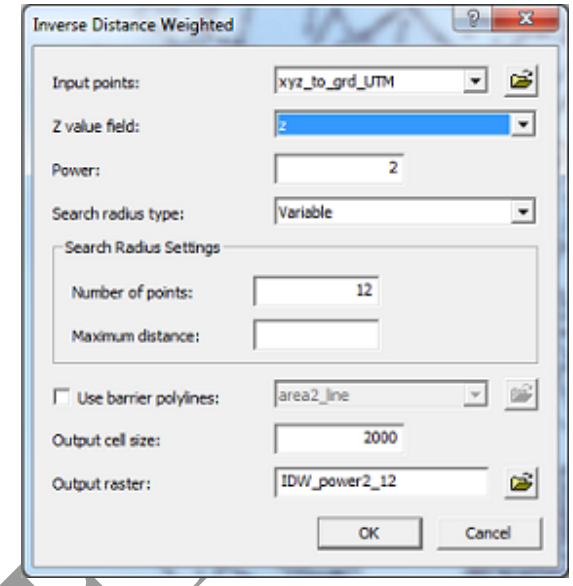
- Here we will present how to create a gridded surface using Inverse Distance Weighting (IDW) algorithm. IDW can be a good choice for fast interpolation of sparse data. IDW interpolation predicts values at unmeasured locations according to the values from the surrounding data points. Points which are closer to the prediction location have more influence (weight) on prediction than those which are further away. For more information on IDW see:

- http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?id=3304&pid=3302&topicname=How_Inverse_Distance_Weighted_%28IDW%29_interpolation_works

DRAFT

- IDW is available through ArcToolbox Spatial Analyst Tools> Interpolation>IDW (or through Spatial Analyst toolbar> Interpolate To Raster> Inverse Distance Weighting). The IDW window will appear.

- As you can see several parameters need to be defined. *Power* defines the distance weight given to each data point in the interpolation process. The higher the power, the faster the weights decrease with distance. *The number of points* defines number of points with known values to be used for prediction of the unknown value.



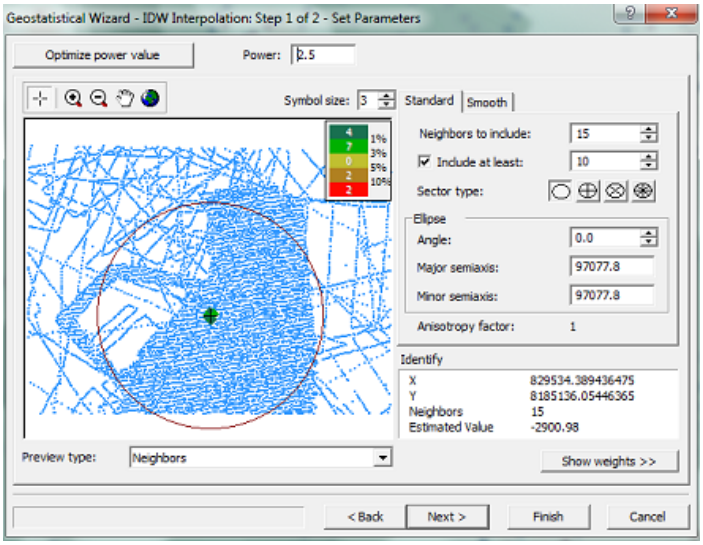
- In order to choose optimal parameters for interpolation, you can use Geostatistical Analyst toolbox. Activate the extension through Tools> Extensions> check Geostatistical Analyst. Also add it to your display through View> Toolbars> Geostatistical Analyst.

- Geostatistical Analyst> Geostatistical Wizard provides cross-validation of your model according to parameters chosen. The wizard allows you to manipulate parameters and to look at the output statistics for the created model. For more information see:

http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?id=3355&pid=3334&topicname=Performing_cross-validation_and_validation

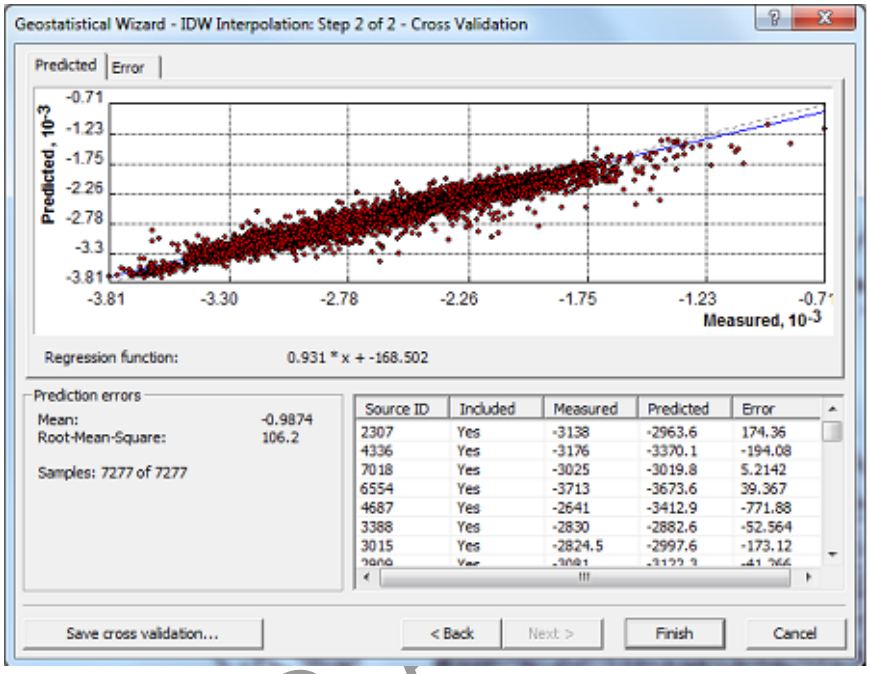
- In Geostatistical Analyst open Geostatistical Wizard. Select input data point layer and attribute (Depth)

- As an example we will use Power 2.5, and minimum number of points 10. Click next.

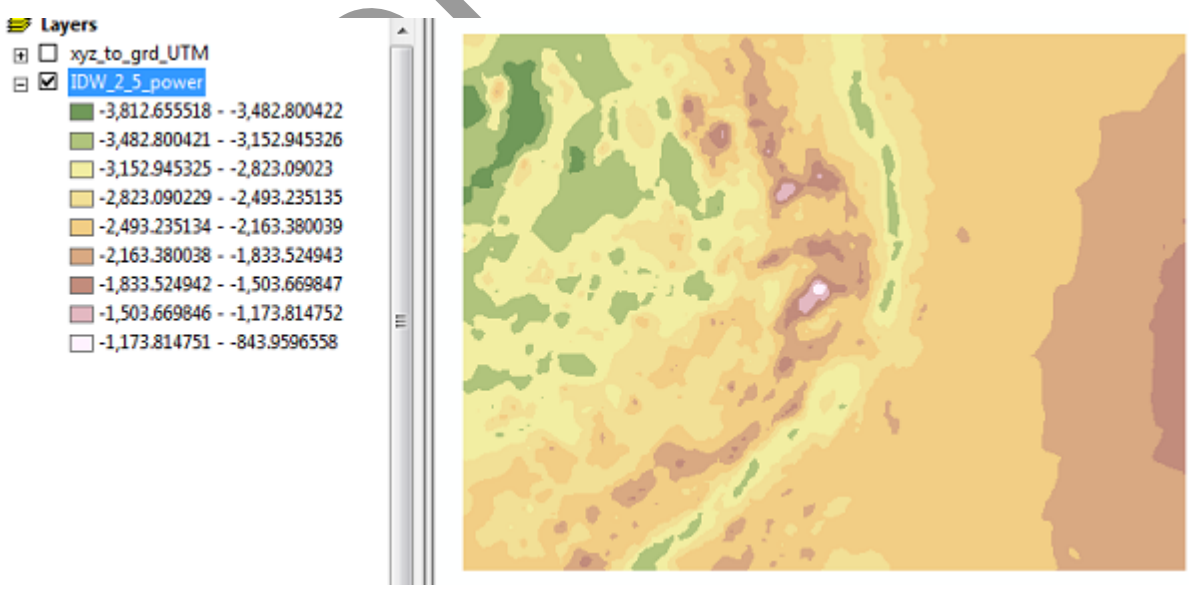


DRAFT

- You can see from the figure below results of our cross-validation. The root mean square error is more than 100 meters. This can be explained by low density of data points (see figure above) and overall uncertainty of depth in the area (average depth of 2000m, with uncertainty of around … water depth). You can vary interpolation parameters and see how they affect on the model fit.



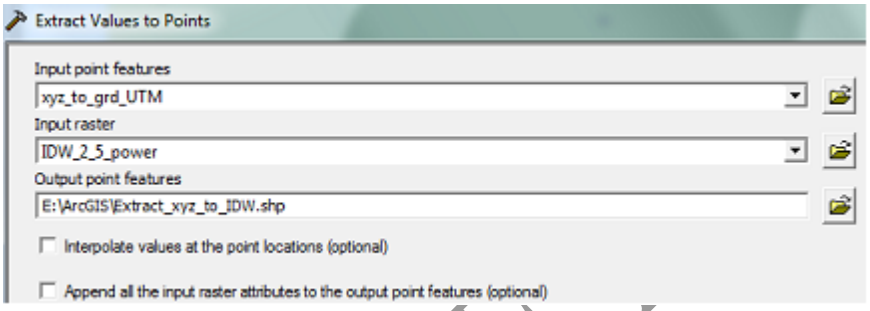
- Our final surface using above described parameters look like that:



Comparison with original data values

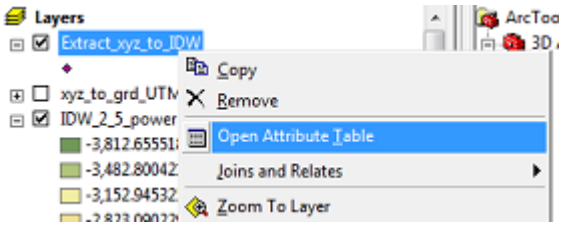
- Here we present how to compare the original data point values with created surface. This could be useful in order to investigate spatial distribution of misfit between model and the measured points. For more information on quality assessment of grids see section 9.2.1.

- You can extract values from the gridded surface corresponding to original data points with Spatial Analyst Tools> Extract Values to Points. Where Input point features are your original .xyz data points and Input raster is your interpolated raster :
-

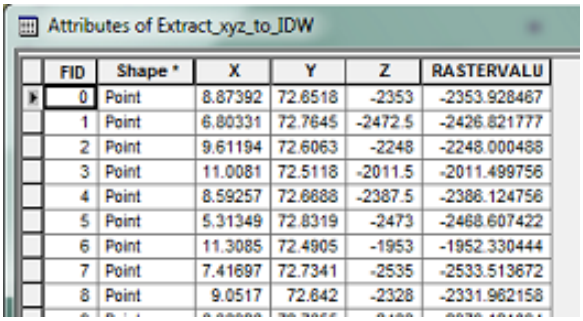


- As a result, a point feature layer is created (we call it Extract_xyz_to_IDW).

- You can view the Attribute table of the layer by right click on the layer> Open attribute table:



- As you can see from the table it contains information for original data points (x,y,z) and corresponding depth values from our gridded surface (RASTERVALU) for each original data point:

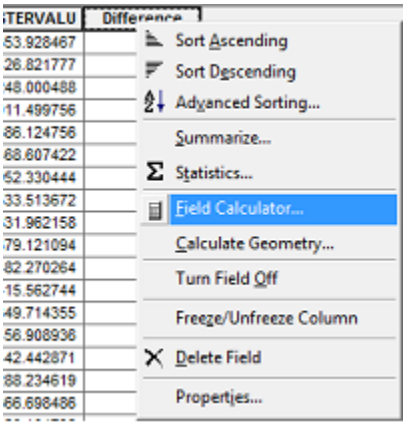

Attributes of Extract_xyz_to_IDW

FID	Shape *	X	Y	Z	RASTERVALU
0	Point	8.87392	72.6518	-2353	-2353.928467
1	Point	6.80331	72.7645	-2472.5	-2426.821777
2	Point	9.61194	72.6063	-2248	-2248.000488
3	Point	11.0081	72.5118	-2011.5	-2011.499756
4	Point	8.59257	72.6688	-2387.5	-2386.124756
5	Point	5.31349	72.8319	-2473	-2468.607422
6	Point	11.3085	72.4905	-1953	-1952.330444
7	Point	7.41697	72.7341	-2535	-2533.513672
8	Point	9.0517	72.642	-2328	-2331.962158

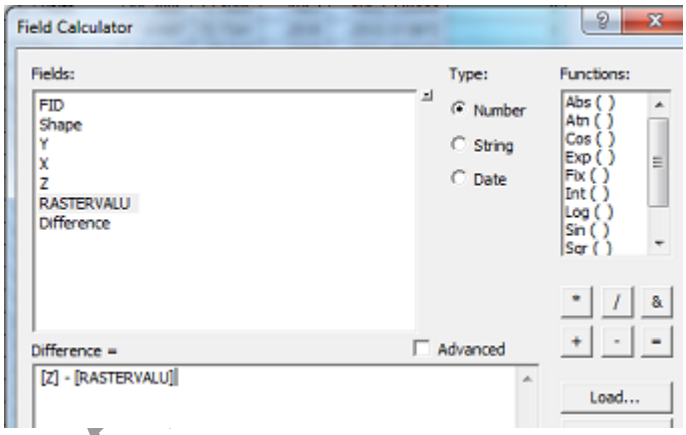
DRAFT

- You can calculate differences between these values and visualize them.

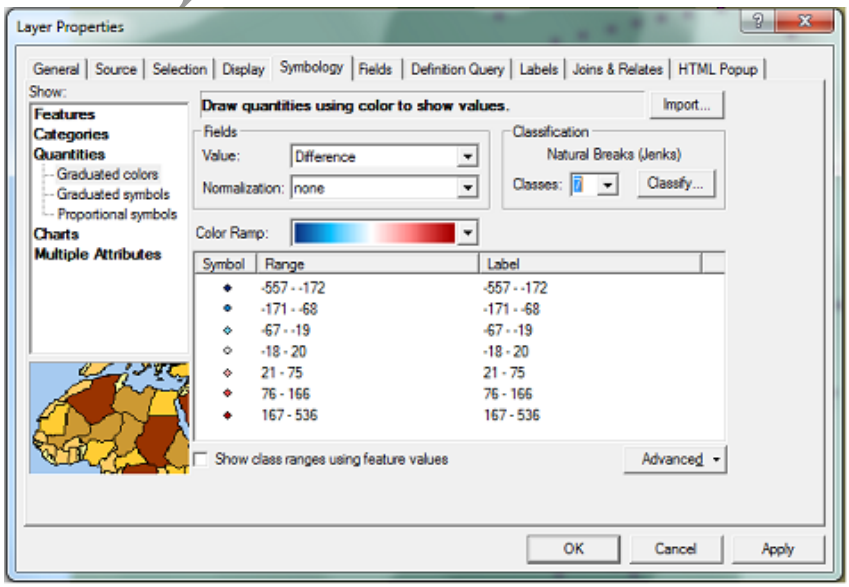
- Go to Options inside the Attribute table and select Add Field option. We will call the new column as "Difference". After the column is created, right click on the header of new created column and select Field Calculator



- Inside the Field calculator we can define expression for the new column values: a difference between Z and RASTERVALUE columns. After clicking Ok, the values are calculated. Close the attribute table

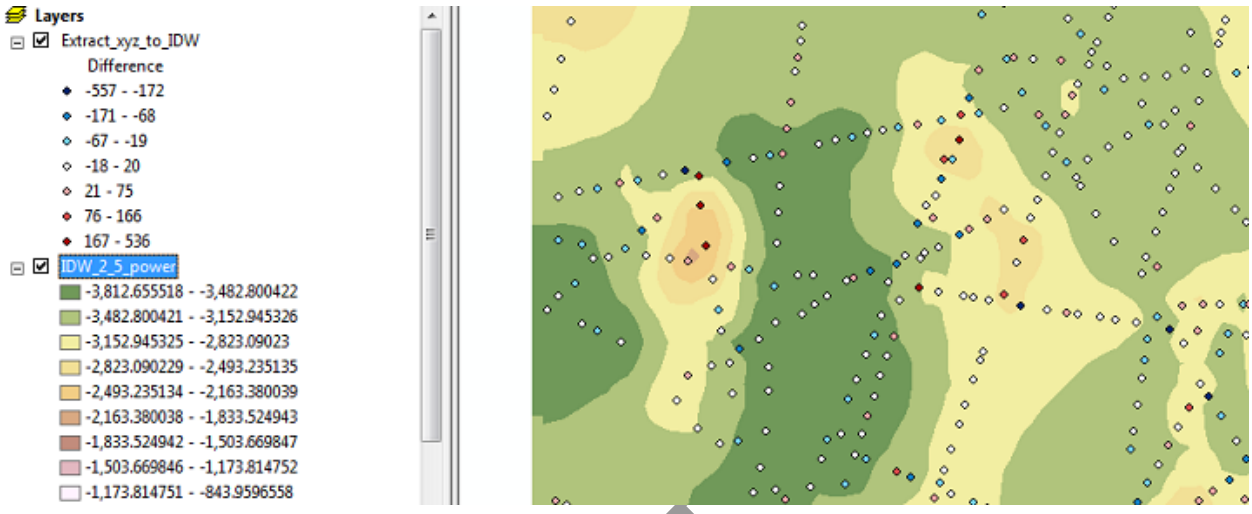


- Now you can visualize the Difference by right clicking on the point layer (Extract_xyz_to_IDW)> Properties. Go to Symbology tab. Select Quantities> Graduated colors. Select the Value Fields as Difference, select the color ramp and number of classes, and click OK.classes, and click OK.



DRAFT

- Now you can see overlaid difference results over the created gridded surface and investigate the regions where differences seem to be significant:



- For more information on analyzing gridded surfaces that you create go to:

http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?TopicName=About_analyzing_raster_data
http://webhelp.esri.com/arcgisdesktop/9.2/index.cfm?id=603&pid=598&topicname=Surface_creation_and_analysis

See more on interpolation error analyses in other sections 8.3.

DRAFT

8.3 Testing of gridding methods

TEXT IN PREPARATION

8.3.1 Distance-to-control

TEXT IN PREPARATION

8.3.2 Histograms

TEXT IN PREPARATION

8.3.3 Statistical tests

TEXT IN PREPARATION

Chapter 9.0 Uncertainty

9.1 Sources of Uncertainty

The following has been extracted from 5th Edition, S-44 (2008) Annex A, section A.4 by Rob Hare and editing for context

Although the following text focuses on uncertainties in data acquired with swath systems, it should be noted that it is in principle applicable to data acquired with any depth measurement system. A single-beam echosounder is just a special case of a multibeam echosounder (i.e. nadir beam only)

With swath systems the distance between the sounding on the seafloor and the positioning system antenna can be very large, especially in deep water. Because of this, sounding position *uncertainty* is a function of the errors in vessel heading, beam angle, refraction correction model and the water depth in addition to the uncertainty of the positioning system itself.

Roll and pitch errors will also contribute to the uncertainty in the positions of soundings. Overall, it may be very difficult to determine the position *uncertainty* for each sounding as a function of depth. The uncertainties are a function not only of the swath system but also of the location of, offsets to and accuracies of the auxiliary sensors.

The use of non-vertical beams introduces additional uncertainties caused by incorrect knowledge of the ship's orientation at the time of transmission and reception of sonar echoes. Uncertainties associated with the development of the position of an individual beam should include the following:

l) Positioning system uncertainty;
m) Range and beam angle uncertainty;
n) The uncertainty associated with the ray path model (including the sound speed profile), and the beam pointing angle;
o) The uncertainty in vessel heading;
p) System pointing uncertainty resulting from transducer misalignment;
q) Sensor location;
r) Vessel motion sensor errors i.e. roll and pitch;
s) Sensor position offset uncertainty; and
t) Time synchronization / latency.

Contributing factors to the vertical *uncertainty* include:

a) Vertical datum uncertainty;
b) Vertical positioning system uncertainty (if relevant);
c) Tidal measurement or prediction uncertainty, including co-tidal uncertainty where appropriate;
d) Range and beam angle uncertainty;

DRAFT

e) The uncertainty associated with the ray path model (including the sound speed profile), and the beam pointing angle;
f) Ellipsoidal / vertical datum separation model uncertainty (if relevant);
g) Vessel motion uncertainty, i.e. roll, pitch and heave;
h) Vessel dynamic draught, including static draft, settlement and squat;
i) Seabed slope (when combined with positioning uncertainty); and
j) Time synchronization / latency.

Agencies responsible for the survey quality are encouraged to provide uncertainty budgets for their own systems.

9.2 Measuring uncertainty

9.2.1 Methods of quality assessment, artifacts

TEXT IN PREPARATION

9.2.2 Comparisons with multibeam

TEXT IN PREPARATION

Local

TEXT IN PREPARATION

Regional

TEXT IN PREPARATION

Global

TEXT IN PREPARATION

DRAFT

9.2.3 Monte Carlo Technique

Contributed by Paul Elmore, Naval Research Laboratory, Stennis Space Center, USA

A peer-reviewed methodology for estimating error on historic data sets using a Monte Carlo technique is published in Jakobsson et al. (Jakobsson et al. 2002); Figure 9.2.3.1 illustrates the procedure. It can be used on either historic or new soundings to provide an error layer for the GMT Splines-In-Tension routine (Surface routine). In order for this technique to work, estimates of the horizontal and vertical uncertainties for the soundings are needed.

In the simulations, the original two-dimensional navigation positions and the one-dimensional soundings are randomly perturbed using a normally distributed random number generator (RNG). Let the number of original soundings be J, the number of surveys be K, the horizontal navigation uncertainty of the Kth survey be H_K, the vertical sonar uncertainty of the Kth survey be V_K, the number of output grid points be I, and the number of Monte Carlo simulations be N, with each loop denoted by n. For the Kth survey, the RNG perturbs the position data ~ $\mathcal{N}(0, H_K^2)$[†] and the vertical position ~ $\mathcal{N}(0, V_K^2)$. The gridded bathymetry surface is constructed from the GMT Surface (or other interpolator of choice) during each loop of the Monte Carlo simulation, resulting in N different interpolated bathymetry surfaces. The gridded uncertainty estimate is then the standard deviation of the N surfaces at each ith grid point, $i \in I$. Assuming that all sounding used have been cleaned and corrected for ship dynamics, sound speed profile errors, etc., positional uncertainty and the bottom slope predominantly influence the bathymetric uncertainty estimated from this method.

A benefit of this method is that it is relatively simple to code: one simply needs the interpolation and RNG routines called within a Monte Carlo shell that perturbs the horizontal and vertical positions. This procedure, however, is computationally intensive and requires the use of original soundings data. Attempting this approach for a large number of soundings data may not be pragmatic.

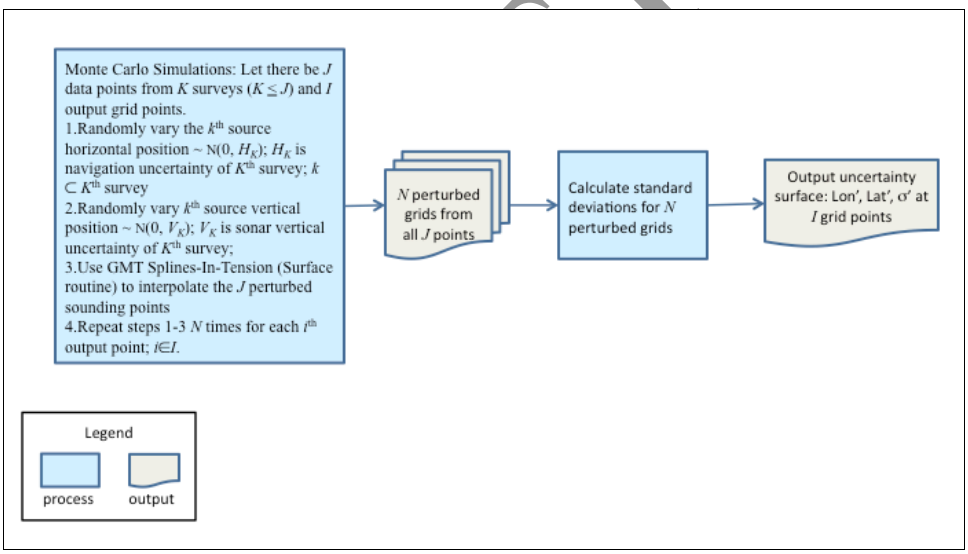


Figure 9.2.3.1 Process flow for the Monte Carlo procedure of Jakobsson et al (2003).

9.2.4 Bayesian Network Technique

Contributed by Paul Elmore, Naval Research Laboratory, Stennis Space Center

When the data sets are large, a Bayesian network (BN; a good discussion is given in Chapter 14 of Russell and Norvig (2003)) can be an alternative to the Monte Carlo method. With

[†] *The notation ~ $\mathcal{N}(\mu, \sigma^2)$ means that the quantity follows a normal, or Gaussian, probability distribution with mean μ and variance σ^2*

this approach, published in Elmore et al. (2009) and illustrated in Figure 9.2.4.1, causal relationships of navigation error and bottom slope to bathymetric uncertainty are quantified by conditional probability densities (CPD's). One uses the Monte Carlo technique on representative sets of soundings data to tabulate the CPD's necessary for the statistical inference. The BN then produces a histogram of this uncertainty estimate for an area given the navigation errors used to survey the region and bottoms slopes that are present. An intermediate uncertainty is obtained from the mean plus one standard deviation of the histogram. The final uncertainty estimate is obtained by first adding the square of this intermediate uncertainty to the square of the vertical error estimate under the assumption of statistical independence between the two, then computing the square root of this sum. This type of computation is much less costly than the Monte Carlo technique both in computation overhead and required input data, potentially leading to a significantly large (two orders of magnitude) increase in computational speed. Estimates of the horizontal and vertical uncertainty for the soundings must be available to both train the network and run the Bayesian Network.

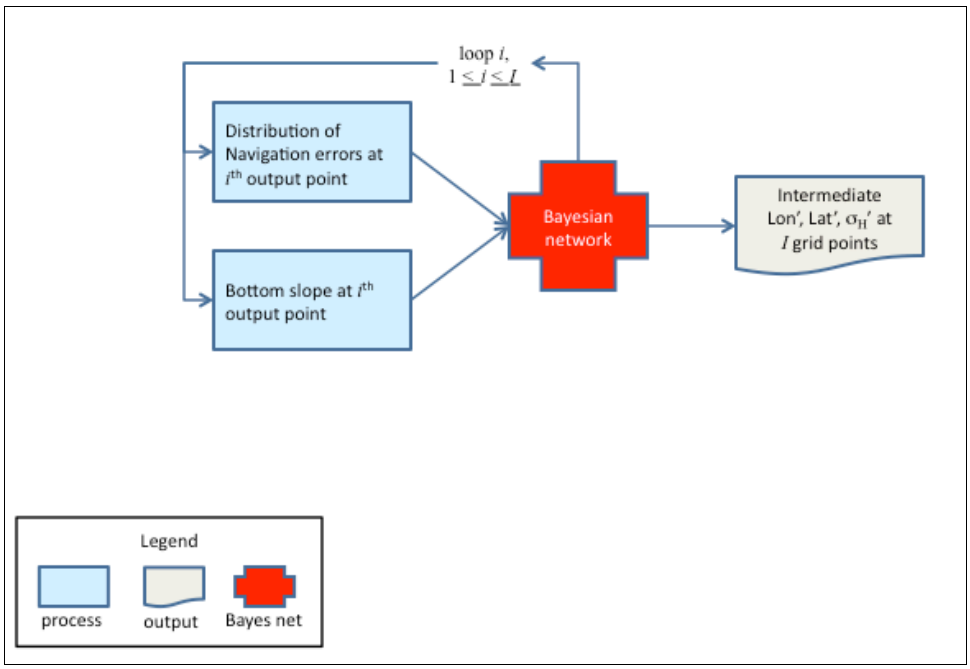


Figure 9.2.4.1 Conceptual flow of the Bayesian Network approach.

Bayesian Network Design and Access

The Bayesian network itself can be programmed using commercial Netica software, Version 4.0.8 (2008), which provides a GUI for construction and programming of the BN and (as of this writing) is free to use for the small Bayesian network created here. Figure 9.2.4.2 shows the network topology as displayed by the Netica GUI from Elmore et al (2009). It consists simply of two main parent nodes, one for navigation error and one bottom slope, and one child node, the intermediate uncertainty estimate. In this example, there are eight different surface navigation uncertainties categories; Table 9.2.4.1 associates historical navigation techniques to uncertainty. The bottom slope and bathymetric uncertainties follow a logarithmic binning

scheme so that one significant digit is maintained in the uncertainty estimate while keeping the BN to a manageable size. Once created, the software stores the BN as either a binary object or ASCII file that can be accessed and manipulated through the use of vendor provided application program interface (API) libraries.

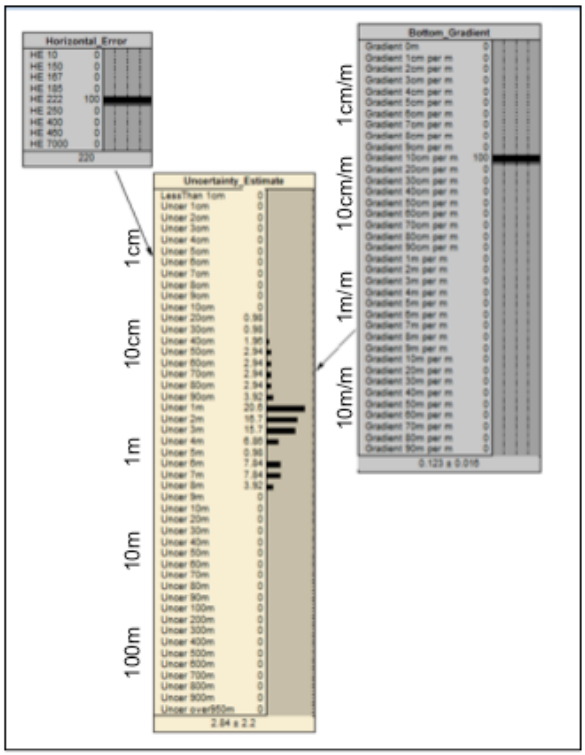


Figure 9.2.4.2 Bayesian network topology as displayed by the Netica software GUI. The "Horizontal_Error" node variables follow the navigation errors provided in Table 1, with the numbers for surface navigation error in meters. The "Bottom_Gradient" and "Uncertainty_ Estimate" nodes follow a logarithmic binning scheme.

Table I: Horizontal Error Categories

Navigation Mode	Accuracy	Navigation Mode	Accuracy
GPS/SINS (3 or more Satellites)	10-15 m	NAVSAT/Single Range LORAN/SINS	250 m
GPS/DR (3 or more Satellites)	10-15 m	NAVSAT/ SINS	250 m
NAVSAT/Range Range LORAN/SINS	150 m	NAVSAT/Single Range LORAN/DR	250 m
NAVSAT/Range Range LORAN/DR	167 m	NAVSAT/DR	400 m
NAVSAT/Hyperbolic LORAN/SINS	185 m	LORAN/SINS LORAN/DR	463 m
NAVSAT/Hyperbolic LORAN/DR	222 m	Satellite Altimetry	7000 m

Table 9.2.4.1 Navigation uncertainty categories used in Elmore et al. (2009) for the Bayesian Network approach to uncertainty estimation. The abbreviations are as follows: GPS – Global Positioning System; SINS – Ship's Inertial Navigation System; DR – Dead Reckoning; NAVSAT – Navy Navigation Satellite System; LORAN – Long Range Navigation. The "/" means that two or more techniques are combined.

DRAFT

Bayesian Network Training Procedures

Using the procedure from Jakobbson et al. (2003), one ascertains the propagation of navigational error and bottom slope into bathymetric uncertainty by Monte Carlo simulations. Let there be a total on M navigation error categories. The results of the simulations (a total of $M \times N$ simulations) to populate the CPT of the BN, or "train" the BN. Figure 9.2.4.3 shows a high level view of the entire training process, bracketed into three main blocks: 1) Monte Carlo simulations, 2) standard deviation and slope calculations, and 3) Bayes network programming.

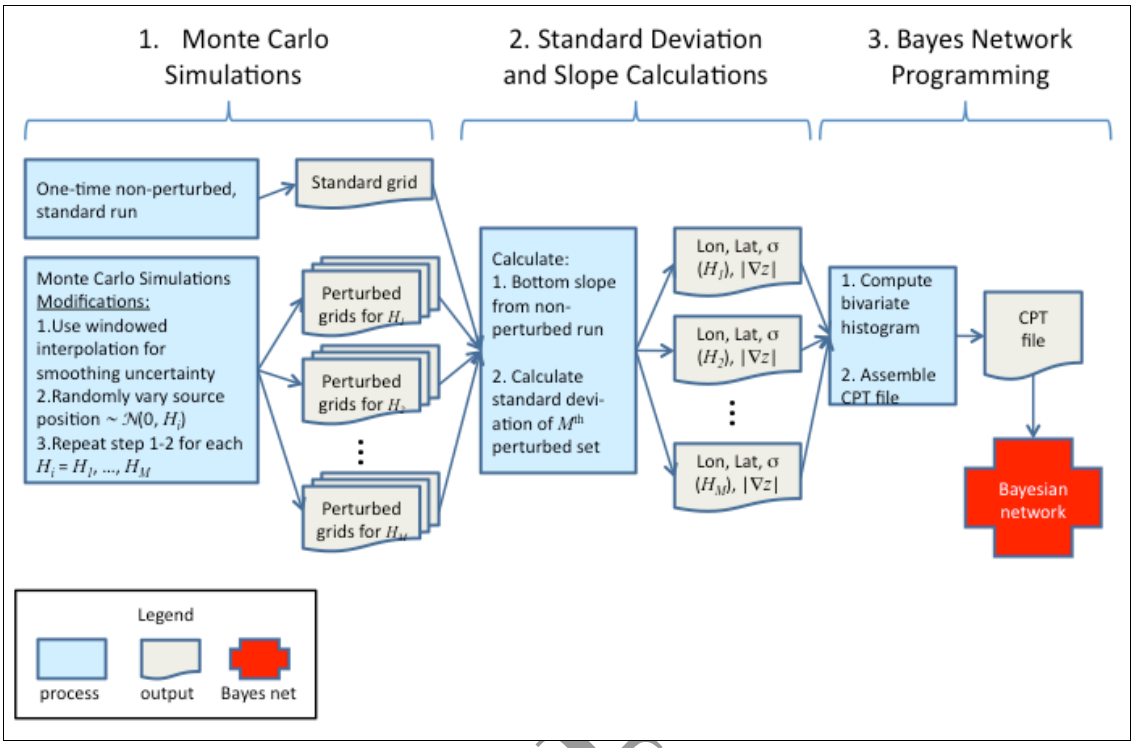


Figure 9.2.4.3 High level view of the Bayesian network training process. The figure illustrates computation of the standard deviations as modified from Jakobsson (2003), and use of these results to construct the conditional probability tables (CPT) for the Bayesian network.

DRAFT

Figure 9.2.4.4 shows a low level flow chart for the Monte Carlo block. In the Monte Carlo simulations, the original locations of the sounding points (or gridded points acting as sounding points) are first moved in random directions. A normal RNG (Marsaglia and Tsang 2000) provides deviations for the north and east positions as specified by the horizontal error category for the mth simulation loop, H_m ($1 \leq m \leq M$), such that the RNG ~ $\mathcal{N}(0, H_m^2)$. The data points are then interpolated to provide a bathymetry surface and stored to an ASCII text file, with the Monte Carlo loop number and navigation error stored in the file name. As in Jakobbson et al. (2003), N = 100 Monte Carlo iterations may be performed for each mth set of simulations; however, a difference between the two is that the set of N Monte Carlo simulations are then repeated M times for each horizontal error category in order to fully populate the CPT of the BN. Another difference from the Monte Carlo procedure is that each sounding is assumed to have the same probability distribution of horizontal errors so that one can obtain a statistical distribution for bathymetric uncertainty for a range of horizontal errors. Although the actual data do not necessarily have these errors, it is assumed they do in order to train the network. Also, the vertical positional will be treated algebraically as discussed in Section 3 below.

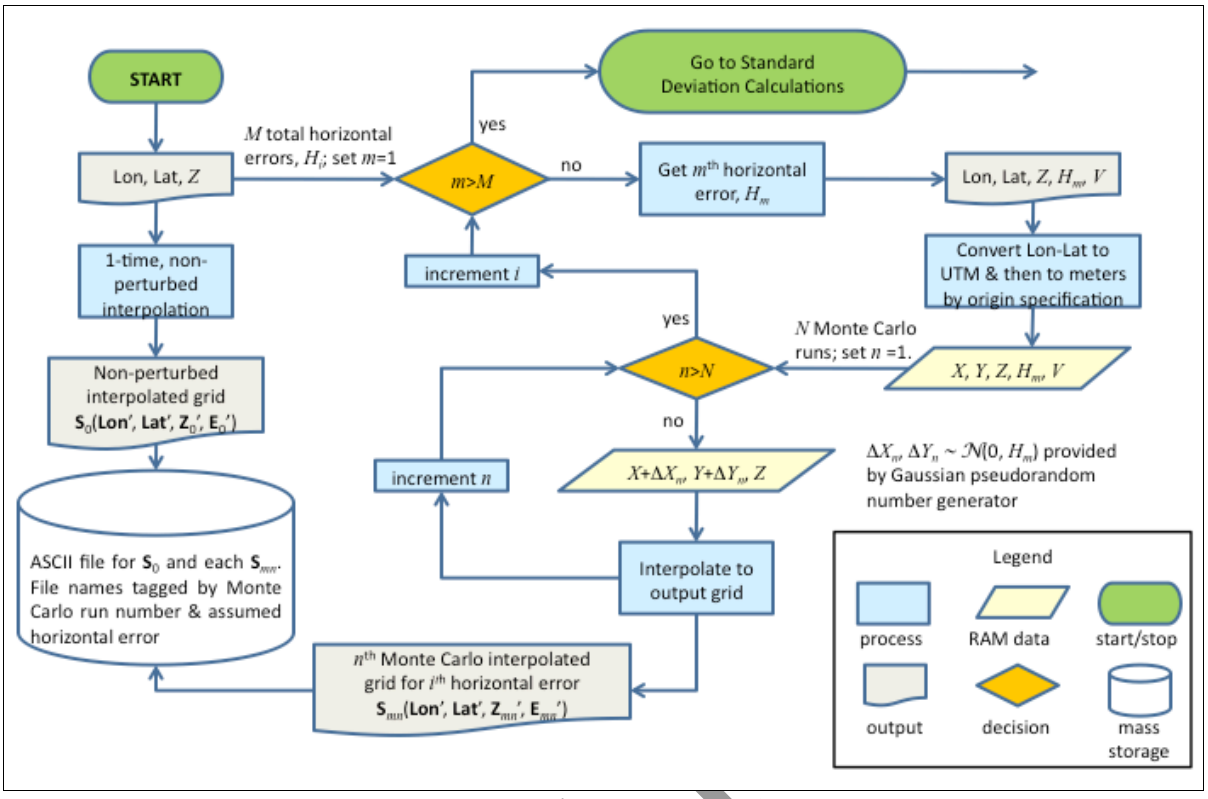


Figure 9.2.4.4 Low level flow chart of the Monte Carlo simulations block for the Bayesian network training process. Here, *X* and *Y* are the east and north positions of the output grid points in meters after conversion to Universal Transverse Mercator (UTM) projection and specification of a UTM origin in the grid, Z is interpolated depth, S_o is an unperturbed interpolation, S_{mn} is the perturbed grid for the nth Monte Carlo loop and m^{th} Horizontal Error Category, E' is error from smoothing. The bold face means that the quantity is a vector.

Figure 9.2.4.5 shows a low level flow chart for the next block of the training process, the standard deviation and slope calculations block. After completion of the final iteration, the uncertainty estimate for the set of *I* output points are computed from the gridded standard deviation of the Monte Carlo bathymetry surfaces and stored to ASCII files for use below (low level flow chart shown in Fig. 9.2.4.6). Specifically, the ASCII files contain longitude, latitude, bathymetry, Monte Carlo standard deviation and magnitude of the two-dimensional slope. To obtain the slopes, an unperturbed calculation of the bathymetry surface is first made, followed by computation of the bottom gradient throughout the grid. One can use finite differences as discussed in Zhou and Liu (2004) and Oksanen and Sarjakoski (2005) to calculate bottom gradient. Longitude and latitude positions are translated to UTM coordinates so that positional differences and gradients can be computed in meters. Linear extrapolation is used to aid computation of the gradient along the computational edges.

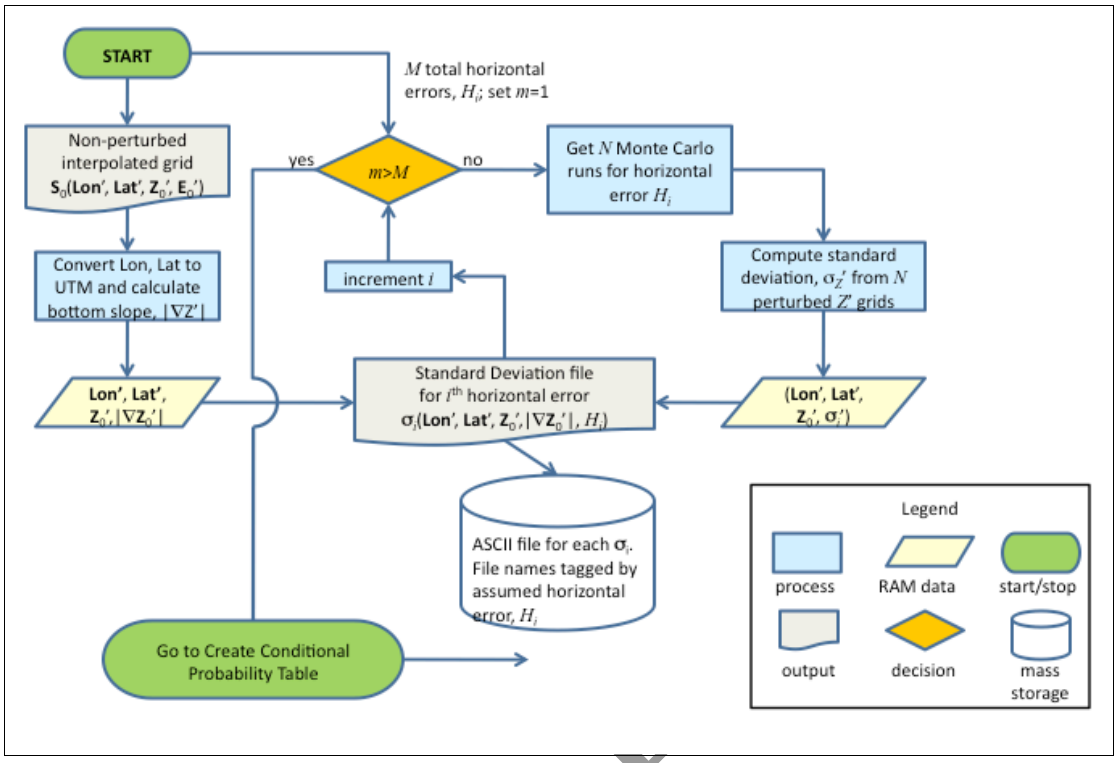


Figure 9.2.4.5 Low level flow chart of the Standard Deviations block for the Bayes net training process.

Figure 9.2.4.6 shows a low level flow chart for the final block of the training process, the Bayes network programming block. The BN is programmed using the graphic user interface (GUI) in Netica on a Windows based PC. The ASCII files containing the Monte Carlo uncertainties and bathymetric slopes are first read from the standard deviation file discussed in the above section for the first navigation uncertainty category. Then, using the logarithmic binning scheme discussed above we compute the bivariate histogram (i.e., three-dimensional histogram) with slopes and bathymetry uncertainty, respectively, indexing the row and column bins of the bivariate histogram. This histogram is normalized and written to an ASCII file. This process is then repeated for subsequent navigation uncertainties with the resultant normalized bivariate histograms appended to the same output file.

DRAFT

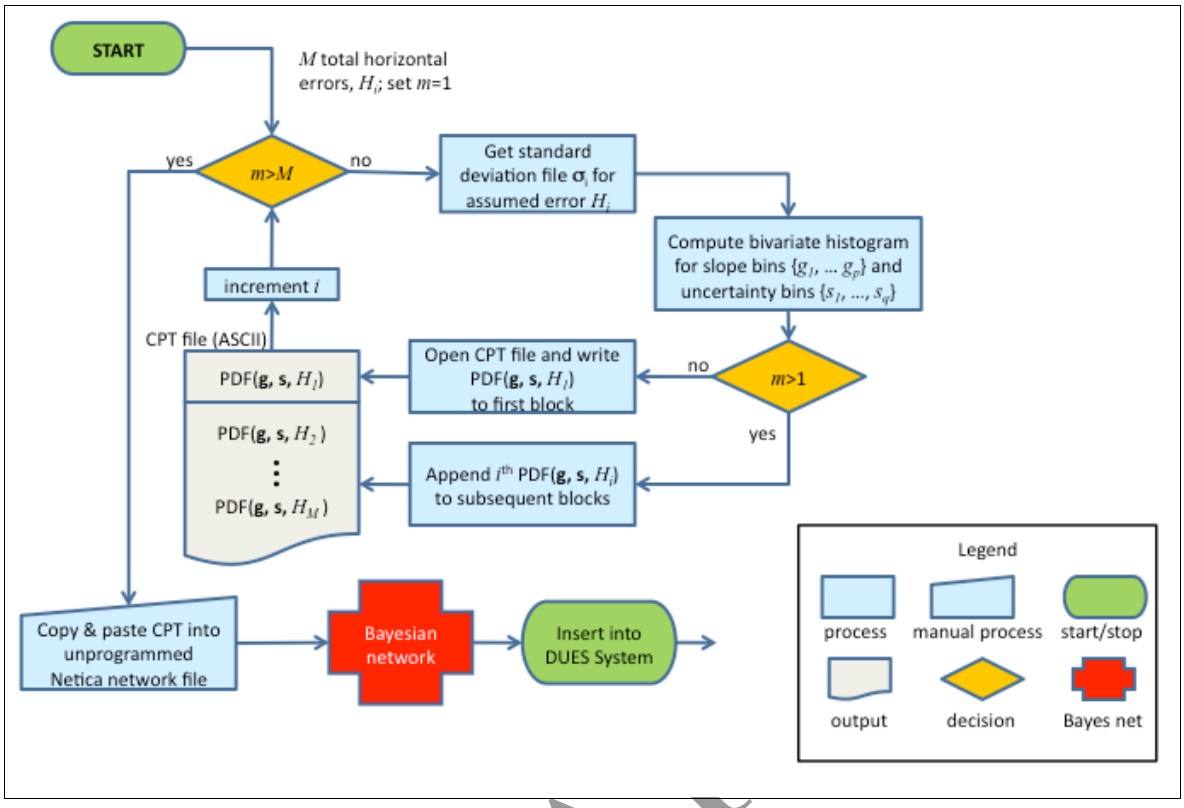


Figure 9.2.4.6 Low level flow chart of the Bayes network programming block for the Bayesian network training process. The network is stored in a library file that is accessed by API.

DRAFT

After completing this process, the Bayesian network *.neta file is opened in the Netica GUI, and the CPT for the "Uncertainty Estimate" node is displayed. The CPT is programmed by simple "copy-and-paste" from the ASCII file into the spreadsheet-like data display for the CPT in the GUI. The CPT is already arranged to have the same stacked structure as the ASCII file.

Operational Flow

Figure 9.2.4.7 shows how the BN could be integrated into the overall operational system, with key components and data flow. The end user of the system first defines a region of interest for uncertainty estimation, followed by query of available navigation and bathymetry for the area. The navigation metadata would indicate survey era and platform used from the navigation data to obtain the types of navigation (for horizontal uncertainty) and sonar systems (for vertical uncertainty) used for the track lines in the area. The fractional lengths of track lines with like uncertainties could be calculated (i.e. sum of all track lengths with a like error divided by the total of all track lengths) and become the weights for the BN's Horizontal Error node. Also extracted is the bathymetry from the same area to obtain the bottom gradient. The system loops through each grid point, assigns the bottom gradient to the appropriate bin of the BN's Bottom Gradient node, and extracts the resultant histogram. Calculation of mean plus one standard deviation for this histogram provides the intermediate uncertainty. The final uncertainty estimate

is obtained by first adding the square of the intermediate uncertainty to the square of the vertical uncertainty estimate.

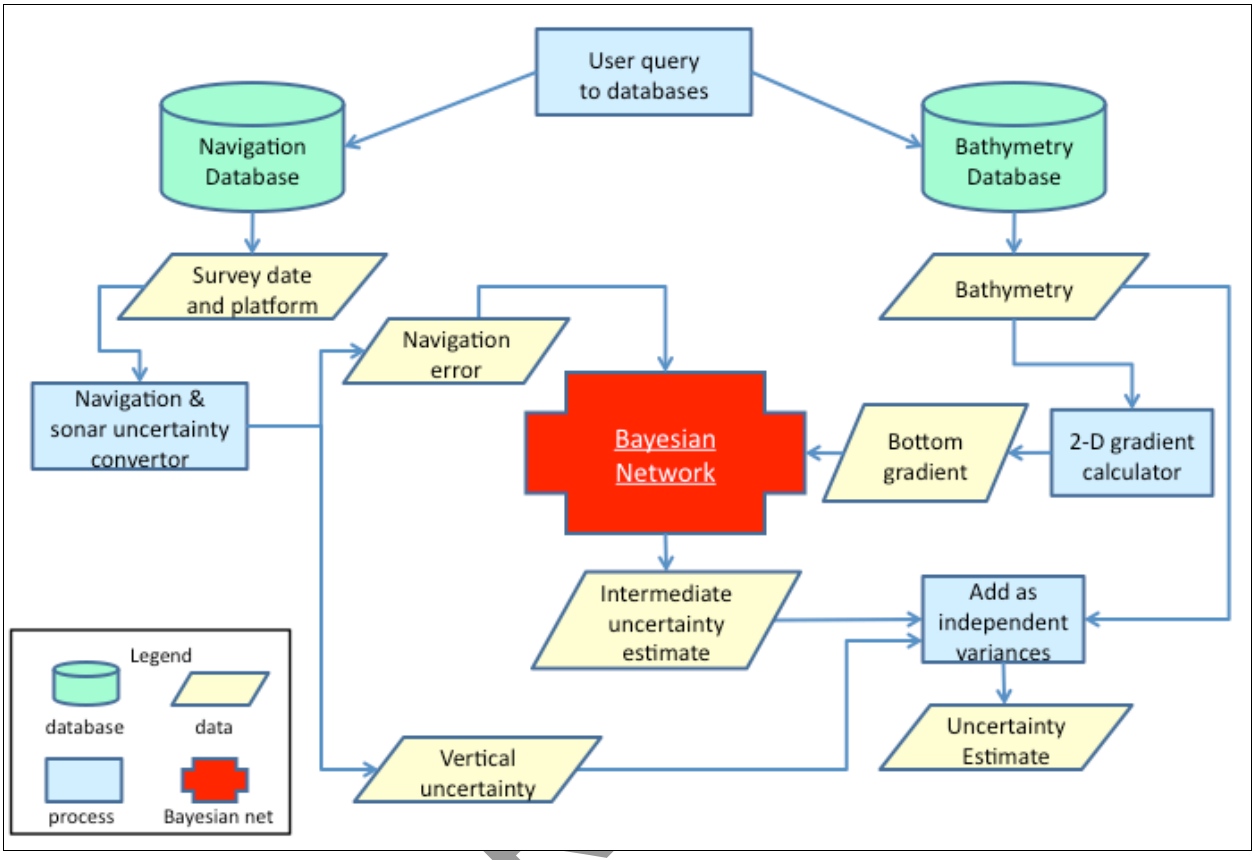


Figure 9.2.4.7 Integrated flow chart of the DBDB-V Uncertainty Expert System (DUES) after programming of the Bayes network component is complete.

9.2.5 Windowed Regression with Kriging Corrections

Contributed by Paul Elmore, Naval Research Laboratory, Stennis Space Center

As an alternative to using GMT surface, localized regression techniques of Cleveland (1979) have been applied to bathymetry problems in Plant et al. (2002). This technique was then refined by Calder (2006) to provide refinements to the interpolated surface and uncertainty estimate using ordinary kriging and the Monte Carlo technique discussed in Section A above. An advantage of this technique is that it provides an uncertainty estimate with the interpolated surface. A disadvantage is that it requires dense data sounding to be meaningful. The discussion that follows below closely to the discussions in the papers by Cleveland (1979) and Calder (2006).

Linear Smoothing by Locally Weighted Regression (Loess)

Linear smoothing techniques construct the interpolated surface, $\hat{z}(\mathbf{s})$, $\mathbf{s} \equiv$ matrix of (x, y) coordinates, from a linear weighted average of known data values. Unlike the splines-in-tension technique that finds a global solution to the available data, this technique obtains an interpolated value at a grid point by using only a subset of neighboring points. Mathematically,

$$\hat{z}(\mathbf{s}) = \sum_i \alpha(\mathbf{s} - \mathbf{s}_i) z(\mathbf{s}_i), \tag{(a)}$$

where $\alpha(\mathbf{s} - \mathbf{s}_i)$ specifies the smoother coefficients and the index i corresponds to the subset of local points, $z(\mathbf{s}_i)$, to be used for the interpolated value at $\mathbf{s}$, $\hat{z}(\mathbf{s})$. A common technique used to specify these coefficients is locally weighted regression (or "loess"), first published in Cleveland (1979) and further developed in Cleveland and Devlin (1988). A textbook by Givens and Hoeting (2005) also discusses the technique. The methodology determines the smoother coefficients from a weighted least squares polynomial (linear or quadratic) fit of windowed data.

1. Methodology

To summarize Cleveland's methodology, the two-dimensional case is considered first. These equations to the three-dimensional case in Section B, but the methodology will remain the same. Let data points x_i and y_i be related as

$$y_i = g(x_i) + \varepsilon_i, \tag{(b)}$$

where $g(x)$ is a smooth function and ε_i is Gaussian noise with zero mean and variance σ^2. Define $\hat{y}_i$ to be the estimate of $g(x_i)$ (i.e. $\hat{y}_i \approx g(x_i)$). Let

- the number n be the predetermined number of data points to be used for estimating $\hat{y}_i$
- the set x_k, $k = 1,\ldots,n$, be the subset of x_j's ($j = 1,\ldots,N$; N = total number of data points, $n < N$) that are closest to x_i
- the distance h_i be the distance from x_i to the furthest x_k.
- the windowing weights to be used for the regression, $w_k(x_i) = W([x_k - x_i]/h_i)$, where $W(x)$ is the tricube function, defined as

$$W(x) \equiv \begin{cases} \left(1 - |x|^3\right)^3, & x < 1 \\ 0 & , \text{ otherwise} \end{cases} \tag{(c)}$$

With these definitions, the loess procedure calculates the set of polynomial coefficients, $\hat{\beta}_l(x_i)$, that are the values for the β_l's that minimize

$$q(x_i) = \sum_{k=1}^{n} w_k(x_i)\left(y_k - \beta_0 - \beta_1 x_k - \cdots - \beta_d x_k^d\right)^2, \tag{(d)}$$

where $q(x_i)$ is the error function. The interpolated value for $g(x_i)$ is then

DRAFT

$$\hat{y}_i = \sum_{\ell=0}^{d} \hat{\beta}_\ell(x_i) x_i^d = \sum_{k=1}^{n} \alpha_k(x_i) y_k \,. \qquad \text{((e)}$$

Some points of interest are as follows (Cleveland 1979).

- The coefficients obtained have a decreasing trend to the edge of the window so that the data centered in the window generally have the greatest influence.
- The size of the polynomial typically used in Eqn. (((d) is d = 1 or 2. For d = 0, the result is a simple moving average. The d = 1 case is called linear loess or "lowess" smoothing, and d= 2 is quadratic "loess" smoothing.
- Cubic and higher fits typically are not used as the fits can become over fitted and numerically unstable.
- This technique also has a robustness option, so that the interpolation can be shielded from the effects of outliers in the data.
- Cleveland choose the tricube weighting function because it allowed the estimate of the error variance to be approximated by a chi-square distribution and usually lowered the variance of the estimate surface as the number of points used for the estimate increased.

Errors propagated into the interpolation by the technique are straight forward to compute. Under the assumption the data follow Eqn. (((b), the estimate of the variance for $\hat{y}_i$, $\hat{\sigma}_i^2$, is (Plant et al. 2002)

$$\hat{\sigma}_i^2 = \sigma^2 \sum_{k=1}^{n} [\alpha_k(x_i)]^2 \,, \qquad \text{((f)}$$

which is derivable from independent error propagation.

2. Convolution Approach

Figure 9.2.5.1 plots the set of α_k's as determined for a centered impulse response using the linear and quadratic loess interpolators. These coefficients are the smoother weights in Eqn. (((e) and are also called the "equivalent kernel". They depend only on the grid points, and, due to their finite width, act as a window or low-pass filter function on the spatial data (i.e. the smoothing weights and the data undergo convolution). Hence, one could bypass solving a weighted least square problem and simply compute the convolution of the smoothing window with the data, which should be computationally faster. In addition, other windowing functions could be used to perform the smoothing.

DRAFT

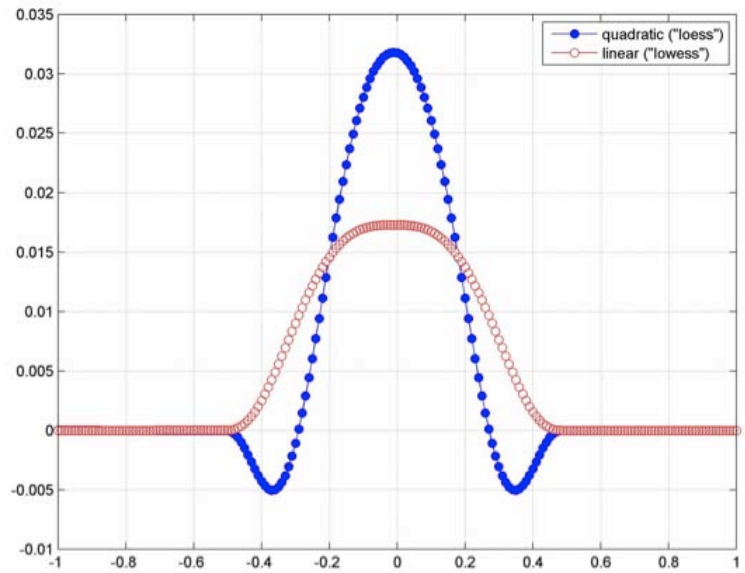


Figure 9.2.5.1 Equivalent kernel weights for linear and quadratic loess windows.

The use of other weighting functions would also allow one to interpolate over data points with differing accuracy. In this case, Eqns. (((e) and (((f) become

$$\hat{y}_i = \sum_{k=1}^{n} \alpha_k(x_i)\left(\sigma_k^{-2} \Big/ \sum_{k=1}^{n} \sigma_k^{-2}\right) y_k = \sum_{k=1}^{n} \tilde{\alpha}_k(x_i) y_k \,, \qquad \text{((g)}$$

$$\hat{\sigma}_i^2 = \sum_{k=1}^{n} \sigma_k^2 \tilde{\alpha}_k^2(x_i) \qquad \text{((h)}$$

where σ_k^2 is the variance of y_k and $\tilde{\alpha}_k(x_i) \equiv \alpha_k(x_i)\left(\sigma_k^{-2} \Big/ \sum_{k=1}^{n} \sigma_k^{-2}\right)$.

Combining windowing with ordinary kriging corrections

Since the above regression technique smooths the data, Calder (2006) provides an ordinary kriging step for correcting at the sounding points with the Monte Carlo technique to accounts for navigation uncertainty. The summary of the complete methodology is as follows.

1. Following Plant et al. (2002), interpolate the data with the quadratic loess interpolation technique of Cleveland (1979) to provide a trend surface for the bathymetry and uncertainty layer.
2. Restore finer details smoothed by the interpolation from ordinary kriging of the residuals; add the errors associated with ordinary kriging to the uncertainty layer from Step 1

DRAFT

assuming statistical independence (i.e. the variances add). The surface generated from kriging the residuals is the residual surface.

3. Estimate additional uncertainty caused by positional errors from the Monte Carlo technique of Jakobsson et al. (2002), but repeat steps 1 and 2 above instead of using the splines-in-tension algorithm at each iteration. As before, add the estimated error to the uncertainty layer by assuming statistical independence.

In equation form, the final bathymetry surface, $Z(\mathbf{s})$, is the sum of the trend surface, $\mu(\mathbf{s})$, and the residual surface, $R(\mathbf{s})$.

$$Z(\mathbf{s}) = \mu(\mathbf{s}) + R(\mathbf{s}) \qquad \text{((i)}$$

The uncertainty layer, $\sigma_Z(\mathbf{s})$, is

$$\sigma_Z(\mathbf{s}) = \left[\sigma_\mu^2(\mathbf{s}) + \sigma_R^2(\mathbf{s}) + \varepsilon^2(\mathbf{s})\right]^{1/2} \ , \qquad \text{((j)}$$

where $\sigma_\mu(\mathbf{s})$ and $\sigma_R(\mathbf{s})$ are the uncertainties for the trend and residual layers, respectively, and $\varepsilon(\mathbf{s})$ is the uncertainty layer associated with positional errors.

1. Quadratic Loess Interpolation for Trend Surface

Extending Eqn. (((c) through (((f) to two dimensions, the equations for calculating the trend surface, $\mu(\mathbf{s})$ in Eqn. (((i) are as follows.

$$\mu(\mathbf{s}) = \mathbf{p}^T(\mathbf{s})\hat{\boldsymbol{\beta}}(\mathbf{s}), \qquad \text{((k)}$$

where $\mathbf{p}^T(\mathbf{s}) \equiv [x^2, y^2, xy, x, y, 1]$ and vector $\hat{\boldsymbol{\beta}}(\mathbf{s}) \equiv [\hat{\beta}_5(\mathbf{s}), \ldots, \hat{\beta}_0(\mathbf{s})]^T$ being the set of $\beta_n(\mathbf{s})$'s that minimize the weighted least squares

$$q(\mathbf{s}) = \sum_{k=1}^{n} w_k(\mathbf{s})\left(z_k - \beta_0(\mathbf{s}) - \beta_1(\mathbf{s})y_k - \beta_2(\mathbf{s})x_k - \beta_3(\mathbf{s})x_k y_k - \beta_4(\mathbf{s})y_k^2 - \beta_5(\mathbf{s})x_k^2\right)^2 \qquad \text{((l)}$$

using the two-dimensional tri-cube weighting function

$$w_k(\mathbf{s}) = \begin{cases} \left(1 - \left|\dfrac{\mathbf{s} - \mathbf{s}_k}{d_0}\right|^3\right)^3, & \left|\dfrac{\mathbf{s} - \mathbf{s}_k}{d_0}\right| < 1 \\ 0 & , \text{ otherwise} \end{cases}, \qquad \text{((m)}$$

and the user provided value for d_0, which Calder's paper suggests that d_0 be ten times the largest sample spacing. Information lost by oversmoothing is regained when the residuals are calculated in the kriging portion.

DRAFT

Estimates for the variances follow Eqn. (((f) for like variances in all the data or $\hat{\sigma}^2(\mathbf{s}) = \sum_{k=1}^{n}\left[\alpha_k(\mathbf{s})\sigma_k\right]^2$ for differing variances. The $\alpha_k(\mathbf{s})$'s now have azimuthal symmetry and the same radial dependence as the 1-D case.

In addition, the estimate is made more robust (i.e. eliminate outliers) by flagging estimates greater than three Mahalanobis units as "no data". The Mahalanobis distance (Mahalanobis 1936), $\mathbf{M}(\mathbf{p},\mathbf{m},\mathbf{C})$, is

$$\mathbf{M}(\boldsymbol{\mu},\mathbf{m},\mathbf{C}) \equiv (\boldsymbol{\mu}-\mathbf{m})^T\,\mathbf{C}^{-1}(\boldsymbol{\mu}-\mathbf{m}) \qquad \text{((n)}$$

where $\boldsymbol{\mu}(\mathbf{s})$ is the vector of all estimated depths and $\mathbf{m}(\mathbf{s})$ and $\mathbf{C}(\mathbf{s})$ are the corresponding mean depth measurements and covariances for the windowed (c.f. Eqn. (((m)) data sets. To intuitively explain Eqn. (((n), suppose that the data points are all independent and have mean m and variance σ^2. The $\mathbf{C}$ matrix is all zeros except for the diagonal elements, which are all σ^2. Then, if an estimate differs from its corresponding windowed mean by more than three variances (i.e. $(\mu-m)^2/\sigma^2 > 3$), that estimate receives the "no data" mark. Eqn. (((n) generalizes this simpler scenario to account for covariance between the data.

2. Kriging the Residuals

Since loess interpolation is not exact at the data points, residuals, $R(\mathbf{s}_i)$, exists between actual measurements, $z(\mathbf{s}_i)$, and the corresponding estimated depths along trend surface, $\mu(\mathbf{s}_i)$, such that $R(\mathbf{s}_i) = z(\mathbf{s}_i) - \mu(\mathbf{s}_i)$. The residual surface in Eqn. (((i) is found from interpolation of the $R(\mathbf{s}_i)$set using ordinary kriging, as an overall but constant unknown bias may exist in the residuals. The variogram will likely have directional anisotropy, so that $2\gamma(\mathbf{s}_i,\mathbf{s}_j) = 2\gamma(h,\theta)$, where h is the separation distance between the two points and θ is the heading angle (clockwise from the north).

To account for this anisotropy, the following semivariance, $\gamma_D(\mathbf{s}_i,\mathbf{s}_j)$, is used and constructed (unconditionally valid for two-dimensions (Brandt 1998)) in the following manner:

1) Obtain an empirical estimate of the variogram
2) Compute the average azimuthal variogram from the empirical estimate to detect the directions of minimum and maximum variation,
3) Fit the variograms along these two axes to the spherical model for variograms
4) Using parameters from the fits to the spherical model and the direction for minimum variance, construct $\gamma_D(\mathbf{s}_i,\mathbf{s}_j)$ from Eqns. (((t) – (((w) below. These steps are now discussed in more detail.

DRAFT

Step 1: Following page 69 in Cressie (1993), the variogram for the residuals is approximated using the methods-of-moments (or classical) estimator in blocks sizes = $2d_0$ (c.f. Eqn. (((m)), so that

$$2\gamma(h,\theta) \approx 2\hat{\gamma}(h_i,\theta_i) = \frac{1}{|N(h_i)||N(\theta_j)|} \sum_{(a,b)\in N(h_i)\cap N(\theta_j)} \left(R(\mathbf{s}_a) - R(\mathbf{s}_b)\right)^2 . \qquad \text{((o)}$$

The sets $N(h_i)$ and $N(\theta_i)$ contain binned separation distances and heading angles as defined by the equations

$$N(h_i) \equiv \left\{(a,b): h_i - \frac{\Delta h_i}{2} \le d(\mathbf{s}_a,\mathbf{s}_b) \le h_i + \frac{\Delta h_i}{2}\right\} \qquad \text{((p)}$$

$$N(\theta_j) \equiv \left\{(a,b): \theta_j - \frac{\Delta\theta}{2} \le \angle(\mathbf{s}_a,\mathbf{s}_b) \le \theta_j + \frac{\Delta\theta}{2}\right\}, \qquad \text{((q)}$$

where $d(\mathbf{s}_a,\mathbf{s}_b)$ and $\angle(\mathbf{s}_a,\mathbf{s}_b)$ are the Euclidian distance and heading angle between $\mathbf{s}_a$ and $\mathbf{s}_b$ and $|N(h_i)|$ and $|N(\theta_i)|$ are the number of bins in these sets. For the data set analyzed in Calder (2006), $\Delta h_i = 0.05d_0$, $\Delta\theta_i = 45^\circ$, $h_i = i\Delta h + \Delta h/2$ and $\theta_j = j\Delta\theta$.

Step 2: The average azimuthal variogram, $2\bar{\gamma}(\theta_j) = |N(h_i)|^{-1}\sum_i 2\hat{\gamma}(h_i,\theta_j)$ is then calculated. In Calder (2006), $2\bar{\gamma}(\theta_j)$ had two sets of maximum and minima (caused by trending ridges), so it could be modeled by the function

$$2\bar{\gamma}(\theta_j) \approx 0.5g_0 + g_2\cos(4\pi\theta_j + \phi_2), \qquad \text{((r)}$$

which is the second Fourier eigenfunction, Equation (((r) is fitted to the data by evaluating the second discrete Fourier transform coefficient. The phase constant, ϕ_2, is the radian angle (which goes counterclockwise from the east) where the first minimum is found. It is changed to a heading angle (again, clockwise from the north), θ_m, from the transformation $\theta_m \equiv -\phi_2/2 + \pi/2$. This angle is obtained for each $2d_0$ block, then interpolated to get $\theta_m(\mathbf{s}_i)$ for Step 4 below.

Step 3: Two finer-scaled directional variograms are then calculated; one in the θ_m direction, the other in the perpendicular direction, $\theta_m^\perp$; using the methods-of-moments estimator with angle bins of $\pm\pi/2$ about each direction. In this case, $\Delta h = 0.02d_0$ and the data were truncated to those within 95% of the mean to reduce outliers. These empirical variograms are fitted to the standard spherical variogram model (Cressie (1993), Eqn. (2.3.8); Davis (2002), Eqn. (4.98))

DRAFT

$$2\gamma(h) = \begin{cases} 0, & h = 0 \\ a_0 + a_1\left(1.5(h/a_2) - 0.5(h/a_2)^3\right), & 0 \le h/a_2 \le 1 \\ a_0 + a_1, & h/a_2 \ge 1 \end{cases} \qquad \text{((s)}$$

using the Levenberg-Marquardt algorithm (Marquardt 1963; Brandt 1999) for fitting (i.e. solving for a_0, a_1 and a_2). Let the modeled variogram along the θ_m direction be $2\gamma_0(h)$ and the fitting constants be a_0, a_1 and a_2. Similarly, let the modeled variogram along the $\theta_m^{\perp}$ direction be $2\gamma_0^{\perp}(h)$, and the fitting constants be $a_0^{\perp}$, $a_1^{\perp}$, and $a_2^{\perp}$. When solving for both sets of constants, the variograms are constrained to equal the same sill at large distances so that $a_0^{\perp} = a_0$ and $a_1^{\perp} = a_1$, but $a_2^{\perp} \neq a_2$ in general.

Step 4: Define the anisotropy parameter, $\alpha_{aniso} \equiv a_2 / a_2^{\perp}$. Then, $\gamma_D(\mathbf{s}_i, \mathbf{s}_j)$, is evaluated in the following manner.

$$\gamma_D(\mathbf{s}_i, \mathbf{s}_j) = \gamma_0\left(d'(\mathbf{s}_i, \mathbf{s}_j)\right), \qquad \text{((t)}$$

where

$$d'(\mathbf{s}_i, \mathbf{s}_j) \equiv \left\| \mathbf{A}(\theta_m(\mathbf{s}_i), \alpha_{aniso})(\mathbf{s}_i - \mathbf{s}_j) \right\|, \qquad \text{((u)}$$

$$\mathbf{A}(\theta_m(\mathbf{s}_i), \alpha_{aniso}) = \mathbf{R}(-\theta_m(\mathbf{s}_i))\, diag(1, \alpha_{aniso})\, \mathbf{R}(\theta_m(\mathbf{s}_i)), \qquad \text{((v)}$$

and $\mathbf{R}(\theta_m(\mathbf{s}_i))$ is the standard two-dimensional rotational matrix (Goldstein 1980)

$$\mathbf{R}(\theta_m(\mathbf{s}_i)) = \begin{bmatrix} \cos(\theta_m(\mathbf{s}_i)) & \sin(\theta_m(\mathbf{s}_i)) \\ -\sin(\theta_m(\mathbf{s}_i)) & \cos(\theta_m(\mathbf{s}_i)) \end{bmatrix}. \qquad \text{((w)}$$

For applications where only the bathymetry is needed, $\gamma_D(\mathbf{s}_i, \mathbf{s}_j)$, may be sufficient for use. In hydrographic situations, however, where extra caution is required for navigation safety, an additional "hydrographic uncertainty" variogram is added to increase the total uncertainty for safety (see Calder (2006) for details). This variogram is defined to be

$$2\gamma_H(h) \equiv b_0 + b_1 h \qquad \text{((x)}$$

The choice of the coefficients is arbitrary, but Calder uses $b_0 \equiv 0$ and $b_1 \equiv 6.51 \times 10^{-4}$ m so that the 95% confidence interval of the uncertainty increases by 0.05 m for every meter of horizontal separation in soundings. Thus, for hydrography, the final variogram to use is

$$2\gamma_{total}(\mathbf{s}_i, \mathbf{s}_j) = 2\gamma_D(\mathbf{s}_i, \mathbf{s}_j) + 2\gamma_H\left(\left\|\mathbf{s}_j - \mathbf{s}_i\right\|\right) \qquad \text{((y)}$$

DRAFT

3. Monte Carlo Estimation of Depth Error Due to Positional Stability Errors

This last part is of use for fusion with soundings data, either with other soundings data or a historical grid, as the position of the soundings contains errors (this part is not applied when fusing historical gridded data sets). As before (c.f. Section I.A), positioning errors result in errors of depth estimates on the interpolation surface as the position of the soundings affect the interpolated solution; the Monte Carlo technique of Jakobbsson et al. (2002) is used to estimate this error. At the beginning of each iteration, the locations of the soundings are perturbed according to a probability density function appropriate for the sounding, often assumed to be Gaussian unless otherwise known (c.f. Fig. 5 in Calder (2006) for a non-Gaussian example). Loess interpolation and kriging, using the variogram from the unperturbed set, are repeated at each iteration. The standard deviation of the solutions at each interpolation point provides $\varepsilon(\mathbf{s})$ in Eqn. (((j).

Acknowledgements

This work is funded through the Office of Naval Research, Naval Research Laboratory base program. The mention of commercial products or the use of company names does not in any way imply endorsement by the U.S. Navy.

DRAFT

References

(2008). Netica, Version 4.0.8. Vancouver, BC, Norsys Software Corporation.

Brandt, S. (1998). Data analysis : statistical and computational methods for scientists and engineers. New York, Springer.

Brandt, S. (1999). Data Analysis: Statistical and Computational Methods for Scientists and Engineers New York, Springer.

Calder, B. (2006). "On the uncertainty of archive hydrographic data sets." IEEE Journal of Oceanic Engineering **31**(2): 249-265.

Cleveland, W. S. (1979). "Robust locally weighted regression and smoothing scatterplots." Journal of the American Statistical Association **74**(368): 829-836.

Cleveland, W. S. and S. J. Devlin (1988). "Locally weighted regression - an approach to regression-analysis by local fitting
" Journal of the American Statistical Association **83**(403): 596-610.

Cressie, N. A. C. (1993). Statistics for Spatial Data. New York, Wiley.

Davis, J. C. (2002). Statistics and Data Analysis in Geology. New York, Wiley.

Elmore, P. A., D. H. Fabre, et al. (2009). Uncertainty estimation of historical bathymetric data from Bayesian networks. OCEANS 2009, MTS/IEEE Biloxi - Marine Technology for Our Future: Global and Local Challenges.

Givens, G. H. and J. A. Hoeting (2005). Computational Statistics. Hoboken, NJ, Wiley-Interscience.

Goldstein, H. (1980). Classical Mechanics. Reading, MA, Addison Wesley.

Jakobsson, M., B. Calder, et al. (2002). "On the effect of random errors in gridded bathymetric compilations." Journal of Geophysical Research-Solid Earth **107**(B12): Article 2358.
Mahalanobis, P. C. (1936). "On the generalised distance in statistics." Proceedings of the National Institute of Science of India **12**: 49-55.
Marquardt, D. W. (1963). "An algorithm for least-squares estimation of nonlinear parameters." Journal of the Society for Industrial and Applied Mathematics **11**(2): 431-441.
Marsaglia, G. and W. W. Tsang (2000). "The Ziggurat Method for Generating Random Variables." Journal of Statistical Software **5**.
Oksanen, J. and T. Sarjakoski (2005). "Error propagation of DEM-based surface derivatives." Computers & Geosciences **31**(8): 1015-1027.
Plant, N. G., K. T. Holland, et al. (2002). "Analysis of the scale of errors in nearshore bathymetric data." Marine Geology **191**(1-2): 71-86.
Russell, S. J. and P. Norvig (2003). Artificial Intelligence: A Modern Approach. Upper Saddle River, N.J., Prentice Hall/Pearson Education.
Zhou, Q. M. and X. J. Liu (2004). "Error analysis on grid-based slope and aspect algorithms." Photogrammetric Engineering and Remote Sensing **70**(8): 957-962.

DRAFT

9.2.6 Error Growth Model

The following has been extracted from 5th Edition, S-44 (2008) Annex A, section A.5 by Rob Hare and edited for context

TPU is a combination of random and bias based uncertainties. Random and short period uncertainties have to be recognized and evaluated both in horizontal and vertical directions.

The propagated uncertainty may be expressed as a variance (in metres^2) but is more often reported as an *uncertainty* (in metres) derived from variance with the assumption that the uncertainty follows a known distribution. In the latter case, the confidence level (e.g., 95%) and the assumed distribution shall be documented. Horizontal uncertainties are generally expressed as a single value at a 95% level, implying an isotropic (circular) distribution of uncertainty on the horizontal plane.

In the hydrographic survey process it is necessary to model certain long period or constant factors related to the physical environment (e.g. tides, sound speed, vessel dynamic draft, including squat of the survey vessel). Inadequate models may lead to bias type uncertainties in the survey results. These uncertainties shall be evaluated separately from random type uncertainties.

TPU is the resultant of these two main uncertainties. The conservative way of calculating the result is the arithmetic sum, although users should be aware that this may significantly overestimate the total uncertainty. Most practitioners, and the appropriate ISO standard, recommend quadratic summation (i.e., summation of suitably scaled variances).

An example is given in S-44 for fixed (a) and depth-variable (b) uncertainty components:

$$\text{TVU} = \pm\sqrt{a^2 + (b \times d)^2}$$

9.2.7 Split-Sample Approach

Contributed by Chris Amante and Matt Love, NOAA National Geophysical Data Center (NGDC)

There are a number of techniques used to quantify the uncertainty of interpolated elevations, e.g., split-sample, cross-validation, jack-knifing, and boot-strapping. Using a split-sample approach, a percentage of the data is omitted, an interpolation method is applied, and the differences between the interpolated elevations and the original omitted elevations are calculated. This method is often used to assess the stability of various interpolation methods by omitting increasingly greater percentages of the original data and analyzing changes in the uncertainty.

A program for Linux users written in Guile Scheme (http://www.gnu.org/s/guile/) has been developed to quantify the uncertainty of interpolation methods using a split-sample approach. The program, named ss_unc.sch, utilizes the Generic Mapping Tools (GMT; http://www.soest.hawaii.edu/gmt/), and ancillary programs written in the C programming language by Matthew Love of the National Oceanic and Atmospheric Administration (NOAA) National Geophysical Data Center (NGDC; see glossary).

The program omits a percentage of the xyz points, applies an interpolation method, and calculates the differences between the interpolated values and the omitted elevations. In order to quantify the uncertainty of the interpolation method at every data point, the program repeats this process and aggregates the differences between the original xyz file and the interpolated

DRAFT

elevations. The program produces several useful products to be used in assessing the uncertainty of various interpolation methods. The products include a histogram (see Figure 9.2.7.1) of the differences with statistical measurements such as the minimum, maximum, mean, root mean squared error (RMSE), and standard deviation. In addition, the program produces a binary .grd file of the differences between the original and interpolated elevations.

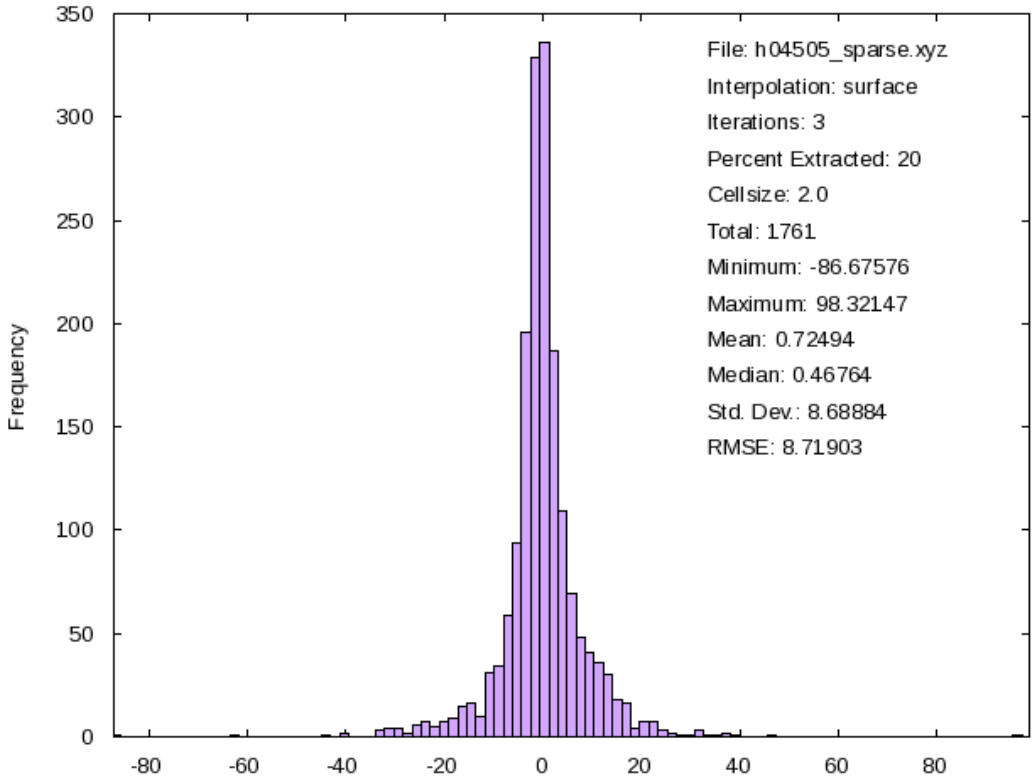


Figure 9.2.7.1 Example histogram output from Split-Sample program SS_UNC.

The program has multiple user-defined parameters. The user defines the input xyz file to be used to evaluate multiple gridding methods, the gridding methods to evaluate, the percentage of points to be omitted, the grid cell size, the input xyz file delimiter, and the number of iterations to repeat the process (see example below). Currently the program can evaluate GMT

‘surface,’ ‘nearneighbor,’ and ‘triangulate’ gridding algorithms. Additional gridding algorithms, such as MB-System ‘mbgrid,’ will be included in the program in the near future.

ss_unc.scm [infile] [-e arg] [-percent arg] [-grd_cells arg] [-delimiter arg] [-dotimes arg] [-engines? arg] [-help]

infile	The input xyz file
-e	Specify which gridding engine(s) to use. If multiple engines are desired, separate them with commas.
-percent	Percentage of xyz data to extract, default is 20.
-grd_cells	The grid cell size in arc-seconds.
-delimiter	The input xyz delimiter
-dotimes	The number of iterations
-engines?	Check available engines
-help	Display the help

Split-sample uncertainty program example:

~] ss_unc.scm h04505_sparse.xyz -e surface, nearneighbor, triangulate -percent 25 -grd_cells 100 -delimiter "," -dotimes 10 > ss_unc_example.sh

~] chmod +x ss_unc_example.sh

~] ./ss_unc_example.sh

Another option currently in progress is to evaluate the stability of the interpolation method by omitting an increasingly greater percentage of original data. For example, omit 20% of the original data and increase the percentage of omitted data at 10% intervals until 80% of the original data has been omitted. This option will produce a line graph depicting the change in RMSE as a function of the percentage of points omitted for various interpolation methods.

DRAFT

SS_UNC

Compare interpolation gridding methods to help determine uncertainty. Manual

DRAFT

Matthew Love

DRAFT

Copyright © 2011 CIRES.

Permission is granted to make and distribute verbatim copies of this manual provided the copyright notice and this permission notice are preserved on all copies. Permission is granted to copy and distribute modified versions of this manual under the conditions for verbatim copying, provided that the entire resulting derived work is distributed under the terms of a permission notice identical to this one.

1 Copying

SS UNC is copyright (C) 2011 CIRES and is released under the terms of the GNU General Public License. See the included file 'COPYING' for the full text of the license (or see section "Copying" in The GNU Emacs Manual).

> This is free software – you are welcome to redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2, or (at your option) any later version.
>
> SS UNC is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

DRAFT

2 Introduction

The Split Sample UNCertainty program (SS UNC) is a Guile scheme command-line program designed to run on the GNU/Linux Operating System. SS UNC is highly configurable, allowing the user to add gridding engines, change and add command-line options, and even change the way SS UNC runs.

DRAFT

3 Installing SS UNC

3.1 Dependencies

For SS UNC to work, there are a number of required dependecies:

- GNU/Linux
- Guile
- GMT
- rpslit
- ncquery

3.2 Autotools

Installation through autotools is fairly straight forward.

1. Unpack the package
   ~] tar xfvz ss_unc-0.1.tar.gz
2. Enter the package root directory
   ~] cd ss_unc-0.1
3. Configure and Make the package
   The '--prefix' switch to ./configure specifies where to install files.
   ~/ss_unc-0.1] ./configure --prefix=~/ss_unc && make
4. Install the built package
   ~/ss_unc-0.1] make install

3.3 Manual Installation

1. Unpack the package
   ~] tar xfvz ss_unc-0.1.tar.gz
2. Enter the package root directory
   ~] cd ss_unc-0.1
3. Enter the package 'src directory
   ~] cd src
4. Copy ss_unc.scm to a directory locatable with PATH
   ~] cp ss_unc.scm ~/bin

DRAFT

4 Using SS UNC

SS UNC is a command-line program designed to be run in a shell program such as bash.

4.1 Basic Usage

For general usage help from the command-line, the '-help' swich can be used. The usage display will display the default command-line switches as well as any additional switches defined by the user, see See Section 5.2 [Adding Command Line Options], page 6. The '-e' switch specifies the gridding engine to use in gridding comparisons, see See Section 5.1 [Adding Engines], page 6.

```
~] ss_unc.scm -help
ss_unc.scm version 0.2.0
Usage: ss_unc.scm [infile] [-e arg] [-percent arg] [-grd_cells arg]
                  [-delimiter arg] [-dotimes arg] [-engines? arg] [-help]

infile          The input xyz file.
-delimiter      The input xyz record delimiter, default is space (‘‘ ‘‘)
-e              Specify which gridding engine(s) to use,
                if multiple engines are desired, separate them with commas.
-percent        Percentage of xyz data to extract, default
                is 20, separate values with a comma.
-grd_cells      The grid cell size -dotimes The number of iterations -engines?
                Check available engines, optionally specify
                which engine to get more information about..
-help           Display the help
```

infile

The infile is the input xyz file to perform the gridding algorithms on.

-e

The -e switch specifies which gridding engine(s) to use. To use multiple engines, separate the engine names with a comma, this will produce histograms and stat files for each of the engines mentioned.

-percent

The -percent switch specifies the percentage of random points to extract from infile. If the argument is a list of comma separated values, ss unc.scm will run for each of the specified percentages, and append the statistics for each into a separate file.

-grd cells

The -grd_cells switch specifies the cell-size to use for gridding operations.

-dotimes

The -dotimes switch allows the user to specify how many iterations of gridding and comparing will be performed.

-engines?

The -engines? switch without arguments will display a list of the available gridding engines, optionally specify an engine as an argument to get more details about that engine.

DRAFT

-help

This switch will display the help text and exit.

When ss unc.scm is run, it will immediately output a number of commands to stdin. If these commands are redirected to a file, e.g. > ss_unc_test.sh, then that file can be used to generate the output files. The generated script can be edited to fine-tune some of the commands if needed and can be saved for future use.

4.2 Default Engines

SS UNC comes with a number of gridding engines for general use. To check which engines are available on the running version of SS UNC, use the '-engines? swich, which will display all the engine keys which would be used with the '-e option, including those added by the user, see See Section 5.1 [Adding Engines], page 6.

Each of the search engines has a description, which can be displayed by specifying which engine to display more information about after the '-engines? swich:

~] ss_unc.scm -engines?

surface

nearneighbor

triangulate

~] ss_unc.scm -engines? surface

surface

GMT surface (SPLINE)

DRAFT

4.3 Output

SS UNC will generate a number of output files. The type and number of output files will depend on the given command-line arguments.

- Histogram(s)
- Difference Grid(s)
- Statistics file(s)

5 Extending SS UNC

SS UNC allows for complete extendibility. By use of hooks and definitions, one could rewrite SS UNC to ones own desires.

5.1 Adding Engines

SS UNC allows for the adding of new gridding engines. This is accomplished through the configuration file (~/.ss unc). The scheme function ’ci-add-engine is used to add a new gridding engine into SS UNC:

```
(si-add-engine ’(engine-name .
                    (’engine-function ‘‘engine-description’’)))
```

where engine-function is a function which should accept 2 arguments (input-xyz-file and gridding-cell-size) and engine-description is a string describing the engine. Adding an engine in this manner will add the engine to the output of the -engines? command-line switch, seemlessly integrating the added engines into the SS UNC interface.

5.2 Adding Command Line Options

SS UNC allows for the adding of new command-line options. There are two steps involved in adding command-line options, adding the option to the help menu and adding a handling function to the command-line hook.

```
(define switch-var #f)

    (sc-add-option ’(-switch . (takes-options?(#t or #f)
    "Description"))) (add-hook! ci-command-line-hook
    (lambda (a b)
            (if (equal? (car b) "-show_url")
  (begin
     (set! show-url #t)
     (parse-options a (cdr b))))))
```

DRAFT

5.3 Hooks

Hooks are a usefull tool in extending scheme code. SS UNC takes advantage of hooks to allow the user the ability to add command-line options (including in the -help output).

SS UNC provides one hook, listed and described below:

ci-command-line-hook

The ’sc-command-line-hook is the hook that allows adding to the default command-line options. The hook is run before the general command-line parsing begins, allowing the user to have control over the command-line environment. This hook is passed 2 options, the first is the command that ran the program (e.g. /usr/bin/sicl) and the second is the remaining command-line options. The hook will be run for each command-line option, so the user doesn’t have to write any loops in the configuration files to take advantage of adding command-line options in this manner.

6 Examples

Running ss unc.scm

Run ss unc.scm on the comma delimited ascii xyz file test.xyz, randomly extracting 25 percent of the datapoints, gridding using surface and triangulate at 10 arc-seconds.

```
#run the command
~]   ss_unc.scm test.xyz -percent 25 -e surface,triangulate \

-grd_cells 10 -delimiter ‘‘,’’ > test.sh

#Allow the output script to be executable
~]   chmod +x test.sh

#Run the output script
~] ./test.sh
```

Re-Define the ’surface engine function

Redefine the ’surface engine function in the SS UNC configuration file, adding a tension parameter.

```
(define (run-surface in cellsize)
   (display (string-append
                  "surface "
                  in
                  " -I"
                  cellsize
                  "c   -T.4 -G"
                  (basename in ".xyz") "_surface.grd"
                  " $(minmax " in " -I" cellsize "c)"))
   (newline))
```

DRAFT

SS_UNC.SCM


```
#!/usr/bin/guile \
−e main −s
!#

;;−−

(use−modules (ice−9 rdelim))

(define percent "20")
(define inxyz '())
(define grdcellsize "1")
(define do−times "2")
(define outps "out.ps")
(define delim " ")
(define version−wanted? #f)
(define help−wanted? #f)

;;−−−−−−−−−−;;
;; Engines ;;
;;−−−−−−−−−−;;
;;−−−
;; Define some hooks.

(define ci−command−line−hook (make−hook 2))

;; The User Config File, this can be set in the .guile config file.

                (if (not (defined? 'ci−config))
                        (define ci−config
                        (cond ((file−exists? (string−append (getenv "HOME") "/.ss_unc.conf"))
                        (string−append (getenv "HOME") "/.ss_unc.conf"))
                        ((file−exists? (string−append (getenv "HOME") "/.ss_unc"))
                        (string−append (getenv "HOME") "/.ss_unc")))))

;; The list of default engines. Formatted
;; thusly: '(name . ('function "description")
;; The 'function should accept 2 arguments,
;; the input xyz file and the gridding cellsize

(define ci−engine−alist

    '((surface . ('run−surface "GMT surface (SPLINE) \n"))
        (triangulate . ('run−triangulate "GMT triangulate (Triangulation) \n"))
        (nearneighbor . ('run−nn "GMT nearneighbor (NEAREST NEIGHBOR) \n"))))

(define (ci−add−engine specs)

    "Add a gridding engine to the 'ci−engine−alist in the format
'(name . 'function \"description\")"

    (append! ci−engine−alist (list specs)))

(define (ci−display−engines engine−alist)
```

DRAFT

```
  "Display the available engines in the given 'engine−alist."

   (if (pair? engine−alist)

        (begin
          (display (caar engine−alist))
          (newline)
          (ci−display−engines (cdr engine−alist)))))

(define (ci−describe−engine ci−eng engine−alist)

  "Describe 'ci−eng using the given 'engine−alist"

   (display ci−eng)
   (newline)
   (display "−−−−−−")
   (newline)
   (if (and (assq (string−>symbol ci−eng) engine−alist)

              (not (null? (cddr (assq (string−>symbol ci−eng) engine−alist)))))
        (display (caddr (assq (string−>symbol ci−eng) engine−alist)))
        (display "Sorry, this engine has no description available."))

   (newline))

;;−−−

;;−−−Engine functions:

(define (run−surface in cellsize)
   (display (string−append

              "surface "

              in

              " −I"

              cellsize

              "c −G"
              (basename in ".xyz") "_surface.grd"
              " $(minmax " in " −I" cellsize "c)"))

   (newline))

(define (run−nn in cellsize)
   (display (string−append

              "nearneighbor "

              in

              " −I"
```

DRAFT

```
                cellsize

                "c −S20c −N8 −V −G"
                (basename in ".xyz") "_nearneighbor.grd"
                " $(minmax " in " −I" cellsize "c)"))

    (newline))

(define (run−triangulate in cellsize)
    (display (string−append

                "triangulate "

                in

                " −I"

                cellsize

                "c −G"
                (basename in ".xyz") "_triangulate.grd"
                " $(minmax " in " −I" cellsize "c) > tmp.idk"))

    (newline))

;;−−−

;;−−−PROC:

(define (run−rsplit in percentage)
    (display (string−append

                "rsplit −−percent "

                percentage
                " −r " in " > rand_"
                (basename in)

                " 2>base_"

                (basename in)))
    (newline))

(define (run−ncquery ingrd inxyz delim d_format outxyz)
    (display (string−append

                "ncquery −r "

                d_format

                " −g "
```

DRAFT

```
                ingrd

                " −i "

                inxyz

                " −d "

                delim

                " | grep −v \"nan\" >> "

                outxyz))
    (newline))

(define (run−xyz2grd inxyz outgrd cellsize)
    (display (string−append

                "awk −F"

                delim

                " '{print $1,$2,$5}' | "
                "xyz2grd "

                inxyz

                " −V −I"

                cellsize

                "c −G"

            outgrd
            " $(minmax " inxyz " −I" cellsize "c)"))
             (newline))

;;−−−STATS:

    (define (run−pshistogram inxyz outps rperc ptimes engine)
        (display
       (string−append

      "pshistogram "

       inxyz
```

DRAFT

```
" −V −Gblue −W$(minmax −C "
 inxyz " | awk '{print ($10−$9)*.10}') −L1 −Z0 −T4
−B:\"Value\":/:\"Frequency\"::.:WSne −U/−.2
5i/−0.75i/\"ss_unc.scm\" −K > "
 outps " 2>" outps ".info"))
 (newline))

(define (run−pstext inxyz outps
    rperc ptimes engine)
     (display
    (string−append


   "pstext −R0/3/0/5 −JX3i −O −N −V << EOF >> "


    outps

   "\n.25 11.5 10 0 0 LT Iterations: " ptimes

   "\n.25 11 10 0 0 LT Percent: " rperc

   "\n.25 10.5 10 0 0 LT Cell−Size: " grdcellsize

   "\n1.5 11.5 10 0 0 LT Total: $(wc −l " inxyz " | awk '{print $1}')"

   "\n1.5 11 10 0 0 LT Min: $(grep min/max " outps ".info | awk '{print $6}')"

   "\n1.5 10.5 10 0 0 LT Max: $(grep min/max " outps ".info | awk '{print $7}')"

   "\n2.75 11.5 10 0 0 LT Mean: $(tail −1 "

    (basename inxyz ".xyzgd") ".stat | awk '{print $4}')"

   "\n2.75 11 10 0 0 LT RMSE: $(tail −1 "

    (basename inxyz ".xyzgd") ".stat | awk '{print $5}')"

   "\n4 11.5 10 0 0 LT File: " inxyz

   "\n4 11 10 0 0 LT Interpolation: " engine

   " \nEOF"))

(newline))

(define (set−gmt−defaults)

(display "gmtset COLOR_NAN 255/255/255 DOTS_PR_INCH 300 ANNOT_FONT_PRIMARY Times −Roman\
ANNOT_FONT_SIZE_PRIMARY 8 ANNOT_FONT_SIZE_SECONDARY 8 HEADER_FONT Times−Roman \
LABEL_FONT Times−Roman LABEL_FONT_SIZE 8 HEADER_FONT_SIZE 10 PAPER_MEDIA letter \
UNIX_TIME_POS 0i/−1i PS_COLOR CMYK Y_AXIS_TYPE ver_text PLOT_DEGREE_FORMAT ddd:mm:ss \
D_FORMAT %.1f COLOR_BACKGROUND 0/0/0 COLOR_FOREGROUND 255/255/255 COLOR_NAN 255/255/255")

(newline))

(define (run−stats inxyz outxyz rperc ptimes grdcellsize)

(display (string−append
            "awk −F" delim " '{sum+=$5} {sum2+=($5*$5)} END { print \""
            rperc "\",\"" ptimes "\",\"" grdcellsize "\",sum/NR,sqrt(sum2/NR)}' "
            inxyz " >> " outxyz))

(newline))

    (define (run−ps2raster psfile outf)
        (display (string−append
```

DRAFT

```
        "ps2raster −T" outf " " psfile))
        (newline))

;;−−−

;;−−−−−−−−−−−−−−−−;;
;; Run−Commands ;;
;;−−−−−−−−−−−−−−−−;;
;;−−−Run through the commands

    (define (run−commands
      run−times engines p)
       (let ((engine−list
      (string−split engines #\,)))
       (if (> run−times 0)
      (begin
      (let ((base−xyz
      (string−append "base_"
      (basename inxyz)))

            (rand−xyz (string−append "rand_" (basename inxyz))))
          (run−rsplit inxyz p)
          (map (lambda (x)

              (primitive−eval
                (list (cadr (cadr (assq
                (string−>symbol x)
                ci−engine−alist)))
                base−xyz grdcellsize)))
                engine−list)
                (map (lambda (x)

              (run−ncquery
                (string−append (basename base−xyz ".xyz") "_" x ".grd")
                rand−xyz
                delim

                "xyzgd"
                (string−append "rand_" p "_"

                                        (basename
            inxyz ".xyz") "_" x ".xyzgd")))
            engine−list)
            (run−commands (− run−times 1)
            engines p)))
            (begin
            (map (lambda (x)
            (run−xyz2grd
            (string−append "rand_" p "_"
            (basename inxyz ".xyz") "_" x
            ".xyzgd")
            (string−append "rand_" p "_"
            (basename inxyz ".xyz") "_" x
            ".xyzgd.grd")
            grdcellsize))
            engine−list)
            (map (lambda (x)
            (run−stats (string−append "rand_" p
            "_"
            (basename inxyz ".xyz") "_" x
            ".xyzgd")
            (string−append "rand_"
            (basename inxyz ".xyz") "_" x ".stat")
```

DRAFT

```
                                p do−times grdcellsize)
                                (run−stats (string−append "rand_" p
                                "_"
                                (basename inxyz ".xyz") "_" x
                                ".xyzgd")
                                (string−append "rand_" p "_"
                                (basename inxyz ".xyz") "_" x ".stat")
                                p do−times grdcellsize))

                engine−list)
          (set−gmt−defaults)
          (map (lambda (x)

                   (run−pshistogram
                    (string−append "rand_" p "_" (basename inxyz ".xyz") "_" x ".xyzgd")
                    (string−append "rand_" p "_" (basename inxyz ".xyz") "_" x ".ps")
                    p do−times x))

                engine−list)
          (map (lambda (x)

                   (run−pstext
                    (string−append "rand_" p "_" (basename inxyz ".xyz") "_" x ".xyzgd")
                    (string−append "rand_" p "_" (basename inxyz ".xyz") "_" x ".ps")
                    p do−times x))

                engine−list)
          (map (lambda (x)

                   (run−ps2raster
                    (string−append "rand_" p "_" (basename inxyz ".xyz") "_" x ".ps")
                    "G"))

                engine−list)))))

;;−−−

;;−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−;;
;; Command−line options ;;
;;−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−−;;

;;−−−Options
;; '(switch . (args? "desc"))

(define ci−option−alist '((infile . (#f "The input xyz file."))
                                  (−e . (#t "Specify which gridding engine(s) to use, \

if multiple engines are desired, separeate them with commas."))
                                  (−percent . (#t "Petcentage of xyz data to extract, default is 20"))
                                  (−grd_cells . (#t "The grid cell size"))
                                  (−delimiter . (#t "The input xyz delimiter"))
                                  (−dotimes . (#t "The number of iterations"))
                                  (−engines? . (#t "Check available engines, optionally specify \

which engine to get more information about.."))
                                  (−help . (#f "Display the help"))))

(define (ci−add−option specs)
    (append! ci−option−alist (list specs)))

        (define (ci−display−options
            option−list)
```

DRAFT

```
        (if (pair? option−list)
        (begin

          (if (cadar option−list)
                (display (string−append " [" (symbol−>string (caar option−list)) " arg]"))
                (display (string−append " [" (symbol−>string (caar option−list)) "]")))

          (ci−display−options (cdr option−list)))
        (display "\n\n")))

          (define (ci−describe−options option−list)
              (if (pair? option−list)
              (begin
             (let ((cname (symbol−>string (caar option−list)))
             (cdesc (caddr (car option −list))))
             (if (> 20 (string−length cname))

                    (display (string−append

                                    (string−append
                                      cname
                                      (make−string (− 20 (string−length cname)) #\ ))

                                  "\t" cdesc "\n"))))
         (ci−describe−options (cdr option−list)))
          (display "\n")))

(define (display−help command−name option−list)
    (if (pair? option−list)

        (begin
          (display (basename command−name))
          (display " version 0.2.1")
          (newline)
          (display "Usage: ")
          (display (basename command−name))
          (ci−display−options option−list)
          (ci−describe−options option−list))))

;; Read the given port (’p) and put it into the string ’pstring.


(define (read−port p pstring)

    (if (not (eof−object? (peek−char p)))
        (read−port p (string−append pstring (read−line p)))
        pstring))

;; Parse the command −line


         (define (parse−options cca ccl)
             (if (pair? ccl)
             (begin
            (if (not (hook−empty? ci−command−line−hook))
            (run−hook ci−command−line−hook cca ccl))
            (cond

            ;; Help? ;;

            ((or (equal? (car ccl) "−h")
                  (equal? (car ccl) "−help")
                  (equal? (car ccl) "−?"))

             (display−help cca ci−option−alist)
```

DRAFT

```
        (exit 1))

      ((equal? (car ccl) "−percent")

        (begin
          (set! percent (cadr ccl))
          (parse−options cca (cddr ccl))))

      ;; Engines ;;

      ((equal? (car ccl) "−engines?")

        (if (not (null? (cdr ccl)))
              (ci−describe−engine (cadr ccl) ci−engine−alist)
              (ci−display−engines ci−engine−alist))

        (exit 1))

      ;; Gridding Engine ;;

      ((equal? (car ccl) "−e")

        (begin
          (set! ci−engine (cadr ccl))
          (parse−options cca (cddr ccl))))

      ((equal? (car ccl) "−delimiter")

        (begin
          (set! delim (cadr ccl))
          (parse−options cca (cddr ccl))))

      ((equal? (car ccl) "−grd_cells")
        (begin

          (set! grdcellsize (cadr ccl))
          (parse−options cca (cddr ccl))))

      ((equal? (car ccl) "−dotimes")

        (begin
          (set! do−times (cadr ccl))
          (parse−options cca (cddr ccl))))

      ;; Input xyz file ;;

      ((null? inxyz)
        (set! inxyz (car ccl))
        (parse−options cca (cdr ccl)))

      (else
        (parse−options cca (cdr ccl)))))))

;; Load the User Config File
;(display ci−config)

    (if (file−exists? ci−config)
          (begin
          (load ci−config)))
```

DRAFT

```
(if (not (defined? 'ci-engine))
        (define ci-engine "surface"))

(define (main args)

    ;; Parse the command-line

     (parse-options (car (command-line)) (cdr (command-line)))

    ;; If 'inxyz was not given, get it from 'current-input-port

     (if (null? inxyz)
            (set! inxyz (read-port (current-input-port) "")))

    ;(display inxyz)

                (let ((percent-list (string-split percent #\,)))
                     (map (lambda (p)
                    (run-commands (string->number do-times) ci-engine p))
                    percent-list)))

;;--END
```

DRAFT

RSPLIT

Randomly extract lines from a text file.
Manual

DRAFT

Matthew Love

Copyright © 2011 Matthew Love.
Permission is granted to make and distribute verbatim copies of this manual provided the copyright notice and this permission notice are preserved on all copies. Permission is granted to copy and distribute modified versions of this manual under the con
ditions for verbatim copying, provided that the entire resulting derived work is distributed under the terms of a permission notice identical to this one.

DRAFT

1 Copying

RSPLIT is copyright (C) 2011 Matthew Love and is released under the terms of the GNU General Public License. See the included file 'COPYING' for the full text of the license (or see Section "Copying" in The GNU Emacs Manual).

> This is free software – you are welcome to redistribute it and/or modify it under the terms of the GNU General Public License as published by the Free Software Foundation; either version 2, or (at your option) any later version.
>
> rsplit is distributed in the hope that it will be useful, but WITHOUT ANY WARRANTY; without even the implied warranty of MERCHANTABILITY or FITNESS FOR A PARTICULAR PURPOSE. See the GNU General Public License for more details.

DRAFT

2 Introduction

Randomly extract a percentage of lines from a text file. The randomly extracted line gets printed to stdout, optionally, the remaining lines can be printed to stderr, allowing for the splitting of text files.

DRAFT

3 Installing rsplit

Installation through autotools is fairly straight forward.

1. Unpack the package
   ~] tar xfvz rsplit-0.1.3.tar.gz

2. Enter the package root directory
   ~] cd rsplit-0.1.3

3. Configure and Make the package
   The '--prefix' switch to ./configure specifies where to install files.
   ~/rsplit-0.1.3] ./configure --prefix=~/rsplit && make

4. Install the built package
   ~/rsplit] make install

DRAFT

4 Using rsplit

RSPLIT is a command-line program designed to be run in a shell program such as bash. For general usage help from the command-line, the '--help' swich can be used.

```
~] rsplit --help
Usage: rpslit [OPTIONS] [infile]
Randomly extract a percentage of lines from a text file.
                                    Options:
  -p, --percent The percent of lines to randomly extract
                                    from the input.
  -l, --line-num Include line numbers in output.
  -r, --remaining Send the lines that aren't extracted to stderr.
  --version                         Print version information and exit.
```

Example: rsplit test.xyz -p 20 -lr > test_rand20.xyz 2>test_base.xyz

- The '-p, --percent' switch specifies the percentage of randomly selected points to extract from the given text file.
- The '-l, --line-num' switch prints the line number along with the original line from the input text file.
- The '-r, --remianing' switch prints the remianing lines after random extraction to stderr.

DRAFT

5 Examples

- Extract 20 percent of the lines in the input text file and print the results to stdout ~] rsplit -p 20 input.txt

- Extract 15 percent of the lines in the input text file and write the results to a file, include the line numbers ~] rsplit -p 20 --line-num input.txt > rand20_input.txt

- Extract 40 percent of the lines in the input text file, write the results to a file and send the remaining lines to another file. ~] rsplit test.xyz -p 40 -r > test_rand40.xyz 2>test_base60.xyz

9.2.8 Spectral analysis

9.2.9 Slope

9.2.10 Kriging discussion and cautionary note

9.3 Sharing uncertainty results

TEXT IN PREPARATION

9.3.1 ASCII files

TEXT IN PREPARATION

9.3.2 netCDF files

TEXT IN PREPARATION

9.3.3 “BAG” files

Contributed by Rob Hare, Canadian Hydrographic Service, Canada

Bathymetric Attributed Grid (BAG) is a non-proprietary file format for storing and exchanging bathymetric data developed by the Open Navigation Surface Working Group. BAG files are gridded, multi-dimensional bathymetric data files that contain position and depth grid data, as well as position and uncertainty grid data, and the metadata specific to that BAG file. Information about BAG files may be obtained from the Open Navigation Surface website:

http://www.opennavsurf.org/whitepapers.html

DRAFT

Chapter 10.0 Additional Resources

10.1 Multibeam Data

Japan Agency for Marine-Earth Science and Technology (JAMSTEC):
http://www.godac.jamstec.go.pj/cruisedata/e/

National Geophysical Data Center (NGDC)
http://www.ngdc.noaa.gov/mgg/bathymetry/multibeam.html

Center for Coastal & Ocean Mapping, Joint Hydrographic Center, University of New Hampshire
http://ccom.unh.edu/index.php?p=53|58&page=law_of_the_sea.php

Earthref.org Seamount Catalog
http://earthref.org/cgi-bin/sc-s0-main.cgi

Lamont Doherty Earth Observatory Marine Geoscience Data System
http://www.marine-geo.org/portals/gmrt/

Geoscience Australia
http://www.ga.gov.au/index.html

University of Hawaii
http://www.soest.hawaii.edu/HMRG/Multibeam/index.php

European Marine Observation and Data Network
http://www.emodnet-hydrography.eu/

DRAFT

10.2 Global Bathymetry Models

Altimetric Bathymetry Grid (Smith and Sandwell, 1997)
http://topex.ucsd.edu/WWW_html/mar_topo.html

SRTM30_Plus
http://topex.ucsd.edu/WWW_html/srtm30_plus.html

GEBCO_08 Bathymetry Grid
http://www.gebco.net

Chapter 11.0 Glossary

11.1 S-44 Terms

The following has been extracted from "IHO Publication S-44 Standards for Hydrographic Surveys" and edited for context.

Note: The terms defined below are those that are most relevant to this publication. A much larger selection of terms are defined in IHO Publication S-32 (Hydrographic Dictionary) and this should be consulted if the required term is not listed here. S-32 is available as an on-line Wiki at: http://hd.iho.int/en/index.php/Main_Page.

Accuracy: The extent to which a measured or enumerated value agrees with the assumed or accepted value (see: *uncertainty*, *error*).

Bathymetric Model: A digital representation of the topography (bathymetry) of the sea floor by coordinates and depths.

Blunder: The result of carelessness or a mistake; may be detected through repetition of the measurement.

Confidence interval: See *uncertainty*.

Confidence level: The probability that the true value of a measurement will lie within the specified *uncertainty* from the measured value. It must be noted that confidence levels (e.g. 95%) depend on the assumed statistical distribution of the data and are calculated differently for 1 Dimensional (1D) and 2 Dimensional (2D) quantities. In the context of this standard, which assumes Normal distribution of error, the 95% confidence level for 1D quantities (e.g. depth) is defined as 1.96 x sigma and the 95% confidence level for 2D quantities (e.g. position) is defined as 2.45 x sigma.

Correction: A quantity which is applied to an observation or function thereof, to diminish or minimize the effects of errors and obtain an improved value of the observation or function. It is also applied to reduce an observation to some arbitrary standard. The correction corresponding to a given error is of the same magnitude as the computed error but of opposite sign.

Error: The difference between an observed or computed value of a quantity and the true value of that quantity. (N.B. The true value can never be known, therefore the true error can never be known. It is legitimate to talk about error sources, but the values obtained from what has become known as an error budget, and from an analysis of residuals, are *uncertainty* estimates, not errors. See *uncertainty*.)

Metadata: Information describing characteristics of data, e.g. the *uncertainty* of survey data. ISO definition: Data (describing) about a data set and usage aspect of it. Metadata is data implicitly attached to a collection of data. Examples of metadata include overall quality, data set title, source, positional *uncertainty* and copyright.

DRAFT

Quality assurance: All those planned and systematic actions necessary to provide adequate confidence that a product or a service will satisfy given requirements for quality.

Quality control: All procedures which ensure that the product meets certain standards and specifications.

Reduced depths: Observed depths including all *corrections* related to the survey and post processing and reduction to the used vertical datum.

Sounding datum: The vertical datum to which the soundings on a hydrographic survey are reduced. Also called 'datum' for sounding reduction.

Total horizontal uncertainty (THU): The component of *total propagated uncertainty* (TPU) calculated in the horizontal plane. Although THU is quoted as a single figure, THU is a 2D quantity. The assumption has been made that the uncertainty is isotropic (i.e. there is negligible correlation between errors in latitude and longitude). This makes a Normal distribution circularly symmetric allowing a single number to describe the radial distribution of errors about the true value.

Total propagated uncertainty (TPU): the result of uncertainty propagation, when all contributing measurement uncertainties, both random and systematic, have been included in the propagation. Uncertainty propagation combines the effects of measurement uncertainties from several sources upon the uncertainties of derived or calculated parameters.

Total vertical uncertainty (TVU): The component of *total propagated uncertainty* (TPU) calculated in the vertical dimension. TVU is a 1D quantity.

Uncertainty: The interval (about a given value) that will contain the true value of the measurement at a specific *confidence level*. The *confidence level* of the interval and the assumed statistical distribution of errors must also be quoted. In the context of this standard the terms uncertainty and *confidence interval* are equivalent.

Uncertainty Surface: A model, typically grid based, which describes the depth uncertainty of the product of a survey over a contiguous area of the skin of the earth. The uncertainty surface should retain sufficient *metadata* to describe unambiguously the nature of the uncertainty being described.

DRAFT

Chapter 12.0 Acronyms and Abbreviations

A

ASCII	American Standard Code for Information Interchange

B

BAG	Bathymetry Attributed Grid
BODC	British Oceanographic Data Centre (UK)

C

CHM	Center of Hydrography and Navigation
CHS	Canadian Hydrographic Service
CIOH	Centro de Investigáciones Oceanograficas e Hidrográficas (Columbia)
CIRES	Cooperative Institute for Research in Environmental Sciences (USA)

D

DEM	Digital Elevation Model
DHN	Directoria de Hidrografia e Navegaçao (Brazil)
DOC	Department of Commerce (USA)
DTM	Digital Topographic Model

E

F

G

GEBCO	General Bathymetric Chart of the Oceans
GEODAS	Geophysical Data System
GGC	GEBCO Guiding Committee
GIS	Geographic Information System
GMT	Generic Mapping Tools
GPS	Global Positioning System

H

HIPS	Hydrographic Information Processing System

I

IHB	International Hydrographic Bureau (Monaco)
IHO	International Hydrographic Organization (Monaco)
IOC	Intergovernmental Oceanographic Commission

J

JAMSTEC	Japan Agency for Marine-Earth Science and Technology (Japan)
JODC	Japan Oceanographic Data Centre (Japan)

DRAFT

K

km	kilometer

L

LDEO	Lamont-Doherty Earth Observatory

M

m	meter
MBES	Multi-beam echo sounder

N

NERC	Natural Environment Research Council (UK)
NESDIS	National Environmental Satellite, Data, and Information Service
NGDC	National Geophysical Data Center (USA)
NOAA	National Oceanic and Atmospheric Administration (USA)
NOC	National Oceanography Centre (UK)
NRL	Naval Research Laboratory

O

OGC	Open Geospatial Consortium

P

Q

R

RAS	Russian Academy of Sciences (RUSSIA)
R/V	Research Vessel

S

SHOM	Service Hydrographique et Oceanographique del la Marine (France)
SOEST	School of Ocean and Earth Science and Technology (USA)
SRTM	Shuttle Radar Topography Mission

T

TAR	Tape Archive software utility
TIN	Triangular Irregular Networks

U

UNH	University of New Hampshire (USA)
URL	Universal Resource Locator
UTM	Universal Transverse Mercator

V

DRAFT

W

WWW World Wide Web

X

Y

Z

DRAFT