**Bathymetry Fusion Algorithm:**

**MergeBathy C++ User’s Guide**

Version 5.0.0

16 December 2014

Kevin Duvieilh, Paul Elmore, Samantha Zambo

*Naval Research Laboratory, Code 7440, Stennis Space Center, MS 39529*

*{ kevin.duvieilh; paul.elmore, samantha.zambo }@nrlssc.navy.mil*

*228-688-4613*

***WARNING!!!***

This is for the C++ version. There are some notable differences. The defaults in Matlab at the time of writing were **mgrid** 50, **msmooth** 200, and to keep all final depths. These were changed to **mgrid** 100, **msmooth** 100, and the removal of all bad depths by default unless **modelflag** specified in MergeBathy 5.0 for Matlab and C++. Furthermore, the Matlab version computes offsets by default and can be turned off with **nooffset**. This is opposite in C++ where by default they are not computed but can be with **computeOffset**. The options **local**, **rect**, **smoothplots**, **errorplots**, **saveplots** and **plotpts** are for displaying figures in Matlab only. The options **itersmooth** and **calcsmooth** are Matlab only; there is no findLambda in C++. Custom file formats are not in the current Matlab or C++ but may be added in. This existed in an older Matlab. Tide corrections and maximum allowable offsets require delimiters in C++. Reference lat and lon are reversed between Matlab and C++.

1. **Purpose**

The purpose of this guide is to familiarize users with the MergeBathy bathymetric smoothing algorithm application. This is the user guide for MergeBathy C++. For the Matlab version and through documentation with examples, see mergeBathyUser.doc.

1. **Using MergeBathy**

MergeBathy is a command line based bathymetric processing utility used to merge and smooth bathymetric input data according to a user defined weighting window and node spacing. End users must have input data in one of the supported file formats and know the center Longitude and Latitude of that input data. All processing performed by MergeBathy can be completed successfully if these three items are known. MergeBathy also has the ability to pre-spline interpolate input data to reduce aliasing that may be incurred from very sparsely sampled data sets. If aliasing is seen in the output data then it is recommended that pre-splining be done using either the GMT Surface or MB ZGrid options to correct the aliasing problems. MergeBathy is also able to leverage modern, multi-core CPU architectures through the multi-threading of its computational routines. When run as a multi-threaded application the runtime of MergeBathy is substantially reduced; multi-threaded bathymetric computation is a feature unique to MergeBathy.

1. **Supported Input File Formats**

Input files are passed to MergeBathy through an input list file. This list file is a flat text file with each new line of the file containing the full path to an input data file. The paths to individual data files may be absolute or relative to the working location of the command window. Either way, only one input data file may appear on each line of the input file list. The supported input file types that can be listed in the input file list are as follows:

**\*\_xyz.dat** – Longitude, Latitude, Depth – Tab delimited between data fields.

**\*\_xyd.txt** – Longitude, Latitude, Depth – Tab delimited between data fields.

**\*\_xyze.dat** – Longitude, Latitude, Depth, Error – Tab delimited between data fields.

**\*\_xyde.txt** – Longitude, Latitude, Depth, Error – Tab delimited between data fields.

**\*\_xyzhv\_mc.dat** – Longitude, Latitude, Depth, Horizontal Error, Vertical Error – Tab delimited between data fields.

**\*\_xydhv\_mc.txt** – Longitude, Latitude, Depth, Horizontal Error, Vertical Error – Tab delimited between data fields.

**\*.d** – Standard GSF file format.

**\*.gsf** – Standard GSF file format.

**\*.grd** – Standard ARC ASCII Raster file format.

**\*.asc** – Standard ARC ASCII Raster file format.

In addition to listing the input files to be used by MergeBathy for interpolation, the input list file may also contain input data set specific weight factors. Weight factors are useful if certain data sets are known to be more accurate than other data sets and need to have greater weight in the interpolation. These weight factors appear after the file name in the following format:

*{path}/{filename} & {weightValue} ^ {tideCorrection} # {maxOffset}*

Where *“{path}”* is the path to an input file, *“{filename}”* is the name of the input file, *“&* *{weightValue}”* is the weight value to be applied to the data set, *“^ {tideCorrection}”* is the tidal offset to be applied to the data set, and *“# {maxOffset}”* is the maximum vertical offset that can be applied to a data set when calculating offsets between data. Higher weight values mean that a data set will be weighted higher than other data sets with lower weight values. If no *“& {weightValue}”* is provided then all input data sets will be weighted equally. If no *“^ {tideCorrection}”* is provided then no tide factors are used. If no *“# {maxOffset}”* is provided then the data set can be offset by any value based on calculations performed in MergeBathy. These three options can appear in any order in the string and each must be separated by a space between the value and the next character delimiter *“& ^ #”*.

1. **Supported Output File Formats**

Five types of output formats are supported by MergeBathy. The first is a flat text output file format using:

*Longitude, Latitude, Depth, Uncertainty, Normalized Uncertainty, Residual Uncertainty*

Each value is tab delimited and the last three values are optional and may be included based on command line arguments.

The second type of output file is an ARC ASCII Raster output grid file. This grid file is output in UTM meters and follows the standard ARC ASCII Raster header and formatting protocols. The standard ARC ASCII Raster projection file is also output for the associated data file when the user specifies ARC ASCII Raster.

The third type of output file is a Bathymetric Attributed Grid (BAG) Object. This BAG file follows the standard BAG-file formatting protocols. The standard BAG XML metadata file is also output for the associated data file when the user specifies BAG.

The fourth is a flat text output file format using:

*Longitude, Latitude, Depth, Uncertainty, Normalized Uncertainty, Residual Uncertainty, Kalman Depth, Kalman Uncertainty*

As with the first file format, each value is tab delimited and the fifth, sixth, and seventh values are optional and may be included based on command line arguments. This differs from the first by appending the Kalman calculated depth and uncertainty to an MSE computed output file.

The fifth is a flat text output file format using:

*Longitude, Latitude, Depth, Depth (ft), X, Y, Normalized Uncertainty, Residual Uncertainty, Uncertainty, Kalman Depth, Kalman Uncertainty*

As with the other flat file formats, each value is tab delimited and the seventh, eigth, and ninth values are optional and may be included based on command line arguments. This format rearranges the output to match that generated by the MATLAB MergeBathy version.

The last two output file formats maybe removed in future versions.

1. **Command Line Arguments**

The computations performed by MergeBathy are determined through a series of command line arguments that are passed to the application at runtime. There are two types of command line arguments – required arguments and optional arguments. These two types of arguments are detailed below.

* 1. **Required Command Line Arguments**

|  |  |
| --- | --- |
| Output File Name | This argument contains the path and name of the output file to be generated by MergeBathy. The path may be relative or absolute. |
| Grid Spacing | This argument is an interpolation grid spacing defined in meters between interpolation data points.  **If one value is given it will be applied in both the X and Y directions. If two values are given (separated by a space) then the first value will be the spacing in the X direction and the second value will be the spacing in the Y direction.**  This value will also be used as the smoothing scale unless otherwise specified with the **–msmooth** or **-llsmooth** options. |
| Kernel Name | The spacing window weighting methodology to use in interpolating data points. The available options are **hanning** or **hann**, **boxcar**, **loess**, and **quadloess.** |
| Input File List | This argument contains the path and name to the input file list that contains a listing of all of the input files to be used by the MergeBathy interpolation. The path to this file may be relative or absolute. |
| Reference Longitude | The center reference Longitude of the input data sets. If using  –preInterpolatedLocations, this value must be in Longitude for ll2utm conversion of locations.. |
| Reference Latitude | The center reference Latitude of the input data sets. If using  –preInterpolatedLocations, this value must be in Latitude for ll2utm conversion of locations. |
| Rotation Angle | The rotation angle of the sensing beam used in collecting data for the input data sets. Typically 0. |
| Number or Monte Carlo Data Runs | Sets the number of Monte Carlo data runs to perform. If **-1** is used then only a single interpolation routine is run. If any number greater than 0 is used then that number of Monte Carlo interpolations will be run.  WARNING: Monte Carlo is unsafe in version 4.0.0! It is advised to always use -1. The function was not updated and its behavior is unknown. |

* 1. **Optional Command Line Arguments**

The following are the optional command line arguments that can be used to modify the output and processing performed by MergeBathy. Some optional arguments have mandatory parameters that follow them. The mandatory parameters will be designated in *italic* font and right aligned below the optional arguments that they must follow. All optional arguments and mandatory parameters are separated by a space.

|  |  |
| --- | --- |
| -noerr | Do not output the combined uncertainty estimate in the output file. |
| -nmsei | Output the normalized uncertainty estimate in the output file. |
| -msri | Output the residual uncertainty estimate in the output file. |
| -inputInMeters | Signal that the input files use UTM X and Y coordinates instead of Longitude and Latitude coordinates. |
| -kriging | Perform Kriging on the data once the interpolation is complete. Kriging will restore additional details to the data but will cause the interpolation routine to run substantially slower. For more information on the benefits of Kriging refer to the MergeBathy White Paper.  WARNING: Kriging is unsafe in version 4.0.0! It is advised to never use -kriging. The function was not updated and its behavior is unknown. |
| -ZGrid | Use the MB\_ZGrid interpolator to pre-spline the input data. This will create a smoother input surface for MergeBathy and reduce aliasing for very sparsely sampled input data. |
| *Spacing X* | Computational grid spacing in the X directions. Defined in meters. |
| *Spacing Y* | Computational grid spacing in the Y directions. Defined in meters. |
| *Output File Name* | The name of the output file to be used by MB\_ZGrid to output its interpolated coordinates. |
| *Tension Factor* | Sets the tension of the interpolation. A value of 0.0 yields a pure Laplace (minimum curvature) solution and a value of infinity yields a pure thin plate spline solution. A value of 1e10 value has commonly been used to yield spline solutions. |
| *Usage* | A value of 1 will perform the MB\_ZGrid interpolation and write the results X, Y, and Z to the specified output file.  A value of 2 will perform the MB\_ZGrid interpolation, write the results X, Y, and Z to the specified output file, compute the associated uncertainty, and use the pre-splined X,Y, Z and E values as input for MergeBathy.    If used as input then the data will take the place of the data read from the input files. This allows for a pre-smoothing effect before MergeBathy is run.  Negate the usage values (-1 or -2) to write to the file the computed error associated with the pre-smoothing. |
| -GMTSurface | Use the GMT Surface interpolator to pre-spline the input data. This will create a smoother input surface for MergeBathy and reduce aliasing for very sparsely sampled input data. This pre-spline routine is an industry standard and is the recommended pre-spline interpolator for MergeBathy. |
| *Spacing X* | Computational grid spacing in the X directions. Defined in meters. |
| *Spacing Y* | Computational grid spacing in the Y directions. Defined in meters. |
| *Output File Name* | The name of the output file to be used by GMT Surface to output its interpolated coordinates. |
| *Tension Factor* | The Laplacian tension operator between 0 and 1. Typically 0.1. |
| *Scale Factor* | The multiplier value for a Confidence Interval to be used in error calculation. A value of 1.96 is typically used for a 95% Confidence Interval. This value is used in estimating the uncertainty associated with the GMT Surface interpolation. |
| *Alpha* | This is the alpha value for error computation. Typically 2.0. This value is used in estimating the uncertainty associated with the GMT Surface interpolation. |
| *Usage* | A value of 1 will perform the GMT Surface interpolation and write the results X, Y, and Z to the specified output file.  A value of 2 will perform the GMT Surface interpolation, write the results X, Y, and Z to the specified output file, compute the associated uncertainty, and use the pre-splined X,Y, Z and E values as input for MergeBathy.    If used as input then the data will take the place of the data read from the input files. This allows for a pre-smoothing effect before MergeBathy is run.  Negate the usage values (-1 or -2) to write to the file the computed error associated with the pre-smoothing. |
| -preInterpolatedLocations | Allows a file to be provided to determine the exact Longitude and Latitude locations for the interpolation. This allows irregularly gridded output at set points instead of points determined from a user defined grid spacing. Must be in lat/lon. |
| *File Name* | The path and file name to a file containing Longitude and Latitude coordinates for interpolation nodes. The path may be relative or absolute. |
| *Usage* | A value of 1 will read the values as latitude followed by longitude. Negate the usage value (-1) to reverse and read as longitude, latitude. Use (-1) if using output from –zgrid or –surface. A value of 2 will use the values for locations when presplining instead of computing locations. |
| -computeOffset | Pre-interpolates input data files to determine the offset associated with each data set. This offset is then removed to normalize the input data sets along a single plane. |
| -msmooth | This option sets a smoothing scale to be applied during the interpolation in meters. If no smoothing scale is applied then the smoothing scale is the same as the grid spacing. |
| *Smoothing scale x* | Smoothing scale in meters in the X direction. |
| *Smoothing scale y* | Smoothing scale in meters in the Y direction. |
| -llsmooth | This option sets a smoothing scale to be applied during the interpolation in terms of longitude and latitude. If no smoothing scale is applied then the smoothing scale is the same as the grid spacing. |
| *Smoothing scale longitude (x)* | Smoothing scale in longitude in the X direction. |
| *Smoothing scale latitude (y)* | Smoothing scale in latitude in the Y direction. |
| -llgrid | This option specifies that the grid spacing passed to MergeBathy is defined in terms of longitude and latitude instead of the standard meters. |
| -boundingBox | This option allows for the input data to be cut so that only points that fall within a certain bounding box are used in the interpolation. |
| *Upper bound* | The upper bound of the bounding box. |
| *Lower bound* | The lower bound of the bounding box. |
| *Right bound* | The right bound of the bounding box. |
| *Left bound* | The left bound of the bounding box. |
|  |  |
| -outputRasterFile | Output an ARC ASCII Raster file format. If this option is selected then the flat text output file will not be generated. |
| -outputBagFile | Output a Bathymetry Attributed Grid (BAG) Object file format. If this option is selected then the flat text and ARC ASCII Raster output files will not be generated. The sample.xml file must be in the same location from where mergeBathy is called. |
| -multiThread | Signal that the MergeBathy application is being run on a system with multiple processing cores. If this flag is thrown then the MergeBathy application will break its interpolation processing algorithm up so that it is run on a specified number of cores. Using this flag substantially reduces the processing time taken by MergeBathy. It is cross platform and can be used in Windows or Linux. If unexpected behavior occurs, disable multi-threading by removing the option and run again. |
| *Number of Threads* | The number of processing cores available to MergeBathy. It is recommend that this number be greater than or equal to 2. The number should not be greater than the total number of processing cores available to the system. There is no software maximum limit to the number of cores that can be used during interpolation; the only limitation is hardware capability. |
| -modelflag | Use all points. Default removes bad points on output. |
| -nonegdepth | Do not use negative depths; only water. Default keeps all depths on input and output. |
| -useUnScaledAvgInputs | This option allows an average error e, h, or v to be to be used as input. By default, when the same e, h, or v is provided for all depths (i.e. a standard deviation is 0), it is assumed to be the average and each point will scale it by its depth z. |
| -nnInterp | Performs nearest neighbor interpolation for output depth, pre-splining uncertainty, and output uncertainty surfaces. Default is bilinear for depth and IDW for uncertainty.  Interpolations performed by MergeBathy:  *Depth trend surface analysis*: least-squares regression.  *Smoothing*: scaled-controlled interpolation with hann, boxcar, loess, or quadloess.  *Pre-splining depth surface interpolation*: GMT surface or MB-System zgrid.  *Output depth surface*: nearest neighbor or 3 Delaunay Triangle nearest-neighbors with bilinear (Extrapolation is handled by DT algorithm.).  *Pre-splining and output uncertainty surface*: nearest neighbor or 3 Delaunay Triangle nearest-neighbors with IDW (Extrapolation is handled by DT algorithm.)  *Raster repositioning depth and uncertainty surface*: Rotates output surface back to original angle. Bilinear with no extrapolation. |
| -mse | Use Mean Square Error (linear) Estimator. |
| *Print* | A (default) value of 1 will print the Mean Square Error Estimator results to an output file. Negate the print value (-1) to suppress the output file. |
| -propUncert | Use Propagated Uncertainty Estimator. |
| *Print* | A value of 1 will print the Propagated Uncertainty Estimator results to an output file denoted with “\_P”. Negate the print value (-1) to suppress the output file.  WARNING: The Propagated Uncertainty Estimator is unsafe in version 4.0.0! It is advised to never use -propUncert. The function was not fully verified and its behavior is unknown. |
| -kalman | Use Kalman Estimator. |
| *Print* | A value of 1 will print the Kalman Estimator results to an output file denoted with “\_K”. Negate the print value (-1) to suppress the output file. |
| -printMSEwK | Print the output file for MSE with Kalman results tagged on at the end. |
| -printMatlabMatch | Print an output file formatted to match output from the MATLAB MergeBathy version and denoted with “\_MATLAB”. Can not be used in conjunction with –outputRasterFile or –outputBagFile.  Output format:  latitude,longitude,depth\_m,depth\_ft,X,Y,Nerr,Rerr,Error,KZ,KVAR  Zero column placeholders are used for depth\_ft, X, and Y.  KZ and KVAR are the Kalman Estimator computed depths and errors, respectively. |
| -appendFilename | Appends characters to the user defined filename to identify the various output files with the smoothing window and whether nearest neighbor interpolation was used. |

1. **Standard Defaults**
   1. **Horizontal and Vertical Errors**

The horizontal and vertical errors are only used to perturb the input data locations when Monte Carlo simulations are being performed. The Kalman Estimator and the Propagated Uncertainty Estimator use the horizontal and vertical errors to computed updated uncertainties. If no horizontal and vertical error is specified then the horizontal error is set to 0.000001 meters and the vertical error is set to 0.000001 meters for each observed input location.

In previous versions, if no horizontal and vertical error was specified and a Monte Carlo simulation was performed then the horizontal error was set to 10.00 meters and the vertical error was set to 5% of the observed depth for each observed input location.

* 1. **Individual Weighting of Input Data Sets**

The option used to provide weight to individual data sets in the input list file applies weights to the individual data sets by modifying the errors associated with each data point. Since data points with higher errors are weighted less by MergeBathy these data points will contribute less to the overall smoothed surface. If no errors are provided in an input and the user specified weighting scale is not defined then the errors associated with a data file are set to 0. If a weighting scale is provided by a user and there is no error associated with each data point (i.e. a .xyd.txt file) then the weights associated with all data points in the input file are calculated to:

If a weighting scale is provided and there is an error value associated with each data point of the input file then the error at each point is computed as:

* 1. **Grid Spacing vs. Smoothing Scale**

The grid spacing is defined by a required argument – it is the second argument that defines both the X and Y spacing to MergeBathy; or the second and third argument if a separate spacing needs to be defined in the X and Y directions. The grid spacing values contribute to the overall size and spacing of the output grid. The grid spacing values are also used as the default smoothing scale that is used in weighting and interpolating data between input data points. The smoothing scale is different from the grid spacing in that the smoothing scale does not affect the size of the output grid but instead determines how far from interpolated points MergeBathy should look for input data. While the grid spacing is typically used as the default smoothing scale it is possible, and sometimes desired for sparsely populated input data, to have a larger smoothing scale than grid spacing. To modify the smoothing scale use either the **–msmooth** or **–llsmooth** options to specify a smoothing scale that is different from the grid spacing.

* 1. **Offset Computation**

By default no offset is computed between data sets. Should the offset need to be computed the **–computeOffset** option must be passed on the command line. Offsets between data sets are computed using the MergeBathy interpolation routine by sparsely sampling the data, interpolating the data, and then computing the difference between the overlap of each data set. Each data set is then offset by the difference between data sets up to the maximum offset allowed for the data set. This maximum offset is either Infinite or defined by the *#max\_offset* option in the input list file. If a data set has an offset greater than its maximum offset then all other data sets will receive a greater offset to compensate for the other data set not being able to move past its maximum offset.

* 1. **Multi-Threading**

By default MergeBathy will not use multi-threading in its interpolation. In this case the entire grid is interpolated in a single thread on a single processor. Should multi-threading be desired it must be called with the **–multiThread** command line option. The number of cores specified after this option should always be less than or equal to the total number of processing cores available on a system. Optimally MergeBathy should not use all of the processing cores on an individual workstation since the operating system and background processes will require one processing core for normal system operation. If MergeBathy is using every core available on an individual workstation then at least one thread of MergeBathy will need to be suspended by the operating system so that normal system processes can be performed. If undefined behavior occurs, disable multi-threading and run again.

1. **Supported Operating Systems**

MergeBathy was developed in Microsoft Visual Studio 2010 on Windows 7 Professional (x64). All of the source code was then ported to Red Hat Enterprise Linux 5 and 6 (x64) to ensure cross compatibility between environments. The source for MergeBathy is designed to be platform independent and therefore will work on both Microsoft Windows and Linux without altering the source code. The following environments are known to support MergeBathy, although they were not used in its development.

Microsoft Windows XP (32 and 64 Bit)

Microsoft Windows Vista (32 and 64 Bit)

Microsoft Windows 7 (32 and 64 Bit)

Ubuntu Linux (32 and 64 Bit)

Red Hat Enterprise Linux (32 and 64 Bit)

1. **Appendix**
   1. **Validation Report (\*Not Updated)**

For reference to command line execution and expected results reference the MergeBathy Validation Test Report.

* 1. **Technical Report(\*Not Updated)**

The developers’ technical report describes the files, classes, methods, and functions that provide the functionality of MergeBathy. This technical report exists as a PDF file and as a hyperlinked HTML webpage (The recommended method of viewing the technical report). Should the MergeBathy source need to be modified this report should act as a starting point for introducing developers to the MergeBathy code and functionality.

* 1. **Flow Diagrams (\*Not Updated)**

The developers of MergeBathy have included UML class diagrams and UML sequence diagrams for developers who wish to extend the capabilities of MergeBathy. These UML diagrams provided details on how the normal flow of execution is performed in MergeBathy and are a good point of reference for additional developers.

* 1. **Known Limitations**

As with many technical products one key limitation is the user. Inputs must follow the appropriate format and users are expected to have some degree of understanding on how to operate MergeBathy. It is also important to note the limitations of the computer system running MergeBathy, as some systems may be faster or slower than the system used in MergeBathy’s development. Understanding this hardware limitation is also extremely important in multi-core processing and the use of Kriging in MergeBathy since some systems will perform better than others. Beyond this key user based requirement there are no known or artificially created limitations on file size, data resolution, number of processing cores, output grid resolution, or total number of data points to process that exist in MergeBathy.