

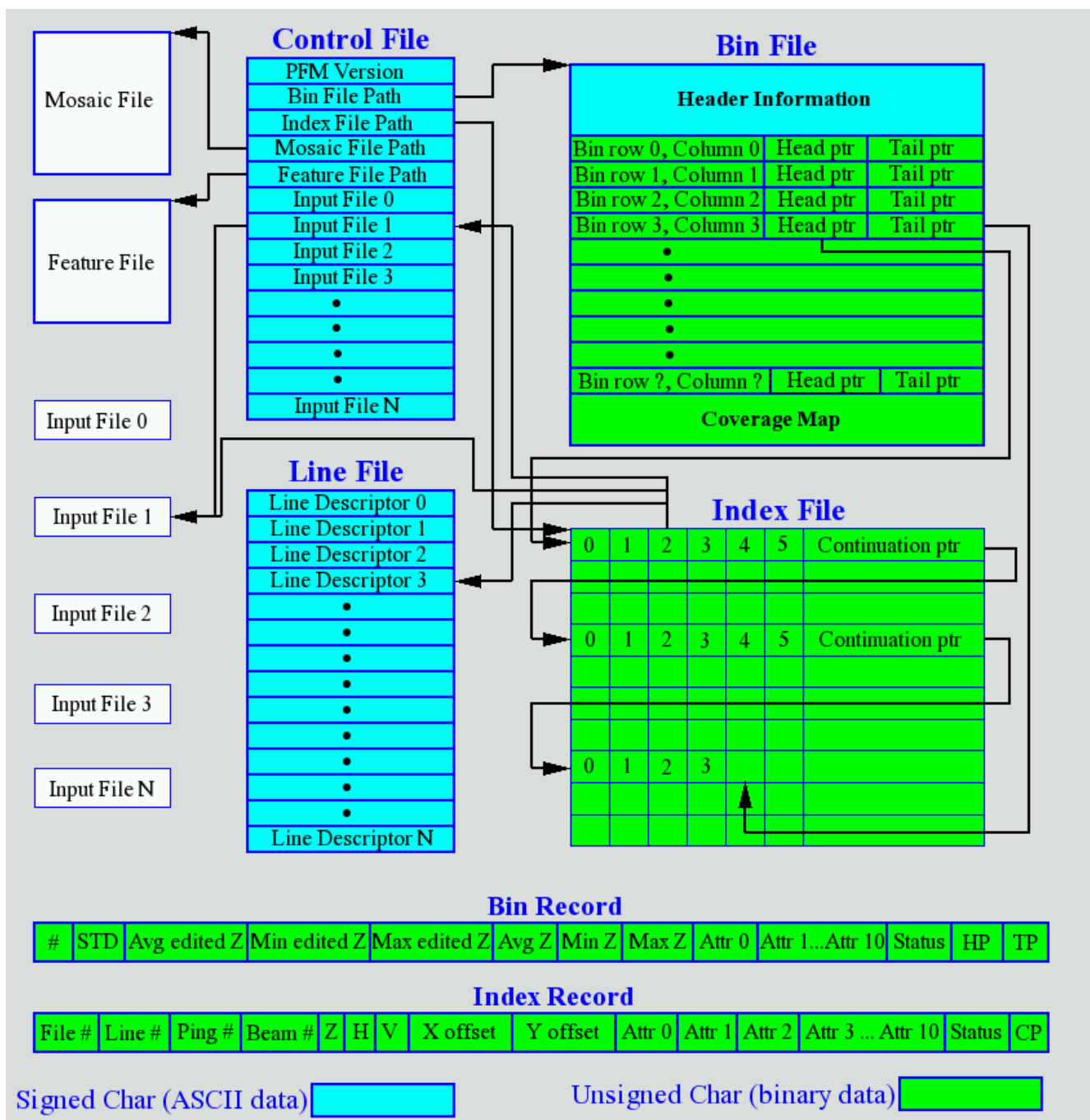
Pure File Magic (PFM) Data Structure

Introduction

The Pure File Magic data structure was originally designed to allow hydrographers to geographically view minimum, maximum, and average binned surfaces, of whatever bin size they chose, and then allow them to edit the actual depth values that contributed to that bin. After editing the depth data, the bins would be recomputed and the binned surface redisplayed. The idea being that the hydrographer could view the min or max binned surface to find outliers and then edit just those areas that required it. In addition to manual viewing and hand editing, the PFM format is helpful for automatic filtering. After editing of data is complete the status information can be transferred back into the original data files. At present there are 41 different file types that are supported either by the Naval Oceanographic Office, SAIC, or IVS in their respective loader/unloaders.

The File Structure

The PFM data structure consists of a number of directories and files. At it's simplest it can be viewed as a single .pfm handle file and a .pfm.data data directory. Inside the data directory are a number of other files and directories. In general terms, the PFM structure consists of an ASCII control file (.ctl) containing the names of all of the associated PFM and non-PFM files and directories, a binned surface file (.bin) containing all of the binned surfaces and links to the indexed data, an indexed file (.ndx) containing the original input data and status information, and an optional line file (.lin) containing line names. A simple graphical overview of the structure follows. Note that the diagram does not show the handle file. Also, all binary data in the bin and ndx files are stored as bit-packed, scaled/offset, unsigned integers (the number of bits, scale, and offset used for each field are stored in the bin header).



You can think of the bin files as a giant grid where each grid cell has a bin record associated with it. The bin records are stored west to east, south to north. That is, the first bin record stored in the file (0, 0) is the bin in the lower left (southwest) corner of the area. The next record is the bin to the immediate right (east). This continues to the eastern end of the area and then moves up one row. As can be seen by the above diagram, PFM allows for very quick access to any data based on its geographic location within the binned surface. In addition, the original input file/record/sub-record can be accessed easily by using the file

number, ping (record) number, and beam (sub-record) number from the index file in combination with the input file name from the ctl file.

The Handle File and Data Directory

The handle file (.pfm) is a very small file that is created when we build the PFM structure. Its purpose is to work as a handle for the data directory. There is only one line that is required in the handle file and that is the version line. The version line is always the first line in the file. All other lines must be comments and must begin with the # character. The data directory contains all of the actual PFM data files/directories (see below). We use a handle file to make file dialogs easier to build and comprehend. The data directory is given the same name plus a .data extension. For example, if the handle name was */data1/datasets/okinawa/fred.pfm*, then the data directory would be */data1/datasets/okinawa/fred.pfm.data*.

The ctl File *<pfm handle file>.ctl*)

The ctl file is a text file containing the full path names of the output and input files. The advantage of this text file is that it can be edited if the location of the files changes after the initial PFM structure creation. This text file is placed in the data directory. It will be created with the same name as the PFM handle file, for instance, if you named the PFM handle file */data1/datasets/sftf.pfm* then the ctl file will be */data1/datasets/sftf.pfm.data/sftf.pfm.ctl*. The bin and ndx file names are saved in the PFM .ctl file in order to maintain backward compatibility with earlier versions (pre-4.5) of PFM that didn't place all of the files in the same directory. Following the bin and ndx file names is a mosaic file name. In the case of the PFM Area-Based Editor (ABE) this is the name of an associated GeoTIFF file but other software may use it for different purposes. Following that is a feature file. This was originally referred to as a target file. The type of file placed here depends on the type of software being used. The ABE expects a Binary Feature Data (BFD) file. The mosaic and feature file slots may contain the word **NONE** to indicate that there is no associated mosaic or feature file.

Sample Ctl file:

```
PFM Software - PFM I/O library V4.71 - 04/23/06
/data1/tags1/test_990508_2/em3000/test.pfm.data/test.pfm.bin
/data1/tags1/test_990508_2/em3000/test.pfm.data/test.pfm.ndx
/data1/tags1/test_990508_2/em3000/liberty.tif
/data1/tags1/test_990508_2/em3000/liberty.bfd
+ 00000 02 /data1/tags1/test_990508_2/em3000/99mbg991281726.d01
+ 00001 02 /data1/tags1/test_990508_2/em3000/99mbg991281738.d01
+ 00002 02 /data1/tags1/test_990508_2/em3000/99mbg991281742.d01
+ 00003 02 /data1/tags1/test_990508_2/em3000/99mbg991281753.d01
- 00004 02 /data1/tags1/test_990508_2/em3000/99mbg991281813.d01
+ 00005 02 /data1/tags1/test_990508_2/em3000/99mbg991281816.d01
```

The + sign in the first column signifies a file that has not been marked as "deleted" in the PFM structure. The - sign signifies that a file has been "deleted". Files that are marked as

“deleted” have had PFM_DELETED set in the status information for each data point for that file in the indexed file. These points should be completely ignored by any applications that use the PFM data. The first number is a sequence number. This is used to make sure that a user has not inadvertently (or verterntly for that matter ;-)) deleted a line from the list file. File names must never be deleted or added to this file nor may the order of the files be changed manually (using an editor). The second number is the data type. Do not ever manually change the +/- or the data type. The currently defined data types are as follows:

```
PFM_UNDEFINED_DATA = Undefined
PFM_CHRTR_DATA = NAVOCEANO CHRTR format
PFM_GSF_DATA = Generic Sensor Format
PFM_SHOALS_OUT_DATA = Optech SHOALS .out format
PFM_CHARTS_HOF_DATA = CHARTS Hydrographic Output Format
PFM_NAVO_ASCII_DATA = NAVOCEANO ASCII XYZ format
PFM_HTF_DATA = Royal Australian Navy HTF
PFM_WLF_DATA = Waveform LIDAR Format
PFM_DTM_DATA = IVS DTM data format
PFM_HDCS_DATA = Caris HDCS data format
PFM_ASCXYZ_DATA = Ascii XYZ data format
PFM_CNCBIN_DATA = C&C Binary XYZ data format
PFM_STBBIN_DATA = STB Binary XYZ data format
PFM_XYZBIN_DATA = IVS XYZ Binary data format
PFM_OMG_DATA = OMG Merged data format
PFM_CNCTRACE_DATA = C&C Trace data format
PFM_NEPTUNE_DATA = Simrad Neptune data format
PFM_SHOALS_1K_DATA = Shoals 1K(HOF) data format
PFM_SHOALS_ABH_DATA = Shoals Airborne data format
PFM_SURF_DATA = Altas SURF data format
PFM_SMF_DATA = French Carribes format
PFM_VISE_DATA = Danish FAU data format
PFM_PFM_DATA = NAVOCEANO PFM data format
PFM_MIF_DATA = MapInfo MIF format
PFM_SHOALS_TOF_DATA = Shoals TOF data format
PFM_UNISIPS_DEPTH_DATA = UNISIPS depth data format
PFM_HYD93B_DATA = Hydro93 Binary data format
PFM_LADS_DATA = Lads Lidar data format
PFM_HS2_DATA = Hypack data format
PFM_9COLLIDAR = 9 Column Ascii Lidar data format
PFM_FGE_DATA = Danish Geographic FAU data format
PFM_PIVOT_DATA = SHOM Pivot data format
PFM_MBSYSTEM_DATA = MBSYSTEM data format
PFM_LAS_DATA = LAS data format
PFM_BDI_DATA = Swedish Binary DIS format
PFM_NAVO_LLZ_DATA = NAVO lat/lon/depth data format
PFM_LADSDB_DATA = Lads Database Link format
PFM_DTED_DATA = NGA DTED format
PFM_HAWKEYE_DATA = Hawkeye CSS Generic Binary Output Format
PFM_CZMIL_DATA = Coastal Zone Mapping and Imaging LIDAR Format
PFM_BAG_DATA = Bathymetric Attributed Grid format
```

Note that the exact way that these data types have been loaded is dependent on the loader used. The vast majority of these were defined by IVS. In some cases (GSF, HOF, TOF) the way the data are loaded has been agreed upon by IVS, SAIC, and NAVOCEANO. If you wish to define a new data type you must inform all interested parties so that we do not have a data type “collision”.

The Line File <pfm handle file>.lin)

The line file is a text file containing line descriptor information. As with the ctl file this file can be edited to change the line descriptions. This file is always called *<pfm handle file>.lin*. Accidentally (or purposely) removing this file is not a problem since the contents are only labels. If this file doesn't exist (as with a PFM 3.0 file) lines are listed as UNDEFINED.

Sample Line file:

```
99mbg991281721.d01-000
99mbg991281721.d01-001
99mbg991281726.d01-000
99mbg991281726.d01-001
99mbg991281726.d01-002
99mbg991281726.d01-003
99mbg991281738.d01-000
99mbg991281738.d01-001
99mbg991281742.d01-000
99mbg991281742.d01-001
99mbg991281742.d01-002
99mbg991281742.d01-003
99mbg991281742.d01-004
99mbg991281742.d01-005
99mbg991281742.d01-006
99mbg991281742.d01-007
99mbg991281742.d01-008
99mbg991281742.d01-009
```

These line file entries were automatically generated. If you wanted to change one to be something else just edit the file and change "99mbg991281742.d01-007" to be "This line really sucks!".

The Bin File (*<pfm handle file>.bin*)

The bin file at present is actually a directory. The reason for this madness is that Microsoft, which is one of the target OSes, did not support the POSIX standard LARGEFILE64_SOURCE when this software was first written so it couldn't handle files larger than 2GB. In order to overcome this limitation we use a directory instead of a file. The directory contains a file named "config" and one or more files named "hugefile.XXX". The hugefiles will be numbered from 000 up to the max number of files needed for the dataset (at present the max is 128). For all practical purposes the directory can be considered to be a file (and will actually be a file whenever we move to version 6). The data in each bin record in the bin file includes minimum filtered depth, maximum filtered depth, average filtered depth, minimum depth, maximum depth, average depth, number of soundings, standard deviation, a validity field, and (optionally) up to ten attributes. The average filtered depth field may be replaced with some other surface (such as a minimum curvature spline interpolated surface (MISP)) but the name of the average filtered surface ([AVERAGE FILTERED NAME] or [AVERAGE EDITED NAME]) must be changed in the bin header so that the library doesn't automatically try to insert the average value on recompute. The header of the bin file contains

descriptive parameters for the bin file. This is a sample bin file header:

```
[VERSION] = PFM Software - PFM I/O library V5.39 - 09/16/11
[RECORD LENGTH] = 6
[DATE] = Tue Oct 18 15:20:44 2011
[CLASSIFICATION] = UNCLASSIFIED
[MIN Y] = 30.162500000
[MIN X] = -88.759722222
[MAX Y] = 30.169446352
[MAX X] = -88.744440674
[BIN SIZE XY] = 2.000102817037066
[X BIN SIZE] = 0.000020762972994
[Y BIN SIZE] = 0.000018042473833
[BIN WIDTH] = 736
[BIN HEIGHT] = 385
[MIN FILTERED DEPTH] = 6.600006
[MAX FILTERED DEPTH] = 20.349976
[MIN FILTERED COORD] = 452,140
[MAX FILTERED COORD] = 417,130
[MIN DEPTH] = 6.600006
[MAX DEPTH] = 20.349976
[MIN COORD] = 452,140
[MAX COORD] = 417,130
[COUNT BITS] = 24
[STD BITS] = 32
[STD SCALE] = 1000000.000000
[DEPTH BITS] = 18
[DEPTH SCALE] = 100.000000
[DEPTH OFFSET] = 500.000000
[RECORD POINTER BITS] = 38
[FILE NUMBER BITS] = 13
[LINE NUMBER BITS] = 15
[PING NUMBER BITS] = 31
[BEOAM NUMBER BITS] = 16
[OFFSET BITS] = 10
[VALIDITY BITS] = 18
[POINT] = 30.169444444,-88.759722222
[POINT] = 30.162500000,-88.759722222
[POINT] = 30.162500000,-88.744444444
[POINT] = 30.169444444,-88.744444444
[MINIMUM BIN COUNT] = 1
[MAKIMUM BIN COUNT] = 607
[MIN COUNT COORD] = 502,11
[MAK COUNT COORD] = 583,124
[MIN STANDARD DEVIATION] = 0.000000
[MAK STANDARD DEVIATION] = 2.874570
[CHART SCALE] = 0.000000
[CLASS TYPE] = 0
[PROJECTION] = 0
[PROJECTION ZONE] = 0
[HEMISPHERE] = 0
[PROJECTION PARAMETER 0] = 0.000
[PROJECTION PARAMETER 1] = 0.000
[PROJECTION PARAMETER 2] = 0.000000000000000
[PROJECTION PARAMETER 3] = 0.000000000000000
[PROJECTION PARAMETER 4] = 0.000000000000000
[PROJECTION PARAMETER 5] = 0.000000000000000
```

[PROJECTION PARAMETER 6] = 0.000000000000
[PROJECTION PARAMETER 7] = 0.000000000000
[PROJECTION PARAMETER 8] = 0.000000000000
[PROJECTION PARAMETER 9] = 0.000000000000
[PROJECTION PARAMETER 10] = 0.000000000000
[PROJECTION PARAMETER 11] = 0.000000000000
[PROJECTION PARAMETER 12] = 0.000000000000
[PROJECTION PARAMETER 13] = 0.000000000000
[PROJECTION PARAMETER 14] = 0.000000000000
[PROJECTION PARAMETER 15] = 0.000000000000
[AVERAGE FILTERED NAME] = AVERAGE MISP SURFACE
[AVERAGE NAME] = Average Depth
[DYNAMIC RELOAD] = 1
[NUMBER OF BIN ATTRIBUTES] = 6
[BIN ATTRIBUTE 0] = ###0
[BIN ATTRIBUTE 1] = ###1
[BIN ATTRIBUTE 2] = ###2
[BIN ATTRIBUTE 3] = ###3
[BIN ATTRIBUTE 4] = ###4
[BIN ATTRIBUTE 5] = ###5
[BIN ATTRIBUTE 6] =
[BIN ATTRIBUTE 7] =
[BIN ATTRIBUTE 8] =
[BIN ATTRIBUTE 9] =
[BIN ATTRIBUTE OFFSET 0] = 0.000000
[BIN ATTRIBUTE OFFSET 1] = 0.000000
[BIN ATTRIBUTE OFFSET 2] = 0.000000
[BIN ATTRIBUTE OFFSET 3] = 0.000000
[BIN ATTRIBUTE OFFSET 4] = 0.000000
[BIN ATTRIBUTE OFFSET 5] = 0.000000
[BIN ATTRIBUTE OFFSET 6] = 0.000000
[BIN ATTRIBUTE OFFSET 7] = 0.000000
[BIN ATTRIBUTE OFFSET 8] = 0.000000
[BIN ATTRIBUTE OFFSET 9] = 0.000000
[BIN ATTRIBUTE MAX 0] = 100.000000
[BIN ATTRIBUTE MAX 1] = 100.000000
[BIN ATTRIBUTE MAX 2] = 100.000000
[BIN ATTRIBUTE MAX 3] = 32000.000000
[BIN ATTRIBUTE MAX 4] = 100.000000
[BIN ATTRIBUTE MAX 5] = 100.000000
[BIN ATTRIBUTE MAX 6] = 0.000000
[BIN ATTRIBUTE MAX 7] = 0.000000
[BIN ATTRIBUTE MAX 8] = 0.000000
[BIN ATTRIBUTE MAX 9] = 0.000000
[BIN ATTRIBUTE NULL 0] = 101.000000
[BIN ATTRIBUTE NULL 1] = 101.000000
[BIN ATTRIBUTE NULL 2] = 101.000000
[BIN ATTRIBUTE NULL 3] = 32001.000000
[BIN ATTRIBUTE NULL 4] = 101.000000
[BIN ATTRIBUTE NULL 5] = 101.000000
[BIN ATTRIBUTE NULL 6] = 0.000000
[BIN ATTRIBUTE NULL 7] = 0.000000
[BIN ATTRIBUTE NULL 8] = 0.000000
[BIN ATTRIBUTE NULL 9] = 0.000000
[MINIMUM BIN ATTRIBUTE 0] = 0.000000
[MAXIMUM BIN ATTRIBUTE 0] = 100.000000
[MINIMUM BIN ATTRIBUTE 1] = 0.000000
[MAXIMUM BIN ATTRIBUTE 1] = 100.000000
[MINIMUM BIN ATTRIBUTE 2] = 0.000000
[MAXIMUM BIN ATTRIBUTE 2] = 100.000000

[MINIMUM BIN ATTRIBUTE 3] = 0.000000
[MAXIMUM BIN ATTRIBUTE 3] = 32000.000000
[MINIMUM BIN ATTRIBUTE 4] = 0.000000
[MAXIMUM BIN ATTRIBUTE 4] = 100.000000
[MINIMUM BIN ATTRIBUTE 5] = 0.000000
[MAXIMUM BIN ATTRIBUTE 5] = 100.000000
[MINIMUM BIN ATTRIBUTE 6] = 0.000000
[MAXIMUM BIN ATTRIBUTE 6] = 0.000000
[MINIMUM BIN ATTRIBUTE 7] = 0.000000
[MAXIMUM BIN ATTRIBUTE 7] = 0.000000
[MINIMUM BIN ATTRIBUTE 8] = 0.000000
[MAXIMUM BIN ATTRIBUTE 8] = 0.000000
[MINIMUM BIN ATTRIBUTE 9] = 0.000000
[MAXIMUM BIN ATTRIBUTE 9] = 0.000000
[BIN ATTRIBUTE BITS 0] = 7
[BIN ATTRIBUTE BITS 1] = 14
[BIN ATTRIBUTE BITS 2] = 14
[BIN ATTRIBUTE BITS 3] = 15
[BIN ATTRIBUTE BITS 4] = 14
[BIN ATTRIBUTE BITS 5] = 14
[BIN ATTRIBUTE BITS 6] = 0
[BIN ATTRIBUTE BITS 7] = 0
[BIN ATTRIBUTE BITS 8] = 0
[BIN ATTRIBUTE BITS 9] = 0
[BIN ATTRIBUTE SCALE 0] = 1.000000
[BIN ATTRIBUTE SCALE 1] = 100.000000
[BIN ATTRIBUTE SCALE 2] = 100.000000
[BIN ATTRIBUTE SCALE 3] = 1.000000
[BIN ATTRIBUTE SCALE 4] = 100.000000
[BIN ATTRIBUTE SCALE 5] = 100.000000
[BIN ATTRIBUTE SCALE 6] = 0.000000
[BIN ATTRIBUTE SCALE 7] = 0.000000
[BIN ATTRIBUTE SCALE 8] = 0.000000
[BIN ATTRIBUTE SCALE 9] = 0.000000
[NUMBER OF NDX ATTRIBUTES] = 1
[NDX ATTRIBUTE 0] = Time (POSIX minutes)
[NDX ATTRIBUTE 1] =
[NDX ATTRIBUTE 2] =
[NDX ATTRIBUTE 3] =
[NDX ATTRIBUTE 4] =
[NDX ATTRIBUTE 5] =
[NDX ATTRIBUTE 6] =
[NDX ATTRIBUTE 7] =
[NDX ATTRIBUTE 8] =
[NDX ATTRIBUTE 9] =
[MINIMUM NDX ATTRIBUTE 0] = -64000000.000000
[MAXIMUM NDX ATTRIBUTE 0] = 64000000.000000
[MINIMUM NDX ATTRIBUTE 1] = 0.000000
[MAXIMUM NDX ATTRIBUTE 1] = 0.000000
[MINIMUM NDX ATTRIBUTE 2] = 0.000000
[MAXIMUM NDX ATTRIBUTE 2] = 0.000000
[MINIMUM NDX ATTRIBUTE 3] = 0.000000
[MAXIMUM NDX ATTRIBUTE 3] = 0.000000
[MINIMUM NDX ATTRIBUTE 4] = 0.000000
[MAXIMUM NDX ATTRIBUTE 4] = 0.000000
[MINIMUM NDX ATTRIBUTE 5] = 0.000000
[MAXIMUM NDX ATTRIBUTE 5] = 0.000000
[MINIMUM NDX ATTRIBUTE 6] = 0.000000
[MAXIMUM NDX ATTRIBUTE 6] = 0.000000
[MINIMUM NDX ATTRIBUTE 7] = 0.000000


```

[MAXIMUM NDX ATTRIBUTE 7] = 0.000000
[MINIMUM NDX ATTRIBUTE 8] = 0.000000
[MAXIMUM NDX ATTRIBUTE 8] = 0.000000
[MINIMUM NDX ATTRIBUTE 9] = 0.000000
[MAXIMUM NDX ATTRIBUTE 9] = 0.000000
[NDX ATTRIBUTE BITS 0] = 27
[NDX ATTRIBUTE BITS 1] = 0
[NDX ATTRIBUTE BITS 2] = 0
[NDX ATTRIBUTE BITS 3] = 0
[NDX ATTRIBUTE BITS 4] = 0
[NDX ATTRIBUTE BITS 5] = 0
[NDX ATTRIBUTE BITS 6] = 0
[NDX ATTRIBUTE BITS 7] = 0
[NDX ATTRIBUTE BITS 8] = 0
[NDX ATTRIBUTE BITS 9] = 0
[NDX ATTRIBUTE SCALE 0] = 1.000000
[NDX ATTRIBUTE SCALE 1] = 0.000000
[NDX ATTRIBUTE SCALE 2] = 0.000000
[NDX ATTRIBUTE SCALE 3] = 0.000000
[NDX ATTRIBUTE SCALE 4] = 0.000000
[NDX ATTRIBUTE SCALE 5] = 0.000000
[NDX ATTRIBUTE SCALE 6] = 0.000000
[NDX ATTRIBUTE SCALE 7] = 0.000000
[NDX ATTRIBUTE SCALE 8] = 0.000000
[NDX ATTRIBUTE SCALE 9] = 0.000000
[USER FLAG 1 NAME] = PFM_USER_01
[USER FLAG 2 NAME] = PFM_USER_02
[USER FLAG 3 NAME] = PFM_USER_03
[USER FLAG 4 NAME] = PFM_USER_04
[USER FLAG 5 NAME] = PFM_USER_05
[COVERAGE MAP ADDRESS] = 11917504
[HORIZONTAL ERROR BITS] = 12
[HORIZONTAL ERROR SCALE] = 100.000000
[MAXIMUM HORIZONTAL ERROR] = 20.000000
[VERTICAL ERROR BITS] = 14
[VERTICAL ERROR SCALE] = 100.000000
[MAXIMUM VERTICAL ERROR] = 100.000000
[NULL DEPTH] = 1001.000000

```

The Index File (*<pfm handle file>.ndx*)

The index file contains all of the input data points. This is the file that actually gets edited (either automatically or manually). Each physical record in the index file contains [RECORD LENGTH] logical records and a continuation (overflow) pointer. Each logical record contains the file number, line number, ping/record number, beam/sub-record number, Z value (depth/elevation), horizontal uncertainty, vertical uncertainty, x and y position (projected or unprojected), status, and (optionally) up to ten attribute values. All values are stored as bit-packed, scaled, unsigned integers to save space. In addition, some incidental compression is being done. For example, the x and y positions are stored as offsets from the lower left corner of the bin in units of 1/4095th of the bin size. This allows us to store the position with ridiculous resolution in 24 bits. For comparison, most systems store positions in two 64 bit floating point values.