Description of Greenland Data provided by Sigma Space on 13 January 2010

The attached files contain heavily processed data yielding the surface of Greenland's glaciers over the regions flown in two of three successful flights. This document describes the data products and processing that was used.

The glacier surface was extracted from the full data set in a sequence of steps. These steps were designed to provide a surface as clean of solar noise events as possible so that simulations of space like returns would be as representative of the surface topography as possible. Therefore, if in doubt, the algorithms removed points. An examination of the removed points show that a significant number, perhaps 25% or more, of true surface returns were deleted. Therefore, the attached data must not be used to evaluate the quality or quantity of data gathered during Sigma's flights over Greenland. Refinements to the processing will permit retention of a higher fraction of surface returns without sacrificing solar noise rejection. Fortunately, the density of data collection is so high that in most cases, the absence of the missing data is not noticeable.

The data are in two top folders, one for flight 6 and one for flight 7. Flight 8 data is not included as we are not yet satisfied with the processing. Inside each top folder are folders named for the local date and time at which processing of the contained data was started. Each sub-folder contains two text files (.txt), a '.dat' file, and up to 2000 lidar data files with a .bin extension. One text file and the '.dat' file contain summary information that describes the data stored in the '.bin' files. The 'ReadMe.txt' file contains the ASCII header stored in each '.bin' file. The two summary files, one comma delimitated text for import into a spreadsheet and the other in single precision floats for reading by a program, contain information about the plane's position, the average ground elevation, and number of lidar returns kept, and etc. The name of each file summary file matches the folder name.

The lidar data is stored in files with a .bin extension and named for the GPS second during which the data was gathered. The files contain 1024 byte header followed by single precision floating point numbers organized as (x, y, z) position triplets. The description of the points is shown in the table at right. All values are in meters. To convert the (x, y) values to UTM add the x and y values to the reference point value.

First value (x)	Easting offset from reference point
Second value (y)	Northing offset from reference point
Third value (z)	Height above GPS elipsoid

For all of the Greenland data, a single reference point (550000, 7675000) about 50 km east of Ilulissat was selected near the center of the flight pattern. The flights extended about 115 km north and south and less than 90 km east and west of this points. Tests have shown that single precision floats can represent this distance with sub-centimeter resolution. Because of the fixed reference point, unless UTM position is desired, the user does not need to use the offset.

The 1024 byte header contains 1000 ASCII characters followed by three double precision (8 byte) floats. The first line of the ASCII header contains the easting (x) and northing (y) reference values. These values are repeated as double precision floats at byte offsets 1000 and

1008 so that a computer doesn't need to interpret the ASCII offset values. The lines following the first text line contain values that were used to control the processing. They are not needed to display or use the surface data.

Name of Data	Description
Time	GPS second
East	Easting position of reference point
North	Northing position of reference point
Num	Number of data values in the second,
(x, y, z) plane	Average plane's offset during the second
(x, y, z) x_min	x, y, z, offset of the point with the smallest x value.
(x, y, z) x_max	x, y, z, offset of the point with the largest x value.
(x, y, z) y_min	x, y, z, offset of the point with the smallest y value.
(x, y, z) y_max	x, y, z, offset of the point with the largest y value.
(x, y, z) z_min	x, y, z, offset of the point with the smallest z value.
(x, y, z) z_max	x, y, z, offset of the point with the largest z value.
(x, y, z) av	Average x, y, z offset of the point cloud
(x_sd, y_sd, z_sd)	Standard deviation of the x, y, and z point cloud values

The text summary file stores values as comma delimited text while the binary file stores the data as single precision floats. Each file contains the same data arranged in records of 31 values in the same order. Records in the text file are not terminated by a return. Values shown in the table as (x, y, z) represent three data values.

To process the data, data was gathered in 1 second groups. The data was clipped to the limits of the range gate and then geo-located. The data was than gathered into 64 shots groups with an approximately square footprint. The altitude data for each group was histogrammed into bins 4 meter vertical height. The mean number of returns in bins in the top (elevation) one third, the middle third, and the lowest third was calculated. While it is not possible to know in which third the ground appears, we can assume that it does not appear in all three. So the third with the lowest mean counts was assume to contain only sky noise. Based on the mean sky count, a low and a high thresholds were established. All altitude bins with more than the high threshold of counts were assumed to contain valid ground return data. Altitude bins with more than the low threshold were assumed to contain valid ground return data if they were adjacent to a bin that exceeded the high threshold. In this way, extremes of hills are kept, even if they generate only a few returns.

The above process is able to remove points that are outside the histogram bins that contain the surface. However, it is unable to remove noise points that are in the same *z* bin as the surface. To attempt to remove these, a routine sensitive to local topography was used. Using a Delaunay triangulation of the (x, y) locations in a file of 1 second data, each event's nearest neighbors were identified. Then all of the local extreme points, defined as points either lower than or higher than all of its nearest neighbors, are identified. Ideally, these points would be tested to see if they are somehow out of line with the nearest neighbors; that is well above or below. However, for this processing, applying the premise that when in doubt remove the point, all extreme points were removed. In the glacier data, about 25% of the points are extreme. While this represents a significant fraction of the points remaining from the earlier processing, many of these points are truly noise and there are still a high density of surface returns that describe the surface and from which to sample for space like simulations.