Airborne Topographic Mapper HDF5 Waveform Data Product

User Guide

Version 0.6

July 10, 2017

1) Introduction

The Airborne Topographic Mapper (ATM) waveform data product is a Level-1B data product that is compliant with NASA's Earth Science Data and Information System (ESDIS) approved standards for use in NASA Earth science data systems (https://earthdata.nasa.gov/user-resources/standards-and-references). In order to promote the use of ATM airborne laser altimetry data with science data products from satellite missions, such as ICESat-2, the ATM waveform data product is similar in format and structure to ICESat-2's ATLAS Science Algorithm Software Standard Data Product (Lee et al., 2016). It is implemented using the HDF5 standard as defined by the HDFGroup (http://www.hdfgroup.org).

2) Theory of Measurements

A laser altimeter measures range from the instrument to a target by measuring the elapsed time between emission of a laser pulse and detection of laser energy reflected by the target surface. Range to the target is calculated as half the elapsed emission/return time multiplied by the speed of light. Target range is converted to geographic position by integration with platform GPS and attitude or Inertial Measurement Unit (IMU) information.

3) Data Acquisition Methods

The ATM instrument package includes suites of LiDAR, Global Positioning System (GPS) and attitude measurement subsystems. The instrument package is installed onboard the aircraft platform and calibrated during ground testing procedures. Installation mounting offsets, the distances between GPS and attitude sensors and the ATM LiDARs, are measured using surveying equipment. One or more ground survey targets, usually aircraft parking ramps, are selected and surveyed on the ground using differential GPS techniques. Prior to missions, one or more GPS ground stations are established by acquiring low rate GPS data over long time spans. Approximately one hour prior to missions both the GPS ground station and aircraft systems begin data acquisition. During the aircraft flight, the ATM instrument suite acquires LiDAR, GPS and attitude sensor data over selected targets, including several passes at differing altitudes over the selected ground survey calibration sites. The aircraft and ground systems continue to acquire data one hour post-mission. Instrument parameters estimated from the surveys of calibration sites are used for post-flight calculation of laser footprint locations. These parameters are later refined using inter-comparison and analysis of ATM data where flight lines cross or overlap. A more detailed description can be found in Martin *et al.*, 2012.

4) Derivation Techniques and Algorithms

Each ATM surface elevation measurement corresponds to one laser pulse. The measurements have not been re-sampled. The transmitted laser pulse and the received backscatter pulse from the ground surface are photodetected and captured by a waveform digitizer. Post-flight processing of the waveforms yields the time of flight between transmitted and received signals. This time of flight value is converted to a distance compensated for speed of light through atmosphere. The scan azimuth of the LiDAR scanner mirror together with the aircraft attitude determine the pointing angle of the LiDAR. GPS aircraft position, pointing angle of the LiDAR, and range measured by the LiDAR are used to compute position of laser footprint on the ground.

5) Trajectory and Attitude Data

Aircraft position is determined by Global Navigation Satellite System (GNSS) systems that incorporate NAVSTAR Global Positioning System (GPS) and, for later campaigns, the Globalnaya Navigatsionnaya Sputnikovaya Sistema (GLONASS). Carrier phase measurements are logged by an antenna and receiver on the aircraft. In post-flight processing, these measurements are combined with similar measurements from static ground stations to produce a kinematic differential solution of the aircraft trajectory at 0.5 second intervals, and more recently at 0.1 second intervals.

Aircraft attitude is logged from a commercial Inertial Navigation System (INS), also known as an Inertial Measurement Unit (IMU).

6) Overview of the TX and RX Range Gate Structure

The analog output from the optical detector is captured by an 8-bit waveform digitizer sampling at a constant rate (2 or 4 gigasamples per second). A sequence of samples, or "range bins", is recorded whenever the signal amplitude exceeds a programmable trigger threshold. Each sequence, or "range gate", can contain a variable number of range bins depending on how long the signal exceeds the threshold. Each laser shot generates a laser waveform record, which can contain multiple range gates, each of which contains multiple waveform samples. The laser record contains the starting position of each range gate, from which the time of each range bin can be determined. For example, Figure 1 shows a laser waveform reconstructed from a laser record containing 6 range gates: one for the transmitted pulse (TX), and four return gates (RX) from a complex target (trees). Range gate #6 contains two distinct return pulses.

The transmitted laser pulse travels through an optical window in the nadir view port on the aircraft to the target. The reflection of the transmitted laser pulse on the optical window can exceed the amplitude trigger threshold and is then recorded in a range gate. In order to separate the recorded transmit pulse from the window reflection, the transmit pulse is routed through an optical delay fiber several meters in length so it appears several tens of nanoseconds after the window reflection. The length of the fiber can change with various system configurations and its delay is incorporated in the calibration for range determination. Since the window reflection does not occur on every laser shot, sometimes the recorded transmit pulse is in range gate #1 or range gate #2. Figure 1 shows an example that includes a window reflection. Therefore the recorded transmit pulse that is used for the ATM range determination is in range gate #2.

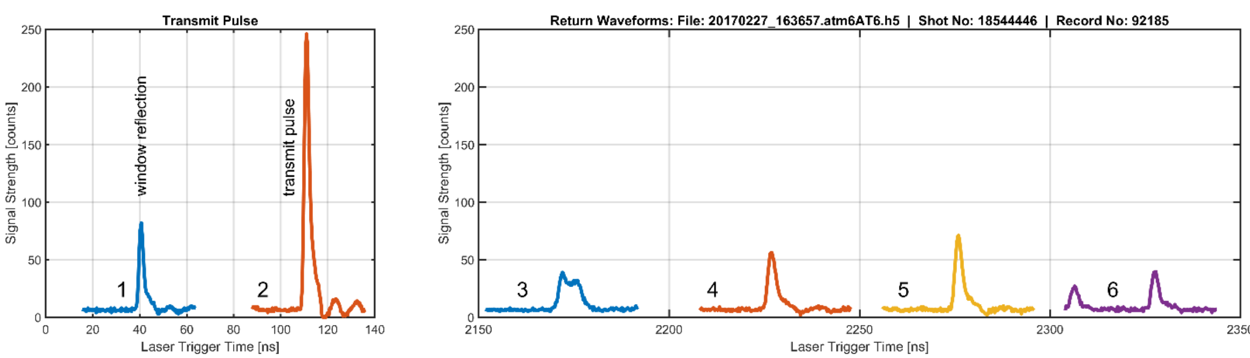

__Figure 1:__ Example of how ATM waveform data are organized into range gates. The example shows a laser shot over a complex target (tree) that has triggered 4 return range gates (#3 - #6) some of which contain multiple return pulses. The transmit pulse is recorded through a delay fiber (range gate #2) and is sometimes preceded by a window reflection (range gate #1).

The capability of recording multiple range gates of varying lengths for each laser shot requires a pointer and indexing scheme to access the waveform data within an HDF5 file that will be described below.

7) Range Determination

The transmit and receive waveforms are captured in separate range gates. The time delay between transmitted and received waveforms includes the delay between the TX and RX range gates and the "tracked" location of the pulse within each gate.

When the laser fires, an electronic trigger starts the digitizer counting the number of elapsed sampling intervals. The gate start position is the value of this counter at the first bin of the gate. The time relative to the trigger, or "trigger time", of any range bin can be computed as the gate start position plus the bin position within the gate, multiplied by the digitizer sampling interval. The time between any range bins in the two gates can be accurately computed as the difference between the two trigger times.

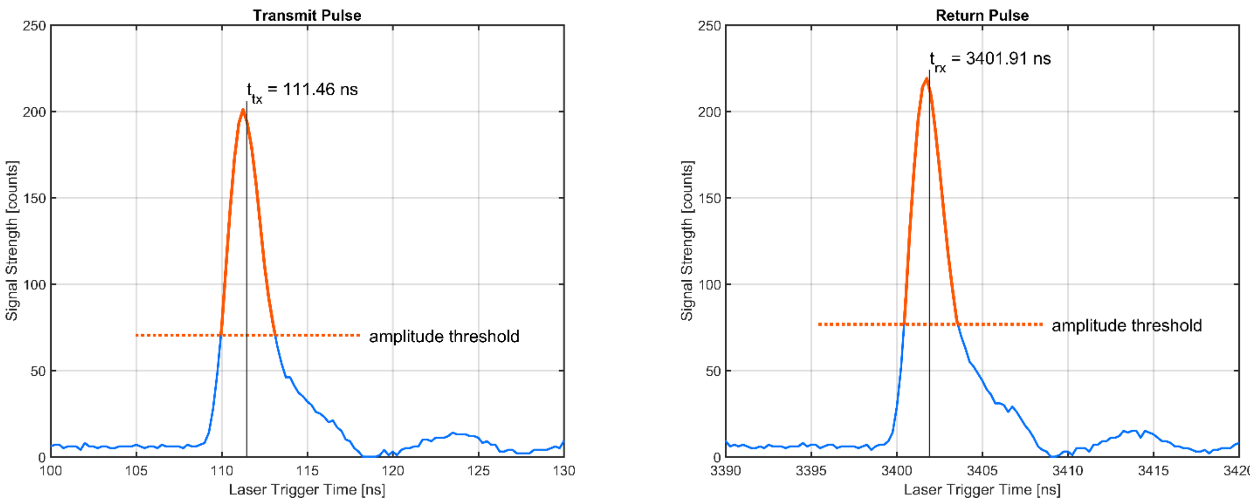

Figure 2: *Example of range determination using a centroid estimate that cuts off data points below 35% of the maximum amplitude of the transmit and return pulse.*

A more complex task is the assignment of a sample time to a waveform. Figure 2 shows the two windows and the current ATM method for calculating the range measurement between the two windows. The ATM range determination is using a centroid estimate that cuts off all values below 35% of the maximum amplitude of the transmit or return pulse. The uncalibrated range between two positions in the TX and RX windows, where c is the speed of the propagation of light through the atmosphere is:

$$range_{uncalibrated}\ [m] = \frac{1}{2} * c * (t_{rx} - t_{tx})$$

The example shown in Figure 2 yields an uncalibrated range of 493.22 meters. As previously described the optical delay fiber and other system components introduce a range bias that is determined in ground tests by shooting the laser at a target at a known distance (Figure 3).

Figure 3: *Range bias determination (a.k.a. ground test) using a calibration target with a known distance. The distance to the calibration target is measured with an electronic distance meter (a.k.a. "total station") with an accuracy of a few mm.*

The range bias is a function of the return signal strength. For the ATM centroid tracker, the bias is almost constant for typical signal strengths encountered in flight, but deviates for weak signals that barely exceed the amplitude trigger threshold and for strong signals that saturate the digitizer. The range bias calibration determined from ground tests depends on the waveform tracking method. Users that wish to determine ranges using a different tracking method will need to develop a suitable calibration from the groundtest data.

8) File Structure and Organization

The ATM waveform product follows the HDF convention of organizing data within groups and subgroups. Some groups contain information for each laser pulse: `/footprint` contains the location of the laser spot on the ground; `/aircraft` contains the aircraft's location and attitude interpolated to the times of the laser shots; `/laser` contains pointing and range information for the laser; and `/time` contains the UTC time for each laser pulse.

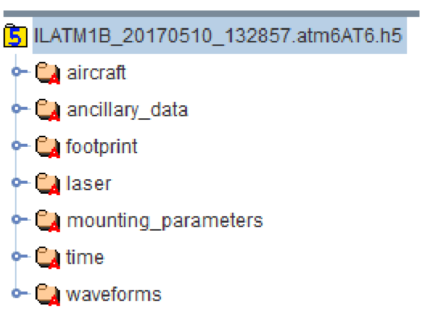

Figure 4: *ATM HDF5 group structure.*

Other groups contain documentation and parameters necessary to use the data in a particular subgroup, or information related to the file as a whole: `/ancillary_data` contains the spatial and temporal limits for the file etc.; `/mounting_parameters` contains information used for computing the footprint location from the laser and aircraft information. For example: `/ancillary_data/epoch_seconds_of_day` would be used to convert `/time/seconds_of_day` into absolute time tags.

Level 1B QFIT Geolocated Spot Elevation Measurements

The primary science data product are the geolocated and filtered spot elevation measurements in the group `/footprint`. These elevations correspond to the separate Level 1B QFIT data product that is available at NSIDC (https://nsidc.org/data/docs/daac/icebridge/ilatm1b/). Data are organized in chronological order. For example: denoting the number of laser shots contained in the file as *N*, the timestamps of the laser shots would be an array of length *N* contained in `/time/seconds_of_day`, and the corresponding measured elevations would be an array of length *N* contained in `/footprint/elevation`.

Waveform Data

Waveform data for surveys beginning with Antarctica 2014 are stored in the subgroup `/waveforms/twv`. As mentioned previously (Section 6) each laser shot can be associated with a varying number of range gates that can also vary in length (number of digitizer samples) (Figure 5). Some elements (*e.g.* `sample_interval`) are single values; some are arrays of *N* values (*e.g.* `shot_number`, others starting with "`shot_...`") corresponding to each laser shot; some are longer arrays (*e.g.* `gate_position`, others starting with "`gate_...`") corresponding to each gate; and the largest array (`wvfm_amplitude`) contains the 8-bit digitizer samples as a concatenation of all the waveform gates recorded in the file.

The link between laser shots, range gates, and digitizer samples/range bins is implemented by a pointer/index scheme. The waveform data for a particular laser shot is found by locating the gates recorded for the laser shot, then locating the waveform samples associated with those gates. Consider a laser shot *j, (1<j<N)*. Using the conventions that ***a****(j)* is the *j*-th element of array ***a*** and ***a****(i:j)* is the sub-array taken from elements *i* through *j*, the time of the shot is given by `shot_seconds_of_day(j)`. The number of gates associated with this laser shot is `shot_gate_count(j)`. Data for the first gate associated with this shot is located in the gate arrays at index `shot_gate_start`(*j*). The first waveform sample for this first gate of this laser shot is located at gate_wvfm_start(shot_gate_start(j)). The value of this first waveform sample is wvfm_amplitude(gate_wvfm_start(shot_gate_start(j))). The number of samples in each gate is stored in the array `gate_wvfm_length`. Therefore the waveform recorded in the first gate of shot j can be read as:

> `wvfm_amplitude`(*k2:k3*)
>
> where *k*=`shot_gate_start`(*j*)
>
> *k2*=`gate_wvfm_start`(*k*)
>
> *k3*=`gate_wvfm_start`(*k*)+`gate_wvfm_length`(*k*)-1

More generally, the digitized waveform for range gate *i* of the laser shot *j* would be:

> `wvfm_amplitude`(*k2:k3*)
>
> Where *k*=`shot_gate_start`(j)
>
> *k2*=`gate_wvfm_start`(*k+i-1*)
>
> k3=`gate_wvfm_start`(*k+i-1*)+`gate_wvfm_length`(*k+i-1*)-1

In order to reassemble all range gates into a time tagged series the offset for the first range bin/sample needs to be known. This information is stored in the field `gate_position` as number of digitizer samples since the laser was triggered. Together with the length of each range bin/sample in seconds (`sampleRate = 2.5e-10 s` for 4 Giga samples per second digitization rate) the time in seconds can be calculated using `gate_position*sampleRate`. In this way, the range gates for a laser shot can be reassembled in order to determine the time of flight between transmit and receive pulse as shown in Figure 1 and Figure 2.

Figure 5 illustrates the indexing scheme using values from an example data file. If the first range gate within a file starts at index 1 (`gate_wvfm_start(1) = 1`) and is 192 range bins or samples long (`gate_wvfm_length(1) = 192`) the second range gate will begin at index 193 (`gate_wvfm_start(2) = 193`).

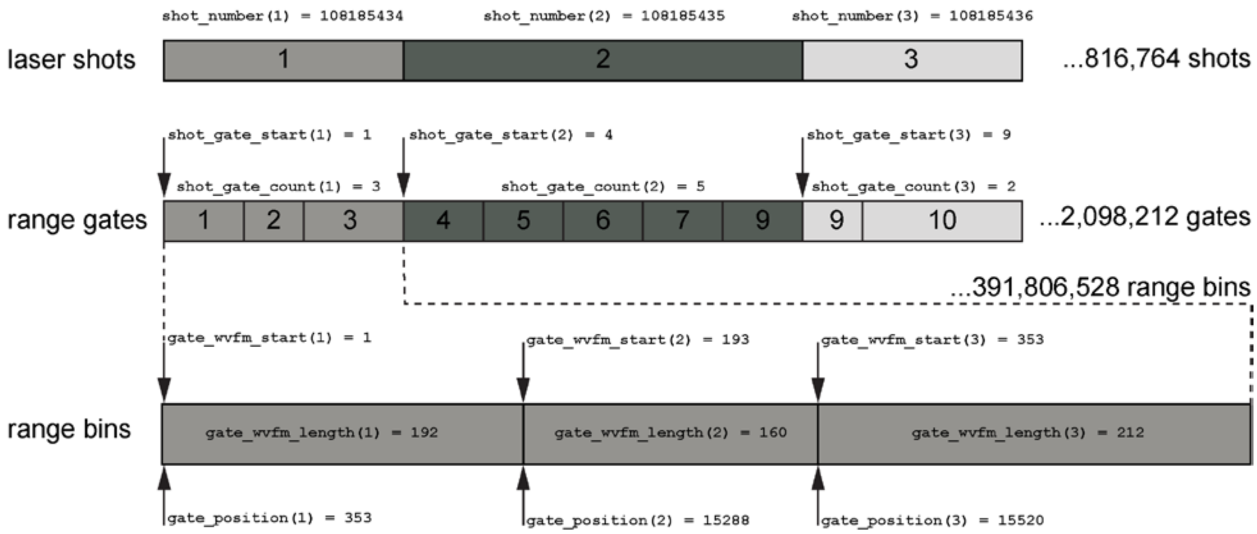

***Figure 5:** Pointer and indexing schema for access of the waveform data (range gates) for a particular laser shot. Range gate numbers are also referred to as record numbers.*

The example file contains 816,764 individual laser shots whose unique shot identifiers are stored in the field `shot_number`. The start index for the first range gate for each shot (`shot_gate_start`) and the number of range gates for each shot (`shot_gate_count`) have the same size as `shot_numbers`. Together the 816,764 individual laser shots contain a total of 2,098,212 range gates, that are comprised of 391,806,528 digitizer samples.

References

Martin, C. F., W. B. Krabill, S. S. Manizade, R. L. Russell, J. G. Sonntag, R. N. Swift, and J. K. Yungel. 2012. Airborne Topographic Mapper Calibration Procedures and Accuracy Assessment. Greenbelt, Md.: National Aeronautics and Space Administration, Goddard Space Flight Center. *NASA Technical Memorandum 2012-215891*. http://ntrs.nasa.gov/archive/nasa/casi.ntrs.nasa.gov/20120008479_2012008321.pdf

Software Tools for Reading ATM HDF5 Data

HDF5 is a stable data format with broad support in the scientific community. There are many free software tools available such as HDFView (https://support.hdfgroup.org/products/java/hdfview/) for browsing, editing and viewing HDF5 files. Common programming languages such as MATLAB® and IDL include native HDF5 support. Open source software libraries for reading and writing HDF5 files are freely available for C, C++, Fortran and Python.

This section contains documentation and user guides for software tools developed by members of the ATM team for reading and viewing ATM HDF5 waveform data. The source code and documentation will be available from the ATM website at https://atm.wff.nasa.gov/.

9) MATLAB® functions

Several MATLAB® functions are provided that can be used to import, select and plot waveform data and range gates based on search criteria.

9.1 ATM_WVFM_READER

The main MATLAB® function that provides basic capabilities for using ATM HDF5 waveform files within MATLAB® is `atm_wvfm_reader`. The `atm_wvfm_reader` function reads ATM waveform and QFIT data from an ATM HDF5 file and imports the data into a structured array (struct) in MATLAB®. The function allows to extract only waveforms and qfit elevations that satisfy spatial and temporal search criteria, i.e. it can be used for subsetting. Typing `help atm_wvfm_reader` in the MATLAB command window shows the 5 different syntax options and includes an example how to call the function from within MATLAB. In its simplest form the function imports an entire HDF5 file (skipping many groups and subgroups):

```
>> atm_wvfm = atm_wvfm_reader('ILATM1B_20170510_132857.atm6AT6.h5');
```

Possible function calls:

```
>> atm_wvfm = atm_wvfm_reader(f_name_inp);

>> atm_wvfm = atm_wvfm_reader(f_name_inp,verbose);

>> atm_wvfm = atm_wvfm_reader(f_name_inp,verbose,poly_lon,poly_lat);

>> atm_wvfm = atm_wvfm_reader(f_name_inp,verbose,t_start,t_end);

>> atm_wvfm = atm_wvfm_reader(f_name_inp,verbose,poly_lon,poly_lat, t_start, t_end);
```

The m-file of this function must be either in your search path or data directory or the folder needs to be added temporarily to MATLAB's search path using the MATLAB addpath() function. If the commands are not run from the folder that contains the HDF5 data file the filename must include the full path.

The output of `atm_wvfm_reader` is a MATLAB® structure array or "struct" consisting of data containers referred to as "fields". Structs are commonly used for storing heterogeneous data. The code can easily be modify to output a cell array instead of a struct. The downside of both, structs and cell arrays is that the contents of just a subset of an ATM HDF5 file will occupy more than an order of magnitude more memory than the size of the compressed HDF5 file. Saving the output struct into a MATLAB® file using MATLAB®'s data compression will result in significantly larger file sizes than the HDF5 input files and is not recommended. Figure 6 shows the basic structure of the output array.

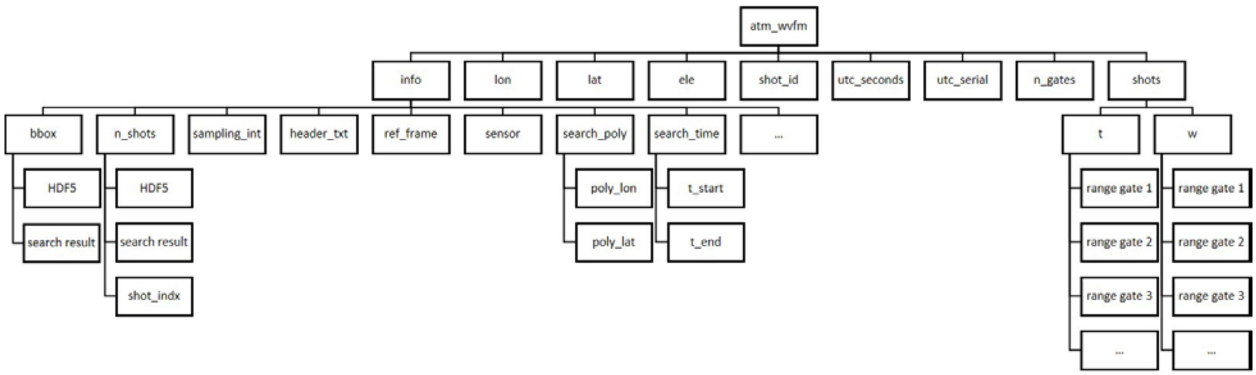

Figure 6: *Contents of output MATLAB® struct. The info field contains metadata about the input file, the search parameters used and the output results. The waveforms are stored in the field "shots" and contain all individual range gates for each shot that has satisfied the spatial and temporal search criteria.*

This basic example shows how to plot the data from all range gates in shot/record #5 in a struct named "atm" using the laser trigger time in nanoseconds as the abscissa:

```
>> shot_nr = 5;
>> figure;
for i = 1:atm.n_gates(shot_nr)
    plot(atm.shots(shot_nr).wf(i).t,atm.shots(shot_nr).wf(i).w,'-'); hold on;
end
>>
```

9.2 ATM_INFO

The MATLAB® function `atm_wvfm_info` can be used to quickly gain information about the contents of an ATM HDF5 waveform file. It takes advantage of HDF5's capability to quickly access only parts of a file. The function returns a polygon with the minimum and maximum longitude and latitude of all laser shots in the file as well as the UTC time for the first and last laser shot. Its primary purpose is to quickly identify HDF5 files for close examination. The function can be called with the following syntax:

```
>>[poly_lon,poly_lat,t_start,t_end,data_fmt] = ATM_WVFM_INFO(f_name_inp,verbose);
```

Typing `help atm_wvfm_info` in the MATLAB® Command Window provides an overview of its functionality.

9.3 PLOT_WVFMS

Also included is a MATLAB® function called `plot_wvfms` that can be used to plot a small number of select waveforms and range gates. Figure 7 shows an example plot of select waveforms over a sea ice pressure ridge north of Ellesmere Island. Here are the commands from the MATLAB® script `make_example_plot.m` that will create the figure:

```
f_name_inp = 'ILATM1B_20170309_133117.atm6AT6.h5';
verbose = 1;
[atm] = atm_wvfm_reader(f_name_inp, verbose);
```

```
shot_list = [548141 549177 549180 549190 549195 549695 561547 562556 562557
562561 563089 563093]; % complex returns over pressure ridge
plt_wvfms(atm,shot_list,200,Inf,'w');
```

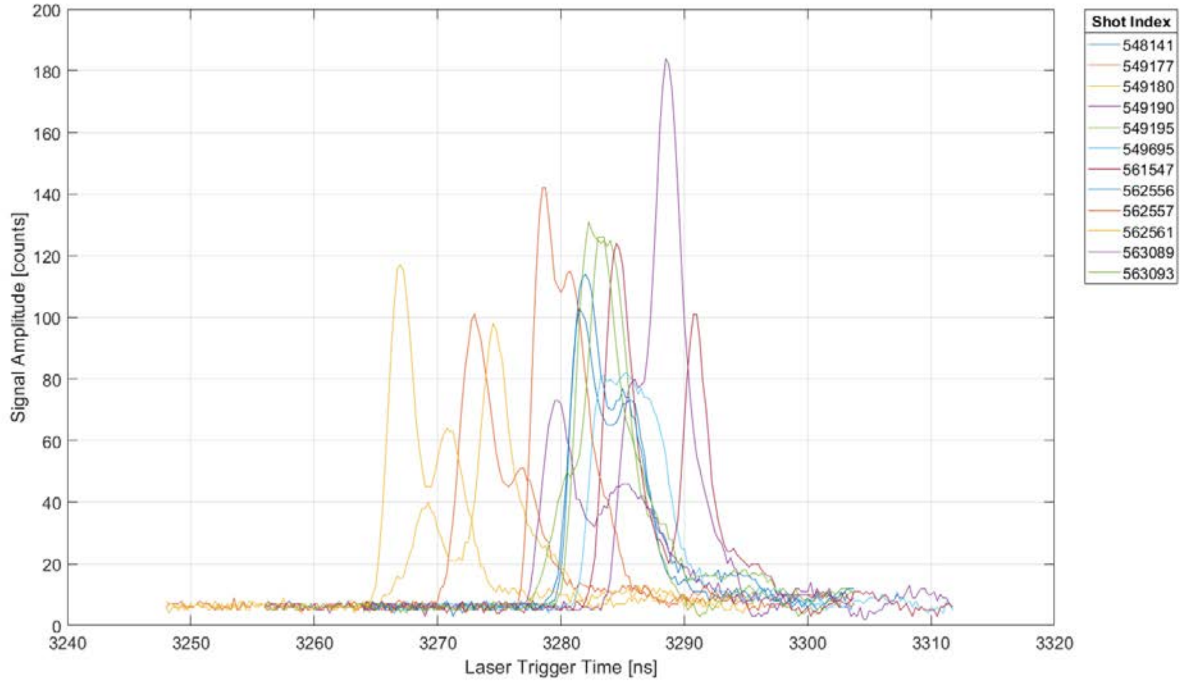

Figure 7: Example plot showing complex waveforms over a pressure ridge north of Ellesmere Island produced with the above MATLAB commands.