OIB Alaska Radar HDF5 Format Description and Processing Information

1 Format

The data are stored in an HDF5 based format with a structure defined in the list below and in Figure 1. Each item in the list and figure is coded by HDF5 object type - group dataset *attribute*. Browse products of the radar data, surface clutter simulation, and airplane track are provided as PNG formatted images. The unit of datasets that have values expressed as physical units are given as a *unit* attribute attached to the dataset. Numeric attributes are stored as compound datatypes where the attribute value is followed by a null termnated ASCII string giving the unit of the attribute value.

- raw
  - rx0 - Raw data acquired by the radar, a real-valued signal. The fast time axis (columns) is two way travel time from the airplane. The data values have not been altered after quantization by the receiver and as such do not have physical units.
    - *numTrace* - Number of traces in the data, will match the number of columns in the dataset.
    - *samplesPerTrace* - Samples per trace in the raw data, will match the number of rows in the dataset.
    - *samplingFrequency* - The fast time (along-trace) sampling frequency.
    - *stacking* - Number of raw traces that are stacked on-board the radar to create each saved trace in rx0.
    - *traceLength* - Time length of each trace in seconds, equal to *samplesPerTrace/samplingFrequency*.
  - tx0 - Information about the transmitted signal.
    - *signal* - The transmitted signal type, will be one of “chirp” (for the IRARES1B data) or “impulse” (for the IRUAFHF1B data).
    - *centerFrequency* - The center frequency of the transmitted signal. Used for all transmitted signal types.
    - *pulseRepetitionFrequency* - The pulse repetition frequency of the transmitted signal. Used for all transmitted signal types. The effective PRF (slow-time sampling rate) is *pulseRepetitionFrequency/stacking*
    - *length* - The length in time of the transmitted signal in seconds. Used only for “chirp” signals.
    - *bandwidth* - The bandwidth of the signal as a percent of its center frequency (0 is 0% and 1 is 100%). Used only for ”chirp” signals. Positive bandwidth is indicative of an up-chirp (low to high frequency sweep) and negative bandwidth indicates a down-chirp (high to low frequency sweep)
  - loc0 - GPS position log from the radar - low position quality and positions are only updated at 1 or 5 Hz. The /ext/nav0 dataset provides high quality positioning information from the GPS system used with the OIB lidar.
    - *CRS* - Coordinate reference system of the data.

- time0 - Time tag for each trace in the raw data.
  - *unit* - The unit used for the time data.
  - *clock* - The standard used for the times.

- drv
  - proc0 - Processed data derived from /raw/rx0. Processing varies based on transmit signal type. The fast time axis (along-trace) is two way travel time from the airplane. Like the rx0 dataset this dataset does not have physical units. The dataset is a complex valued signal. Each sample is a struct with the first item ("r") being the real component of the sample and the second item ("i") being the imaginary component of the sample.
    - *note* - Information about the processing steps applied to create the proc0 dataset.
  - clutter0 - Surface clutter simulation to aid interpretation of the data.
  - pick
    - twtt_surf - Two way travel time to the lidar derived surface in each trace in seconds. No data value is -1. Calculated using /ext/srf0 and /ext/nav0.
      - *unit* - Unit of the two way travel time
    - twtt_bed - Interpreted two way travel time to the bed in each trace in seconds. There are two no data values: -1 indicates that the data has not been interpreted, -9 indicates that the data has been interpreted and there is no observed bed return.
      - *unit* - Unit of the two way travel time
    - thick - Thickness of the glacier in meters, calculated using the two way travel times to the surface and bed: thick = $(c/\sqrt{3.15})\cdot((\text{twtt\_surf} - \text{twtt\_bed})/2)$ where c = 299792458 m/s. The no data values are the same as the twtt_bed no data values.
      - *unit* - Unit of the thickness

- ext
  - nav0 - Positions derived from the GPS used for OIB lidar.
    - *CRS* - Coordinate reference system of the data.
  - srf0 - Surface elevation derived from OIB lidar data in meters.
    - *unit* - Unit of the surface elevations.
    - *verticalDatum* - Vertical datum for the surface elevations.
  - srf0count - Number of lidar points used for each derived surface elevation

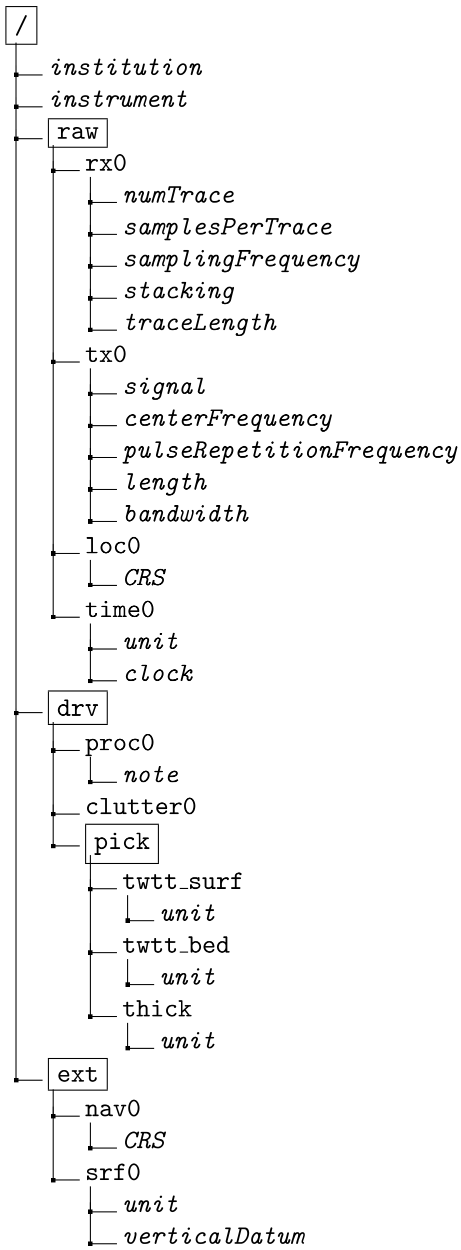

Figure 1: Tree showing structure of HDF5 formatted data files. Each entry is coded by HDF5 object type - group dataset *attribute*

2 Processing

The radar data undergoes the following processing steps to generate the final products:

1. Raw data format to HDF5 conversion
2. Rolling mean removal
3. Pulse compression
4. GPS synchronization
5. LIDAR surface extraction
6. Surface clutter simulation

2.1 Raw data format to HDF5 conversion

Data files recorded by the instrument are converted to initial HDF5 files. Each raw data trace is shifted by a constant offset in order to account for instrument delay so that the first sample in the trace corresponds to when the radar began transmitting. The datasets in the raw group are populated during this step.

2.2 Rolling mean removal

A rolling, windowed mean is subtracted from each trace to remove constant offset and coherent noise. A typical window size is 250 traces, so that trace 500, for example, would have the mean trace of traces 375-625 subtracted from it.

2.3 Pulse compression

All chirped data (Arizona dataset only) is pulse compressed using an ideal reference chirp with a boxcar amplitude window. The processed radar data is saved in the proc0 dataset of the drv group.

2.4 GPS synchronization

Timing data recorded per-trace by the GPS on-board the radar is used to extract high quality positioning data from the GPS record associated with the UAF lidar. The high quality positioning data is saved in the nav0 dataset of the ext group.

2.5 LIDAR surface extraction

A per-trace surface location is extracted from the OIB ILAKS1B data acquired contemporaneously with the radar data during each flight. This is necessary for the chirped data (Arizona dataset) because in most of the data the radar receiver is saturated with the outgoing chirp when the surface reflection is returning to the aircraft due to the low height above ground required by the UAF lidar. For the impulse data (UAF dataset) this extracted surface location is complementary to the surface return normally visible in the data. For each trace the median elevation of all of the lidar points in the first Fresnel zone of the radar is taken to be the surface elevation. The surface elevations are saved in the srf0 dataset of the ext group. A two-way travel time to the surface is calculated using the surface elevations calculated in this step and the nav0 dataset. This two-way travel time is saved in the twtt_surf dataset of the pick group of the drv group.

2.6 Surface clutter simulation

Surface clutter is simulated using a digital elevation model. The importance of considering surface clutter during interpretations and the method for simulating is discussed in more detail in Holt et al. 2006 and Choudhary et al. 2016. This surface clutter simulation is saved in the clutter0 dataset of the drv group. It is meant for direct qualitative comparison with the radar data to help determine if a candidate subsurface reflector is in fact surface clutter.