Full Waveform LiDAR Processing using SPDLib

The Sorted Pulse Data (SPD) format is a HDF5 based format for storing full waveform LiDAR data with a spatial index to allow fast access [1]. Built on top of the SPD format is a library (SPDLib [2]) for converting LiDAR data to/from SPD format and manipulating SPD files to generate common products such as Digital Terrain Models (DTMs), Digital Surface Models (DSMs), Canopy Height Models (CHMs) and other LiDAR metrics. The most recent version of SPDLib (3.3) has the ability to import waveform data from LAS files using LASlib, which is part of LAStools and calculate full waveform metrics.

The aim of this practical is to import full waveform data acquired by NERC-ARF from a Leica ALS50II system in LAS 1.3 format into SPDLib and generate full waveform metrics. The tools within SPDLib are all operated from the command line, which allows easy batch processing.

Import data to SPD format

First open a terminal window and change to the directory containing the data and make a new folder for the SPD files.

```
cd ~/nerc-arf-workshop/lidar_practical/FW10_01-98
mkdir spd
```

Then convert the LAS 1.3 file into an SPD file with no spatial index (UPD):

```
spdtranslate --if LAS --of UPD \
             -x LAST_RETURN \
             --input_proj osgb36.wkt \
             -i las1.3/LDR120726_094707_4.LAS \
             -o spd/LDR120726_094707_4.spd
```

Decompose waveforms

One of the limitations of discrete systems is there are only a given number of 'points' recorded (normally 2 – 4) and the rest of the information is lost. As full waveform data records a digitised version of the received waveform it is possible to extract more returns after data are acquired. A common approach to this is 'Gaussian Decomposition' which involves fitting Gaussian distributions to the peaks, within SPDLib this is available as the `spddecomp` command.

```
spddecomp --all --noise --threshold 25 \
          -i spd/LDR120726_094707_4.spd \
          -o spd/LDR120726_094707_4_decomp.spd
```

This will take a couple of minutes to run through.

Grid data

The SPD format uses a regular horizontal grid to index the data, similar to the structure of a raster. This makes spatial queries on the dataset much faster than an unsorted point cloud. Now the waveforms have been decomposed a gridded SPD file will be generated using:

[1] Bunting, P. J., Armston, J., Lucas, R. M., & Clewley, D. (2013). Sorted pulse data (SPD) library. Part I: A generic file format for LiDAR data from pulsed laser systems in terrestrial environments. Computers & Geosciences, 56(C), 197–206. http://doi.org/10.1016/j.cageo.2013.01.019

[2] Bunting, P. J., Armston, J. D., Clewley, D., & Lucas, R. M. (2013). Sorted pulse data (SPD) library—Part II: A processing framework for LiDAR data from pulsed laser systems in terrestrial environments. Computers & Geosciences.

```
spdtranslate --if SPD --of SPD \
             -x LAST_RETURN \
             -b 2 \
             -t ./LDR120726_094707_4_decomp_tmp \
             -i spd/LDR120726_094707_4_decomp.spd \
             -o spd/LDR120726_094707_4_decomp_grid.spd
```

In the command above, the last return is used to determine which grid cell each pulse is assigned to and a bin size of 2 m is used. As gridding data requires a lot of memory, temporary files are used so the entire dataset doesn't need to be loaded into RAM. After the command has completed the temp files can be removed using:

```
rm -f LDR*tmp*spd
```

Classify ground returns

To derive heights from LiDAR data it is first necessary to determine the ground elevation so heights can be calculated above this. Within SPDLib the ground classification results are achieved using a combination of two classification algorithms: a Progressive Morphology Filter (PMF; [3]) followed by the Multi-Scale Curvature algorithm (MCC; [4]). Both these algorithms use only the discrete points rather than the waveform information.

Progressive Morphology Filter

```
spdpmfgrd --grd 1 -i spd/LDR120726_094707_4_decomp_grid.spd \
          -o spd/LDR120726_094707_4_decomp_grid_pmf_grd.spd
```

Multi-Scale Curvature algorithm

```
spdmccgrd --class 3 --initcurvetol 1 \
          -i spd/LDR120726_094707_4_decomp_grid_pmf_grd.spd \
          -o spd/LDR120726_094707_4_decomp_grid_pmf_mcc_grd.spd
```

Attribute with height

The final step is to attribute each pulse with heights above ground level. An interpolation is used for ground points, similar to generating a Digital Terrain Model (DTM), but rather than using a regular grid the ground height is calculated for the position of each point.

```
spddefheight --interp --in NATURAL_NEIGHBOR_CGAL \
             -i spd/LDR120726_094707_4_decomp_grid_pmf_mcc_grd.spd \
             -o spd/LDR120726_094707_4_decomp_grid_pmf_mcc_grd_defheight.spd
```

Calculate metrics

After all the pre-processing steps to convert the LAS 1.3 file into a gridded SPD format file with a defined height it is possible to generate a number of metrics from the waveform data. The command to calculate metrics within SPDLib (`spdmetrics`) takes an XML file in which the metrics are defined. There are a large number of metrics available and operators (addition, subtraction etc.,) allowing existing metrics to be combined to implement new

[3] Keqi Zhang, Shu-Ching Chen, Whitman, D., Mei-Ling Shyu, Jianhua Yan, & Zhang, C. (2003). A progressive morphological filter for removing nonground measurements from airborne LIDAR data. IEEE Transactions on Geoscience and Remote Sensing, 41(4), 872–882. http://doi.org/10.1109/TGRS.2003.810682

[4] Evans, J.S., Hudak, A.T., 2007. A multiscale curvature algorithm for classifying discrete return lidar in forested environments. IEEE Transactions on Geoscience and Remote Sensing 45 (4), 1029–1038.

metrics. The full list of metrics is available in the SPDMetrics.xml file, distributed with the source of SPDLib. Most metrics have an option to specify the minimum number of returns (`minNumReturns`), setting this to 0 will use the waveform information to calculate the metric, setting to 1 (default) or above will use the discrete data. In this way full waveform and discrete metrics can be created at the same time.

For this exercise we will be calculating Height of Medium Energy (HOME) and waveform distance (WD), a detailed description of these metrics is given in [5].

First, create a file containing these metrics. Open a blank text file using:

```
gedit spd_metrics.xml
```

and paste the text below into it:

```
<!--
    SPDLib Metrics file

    Produced to generate metrics from full waveform LiDAR data for
-->

<spdlib:metrics xmlns:spdlib="http://www.spdlib.org/xml/">
    <!-- HOME -->
    <spdlib:metric metric="home" field="HOME"/>
    <!-- WD -->
    <spdlib:metric metric="maxheight" field="WD" minNumReturns="0"/>
</spdlib:metrics>
```

Save and close this file. Then create a folder to store the output metrics file:

```
mkdir metrics
```

To calculate the metrics and produce an image as an output run.

```
spdmetrics --image -o metrics/LDR120726_094707_4_metrics.bsq \
           -f ENVI \
           -i spd/LDR120726_094707_4_decomp_grid_pmf_mcc_grd_defheight.spd \
           -m spd_metrics.xml
```

Once the command has finished, open the metrics image using:

```
tuiview metrics/LDR120726_094707_4_metrics.bsq
```

Try adding more metrics to the `spd_metrics.xml` file and run the `spdmetrics` command again.

This excercise has demonstrated the steps needed to calculate full waveform metrics on a small subset but the commands can be applied to whole lines or flights although this will take a lot longer to process.

[5] Cao, L., Coops, N., Hermosilla, T., Innes, J., Dai, J., & She, G. (2014). Using Small-Footprint Discrete and Full-Waveform Airborne LiDAR Metrics to Estimate Total Biomass and Biomass Components in Subtropical Forests. Remote Sensing, 6(8), 7110–7135. http://doi.org/10.3390/rs6087110