DIGITIZING 60 METERS OF FULL WAVEFORM LIDAR IN A DENSE TROPICAL RAINFOREST

Martin Isenburg[1] and Hubert Trunzer[2] and Franz Malmer[2]
[1]rapidlasso GmbH, Casparigasse 16, 97286 Sommerhausen, Germany,
Email: martin@rapidlasso.com
[2]Asian Aerospace Services Limited, 23rd Fl, Chamnan Phenjati Bldg, 65/209 Rama 9, Bangkok 10320, Thailand,
Email: {hubert,franz}@asianaerospaceservices.com

KEY WORDS: airborne laser scanning, waveform digitization, ground classification, open standards, PulseWaves

ABSTRACT: The earliest airborne Light Detection And Ranging (LiDAR) scanners delivered just a single elevation return per laser shot. Later systems produced a return for the first and the last interaction between the laser and the landscape below. Today, discrete return LiDAR commonly produces up to 4 or more returns. This allows picking up hits of electricity or telephone wires and captures more information about the vegetation structure. Recently, waveform digitizers have become popular that capture the reflection of the emitted laser pulse with much more detail: The waveform returning to the plane is digitized up to one billion times per second so that the intensity of the laser pulse's reflection is recorded every nanosecond giving a vertical resolution of one sample of digitized amplitude each 15 centimeters for the returning waveform. Because these full waveform LiDAR system are capable of capturing the “stuff below” with much more detail, they are especially interesting for vegetation analysis.

Due to limited bandwidth for storing the data to disk, typical waveform digitizers record only one or two segments of 80 up to 256 samples for those parts of the returning waveform where sufficient signal activity is detected. This covers vertical distances of 12 up to 38 meters which is not always sufficient to record the entire interaction between the laser pulse and all vegetation layers down to the ground for tropical rainforests where the canopy contains tall tree that are 45 meters or higher. Especially in dense tropical forests where discrete return LiDAR systems often do not produce very many ground return, it could be of great advantages to have access to the entire full waveform. It should be possible to extract very faint waveform peaks that are usually not recorded and extract additional details about the topology of the terrain that is often missing from standard LiDAR surveys.

We present full waveform LiDAR capturing dense rainforest canopy in Thailand up to 45 meters tall by digitizing the reflected laser in continuous 60 meter segments. We used a RIEGL LMS Q680i full waveform scanner with a non-standard acquisition setting that always continued to record 400 waveform samples after the first canopy hit.

1. INTRODUCTION

For a number of applications it has become useful to fly airborne laser scanning or LiDAR surveys above densely vegetated areas. In the archaeological community the ability of the laser to penetrate through the canopy has accelerated the rate of discoveries. After digitally removing the trees, archaeologists can literally “see” the bare earth through the vegetation and find smaller and larger structures that have been hiding under the canopy for centuries. In the forestry community one is less concerned with what is on the ground and more interested in automated methods for computing canopy coverage, performing tree inventories, and identifying plant species. Nevertheless, also for foresters it is important to have a detailed model of the bare-earth terrain as an intermediate step to compute the relative elevation of the vegetation returns above ground, which is strongly correlated to the height of the trees. An accurate Digital Terrain Model (DTM) is usually a core deliverable of any LiDAR survey.

The common way to generate a DTM is to find those LiDAR returns that correspond to ground points, compute a Delaunay triangulation in 2D from their x and y coordinates, and then rasterize their linearly interpolated z coordinates onto a grid of elevation samples. The most difficult part is to determine which LiDAR returns are ground and which not, a process known as *ground classification.* Nowadays a number of software packages are available to perform this task. However, this assumes there are sufficiently many LiDAR points that were reflected from the bare earth. In densely vegetated terrains such as tropical rainforests this may not always be the case.

Today's software packages generally operate on a cloud of points that corresponds to those 3D locations where the laser was either hitting or grazing an object. The earliest LiDAR scanners delivered just a single return per laser shot. Later systems produced a return for the first and the last interaction between the laser and the ground below and today's discrete LiDAR deliveries contain up to 4 of more returns. This allows picking up hits of electricity or telephone wires and captures more information about the vegetation structure.

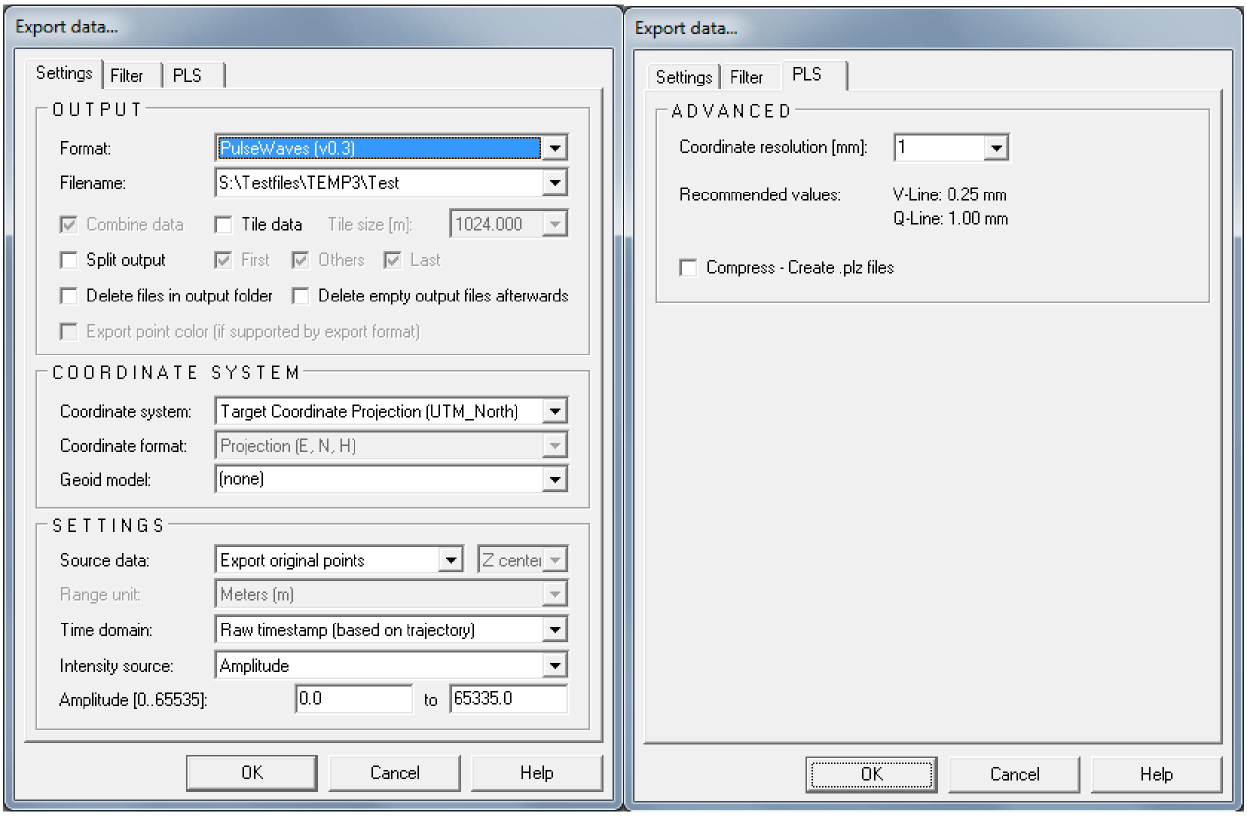
Recently, waveform digitizers have become popular that capture the reflection of the emitted laser pulse with much more detail: The waveform returning to the plane is digitized up to one billion times per second so that the intensity of the laser pulse's reflection is recorded every nanosecond giving a vertical resolution of one sample of digitized amplitude each 15 centimeters for the returning waveform. Because these full waveform LiDAR system are capable of capturing the "stuff below" with much more detail, they are especially useful for vegetation analysis.

Due to data storage bandwidth limitations, typical waveform digitizers record only one or two segments of 80 up to 256 samples whenever sufficient energy is detected in the returning waveform. This covers vertical distances of 12 up to 38 meters, which is not always sufficient to record the entire interaction between the laser pulse and all vegetation layers down to the ground for tropical rainforests, where the canopy can contains tree that are up to 45 meters tall. Especially under conditions of extremely dense vegetation where discrete return LiDAR systems do not always deliver a ground return it could be of great advantage to have access to the entire full waveform. We believe that it should be possible to extract those very faint waveform peaks that correspond to subtle ground returns that are usually not recorded because their returning signal energy it to weak to trigger the digitizer. We think this should make it possible to extract additional details about those terrain features that are missing from standard discrete LiDAR surveys below very dense vegetation.

We present full waveform LiDAR that captures a dense rainforest canopy in Thailand that is up to 45 meters tall by digitizing the reflected laser in continuous 60 meter segments. We used a RIEGL LMS Q680i full waveform scanner with a non-standard acquisition setting that always continues to record 400 waveform samples after the first canopy hit. The resulting full waveform LiDAR was exported to the new PulseWaves format (Isenburg, 2013).

2. PULSEWAVES

The three year long PulseWaves project (Isenburg, 2013) is an effort to create an open specification, an open format, and an open source application programming interface (API) for storing full waveform LiDAR. It was funded in parts by RIEGL, by Flinders University of South Australia, and by Airborne Research Australia. The primary role of this specification is to have a common data exchange format that allows researchers to experiment and commercial software to process full waveform LiDAR without having to struggle with a smorgasbord of proprietary formats. It was developed in an open, transparent, and vendor-neutral process since December 2011 through active involvement of the community via the discussion forum at http://pulsewaves.org. The current release of the format (version 0.1 revision 10) is practically finalized. The specification together with the C++ source code that implements an API and a DLL are LGPL-licensed and available for download. The format is LAS-compatible, offers (optional) compression, and has already been integrated into RIEGL's RiPROCESS software as a new option for exporting full waveform. Figure 1 shows a screen shot of the export window. We are currently cooperating with Carbomap Limited of Nottingham in the UK to also create an IDL reader and writer for the PulseWaves format.



OUTPUT

Format: PulseWaves (v0.3)

Filename: S:\Testfiles\TEMP3\Test

☑ Combine data ☐ Tile data Tile size [m]: 1024.000

☐ Split output ☑ First ☑ Others ☑ Last

☐ Delete files in output folder ☐ Delete empty output files afterwards

☐ Export point color (if supported by export format)

COORDINATE SYSTEM

Coordinate system: Target Coordinate Projection (UTM_North)

Coordinate format: Projection (E, N, H)

Geoid model: (none)

SETTINGS

Source data: Export original points Z cente

Range unit: Meters (m)

Time domain: Raw timestamp (based on trajectory)

Intensity source: Amplitude

Amplitude [0..65535]: 0.0 to 65335.0

ADVANCED

Coordinate resolution [mm]: 1

Recommended values: V-Line: 0.25 mm
Q-Line: 1.00 mm

☐ Compress - Create .plz files

Figure 1: The export window setting in RiPROCESS for storing full waveform LiDAR to PulseWaves

The PulseWaves format consists of two binary files: The Pulse files (*.pls) describe the emitted laser pulses via a geo-referenced origin and target point. The Waves files (*.wvs) contain the actual samples of the outgoing and returning waveform amplitudes for the digitized sections of the emitted and received laser light (e.g. in the vicinity of where something was hit). The compressed counterparts are the Pulze (*.plz) and the Wavez files (*.wvz).

The Pulse file is stand-alone. It describes the geo-referenced locations, directions, and durations for all emitted pulses. For each fired pulse it stores the geo-referenced coordinates of some *anchor* point on the pulse – usually the lasers's optical origin – and the geo-referenced coordinates of a *target* point that is 1000 sampling units past the anchor point in direction away from the scanning system. Together the target and anchor point specify the location and direction of the laser pulse. Also stored is the moment the *first* and *last* sample for the returning waveform was recorded. This alone is, for example, already sufficient to verify coverage or "sweep out" the scanned 3D space.

The Waves file is not stand-alone and depends upon the Pulse file. Each pulse in the Pulse file contains an offset into the Waves file to where the actual digitized samples for the relevant digitized segments of that pulse are stored that describe the shape of the waveform in detail. The format how the waveforms are sampled was kept as flexible as possible allowing each pulse to references a different sampling description.

Via the GPS time stamps the pulses in the Pulse file and their associated Waves may (optionally) be linked to the discrete LiDAR returns stored in corresponding LAS or LAZ files and vice-versa.

3. DATA COLLECTION

The airborne LiDAR data was acquired with a RIEGL LMS Q680i full waveform laser scanner installed into a Diamond Aircraft “Airborne Sensors” DA-42 fixed-wing plane. In Figure 2 you can see this integrated scanning system in front of its hangar. Due to the compact measurements of the LMS Q680i full waveform laser scanner, it is installed into the nose of the aircraft alongside a Hasselblad medium-format camera. The aircraft is owned and operated by Asian Aerospace Services Limited of Bangkok, Thailand and based out of the Don Muang International Airport where it is available “on call” for commercial or research surveys with aerial LiDAR and/or imagery.

Figure 2: The hardware and software of the “Airborne Sensors” being prepared before the full waveform flight.

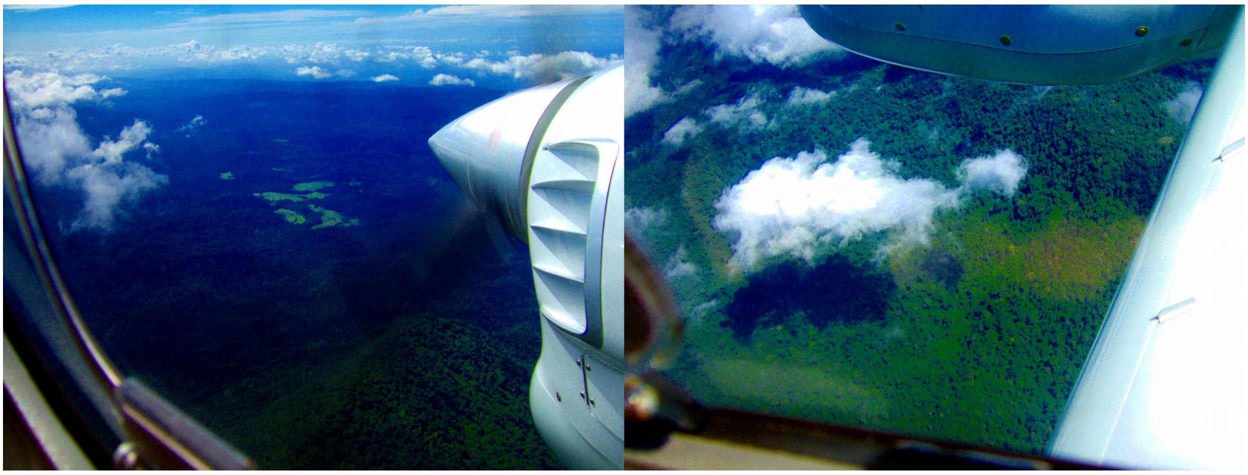


Figure 3: The dense canopy of the tropical rainforest we are scanning is is up to 45 meters tall.

While in air the scanning process is controlled using a standard laptop running RIEGL's RiAQUIRE software. As we were interested to fly with some rather unusual scanner settings, RIEGL supplied us with a special version of RiAQUIRE that had extra capabilities. Usually our full waveform digitizer records segments of 120 samples shortly *before* detecting sufficient signal energy in the returning echo. If needed it restarts – after a brief pause – recording a second or even a third waveform segment on 120 samples when more returning echo signal is detected. For a tall tree canopy of 40 meters, for example, the first segment may get triggered when the reflection of the tree crown is digitized. As a segment of 120 samples covers about 18 meters in distance it may therefore sample the canopy from 42 meters to 24 meters above the ground. A second segments may get triggered when the laser shot reflects off a tree branch at a height of 10 meters. This will sample the canopy from 12 meters above to -6 meters below the ground. Hence, there will always be gaps in the canopy digitization as soon as trees are taller than 18 meters.

Especially for very dense tropical rainforests (see Figure 2), we see three potential disadvantages with this approach: First, when there are small branches across the entire height of the canopy some of them will not be recorded properly, namely those falling into the time after one segment stopped and before the second one was triggered. Second, after the first canopy hit we expect to be glazing parts of the vegetation until the ground was hit. The triggering mechanism may either not be sensitive enough to restart on very faint vegetation returns or be too sensitive and start recording for every bit of dirt or fog in the air already far above the forest. Third, we may loose the so very important peaks that correspond to ground returns even when they are distinct because they may fall between segments but more likely when they are very faint and do not trigger the recording of the waveform.

In this experiment we were trying to avoid these potential pitfalls by operating with a very particular non-standard acquisition setting for the LMS Q680i that *always* continues to record 400 waveform samples after the first canopy hit, which essentially digitizes a distance of 60 meter – more than enough for our tropical rainforest whose trees are maximally 45 meter tall. We were advised, however, to limit the pulse repetition rate to 80kHz to assure there would be enough I/O bandwidth given the increase in digitized samples to record. The acquired full waveform data was stored on the three SSD disks located behind the nose of the “Airborne Sensors” in RIEGL's proprietary SDF format along with the recorded GPS trajectory and the roll, pitch and jaw measurements of the IMU. Back in the office the SDF files were processed with RiPROCESS and – using the latest capabilities of this software – fully geo-referenced full waveform LiDAR was exported compressed to the PulseWaves format (Isenburg, 2013).

4. DATA EXPLORATION

The PulseWaves project has not only produced a specification, a format, a DLL, and an API to easily write and read full waveform LiDAR, but also a few tools for manipulation and exploration. For the time being only three basic tools are available but as this continues to be an active R&D project here at rapidlasso GmbH, more elaborate processing tools should become available eventually. For example, there is the simple 'pulseinfo' tool that prints out a textual description of the contents of a PulseWaves file – akin to the popular 'lasinfo' tool of LAStools (Isenburg, 2014) for discrete LiDAR. Then there is the 'pulse2pulse' tool which, for example, allows to cut out a small segment of pulses file based on their GPS time stamps. And, or course, the 'pulseview' tool for visualizing PulseWaves data. Figure 4 shows a screen shot of inspecting one second worth of pulses. This image requires some explanation:

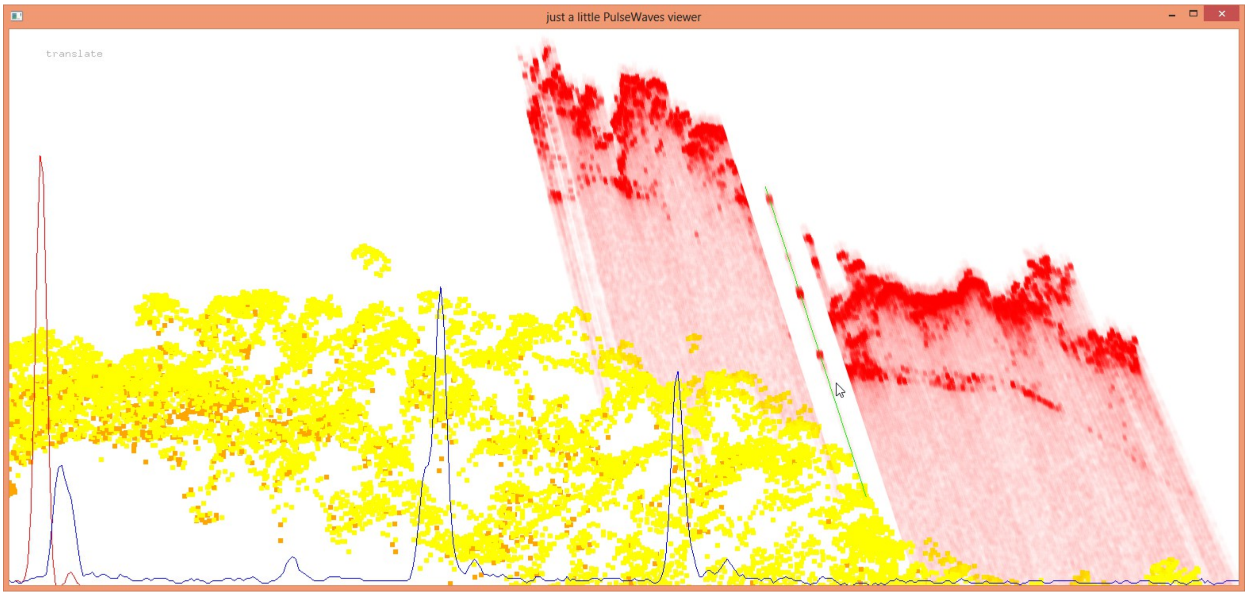


Figure 4: The outgoing (red) and returning (blue) waveform plots belong to the pulse pointed to by the cursor.

The illustrations in Figure 4, 5, 6, and 7 need to be understood as follows: In the foreground there is a 2D plot of the digitized outgoing waveform (short and red) that is recorded as the laser shot leaves the airborne scanning system. It is recorded a few thousand nanoseconds before (depending of the flying altitude) the 2D plot of the 400 sample long returning waveform (long and blue) that is digitized once the reflected canopy echo triggers the recording of a segment. Both plots are from left to right. The returning waveform (long and blue) can also be found in the 3D view where it is represented by a rendering of its intensity going from transparent (weak) to saturated red (strong). There is more than one such waveform rendered in the 3D view (about 70 to 300) and sometimes they overlap and saturate into a deeper red. The waveform that corresponds to the blue 2D plot is separated from the others, pointed to by an arrow, and lined by a thin green segment. The more red areas of the rendering coincide with the peaks of the blue 2D plot. Hence, the red areas are more or less exactly those locations in 3D where the laser was reflected from either the vegetation or the ground and where we would have a return in the discrete LiDAR point cloud.

The only unexplained thing are the yellow and orange points. There is exactly one such point for each pulse that has a returning waveform. Their only raison d'être is to help the user interact with the PulseWaves file. These points serve a visual aid to maneuver and select certain pulses. Here these points are the very *last* sample that was digitized for a pulse and their intensities should – at least in our particular case – just be noise. Since all our segments correspond to 60 meters of waveform their very last sample should always be digitized long *after* the ground reflection. In the rendering these last samples are therefore below the red areas where the actual returns are.

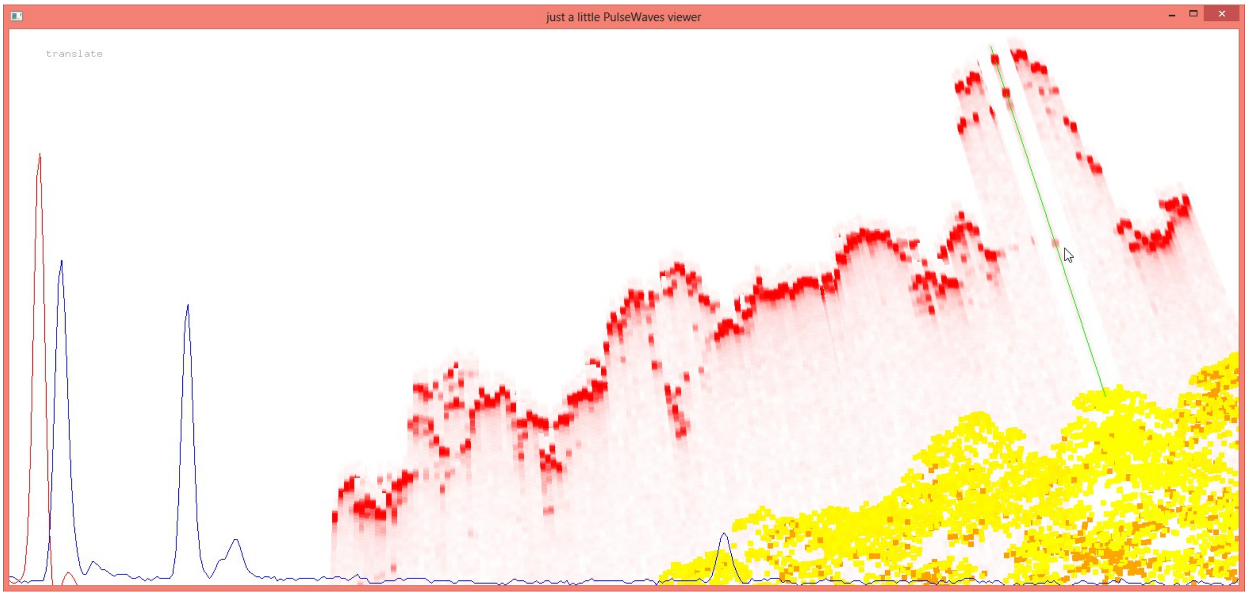


Figure 5: The transparent red lines "standing" on the yellow/orange points plot waveform intensities of 100 pulses.

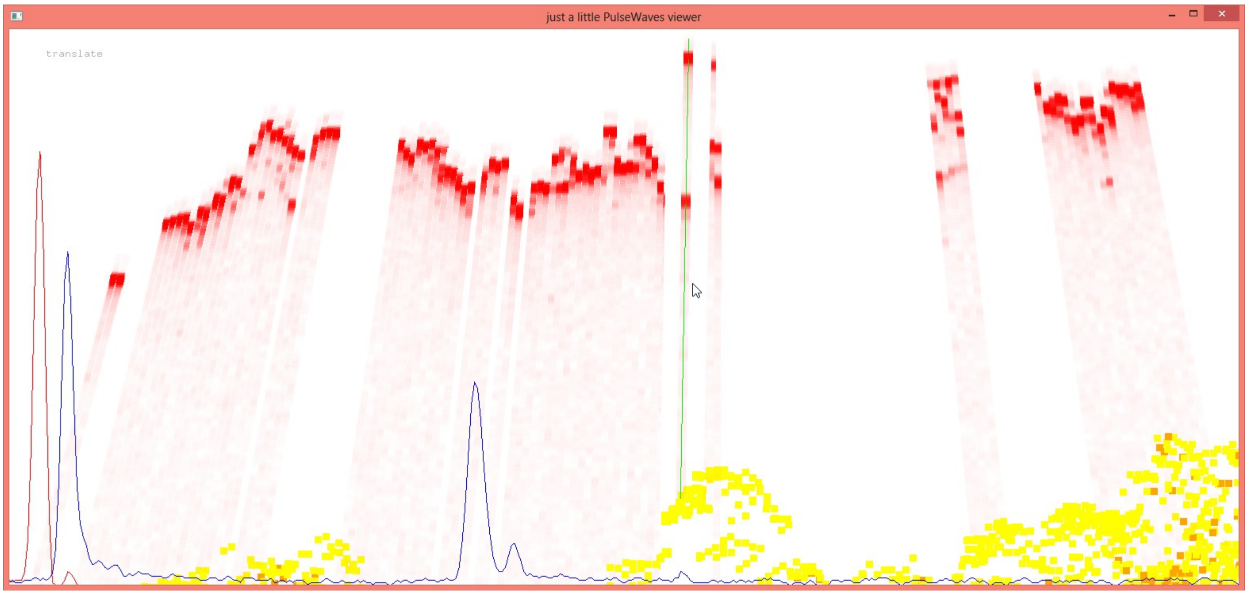


Figure 6: The yellow/orange points are always the last sample that was digitized for the waveform of each pulse.

5. DISCUSSION

We have introduced a unique full waveform data set that was scanned above the dense tropical rainforest of Thailand with a non-standard scanner setting that always recorded the digitized waveform sufficiently long to capture the entire height of the vegetation from the top of the tree canopy down to well below the ground. We believe that with these setting we will be able to capture many subtleties that the standard configuration will often miss, either because the scanning system is in a blind spot where it is transitioning from one segment to the next or because the reflected light energy is to weak to trigger the recording of the digitized waveform. We have described how we have captured this full waveform LiDAR using a RIEGL LMS Q680i and exported to the PulseWaves data exchange format. We have presented this data with intuitive visualizations generated with 'pulseview' and tried to motivate that there is additional value in full waveform LiDAR data that has gone unexploited so far.

One point of this presentation is to spark future research on full waveform LiDAR that focuses on directly processing the digitized waveform. One obvious research direction we like to work on together with interested students or other scientists is the extraction of additional ground returns from full waveform LiDAR. As suggested in Figure 7 we can utilize strong ground returns from nearby waveform to elevate the importance of faint peaks in other waveforms that would be declared noise when looked at in isolation. The ability to exploit full waveform LiDAR in this manner and produce additional elevation samples that can be used to add otherwise missing details to a digital bare-earth terrain model would increase the value of all already collected and future full waveform LiDAR.

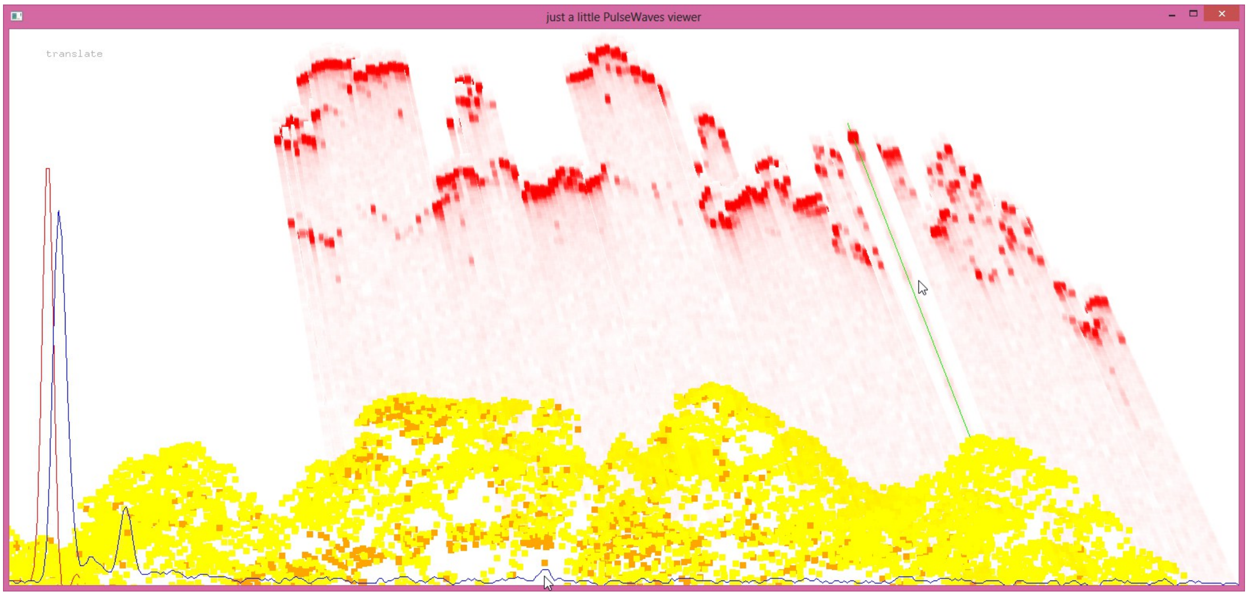


Figure 7: In isolation, the subtle peak in the waveform looks like noise, put in context it is a weak ground return.

ACKNOWLEDGEMENTS

We thank the RIEGL support team, especially Ms. Andrea Spitzer for providing us on very short notice with the custom version of the RiAQUIRE software that allowed us to digitize the returning waveform in segments of 400 samples as well as Mr. Guenther Gruber for adding PulseWaves support to RiPROCESS and helping us in the process to refine the PulseWaves API and DLL into a production strength software package.

REFERENCES

Isenburg, M., 2013. PulseWaves: An Open Data Exchange Format for Full Waveform LiDAR. In: Proceedings of European LiDAR Mapping Forum, Amsterdam, November 2013, available from http://pulsewaves.org

Isenburg, M., 2014. LAStools – efficient LiDAR processing software (version 140831), Retrieved September 3rd, 2014, from http://rapidlasso.com/LAStools