**Waveform (cube) extraction from Riegl SDF files**

July 20-25, 2014

Grzegorz Jozkow

1. Waveform cube

The idea of the waveform cube was discussed several times (Fig. 1) then only short description will be given.

**row**

**band**

**column**



Figure 1. Waveform cube

Since it might be relevant for any subsequent processing of constructed cubes (stored as Matlab mat files), additionally, in green there is shown an origin (first sample – first row, column, and band) and directions of axes showing next row, column, and band as oriented in Matlab array. Note that on the picture the order of waveform sample number is opposite to this which is in the extracted waveform cube. Seeing from the perspective of time of data acquisition, first dimension is band, second is column, and third is row what is opposite to the order dimension of Matlab arrays.

1. Riegl SDF file

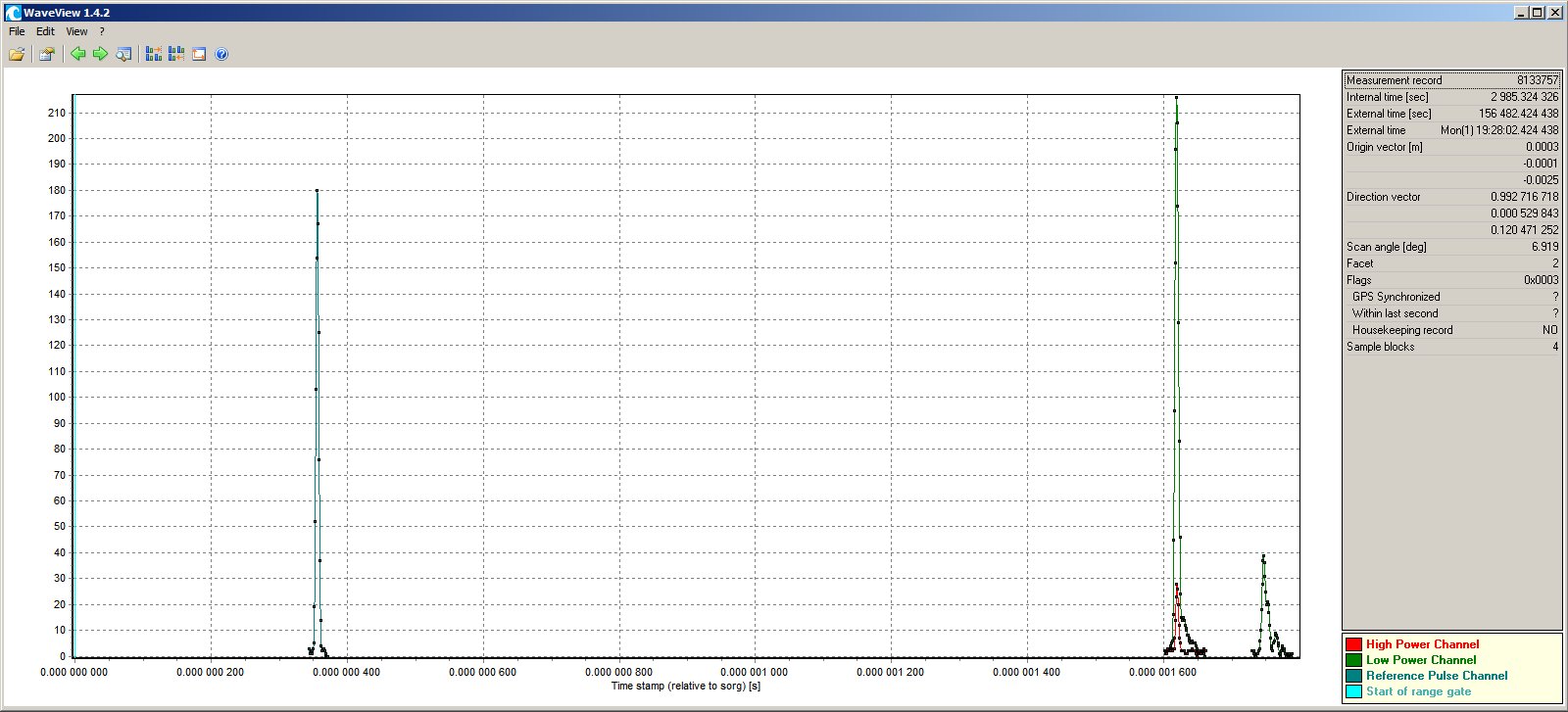
Some information about waveform data and metadata extracted from SDF file using Riegl Waveform Extraction Library is necessary to provide for cube construction and subsequent processing. All data from SDF file is extracted respect in records where one record contains all data (including metadata) of the single waveform. Each extracted record has metadata as follow (used abbreviations are consistent with names of variables of waveform extraction library and Matlab arrays):

* **time\_sorg** – time of start of range gate [s] – it is a time stamp (internal) of instant when data sampling is initiated,
* **time\_external** – external time relative to epoch [s] – it is a time stamp (external) which allows relating samples with external time, e.g. GPS. It is a difference between the instant of initialization of data sampling (need to be checked since that is not mentioned in documentation) and certain epoch (see explanation of the **epoch** metadata for the file),
* **origin** – origin vector [m] – it is 3D position (x, y, z) of laser pulse emission in scanner coordinate system,
* **direction** – unit vector of laser pulse direction – it is given in the scanner coordinate system,
* **flags** – record flag – it tells about GPS data synchronization, synchronization within last second, housekeeping record. Values retrieved during tests showed that usually **flags** value is equal 3 (most of the records). For the last waveform in the scan line it is equal 11, and equal 15 for the housekeeping record (without any useful waveform data) separating scan lines (it proceed as a next record after record with flag 15).
* **facet** – scan mirror facet number starting from 0 – it indicates from which mirror facet of the scanner the pulse was reflected from,
* **sbl\_count** – number of sample blocks – see details of **sbl** and explanation of sample block,
* **sbl\_size** – size of sample block [B] – values retrieved during tests indicated always 32B,
* **sbl** – sample blocks set – it is set of another metadata and waveform samples (see below).

During waveform data acquisition one pulse might be sampled few times using different sampling channel (see below) or in different time using the same channel. Sample block is a consistent group of waveform samples (taken be the same channel) and related metadata. Single sample block contain:

* **time\_sosbl** – time of start of sample block [s] – it indicates time stamp for the first waveform sample of the block. Note that this time could be whole value of internal time or relative to internal time (**time\_sorg**) what was performed during extraction of waveform cubes,
* **channel** – flag of channel at which pulse was sampled – channel flags are as follow: 0 – high power, 1 – low power, 2 – saturation (not observed during tests), 3 – reference,
* **sample\_count** – number of waveform samples – values retrieved during tests showed that for reference channel number of samples was always 24 (Q680i scanner) or 28 (Q780 scanner) and for low and high power channels the number of waveform samples was 60 or 120,
* **sample\_size** – size of one sample value [B] – values retrieved during tests indicated always 2B,
* **sample** – vector of waveform sample values – this is primary data to build the cube.

Description of some parameters of single record is given in the visualization (Fig. 2).



Axis of time\_sosbl relative to time\_sorg

Sample block 1 from reference channel sample\_count = 28

sbl\_count

flags

facet

direction

origin

time\_external

time\_sorg

Record ID

Sample block 2 from low power channel sample\_count = 60

Sample block 4 from low power channel sample\_count = 60

Sample block 3 from high power channel sample\_count = 60

Start of range gate

Fig. 2 Parameters of single record extracted from SDF file

Better point of view on sample block set **sbl** (one record) is given presenting few typical records noticed for test data (Fig. 3). Number of blocks retrieved during tests was equal:

* 0 – for housekeeping records only (Fig. 3a),
* 1 – this block contained always data from reference channel only (Fig. 3b),
* 2 – same as for record containing 1 block plus second block for low power channel (Fig. 3c),
* 3 – same as for record containing 2 blocks plus third block either from low power (Fig. 3d) channel either from high power channel (Fig. 3e),
* 4 – same as for record containing 2 blocks plus third and fourth block both from low power channel (Fig. 3f) or third block from low power channel and fourth block from high power channel (Fig. 3g) or third block from high power channel and fourth block from low power channel (Fig. 3h).

Note that for visualization purposes, time stamp (internal) for some records was changed into range gate index which is the difference between **time\_sosbl** and **time\_sorg** (when **time\_sorg** is non relative) given in the nanoseconds; or into raw range which is the range gate index multiplied by the group velocity of the pulse.

Investigating sample blocks it can be notified that waveform sample blocks from high power channel (Fig. 3e and 3g) have very similar time stamp and shape as one of the waveform sample blocks from low power channel; the amplitude (sample value) is different. That indicates that both blocks representing the same waveform. Since the blocks from high power channel were observed in the test data only occasionally and representing the same data as blocks from the low power channel, they were not considered during waveform cube construction.

|  |
| --- |
|  |
| a) |
|  |
| b) |
|  |
| c) |
|  |
| d) |
|  |
| e) |
|  |
| f) |
|  |
| g) |
|  |
| h) |

Figure 3. Visualization of set of sample blocks (records) of SDF file. Explanation in text.

Beside data for each record, SDF file contains short metadata common for all waveforms (it was stored as Matlab struct array **finfo**):

* **instrument** – instrument type (name),
* **serial** – serial number of the instrument,
* **epoch** – type of external time if known, for the test data it was second of the GPS week,
* **v\_group** – group velocity of the laser pulse [m/s],
* **sampling\_time** – waveform sampling interval [s],
* **flags** – flag indicating GPS synchronization – for the test data flag was equal 1 - data was synchronized,
* **num\_facets** – number of scanner mirror facets.

More information about SDF file can be found in RiWaveLib documentation, particularly in the documentation for Waveform Extraction Library and in Jalobeanu and Gonçalves (2012). Not that according information given in Jalobeanu and Gonçalves (2012), some data stored in SDF file is different than data extracted using Riegl Waveform Extraction Library. It means that Waveform Extraction Library does some kind of computation and not all extracted data and metadata are completely raw!

1. Corresponding point cloud – LAS file

Each SDF file has corresponding LAS file containing point cloud. Since it might be relevant for some subsequent investigation below is the list with available data for each point in the cloud (variables indicates columns of extracted text file containing points of the cloud corresponding to waveform cube):

* X, Y, Z coordinates where horizontal coordinates are in UTM projection (zone 18) and heights are probably geodetic (there is no available information about heights),
* Intensity of 16 bits range,
* Return number,
* Number of returns (given pulse),
* Class of the point – classes according LAS specification,
* Scan angle rank,
* GPS time.

1. Tools

Accessing both SDF and LAS file and creating waveform cube was done using:

1. Riegl Waveform Extraction Library including WaveView binary for visualizing waveforms; two versions of libraries and viewers were tested: 1.3.68 and 1.5.241 (libraries) and 1.1.1 and 1.4.2 (viewers),
2. Matlab scripts allowing use of Riegl Waveform Extraction Library (written by Zoltan Koppanyi, modified by Grzegorz Jozkow),
3. Matlab script for waveform data arrangement in the cube (written by Grzegorz Jozkow),
4. Software allowing visualization of point cloud and reading coordinates,
5. Matlab script for LAS data accessing including reading of GPS time stamp (written by Grzegorz Jozkow).

Beside software for visualization (might be any depending on user preferences), rest of libraries, binaries and scripts can be found in ‘software’ folder.

1. Waveform cube extraction for the specific area

Here is short description of procedure for particular cube extraction including extraction of corresponding point cloud:

* visualizing point cloud of selected scan strip and reading approximated coordinates of two opposite corners of the selected part of the strip,
* reading GPS time stamp for points which are placed close to the selected corners (have similar coordinates),
* in the SDF file reading record IDs having **time\_external** similar to the GPS time stamps as points above,
* since selected above IDs probably will not represent first and last waveform of some scan lines then new IDs were found where the first is smaller than old one (this for the lower time tag) and appears as the next one after the housekeeping record; the second new ID is larger than old one (this with the larger time tag) and appears before the housekeeping record,
* extracting all metadata and waveform sample values for all records in the range of specified above IDs,
* arranging waveform sample blocks as waveforms (based mainly on the **sample\_count** and/or **time\_sosbl** and **time\_sorg**) and constructing waveform cube (arranging scan lines was possible due to housekeeping records and **flags**),
* reading **time\_external** for the first and last waveform of the cube,
* loading corresponding point cloud and sorting it according GPS time stamp and return number as the second sorting level since one time stamp may be valid for more than one point,
* extracting points from the cloud which are in range of **time\_external** of first and last waveform in the cube; note that the exact GPS time stamp might be slightly smaller or larger than **time\_external**, but the end of scan lines should be clearly visible since the difference between GPS time stamps between last point in the scan line and first point in the next scan line is significant.

1. Results

Waveform cubes were extracted for two different sites and two different scanners resulting in 4 cubes (Fig. 4). Detailed specification of extracted cubes and corresponding point cloud is given in the Table 1. Extracted waveform cubes, corresponding metadata, point clouds and other files are attached to this report. Since waveform cubes do not contain useful data (samples are set to be equal 0) at the end or at the beginning of scan lines, therefor they can be cropped for particular investigation as it was performed for waveform data compression where cubes of size (*l, s, w*)504x1200x120 and 488x1170x120 for Site 1, 496x1200x120 and 480x1170x120 for Site 2, for scanners Q680i and Q780, respectively, were used.

|  |
| --- |
| First waveform  Last waveform |
| a) |
| Last waveform  First waveform |
| b) |
| First waveform  Last waveform |
| c) |
| First waveform  Last waveform |
| d) |

Figure 4. Location of test strips (gray) and extracted data (color coded height): a) Site 1, Q680i scanner, b) Site 1, Q780 scanner, c) Site 2, Q680i scanner, d) Site 2, Q780 scanner.

Table 1. Specification of extracted waveform cubes and corresponding point clouds

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Site | Site 1 – Corbin, VA | | Site 2 – Duck, NC | |
| Waveforms | | | | |
| Folder \  file | corb\Scanner\_1\  130617\_192721\_1 | corb\Scanner\_2\  130617\_192721\_2 | duck\Scanner\_1\  130623\_134838\_1 | duck\Scanner\_2\  130623\_134838\_2 |
| Scanner (SN) | Q680i (9997902) | Q780 (9999173) | Q680i (9997902) | Q780 (9999173) |
| Group velocity [m/s] | 299 710 512 .161 254 9 | 299 711 108 .207 702 6 | 299 710 512 .161 254 9 | 299 711 108 .207 702 6 |
| Sampling interval [ns] | 1 | 1 | 1 | 1 |
| Number of facets | 4 | 4 | 4 | 4 |
| Number of records | 12 357 036 | 11 983 916 | 8 840 996 | 8 496 610 |
| Record ID of cube first waveform 1) | 8 391 935 | 8 133 080 | 737 915 | 650 930 |
| Record ID of cube last waveform 2) | 9 179 981 | 8 897 531 | 2 186 530 | 2 054 815 |
| Time external of cube first waveform [s] /day | 156 482 .425 245  (1) 19:28:02 .425 245 | 156 482 .422 688  (1) 19:28:02 .422 688 | 49 724 .516 099  (0) 13:48:44 .516 099 | 49 724 .515 824  (0) 13:48:44 .515 824 |
| Time external of cube last waveform [s] /day | 156 486 .061 432  (1) 19:28:06 .061 432 | 156 486 .059 458  (1) 19:28:06 .059 458 | 49 731 .201 607  (0) 13:48:51 .201 607 | 49 731 .196 978  (0) 13:48:51 .196 978 |
| Blocks of low power channel 3) | 2 | 2 | 2 | 2 |
| Number of reference channel samples | 24 | 28 | 24 | 28 |
| Number of low power channel samples | 60 or 120 | 60 or 120 | 60 or 120 | 60 or 120 |
| Pulses in the scan line (*s* dimension) | 1431-(1432) | 1435-(1436) | 1431-(1432) | 1435-(1436) |
| Scan lines  (*l* dimension) | 550 | 532 | 1011 | 977 |
| Bands (*w* dimension) | 120 | 120 | 120 | 120 |
| Pulses (in the cube) with reference channel only 4) | 1211-1432  at the end of scan line  (negative scan angle) | 1-266  at the start of scan line (positive scan angle) | 1211-1432  at the end of scan line  (negative scan angle) | 1-266  at the start of scan line (positive scan angle) |
| Empty waveforms in the cube | 123 207  (15.64%) | 141 453  (18.52%) | 230 281  (15.91%) | 262 849  (18.74%) |
| Empty waveforms after removing pulses with reference channel only | 1 716  (0.26%) | 534  (0.09%) | 6 866  (0.56%) | 4 113  (0.36%) |
| Corresponding point cloud | | | | |
| Number of points in the whole strip | 19 412 881 | 21 434 764 | 7 862 489 | 7 796 281 |
| Max. no of returns | 7 | 7 | 6 | 7 |
| Time external of the first point (pulse) | 156 482 .425 250 480 | 156 482 .423 384 090 | 49 724 .516 099 560 | 49 724 .516 502 090 |
| Time external of the last point (pulse) | 156 486 .060 863 370 | 156 486 .059 456 070 | 49 731 .201 025 280 | 49 731 .196 975 860 |
| Points in corresponding cloud | 936 285 | 960 664 | 1 378 814 | 1 348 555 |
| Max. no of returns in corresponding cloud | 7 | 7 | 6 | 6 |

1) previous is housekeeping record, 2) next is housekeeping record, 3) bad not included, 4) there is no useful waveform (empty waveform).

Since the cubes might be used in another investigation, there is given explanation of all variables included in the mat files attached to this report:

**lib\_ver** – struct array containing information about used Riegl Waveform Extraction Library,

**finfo** – struct array containing metadata of corresponding SDF file (see section 2),

**id1**, **id2** – IDs of records from corresponding SDF file for first and last waveform in the cube,

**id\_record** – matrix (*l,s* size) containing ID of record from corresponding SDF file, 5)

**time\_sorg**, **time\_external**, **sbl\_count** – matrices (*l,s* size), (see section 2), 5)

**facet** – vector (*l* length), (see section 2),

**origin**, **direction** – 3D arrays (*l, s, 3* size), (see section 2), 5)

**time\_sosbl**, **sample\_count** – 3D arrays (*l, s, 3* size), third dimension is equal 3 since there was no more than 3 sample block, first layer apply to block from reference channel, two last layers apply to blocks from low power channel, (see section 2), 5)

**cube\_ref** – 3D array (*l, s, wref* size), waveform cube constructed from samples from reference channel, 6)

**cube\_lp** – 3D arrays (*l, s, wlp* size), waveform cube constructed from sample from low power channel, since the number of block for low power channel was equal 2 then first block started at first band and second block started directly after first block, 6)

**ns\_start**, **dt\_sbl\_start**, **id\_sbl\_start** – variables left over constructing ‘time dependent’ waveform cube, please ignore since there were issues occurred during this cube type.

5) NaN values appear for artificially created data (not extracted from SDF file) to keep the dimensions,

6) artificially created waveform samples (not extracted from SDF file) are empty having value 0.

1. Issues and remarks notified during cube construction

Few issues occurred during waveform cube construction. Below is the list starting from the less critical.

1. Extra bytes for points in the cloud are unknown **(applies to point cloud, not significant for waveform cube)**

As suggested by the header of LAS file, the point cloud probably was ‘extracted’ from the waveform data using RiPROCESS 1.5.9 software. The point cloud is given in the LAS v 1.2 format with the Point Data Format ID 1 which requires 28 B for each point record according specification (ASPRS, 2008). During LAS files reading, it was noticed that records preserve 32 B for each point record what means that is contain additional 4 B data for each point which is unknown. The name of files suggests extra bytes for points too but there is not any description of it.

1. GPS time stamp for points in the cloud does not match time stamp from the waveform data **(applies to point cloud and waveforms, significant for finding exact matches between points and waveforms)**

It was noticed that points being returns from the same laser pulse have the same GPS time stamp. Since the precision of this stamp is 1 ns then it cannot be rounding issue and this time cannot apply to time stamp to waveform peak (return). Comparing GPS time given for the laser pulse with corresponding **time\_external** of the waveform (record in SDF file) it was noticed that these time stamps are not the same. It does not match **time\_sosbl** for send or received waveform neither.

1. Differences in extracted data (number of sample blocks) using different versions of Riegl Waveform extraction Library **(applies to waveforms, suggests errors in the Riegl Waveform Extraction Library, not significant for waveform cube)**

Using library v. 1.3.68, some of the sample blocks are surely errors as shown in Figure 5a where the raw distance to the fourth sample block is incredibly high. In the newer library v. 1.5.241, distances are reliable, but there are only 3 sample blocks (Fig. 5b), however, waveform samples from fourth block are not missing; they were ‘glued’ to the samples of third block.

|  |
| --- |
|  |
| a) |
|  |
| b) |

Figure 5. Differences in extracted data using different versions of Riegl Waveform Extraction Library: a) 1.3.68, b) 1.5.241.

1. In single records, some sample blocks from low power channel occurred before sample blocks from reference channel **(applies to waveforms, suggests errors in the Riegl Waveform Extraction Library or errors in the data, not significant for waveform cube)**,

As shown in Figure 6, first sample block from low power channel was recorded earlier than sample block from reference channel what means that pulse was received before it was sent which is obviously mistake. This can be error of the data (noise) as well as error of the computation done during signal recording or error of the Riegl Waveform Extraction Library. It is impossible to identify the source of the problem.

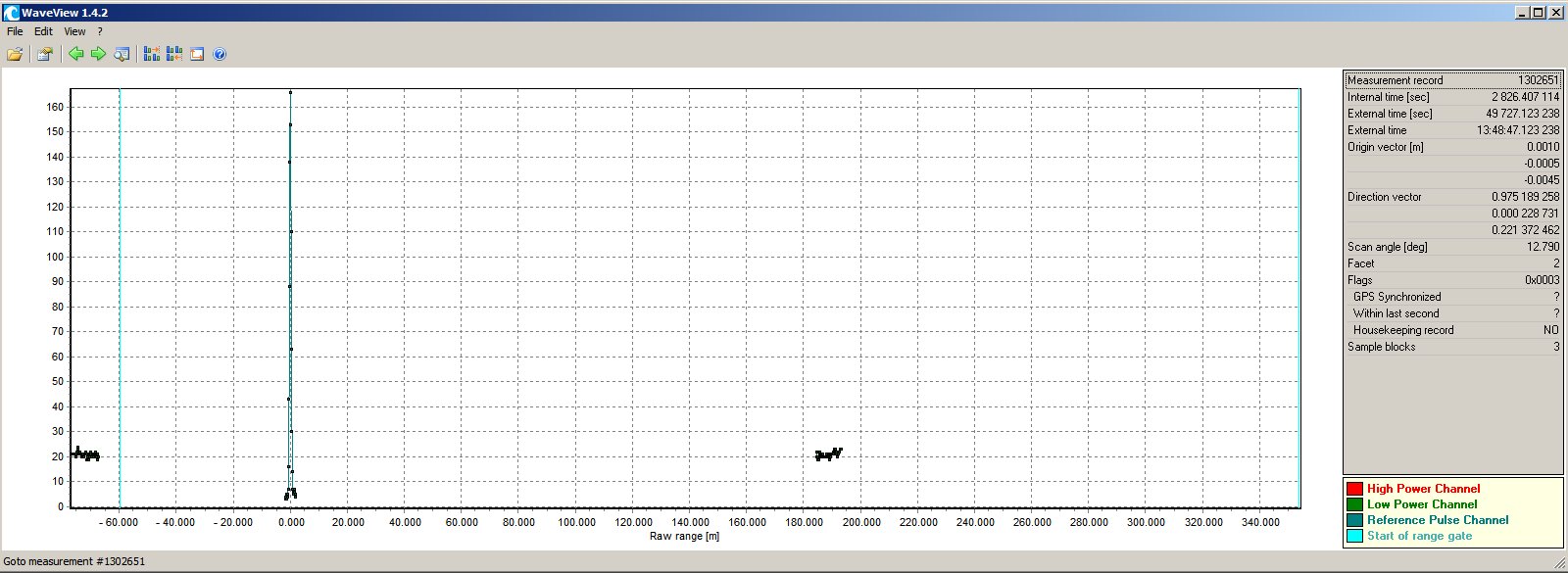


Figure 6. Example of sample block from low power channel occurred before sample block from reference channel.

1. Different number of waveforms for each scan line **(applies to waveforms, significant for waveform cube, solved)**

Number of records (waveforms) between two housekeeping records is not constant meaning that number of waveforms in the scan lines is different, only just by one waveform. This problem was solved assuming *s* dimension of the cube (Fig. 1) as the maximal observed number of waveforms in single scan line and adding artificial empty waveforms (all sample values equal 0) at the end of the scan line where the number of records was smaller than the maximal.

1. Many records with one sample block from reference channel only **(applies to waveforms, significant for waveform cube, solved)**

Depending on the scanner, first or last records of the scan line do not have any sample block from low power channel. Probably that was caused by hardware setup which was not recording data at extreme angle, however, this occur only for one side and blocks from reference channel are present. Additionally, the number of pulse where no low power channel blocks started might be different for each scan line. This problem might be partially solved by cutting some part of the cube for subsequent processing.

|  |
| --- |
|  |
| a) |
|  |
| b) |

Figure 7. Number of sample blocks from low power channel (0 – blue, 1 – green, 2 – red) for scanner: a) Q680i, b) Q780.

1. Length of sample blocks from low power channel **(applies to waveforms, significant for waveform cube, solved)**

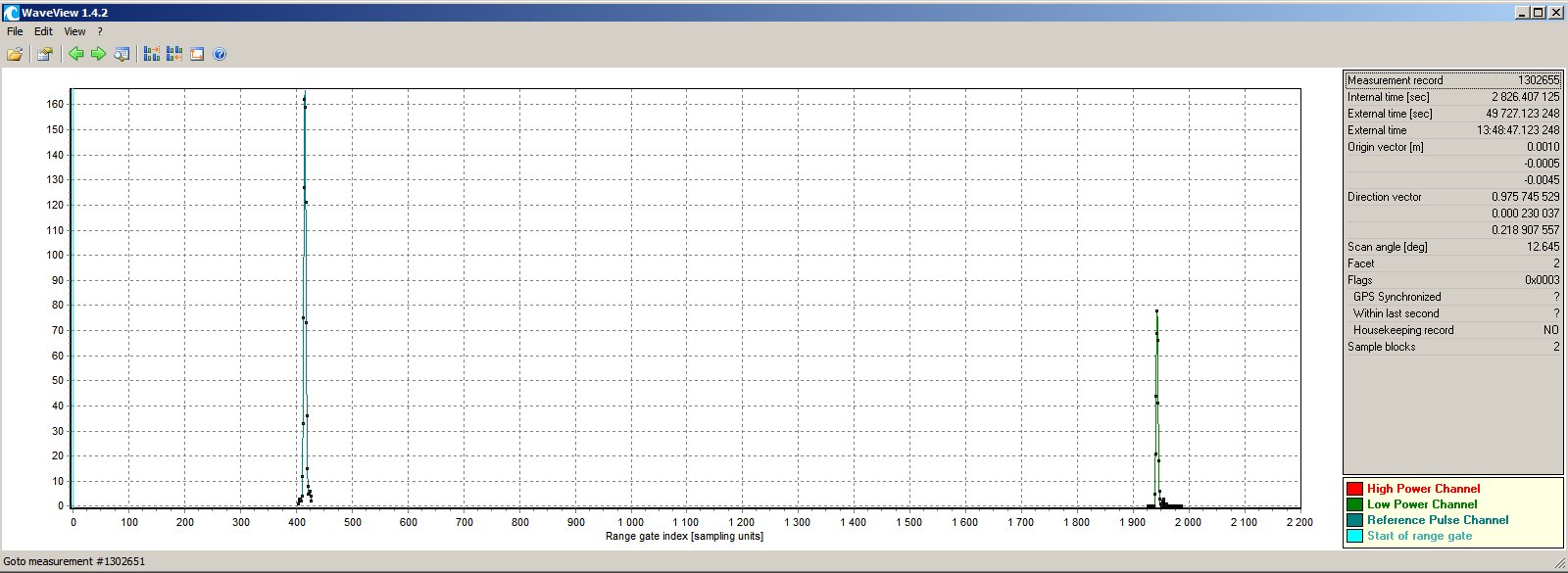
The number of waveform samples from low power channel is not equal for all sample blocks. In the test data this number was equal 60 or 120. After removing sample blocks from high power channel and blocks which are errors, it was notified that for one waveform (record) there are no more than 2 sample blocks from the low power channel, but summarized number of samples is not constant and can be 60 or 120. The *w* dimension of the cube (Fig. 1) was assumed as the maximal summarized number of samples from low power channel i.e. 120. In the case where summarized number of samples was 60, the remaining samples were assumed to have 0 values in order to maintain the cube dimension. Note that time difference between two blocks from low power channel was not considered as ‘empty’ samples in the cube but samples of the second block appeared directly after samples from first block. Visualization of one scan line of the constructed waveform without time consideration is shown in Figure 8. Note empty waveforms for the last pulses mentioned in 7). Note also the 60th band where it is visible which of the waveforms is created from one sample block of the length of 60 samples. Seeing the visualization, this cube seems to be very similar (beside dimensions) to the cube used in previous investigation.



Figure 8. Visualization of intensities of waveform cube cut for single scan line.

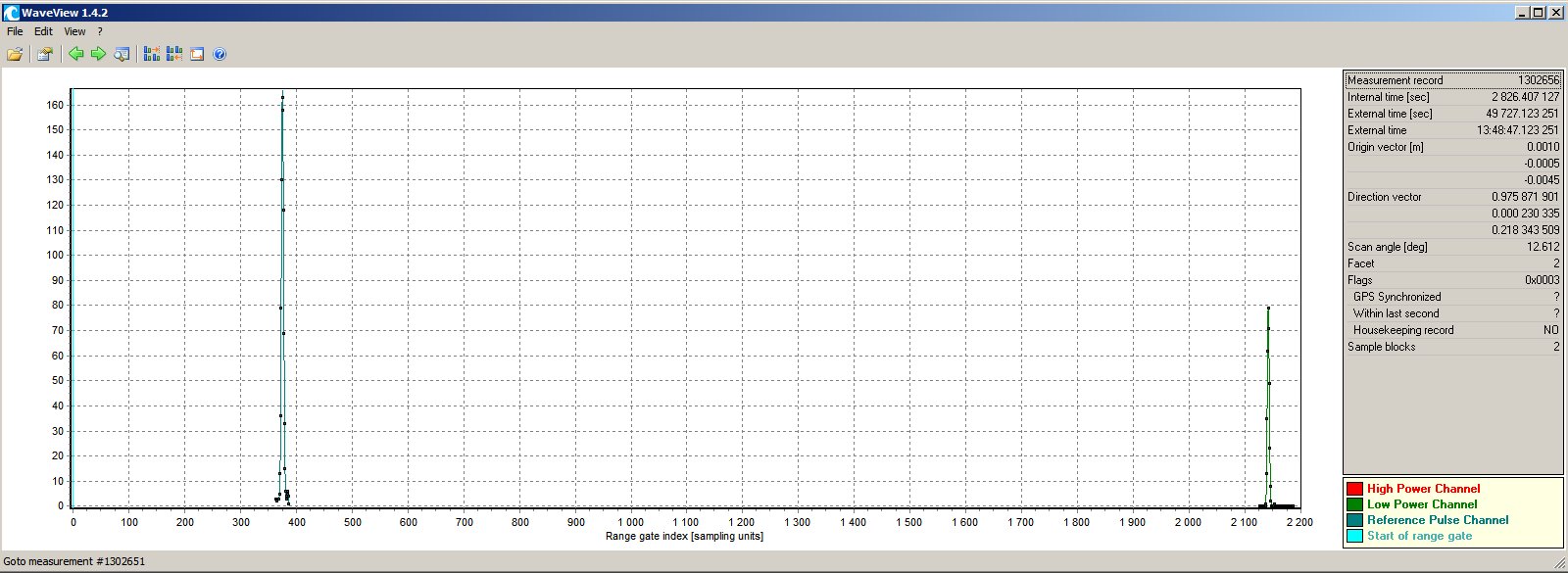
1. Time shifts between waveform **(applies to waveforms, partially significant for waveform cube, suggests errors in the Riegl Waveform Extraction Library, not solved)**

Since the time and/or spatial correlation between bands in the constructed above waveform cube is not preserved (for different waveforms first band might start at different number of nanoseconds after pulse was sent and might apply to top of the tree or ground surface) then it might be relevant to construct another type of waveform cube where this correlation is preserved. The idea was to assign one band for one nanosecond (as it was done before) where first band for each waveform starts at constant time period after pulse was send **(time\_ref)**. Since the time difference might not be integer value of nanoseconds (time stamp **time\_sosbl** and **time\_sorg** are not integer values) then rounding toward 1 ns was performed. According documentation, **time\_ref** should be determined from sample block from reference channel, however it was not discussed how. It is logical that this time should be related with **time\_sosbl** for sample block from reference channel and can be equal this time if the time difference is considered. In the second approach **time\_ref** (for calculating differences only) was assumed to be equal **time\_sorg**. Investigation of waveform times showed that time differences are affected by the shift – differences for neighboring waveforms should be close to 0, but observed values showed even 200 ns (Fig. 9) what corresponds to 30 m of height differences between topographic objects what is practically impossible for the mapped area. Not that presented example (Fig. 9) refers to single (ground) return only.



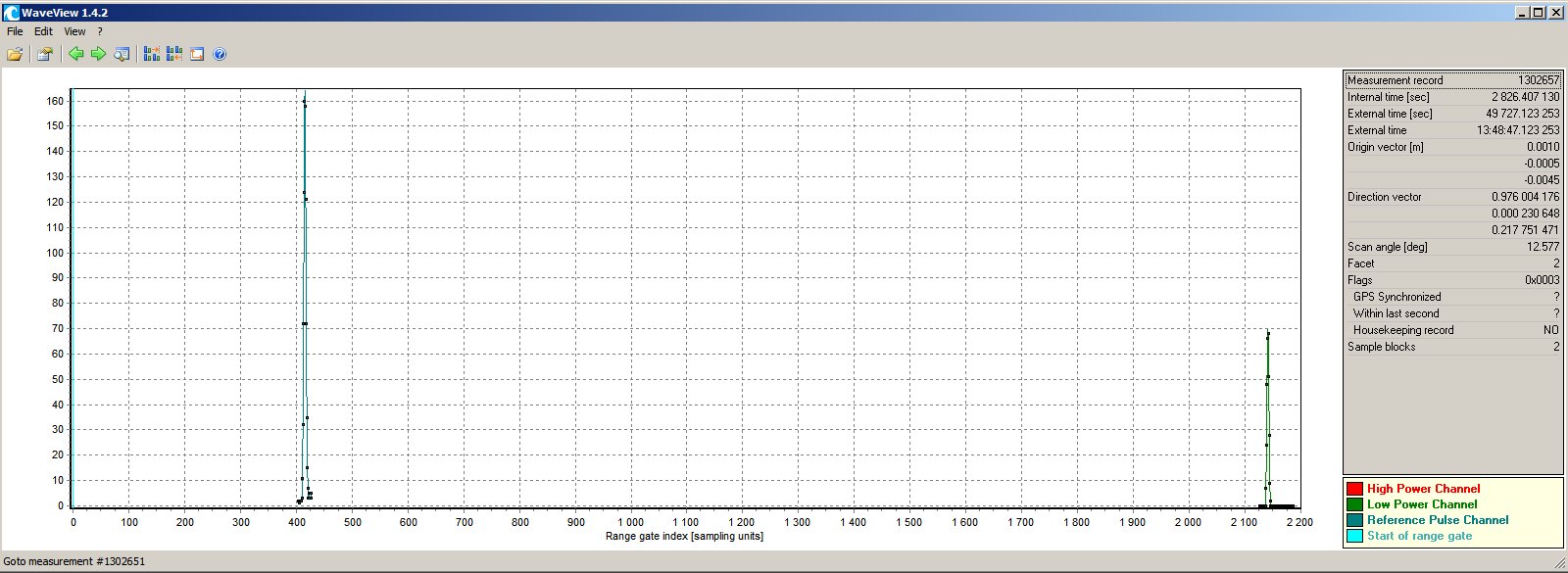
~ 1935 ns

~ 1520 ns



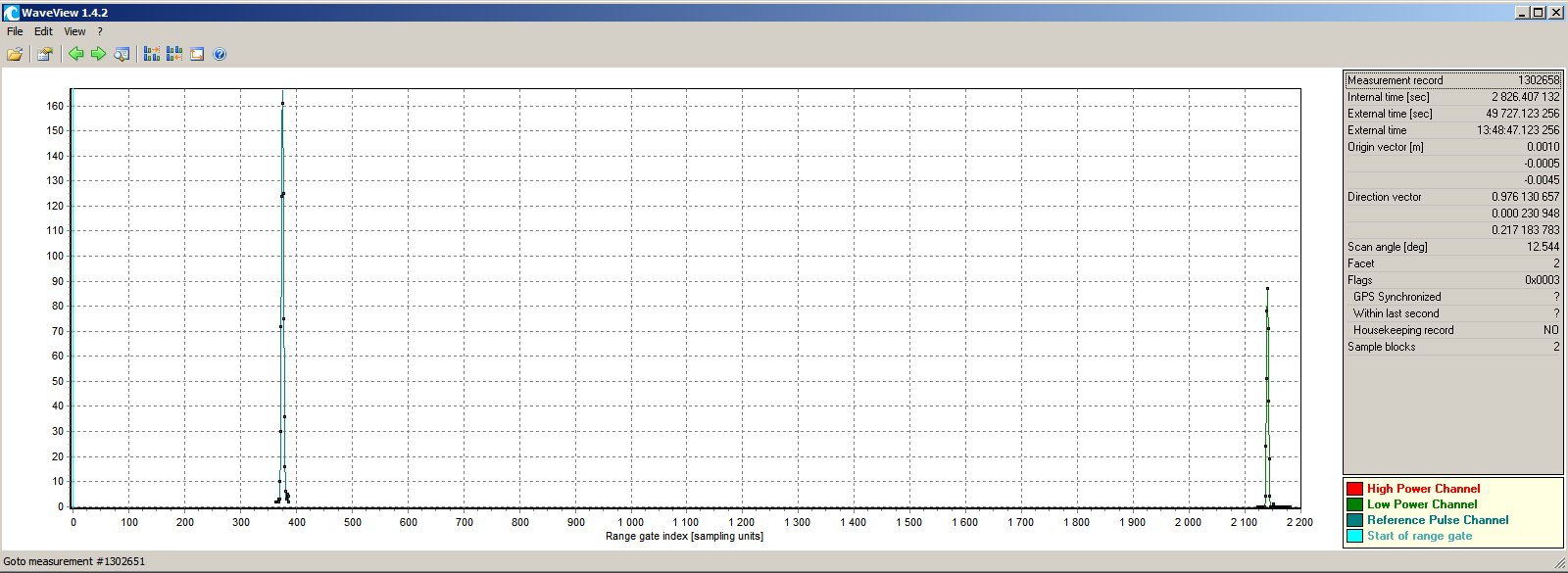
~ 1760 ns

~ 2135 ns



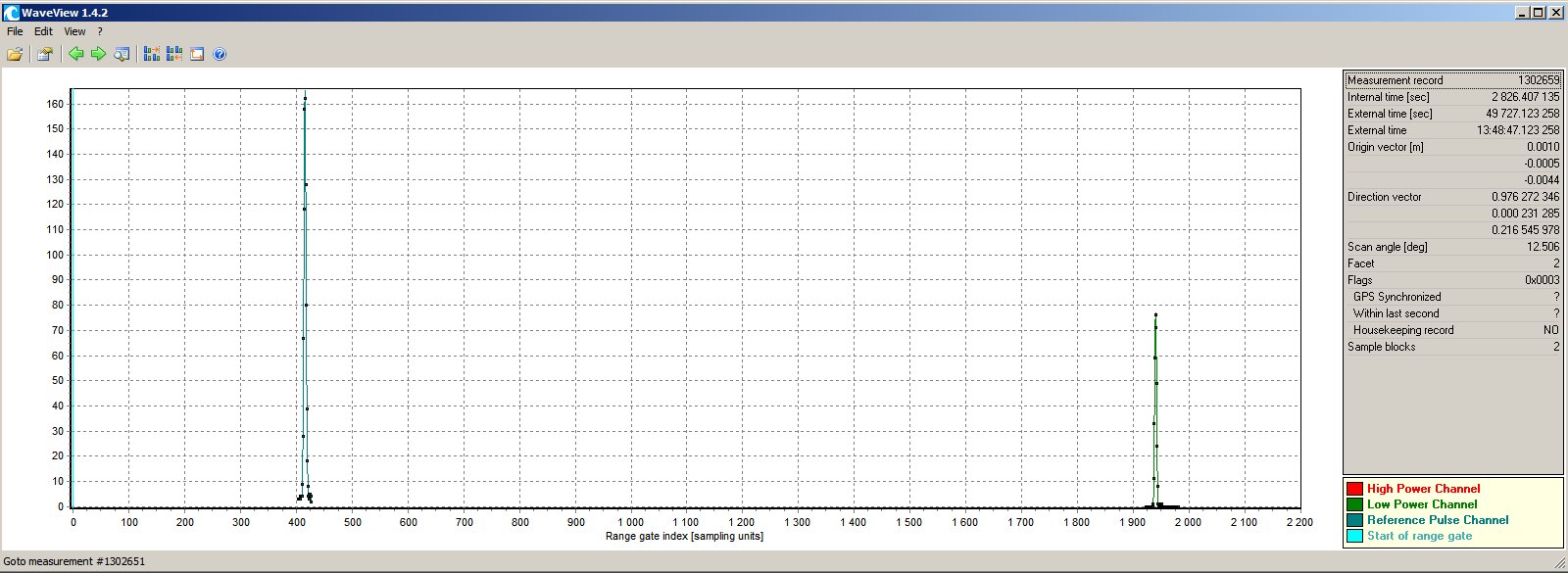
~ 1720 ns

~ 2135 ns



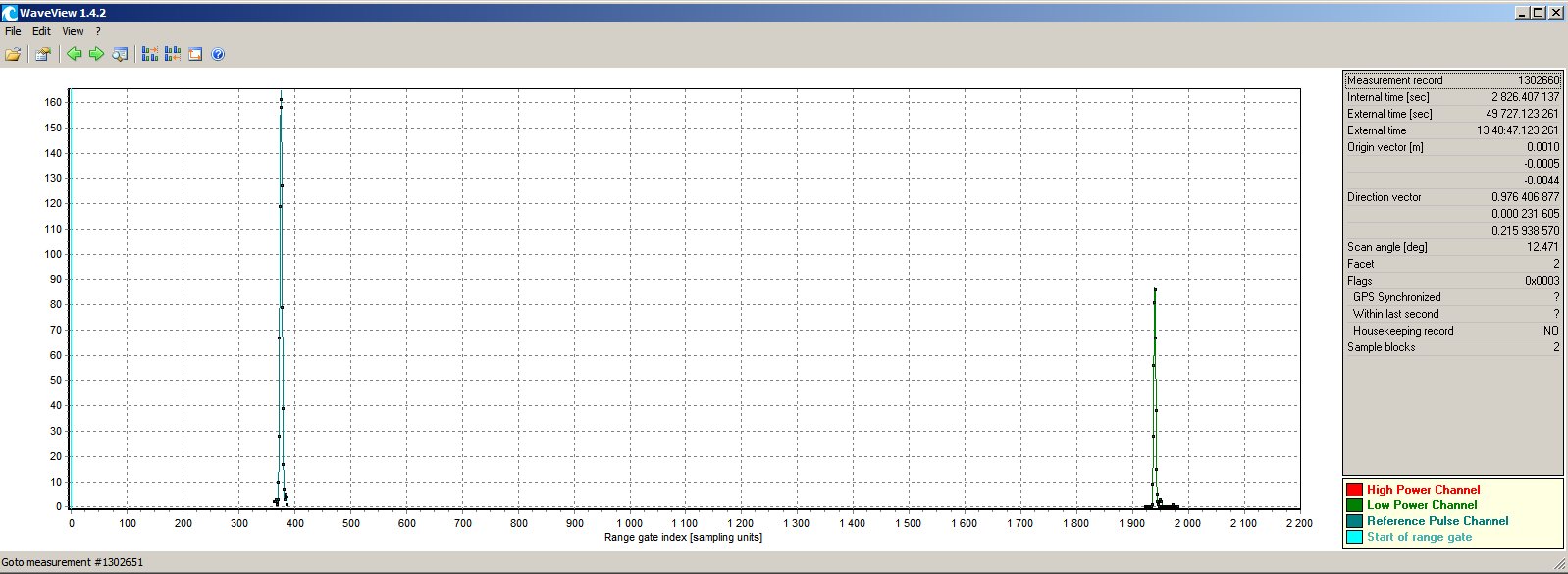
~ 2135 ns

~ 1760 ns



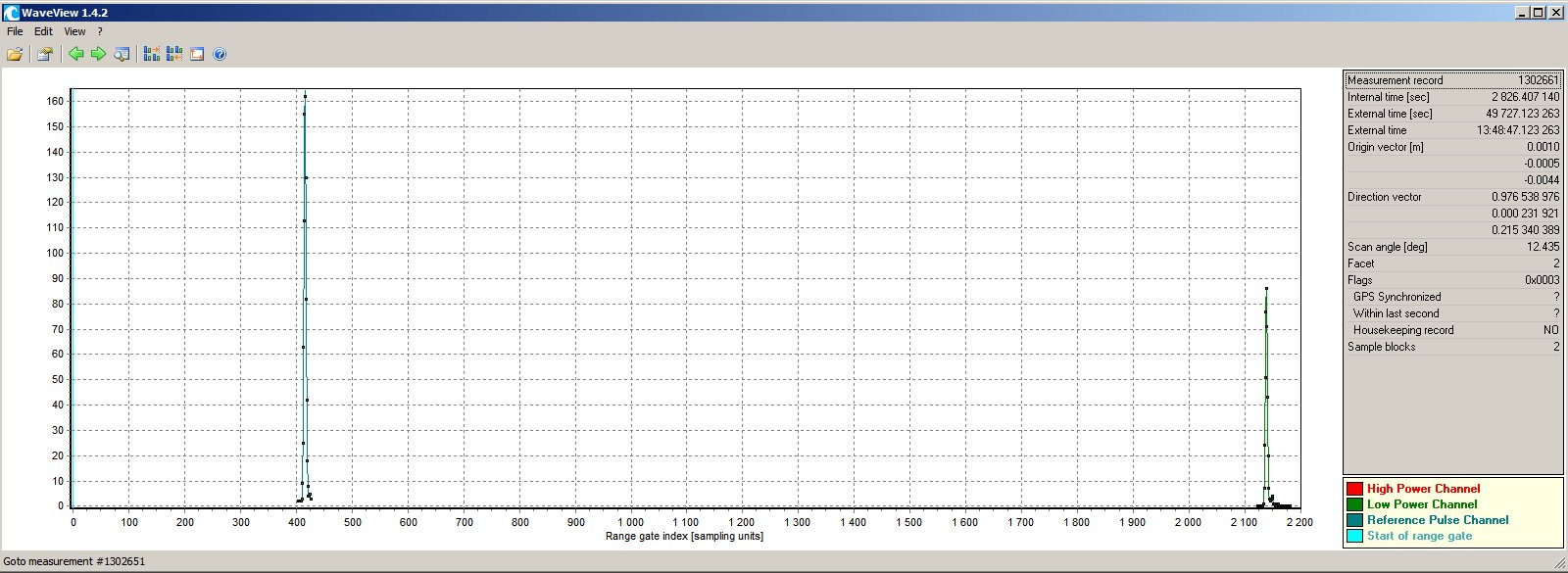
~ 1520 ns

~ 1935 ns



~ 1935 ns

~ 1560 ns



~ 1720 ns

~ 2135 ns

Figure 9. Time shifts between neighboring waveforms.

As presented in Fig. 9, position of reference waveform can be relative to **time\_sorg** by two values of about 375 ns or 415 ns and position of low power waveform of about 1935 ns or 2135 ns. These two values seems to be repeating according regular pattern modulo 2 (reference waveform) and modulo 7 (low power waveform, however, some exceptions occurred making impossible to remove these shifts automatically for the whole cube. Figure 10 shows intensities of time correlated waveform cube sliced along scan line and confirms the constant shifts mentioned above (marked in the Figure 9) since four (Fig. 10a) and two (Fig. 10b) parallel ground surfaces are distinguishable. The curvature of the surface is explained by the longer time of signal travel for the edge of scan line what is correct since the raw distance to the ground is larger when pulses are sent at larger angle from the nadir direction. This is also the reason than more bands are necessary to map all data. Note also empty waveforms for the last pulses in the scan line mentioned in 7). Constant shift occurred at some pattern may indicate systematic error which could be introduced intentionally (into Riegl Waveform Extraction Library).

|  |  |
| --- | --- |
|  |  |
| a) | b) |

Figure 10. Visualization of intensities of time correlated waveform cube, cut along single scan line where reference time is: a) time of start of sample block from reference channel, b) time of start of range gate.

1. References

ASPRS, 2008. LAS Specification Version 1.2. http://www.asprs.org/a/society/committees/standards/asprs\_las\_format\_v12.pdf

Jalobeanu A., Gonçalves G. R., 2012. The full-waveform LiDAR Riegl LMS-Q680i: from reverse engineering to sensor modeling. Proceedings of ASPRS 2012 Annual Conference, Sacramento, California - March 19-23, 2012.