MANAGING FULL WAVEFORM LIDAR DATA:
A CHALLENGING TASK FOR THE FORTHCOMING YEARS

F. Bretar[a], A. Chauve[a,c], C. Mallet[a], B. Jutzi[b]

[a] IGN-Laboratoire MATIS ; 4 av. Pasteur, 94165 Saint Mandé, FRANCE
[b] FGAN-FOM, Gutleuthausstr. 1, 76275 Ettlingen, GERMANY
[c] UMR TETIS Cemagref/Cirad/ENGREF-AgroParisTech, Maison de la Télédétection
500, rue J.F. Breton, 34093 Montpellier Cedex 5, FRANCE

Commission I, WG I/2

KEY WORDS: Full waveform, LIDAR, Researches, Data management, Format, Literature review.

ABSTRACT:

This paper proposes to summarize researches and new advances in full waveform lidar data. After a description of full waveform lidar systems, we will review different methodologies developed to process the waveforms (modelling, correlation, stacking). Applications on urban and vegetated areas are then presented. The paper ends up with recommendations on future research themes.

1. INTRODUCTION

Airborne laser scanning (ALS) is an active remote sensing technique providing range measurements between the laser scanner and the Earth topography by the time-of-flight between the emitted and backscattered laser pulse. Direct georeferencing processes turn such distance measurements into 3D point clouds with high accuracy. Even for small footprints, there may be several objects of different range within the beam corridor of the laser pulse that generate individual backscatter returns (echoes). Consequently, commercial systems typically measure first and last pulse and some are able to record up to six echoes for each emitted laser pulse. Due to the 3D structure of natural and artificial objects, the shape of the received pulses may be extremely complex and most ALS systems only provide the coordinates of these scattering objects.

During the last years, a new generation of airborne laser scanners have been developed which are able to record the signal of the entire backscattered laser pulse. These so called full waveform LiDAR systems give more control to the end user in the interpretation process of the physical measurement. Beside the range information, additional information about the structure of the illuminated surfaces can be determined. It has already been shown that full waveform data post-processing shows enormous potentialities for improving forest and urban mapping.

Full waveform systems can bring two important contributions and advantages for additional investigations possibilities. First the processing of the received signal can be used to recover all individual echoes. This enhances the point cloud density with regard to multiple echos (particularly on forest area) as well as the quality of the extracted features within the laser beam corridor. Second by modelling the received waveforms, additional features can be extracted from them. This is for example the amplitudes (also called intensity) and the pulse width derived by the standard deviation of a Gaussian-based decomposition of the waveform. These two values provide mixed information about the geometry and the reflectance properties of the illuminated surface.

Full waveform data is mostly used for forest research since the first experimental devices, with large footprint, were developed for this particular purpose. They produce more detailed description of vegetation structures. It is then possible to estimate, model and infer forest parameters more reliably.

The potentialities of full waveform data are definitively high and the forthcoming years will be particularly important for the development of methodologies, applications and softwares within the international LiDAR community. However, if the management of multiple echos data made appear standard format (*e.g.* LAS) and software solutions, managing full waveform LiDAR data in terms of efficiency have not been tackled so far. There are three main reasons that could explain this observation: firstly, these data just came out commercially, secondly, their volume is several times larger than multi-echos data and finally, the potentialities of these data have not been turned yet into reality with regard to multi-echo data.

After a brief reminder of the LiDAR systems themselves, we propose in the paper to summarize past researches concerning full waveform LiDAR processing with special emphasis on forest and urban areas. We will then sketch some proposals for a full waveform data format and finally we present new research directions.

2. BACKGROUND ON FULL WAVEFORM LIDAR SYSTEM

The physical principle of ALS consists in the emission of short laser pulses, with an width of 5-10 ns at full-width-half-maximum, from an airborne platform with a high temporal repetition rate of up to 200kHz. They provide a high point density and an accurate altimetric description within the diffraction cone. The two way runtime to the Earth surface and back to the sensor is measured. Then the range between the LiDAR system to the illuminated surface is recorded (Baltsavias, 1999). The emitted electromagnetic wave interacts with the objects surface depending on its wavelength: first with atmospheric components like particles refraction, absorption or scattering, known to have negligible influence if rain is excluded. The main influences on the laser light come from artificial or natural objects belonging to the illuminated surface. For ALS systems, near infra-red sensors are used (typical wavelengths from 0.8 to 1.55μm). Pulse repetition rate depends on the acquisition mode and on the flying altitude. Pulse release is done when the previous pulse recording is effective. Full waveform LiDAR systems record the entire signal of the backscattered laser pulse (Figure 1).

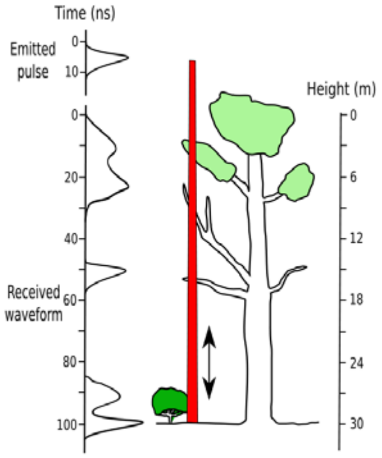


Figure 1: Transmitted and received waveform and the corresponding range in a complex wooded area with a small footprint LiDAR system.

The first operational system Laser Vegetation Imaging Sensor (LVIS), developed by the NASA, appeared in 1999 and demonstrated the value of recording the entire waveform for vegetation analysis (Blair et al., 1999). The first commercial full-waveform LiDAR system was introduced in 2004 (Hug et al., 2004). Today, most of LiDAR-involved companies (*e.g.* Riegl, Optech, Leica, Toposys) propose such an extension to their multiple pulse devices.

Full waveform systems sample the received waveform of the backscattered pulse at a maximum frequency of 1 GHz, which is equivalent to 1 GSamples/s. Such systems differ in sampling rate, in scan pattern and in footprint size. Most commercial systems are small-footprint, typically 0.3 - 1 m diameter at 1km altitude, depending on altitude and beam divergence.

To record full waveform LiDAR data, the main commercial manufacturers have added digitization terminals to their systems and increased the storage media capacity. Whatever the LiDAR system method, the constant digitization sampling period varies between 1 to 10 ns. The waveform is not integrally recorded but only for a predefined maximum number of samples. Indeed, it is necessary to avoid recording too many useless samples because they result into massive storage problems. For example, TopoSys ALTM systems can store up to 440 samples for each pulse. This is equivalent to a continuous vertical section of 66 meters (440 × 0.15 m per sample). The TopEye MarkII system saves 128 samples according to a predefined mode which is either ”first pulse and later ” (127 samples after the first) or ”last pulse and earlier ” (127 samples before the last detected). This means that full-waveform systems will not record both the echoes from the canopy and from the ground within a given waveform if the trees are taller than the ”recording length” of the system.

3. PROCESSING THE WAVEFORMS

The processing of waveform data starts in maximizing the number of relevant detected peaks within each signal.
A first approach consists in deriving a parametric formulation of the received signal. From the local maxima of the fitted function the range value is calculated and 3D points can be determined. Pulse properties (width and amplitude) can be calculated at the same time. Here, waveform processing consists in decomposing the signal *f(x)* into a sum of components $f_j(x)$ so as to characterise the different objects along the path of the laser beam:

$$y = f(x) = \sum_{j=1}^{n} f_j(x)$$

Considered as a sum of Gaussian functions initially in (Hofton et al., 2000) and (Wagner et al., 2006), various formulations of $f_j(x)$ have been tested in Chauve et al. (2007a): the Gaussian function (G), the Log normal function (LG) and the Generalized Gaussian function (GG).

The authors show that the waveform modelling is better using a GG function. More over, it introduces another pulse feature (α) that can be integrated in a segmentation process (Section 4.2).

$$\begin{aligned} f_{\mathrm{G},j}(x) &= a_j \exp\left(-\frac{(x-\mu_j)^2}{2\sigma_j^2}\right) \\ f_{\mathrm{L},j}(x) &= a_j \exp\left(-\frac{(\ln(x-s_j)-\mu_j)^2}{2\sigma_j^2}\right) \\ f_{\mathrm{GG},j}(x) &= a_j \exp\left(-\frac{|x-\mu_j|^{\alpha_j^2}}{2\sigma_j^2}\right) \end{aligned}$$

An other approach consists in applying signal processing methods based on the transmitted and the received waveform. The matched filter is computed by the normalized cross-correlation function R_{sr} between the transmitted waveform *s(t)* of the emitted pulse and the received waveform *r(t)* of the backscattered pulse.

$$R_{sr}(\tau) = \frac{\int_{t=-\infty}^{\infty} s(t)\cdot r(t+\tau)dt}{\sqrt{\int_{t=-\infty}^{\infty} s^2(t)dt \cdot \int_{t=-\infty}^{\infty} r^2(t)dt}}.$$

Here, echoes are local maxima of the correlation functions (Hofton & Blair, 2002; Kirchhof et al., 2008) .

This approach considers strong echoes for further processing, but weak echoes have to be detected, revised and processed again. In order to overcome this problem and retrieve partially occluded objects, waveforms can no longer be processed independently: local neighbourhoods have to be introduced. (Stilla et al., 2007) propose a waveform stacking strategy of several weak echoes within a local environment to increase the signal-to-noise ratio.

Alternatively, the knowledge of the local geometry can be used to improve the cross-correlation techniques. Considering the waveform as a convolution between the transmitted waveform and the surface response (which is unknown at first), the cross-correlation is performed by neglecting the surface response. (Kirchhof et al., 2008) have designed an iterative process to refine the matched filter in so as to integrate a modelled planar response to the matched filter, assuming that laser beam hits a planar surface with a given slope. The process consists in iteratively estimating the objects surface response to first improve range value estimation and discriminate between partially penetrable objects and impenetrable ones, then to extract and select surface primitives in 3D points expected to belong to impenetrable surfaces. Finally the authors estimate the relative surface slope (RANSAC) and compute the surface response, used as prior knowledge in a new step.

4. APPLICATIONS

4.1 Forest areas

Many studies have already been carried out for estimating forest parameters using multiple echo data derived from laser scanning systems: small footprint LiDAR systems, with high point density, can be used to extract trees in small areas (Brandtberg et al., 2003), their height and crown diameter (Persson et al., 2002), their volume (Naesset and Bjerknes, 2001), to classify them according to species (Holmgren and Persson, 2004), estimate their particular characteristics (Andersen et al., 2005) and even to measure the growth of the forest and detect trees that have been felled (Yu et al., 2004). Large footprint systems can reach the ground and the tree tops in dense, tall canopies. Woodland parameters can be estimated at large scale: density of population, coverage, biomass, etc. (Means et al., 1999).

Many algorithms have been developed for forest measurements (Hyyppä *et al.*, 2004). Full-waveform experimental systems with large footprint developed by the NASA have been successfully used in forest environments for measuring the canopy height (Lefsky et al., 1999) or the vertical distribution of canopy material (Dubayah and Blair, 2000).

Furthermore, the modelling of raw LiDAR signal recorded by recent small footprint industrial systems has already proved efficient in increasing the number of detected objects in comparison with data provided by multi-echo LiDAR systems for which real-time point extraction method is unknown to the end user (Persson *et al.*, 2005; Chauve *et al.*, 2007).

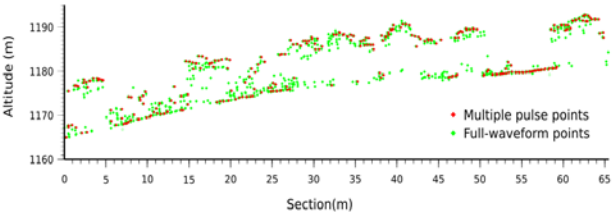


Figure 2: Profil of LiDAR points from multiple pulses data (red) and from full waveform data (green).

For deciduous vegetation, measurements in leaf-off conditions give a direct impression of the internal structure of the canopy (the main levels of vegetation) and are very useful to estimate the real "shape" of the trees (Reitberger et al., 2006).

In forested areas, waveforms can be composed of weak returns from both the top of the canopy and the ground, and also of distributed backscatters from the different layers of the vegetation, which leads to groups of overlapping echoes. Hardware detection algorithms, most of the time using thresholds, can hardly detect or separate such very low peaks or groups of echoes (Figure 1). Therefore, processing the waveforms allows to control the point extraction method and to make it more efficient: an improved peak detection was shown to be very successful to extract additional objects in the received waveforms and to detect from 30% to 130% additional points depending on vegetation density (Chauve et al, 2007b). These points are located mainly within the canopy and in highly dense understory (Figure 2). Detecting such weak echoes allows a better describtion of the 3D vegetation structure and the ground. As a consequence, Digital Terrain Model (DTM), Digital Surface Model (DSM), and the derived Canopy Height Model (CHM) are expected to be significantly improved, but is still to be validated with reference data.

4.2 Urban areas

When processing full waveform data on urban areas, one can observe that the point density is barely improved. Thanks to waveform analysis, more echoes are found in tree canopies, whereas LiDAR pulses cannot penetrate rigid, opaque structures such as buildings and streets. Consequently, a single peak is still present within the waveform and can be detected by signal processing algorithms. Multiple echoes can appear on building edges and superstructures. Indeed, post-processing algorithms enable to extract weak echoes not found by on-line detection techniques. In urban areas, it is observed when the laser beam hits for instance building edges. The resulting waveform is therefore composed of distributed backscatters of the roof and the ground, which can often not be separated by hardware detection algorithm using fixed thresholds.

Modelling LiDAR waveforms permits to extract 3D point clouds featuring supplementary useful parameters in addition to the traditional (x, y, z) coordinates and to perform subsequently point cloud segmentation based on these parameters. The standard features are range, amplitude and width. Nevertheless, analysing a 3D point cloud processed from full waveform data with model parameters does not allow to define particular behaviour for each urban object. The echo is wider on the canopy with regard to roads or grass ones but a wide echo with low amplitude does not necessarily comes from vegetation. High amplitudes are noticed on grass and asphalt and variable amplitude on the roofs of buildings, depending on the roof materials. In fact, roads and building roofs are made from different types of material and, therefore, have different similar values to natural objects (Gross et al., 2007). Moreover, it is visually possible to distinguish between different urban materials hit by a laser beam but a more in-depth processing is required to recognise roof materials (Jutzi and Stilla, 2003).

Consequently, simple algorithms aiming at segmenting building and vegetated areas, *e.g.*, decision tree approach with empirical thresholds, lead to a certain rate of incorrect classification (Duong et al., 2006; Ducic et al., 2006). A more reliable approach to detect vegetated areas is proposed in (Gross et al., 2007), based on the eigenvalues of the covariance matrix computed for each point with the intensity values in a cylindrical and spherical environment. Another method based on a supervised classifier (SVM) has been developed (Mallet et al., 2008). It provides a suitable segmentation of ground, building and vegetation areas but requires classical geometric features in addition to parameters extracted for waveform modelling.
To achieve more advanced point segmentation in urban areas

and to understand how the pulse interacts with the surface, a theoretical knowledge of the influence of the geometric and radiometric properties of the illuminated surface (*i.e.*, the differential laser cross-section) on the shape of the waveforms is required. Quantifying specifically both geometric and radiometric influence of a object on the received waveform is undoubtedly the most promising line of research for that purpose.

5. FUTURE RESEARCH DIRECTIONS

Researches on full waveform data are still at their beginnings, but seem very promising in terms of automatic data extraction. However, if the technology seems well managed by manufacturers, many questions will arise for future work.

Whatever the scenario type will be, if it is clear that full waveform LiDAR data provide range information, it is still expected to be fully proved that they contain physical assessments of the backscattering properties of the illuminated surface. Among other interesting points, two of them seem essential and have to be investigated in priority:

- The influence of the surface geometry and radiometry onto the shape of the waveform.
- The influence of the lasers wavelength onto the measurement itself.

Answering these points needs a modelling step that could be based on a ray-tracing model or even more complex, a electromagnetic model, which bases on the wave characteristic of the light, *e.g.* to investigate speckle effects. Such approaches will consider the LiDAR footprint size, the beam divergence, the sensor FOV, the wavelength, the direction of the laser beam propagation etc., but also the characteristics of the scenario (tree species, building geometry...). It also necessitates an experimental step to build a data base of optical responses in various optical wavelength over various features.

Furthermore, most of topographic LiDAR systems work with a single wavelength. The recording of several waveforms, each of them emitted at a different wavelength could enhance the object description. These so called hyperspectral laser systems could deliver for instance multiple intensity information of surfaces instead of only monochromatic information. Finally, scientists of the ISPRS community would take high benefit of having a better knowledge of commercial LiDAR systems, particularly of its system specifications, which is usually different for each sensor and most of the relevant information is kept secret by the manufacturers.

The extensive use of full waveform LiDAR data leads to think of a new data format and data management system as LAS format and TerraScan© for point clouds. The format could be based on a multiple layer structure in the sensor geometry, each of them linked to the others by pointer arrays (Figure 3). Among important layers, there are a **raw data layer** containing all waveforms, a **georeferencing layer** containing the trajectory and sensor information interpolated at each measurement and a **modelling layer** containing the parameters of the analytic description of the waveform. A XML meta file could describe each layer. An orthorectified geometry should be generated in terms of GUI and linked to the sensor geometry.

6. CONCLUSION

We have presented in this article the main lines of researches on full waveform LiDAR data performed, especially in the ISPRS community. Even if these data have been used for only a short time now, it seems that fruitful and promising results have come out in the recent years. Many problems have to be sorted out before using these data as multi-echo LiDAR data are. Nevertheless, the tasks are extremely challenging since it is a new and wide research area, which begin to deals with physical remote sensing.

REFERENCES

Andersen, H.-E., McGaughey, R., Reutebuch, S., 2005. Estimating forest canopy fuel parameters using LIDAR data. *Remote Sensing of Environment* 94 (6), pp. 441—449.

Baltsavias, E., 1999. Airborne Laser Scanning : Basic relations and formulas. *ISPRS Journal of Photogrammetry and Remote Sensing*, 54 (2-3), pp. 199–214.

Blair, J., Hofton, M., 1999. Modeling Laser Altimeter Return Waveform Over Complex Vegetation Using High-Resolution Elevation Data. *Geophysical Research Letters* 26 (16), 2509—2512.

Brandtberg, T., Warner, T.A., Landenberger, R.E. and McGraw, J.B., 2003. Detection and analysis of individual leaf-off tree crowns in small footprint, high sampling density LiDAR data from the eastern deciduous forest in North America. *Rem. Sens. Environ.*, 85 (3),pp. 290-303.

Chauve, A., Mallet, C., Bretar, F., Durrieu, S., Pierrot-Deseilligny, M. and Puech, W., 2007. Processing full waveform lidar data: modelling raw signals. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*, Vol. 39 (Part 3/W52), Espoo, Finland, pp. 102–107.

Chauve, A., Durrieu, S., Bretar, F., Pierrot-Deseilligny, M., Puech, W., Processing full-waveform LiDAR data to extract forest parameters and digital terrain model: validation in an alpine coniferous forest, 2007b, *Proc. of the ForestSat Conference,* Montpellier, France

Dubayah, R., Blair, J., 2000. LiDAR Remote Sensing for Forestry Applications, *Journal of Forestry* 98 (6), pp. 44–46.

Ducic, V., Hollaus, M., Ullrich, A., Wagner, W., Melzer, T., Feb. 2006. 3D Vegetation mapping and classification using full-waveform laser scanning. In: *EARSeL and ISPRS Workshop on 3D Remote Sensing in Forestry*. Vienna, Austria, pp. 211–217.

Duong, H., Pfeifer, N., Lindenbergh, R., May 2006. Full waveform analysis: ICESAT laser data for land cover classification. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. 36 (Part 7). Enschede, The Netherlands.

Gross, H., Jutzi, B., Thoenessen, U., 2007. Segmentation of Tree Regions using Data of a Full-Waveform Laser. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. 36 (Part 3/W49A). Munich, Germany, pp. 57–62.

Holmgren, J., Persson, A., 2004. Identifying species of individual trees using airborne laser scanner. Remote Sensing of Environment 90 (4), pp. 415—423.

Hofton, M.A., Minster, J.B., Blair, J.B., 2000. Decomposition of laser altimeter waveforms. *IEEE Transactions on Geoscience and Remote Sensing* 38 (4), pp. 1989–1996.

Hug, C., Ullrich, A., Grimm, A., 2004. Litemapper-5600 -A Waveform-Digitizing LIDAR Terrain and Vegetation Mapping System. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. 36 (Part 8/W2). Freiburg, Germany, pp. 24—29.

Hyyppä, J., Hyyppä, H., Litkey, P., Yu, X., Haggrén, H., Rönnholm, P., Pyysalo, U., Juho Pitkänen, J. and Maltamo, M., 2004. Algorithms and Methods of Airborne Laser-Scanning for Forest Measurements. *In International conference "Laser-scanners for Forest and Landscape Assessment"*. Vol. 36 (Part 8/W2), pp. 82-89.

Jutzi, B., Stilla, U., 2003. Laser pulse analysis for reconstruction and classification of urban objects. In: International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences. Vol. 34 (Part 3/W8). Munich, Germany, pp. 151–156.

Kirchhof, M., Jutzi, B., Stilla, U., 2008. Iterative processing of laser scanning data by full waveform analysis. *ISPRS Journal of Photogrammetry & Remote Sensing* 63, pp. 99–114.

Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A. and Harding, D., 1999. LiDAR remote sensing of the canopy structure and biophysical properties of Douglas-fir western hemlock forests. *Remote Sensing of Environment*, 70(3): pp. 339-361.

Mallet, C., Soergel, U., Bretar, F., 2008. Analysis of full-waveform LiDAR data for an accurate classification of urban areas. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. 37 (Part 3A). Beijing, China.

Means, J., Acker, S., Harding, D., Blair, J., Lefsky, M., Cohen, W., Harmon, M., McKee, W., 1999. Use of large-footprint scanning airborne LiDAR to estimate forest stand characteristics in the western cascades of Oregon. *Remote Sensing of Environment* , 67(3), pp. 298—308.

Naesset, E. and Bjerknes, K.-O., 2001. Estimating tree heights and number of stems in young forest stands using airborne laser scanner data. *Remote Sensing of Environment*, 78(3): pp. 328-340.

Persson, A., Holmgren, J. and Soderman, U., 2002. Detecting and measuring individual trees using an airborne laser scanner. *Photogrammetric Engineering & Remote Sensing*, 68(9): 925-932.

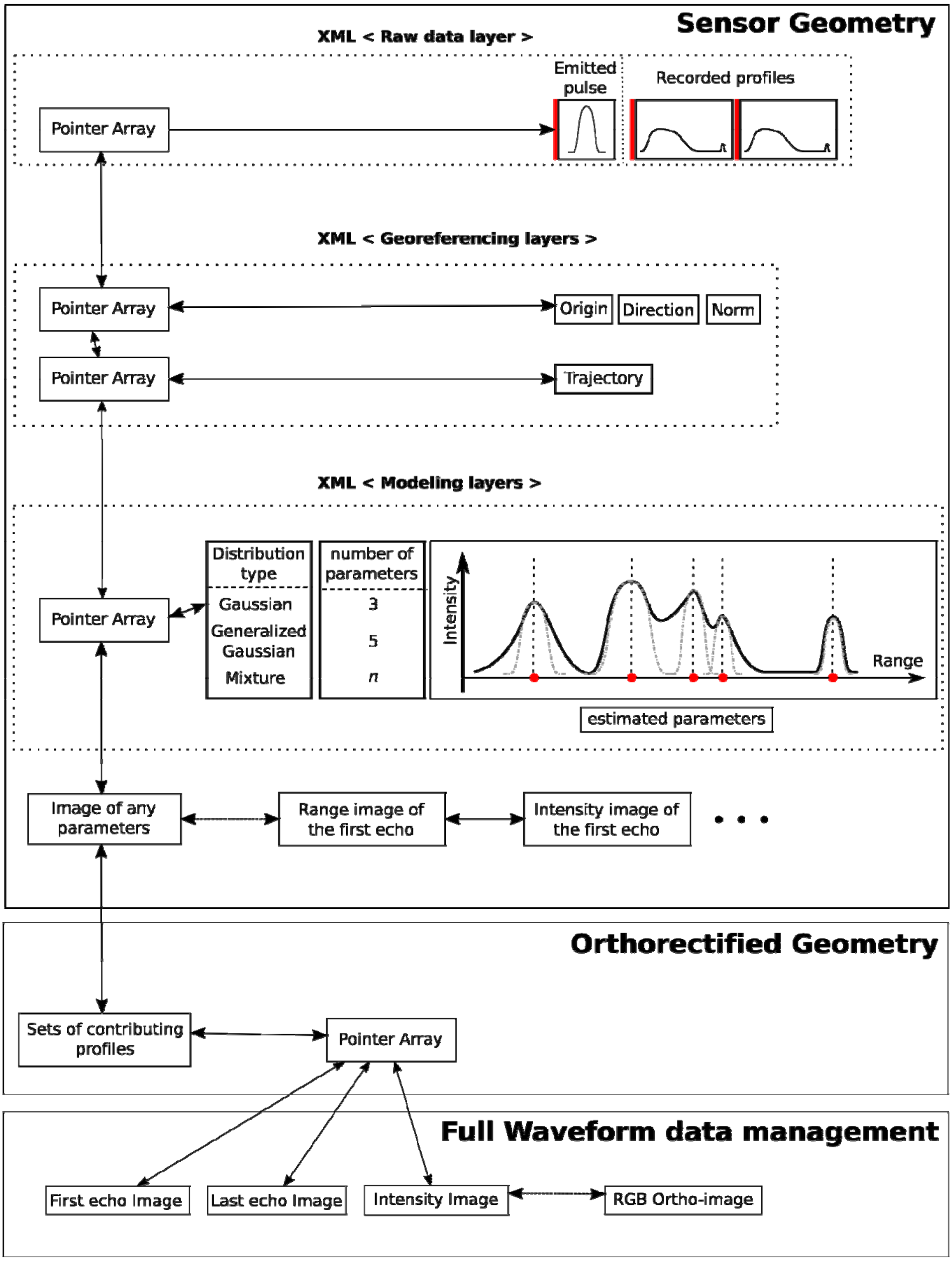
Persson, A., Söderman, U., Töpel, J., Alhberg, S., 2005. Visualization and Analysis of Full-Waveform Airborne Laser Scanner Data. In: *International Archives of Photogrammetry, Remote Sensing and Spatial* Information Sciences. Vol. 36 (Part 3/W19). Enschede, The Netherlands.

Reitberger, J., Krzystek, P., Stilla, U., 2006. Analysis of full-waveform LiDAR data for tree species classification. In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences.* Vol. 36(Part 3). Bonn, Germany, pp. 228—233.

Stilla, U., Yao, W., Jutzi, B., 2007, Detection of weak laser pulses by Full Waveform stacking, In: *International Archives of Photogrammetry, Remote Sensing and Spatial Information Sciences*. Vol. 36 (Part 3/W49A). Munich, Germany, pp. 25—30.

Wagner, W., Ullrich, A., Ducic, V., Melzer, T. and Studnicka, N., 2006. Gaussian decomposition and calibration of a novel smallfootprint full-waveform digitising airborne laser scanner. *ISPRS Journal of Photogrammetry and Remote Sensing* 66, pp. 100–112.

Yu, X., Hyyppä, J., Kaartinen, H., Maltamo, M., 2004. Automatic Detection of Harvested Trees and Determination of Forest Growth Using Airborne Laser Scanning. *Remote Sensing of Environment* 90 (4), pp. 451--462.


Sensor Geometry
XML < Raw data layer >
Pointer Array
Emitted pulse
Recorded profiles
XML < Georeferencing layers >
Pointer Array
Origin
Direction
Norm
Pointer Array
Trajectory
XML < Modeling layers >
Pointer Array
Distribution type
Gaussian
Generalized Gaussian
Mixture
number of parameters
3
5
n
Intensity
Range
estimated parameters
Image of any parameters
Range image of the first echo
Intensity image of the first echo
• • •
Orthorectified Geometry
Sets of contributing profiles
Pointer Array
Full Waveform data management
First echo Image
Last echo Image
Intensity Image
RGB Ortho-image