Challenges in Nationwide LiDAR Datasets: Noise and Outliers

Pedro Llorens – LiDAR Session

Geospatial World Forum 2024

Rotterdam – 15th of May 2024

Presentation

- Pedro Llorens – MSc Aerospace Engineer, expert in LiDAR sensors and operation – Experience selling Leica Geosystems airborne sensors and aerial survey services – Sparked 1st Geiger mode/single photon LiDAR comercial project in Europe.
- Currently Business Development Manager at SPASA.
  - Spanish aerial survey/LiDAR company based in Madrid, in the market for 40 years.
  - Currently working in nationwide imagery and LiDAR projects and digital twins.
- Countries like France, Portugal, Spain, etc. are using post COVID recovery funds like NextGenerationEU to finance increased resolution nationwide LiDAR coverages.
- SPASA has been awarded 70% of the Spanish 3rd LiDAR Coverage: >340,000km² – 11M€ - 3 years, the biggest aerial survey contract awarded to a single company in a public open tender in Europe.

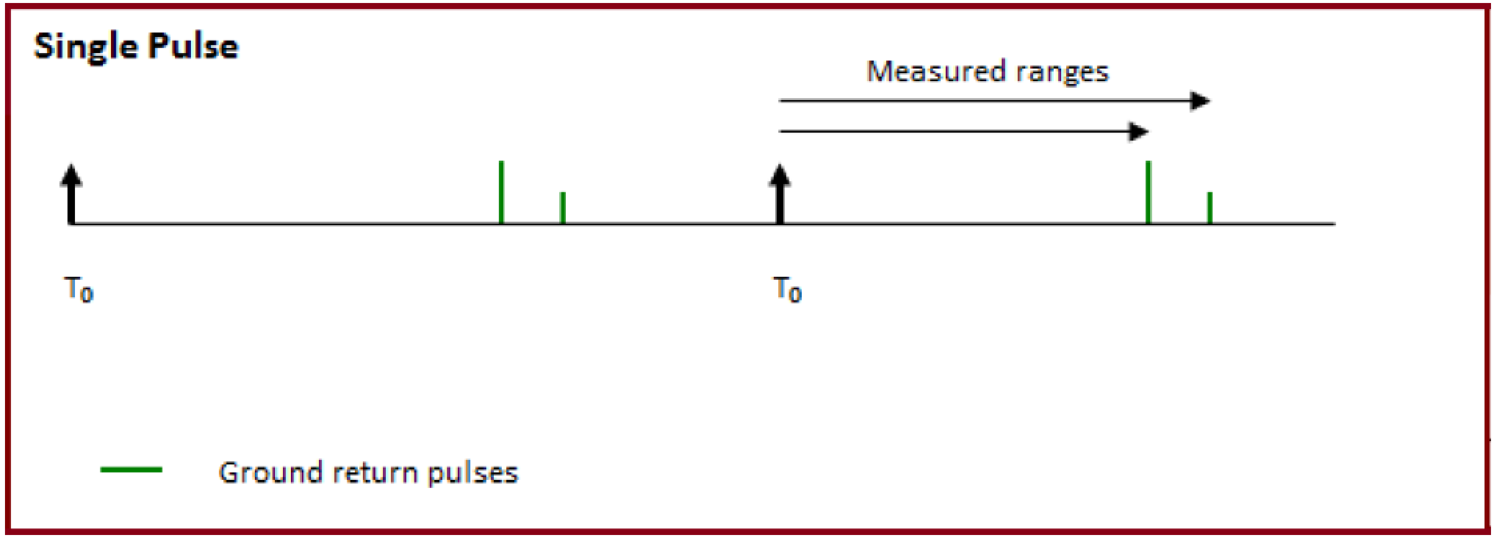
Contents

- Several errors or problems can appear in LiDAR point clouds:
  - Noise – false atmospheric returns
  - Ghost data
  - Outliers
  - Blind zones or pulse transition zones
- Those effects are more intense or exclusive of multipulse systems

TOF Time of Flight – PRF limitation

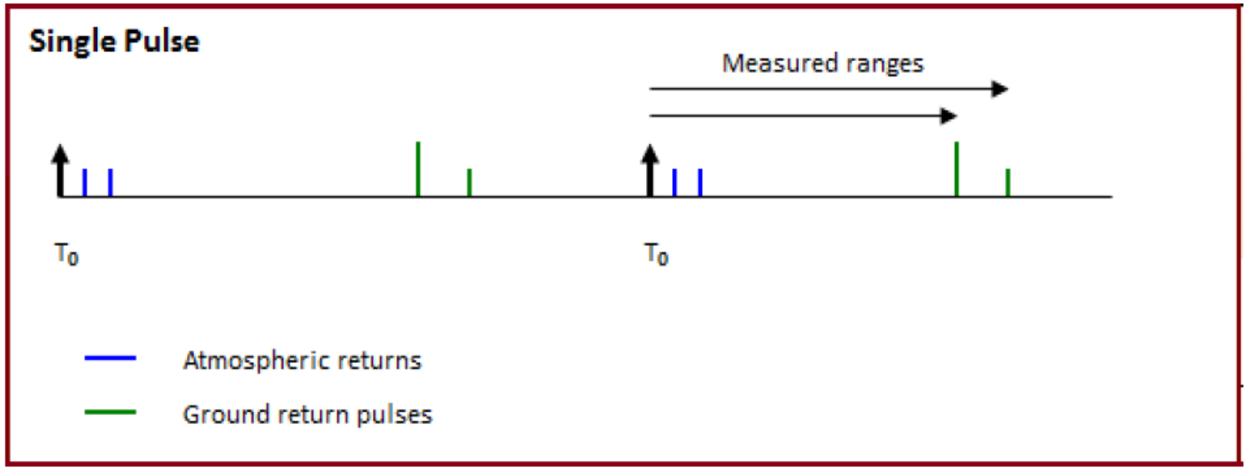
- LiDAR systems determine the distance to the ground or the illuminated object by measuring the time between the emission of the pulse and the return reception TOF (Time Of Flight) travelling through the air at the speed of light.

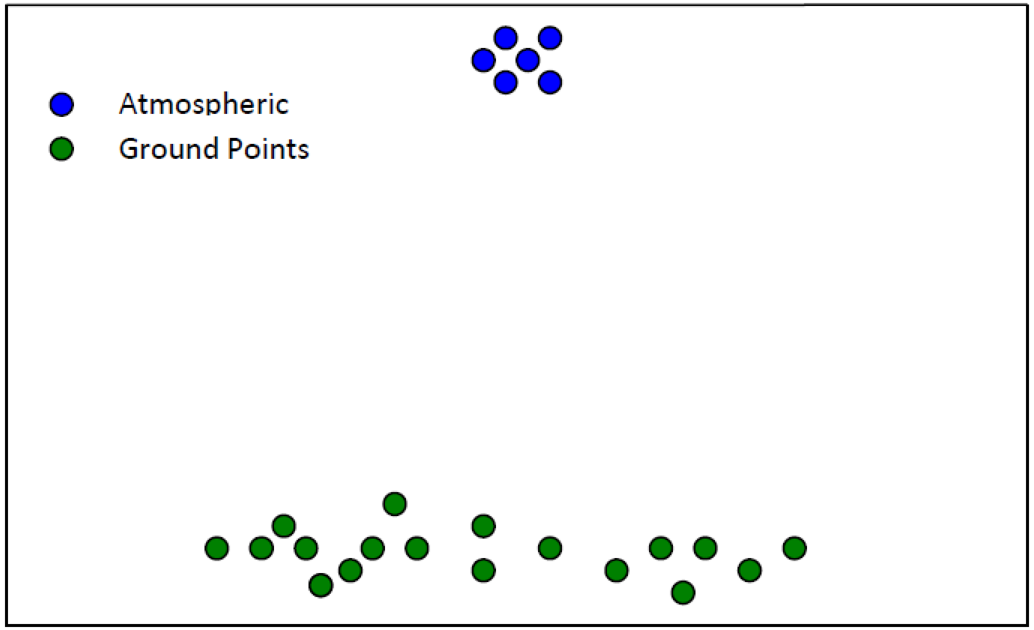
$$D = T \cdot c/2 \quad \text{where } c = 3 \cdot 10^8 m/s$$



Monopulse atmospheric return

- When the pulse is just fired and keeps it maximum power, tiny objects in the air close to the sensor, like dust, water particles or droplets in areas with haze, humidity, pollution, etc. can produce false returns also called atmospheric returns.
- Easily avoided with “gate” closing the reception of returns during an interval after firing the pulse or filtering points above a certain height



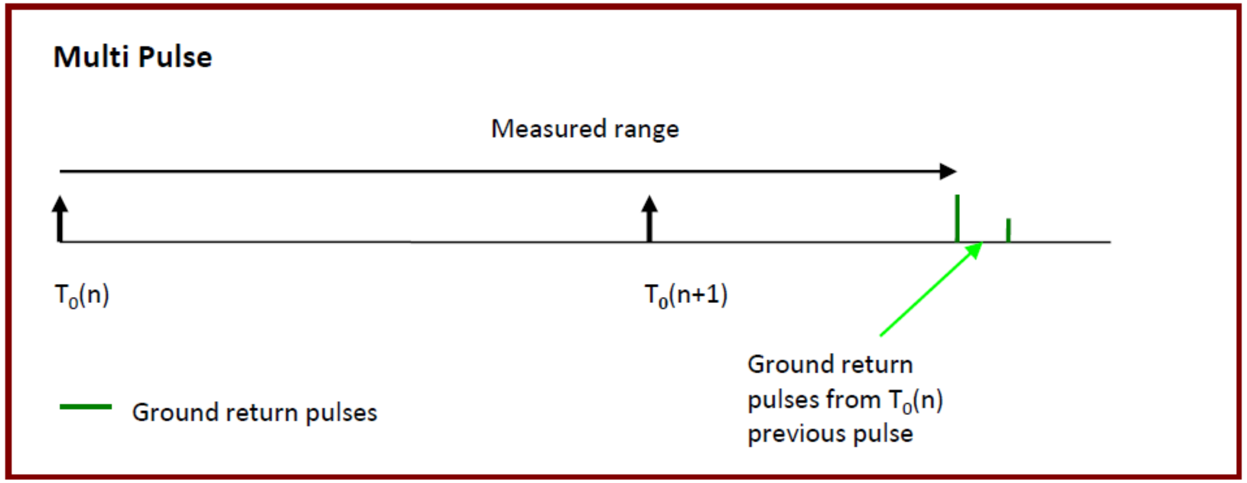


Speed of light PRF limitation

- When flying 1.000m AGL,TOF is $6.67 \cdot 10^{-6}$s or 6.67ns. If the sensor has to wait till the return arrives back to the receptor before emitting the following pulse, PRF would be limited to 150KHz. (145KHz for a 30° FOV).
- Flying at 110kts we would be limited to a point density of 4.76 pts/m2.
- We need higher PRF to achieve higher density but that is not possible. We would be forced to reduce speed > helicopter?
- We would have to reduce PRF to increase flying height.

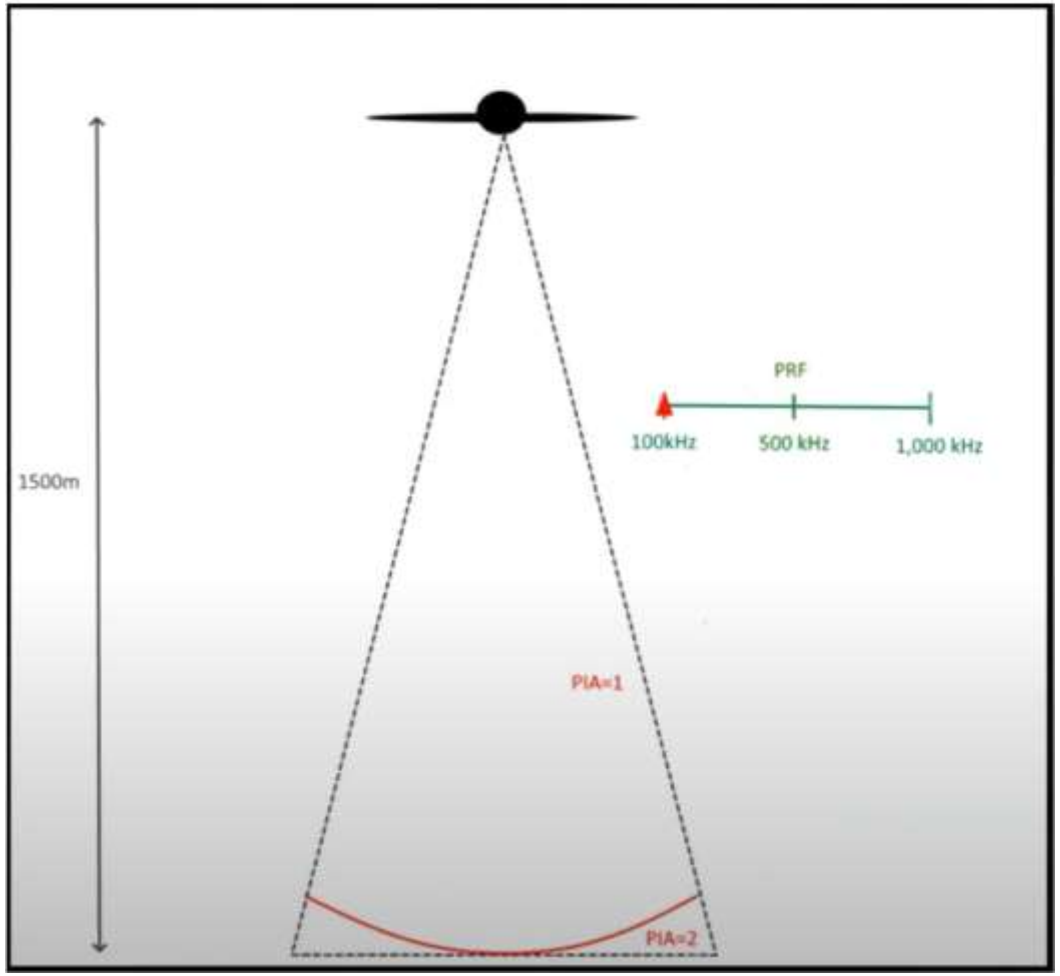
Multipulse technology

- Solution: Firing a 2nd pulse and successive pulses before the return of the 1st one reaches back the sensor. Is this possible?
- The PRF limitation due to speed of light limitation was overcome from 2006, when major LiDAR manufacturers introduced systems that could handle several pulses in the air and their associated returns. Leica (MPiA Multiple Pulse in Air), Optech (CMP™ Continuous Multipulse) and Riegl (MTA).
- The challenge is assigning each return to the right pulse within what are called PiA (Pulse In Air) zones with methods like the analysis of neighbouring ground points.



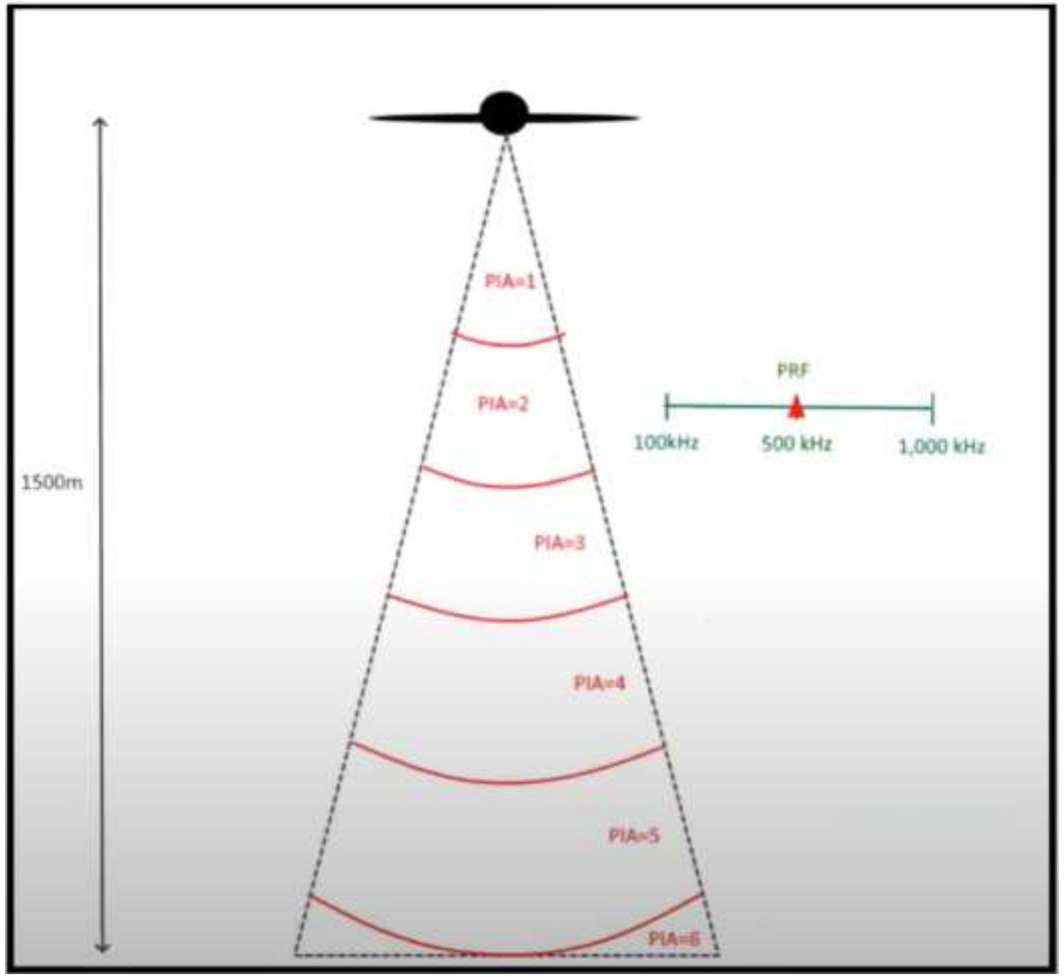
Pulse frequency and PiA zones

- Height 1,500m
- 100 KHz
- 1 zone PiA
- Width 1,500m



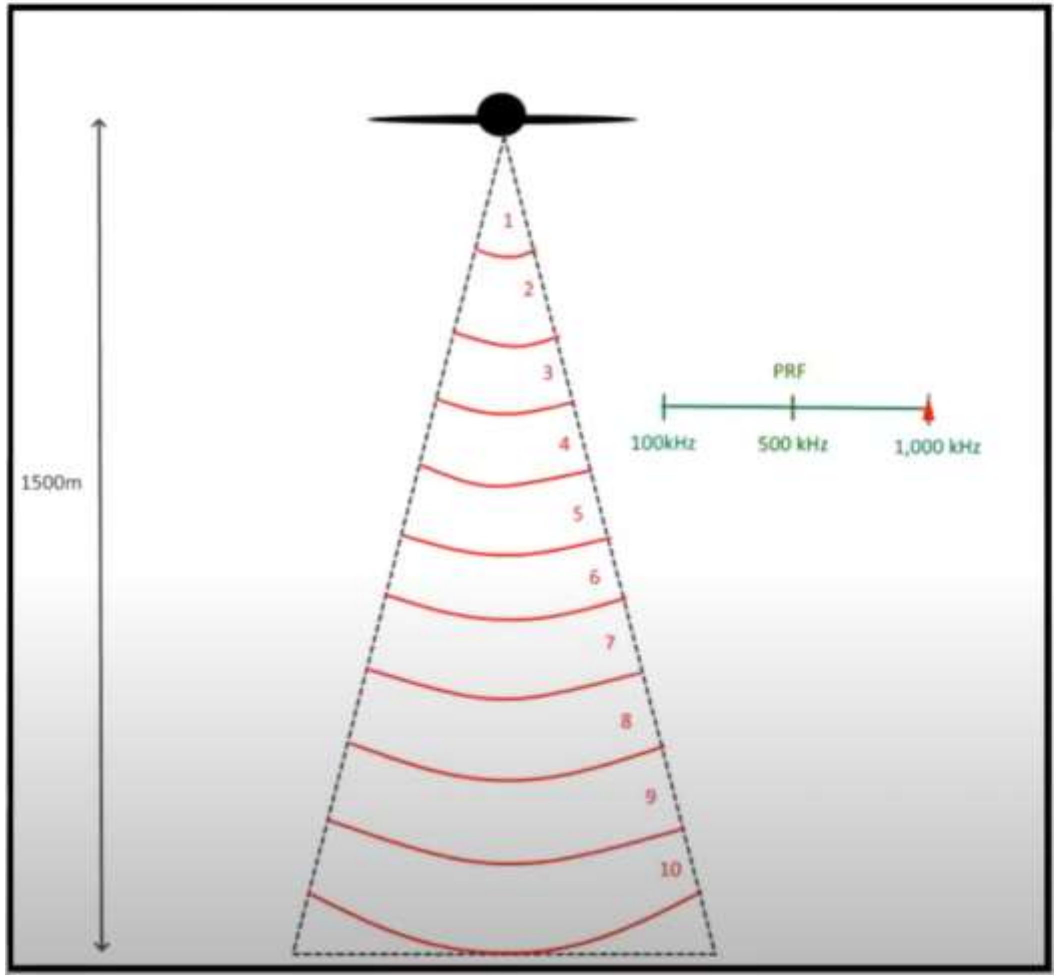
Pulse frequency and PiA zones

- Height 1,500m
- 500 KHz
- 5 zones PiA
- Width 300m



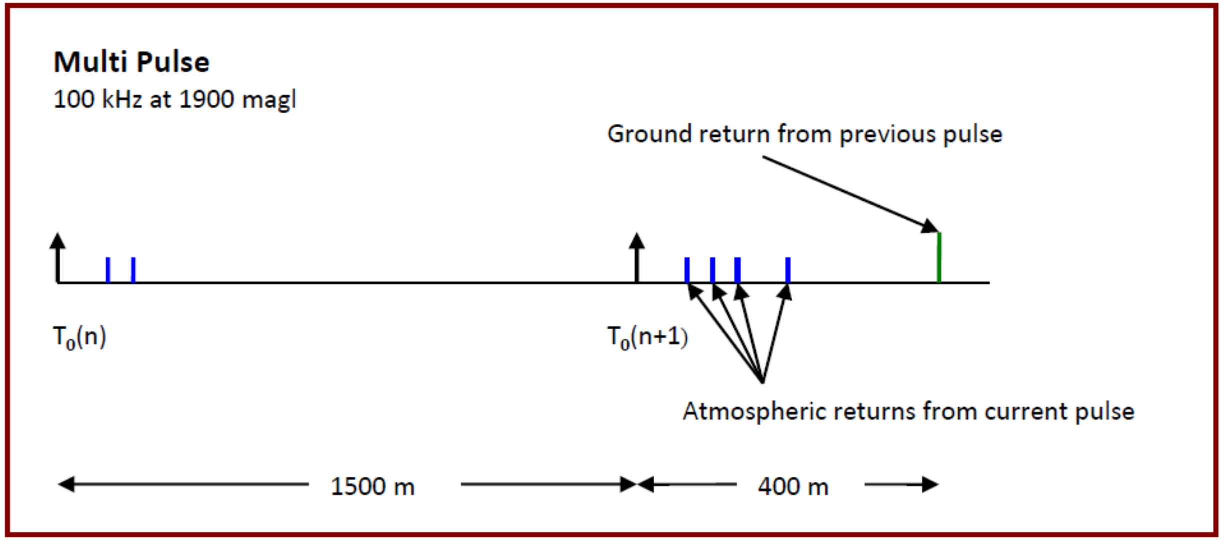
Pulse frequency and PiA zones

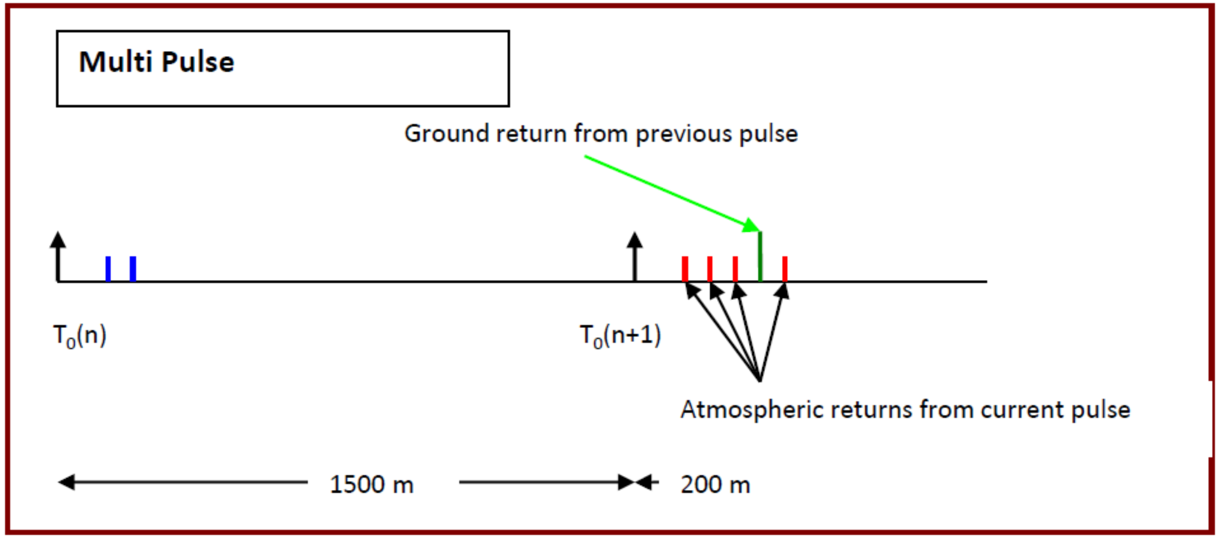
- Height 1,500m
- 1 MHz
- 10 zones PiA
- Width 150m



Multipulse atmospheric return (I)

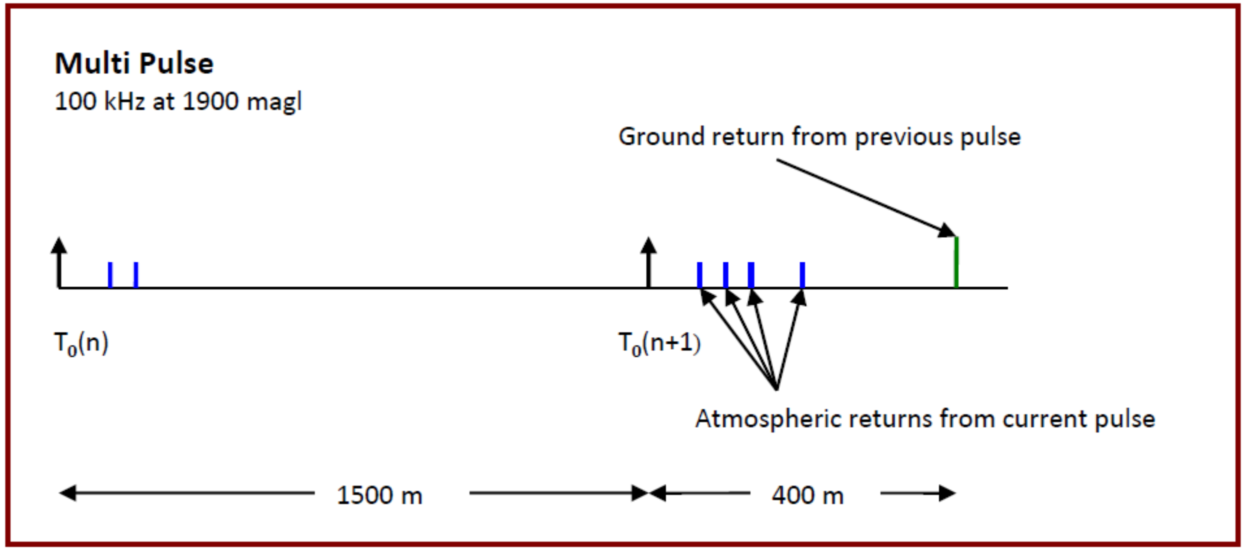
- Using a gate to avoid atmospheric returns could make themultipulse system blind during short intervals (blind zones) so gated interval has to be reduced.
- Increased sensitivity to atmospheric returns

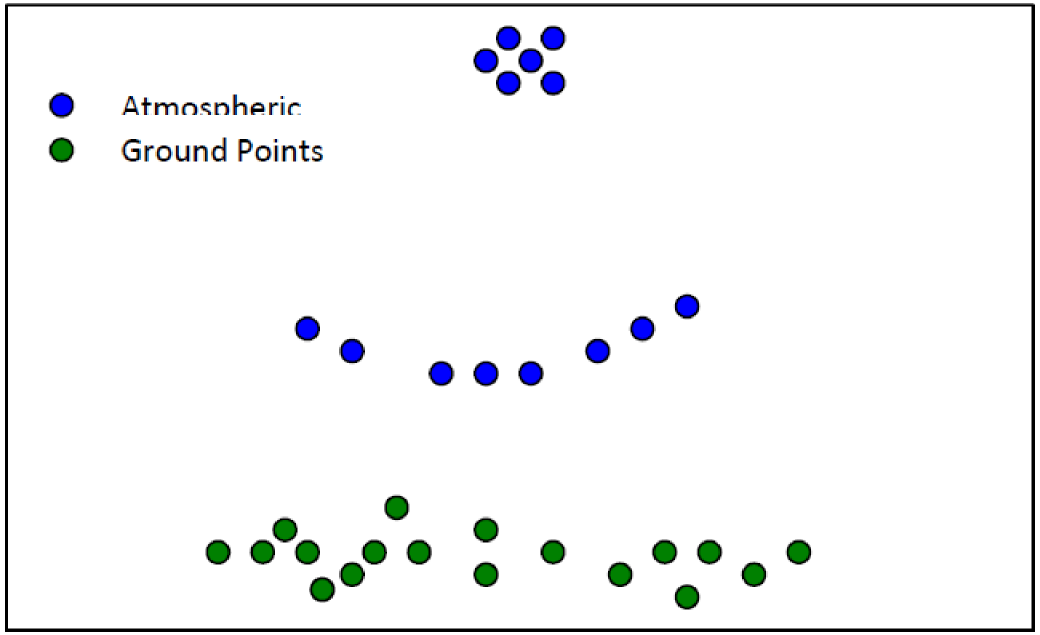




Multipulse atmospheric return (II)

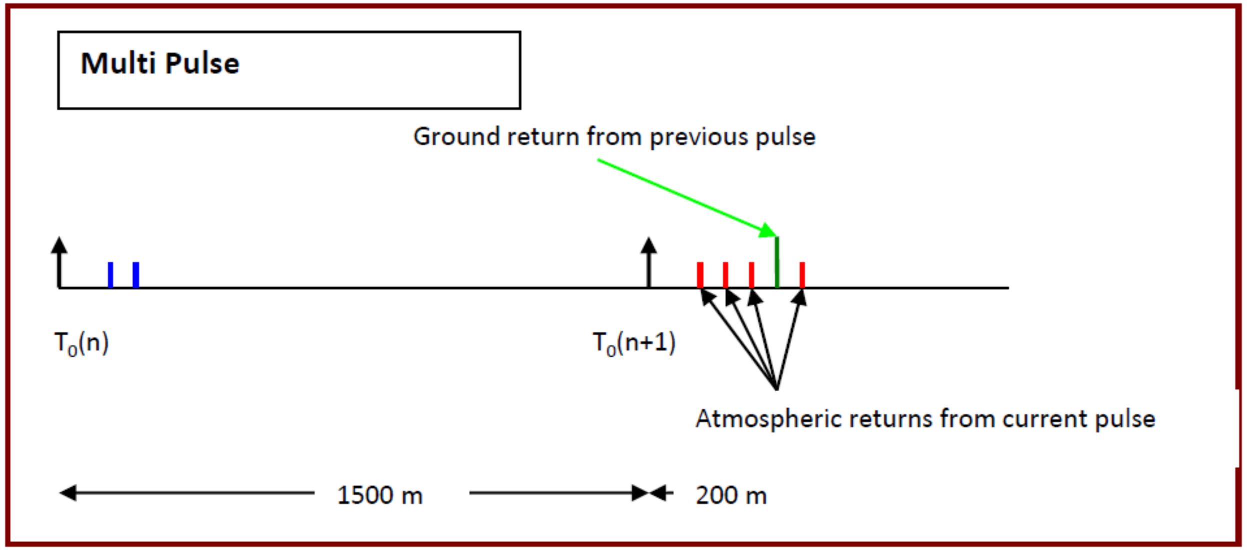
- Atmospheric returns assigned to the wrong pulse/PiA zone without mixing with ground points.

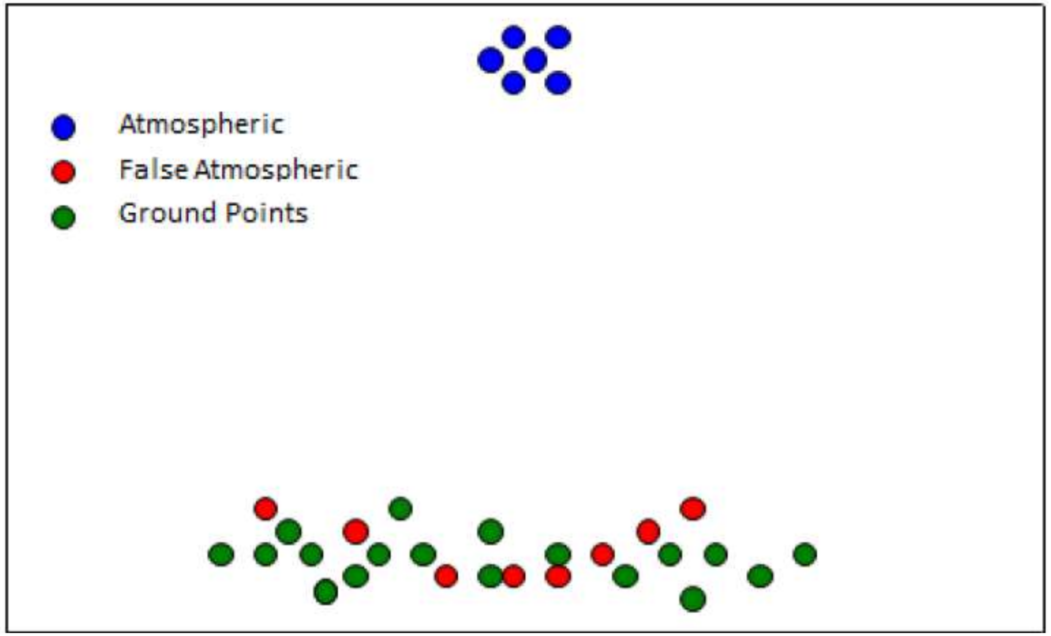




Multipulse atmospheric return (III)

- Atmospheric returns assigned to the wrong pulse/PiA zone mixed with ground points.



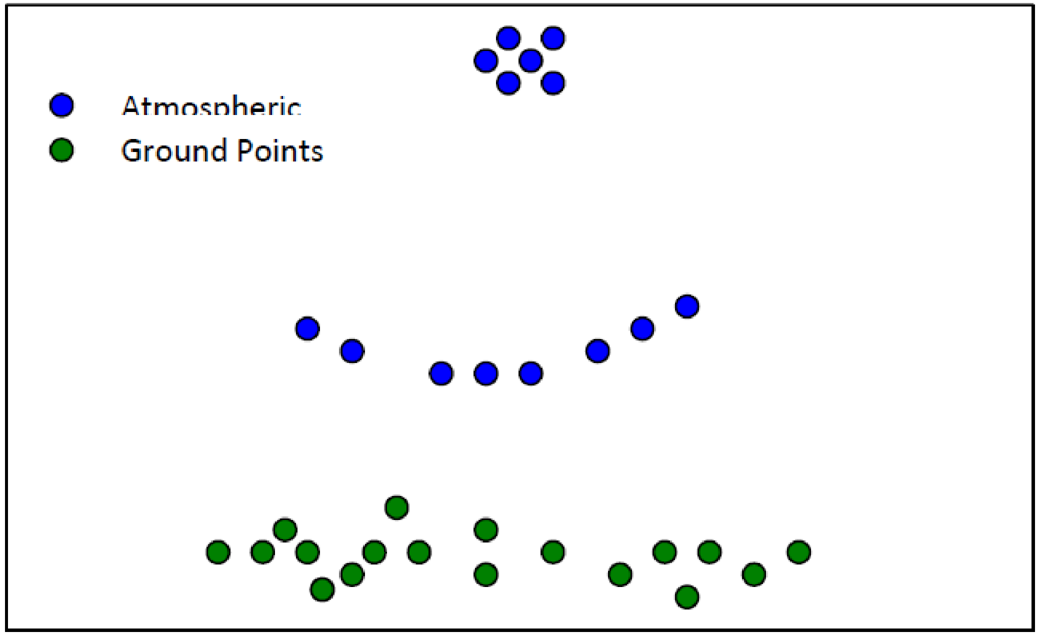


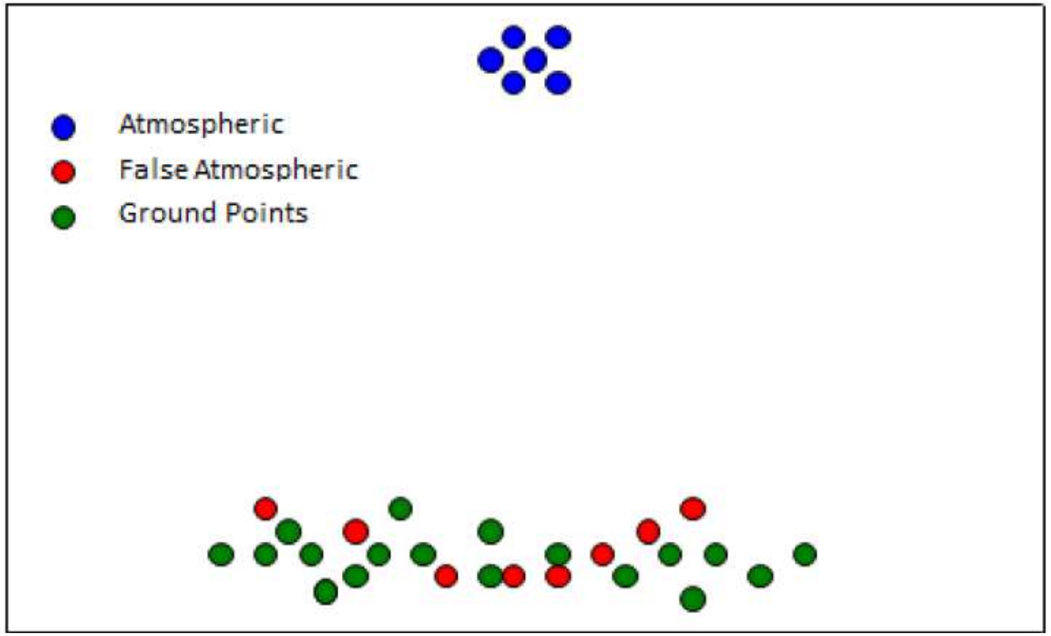
Atmospheric returns - noise

- Due to multipulse technology, atmospheric returns could be projected at a different height than they really are (close to the sensor), in some cases, down near ground points.
- What is usually called “noise” is result of a wrong allocation of atmospheric returns from the vicinity of the sensor to previous pulses by errors in the multipulse system. There is a miscalculation of the source pulse of those atmospheric returns that assigns them to the wrong PiA zone and the wrong height.
- As performance increases, the problema is more pronounced:
  - Higher PRF increases the number or PiA zones meking them shorter
  - Increased flying height requires more sensitive sensors (or higher pulse energy) making them more susceptible to false atmospheric returns.

Noise filtering/classification

- If those false returns are isolated, they can be easily classified or removed but when they are mixed with ground point things get complicated.
- There are several methods and tools to classify or label those false returns.





Traditional classification methods

- Manual selection methods where the operator selects the points to be classified one by one or using fences. Labour-intensive and lengthy. In our case a 2x2km sheet may contain between 25 to 90 million points depending on relief, noise presence, overlap, etc.
- Conventional classification tools, usually based in isolated points filtering to classify sparse noise, analysis of neighbouring points, etc. whose behaviour is defined by parameters that have to be determined heuristically by trial and error methods and only applicable to a determined area of certain flight. Also labour-intensive and lengthy.
- Where noise is closely mixed with ground and object points, it could be added a criteria based on return intensity, much lower in atmospheric returns as they are caused by tiny objects.

Artificial Intelligence

- ML tools are applied to classification tasks (e.g. Teledyne Optech GNC Galaxy Noise Classifier based on AI and ML techniques.
- The user defines a “confidence level” that indicates how convinced has to be the system that a point is noise, to effectively classify it as noise. High values of confidence level will classify as noise only points that fulfil almost all criteria to be considered noise, so some noise points could remain unclassified, and low confidence levels will consider noise points showing less indicators, so ground or object points could be classified as noise.

Thanks

- Thanks for your attention
- Thanks to Sanjay Kumar and GWF2024 organization
- Thanks for those amazing 6 years to SPASA in my last week at the company
- Next Monday I start a new position at Dielmo 3D