**Introduction**

LiDAR has become a mainstream technology for surface data acquisition for a variety of large-scale mapping applications. Since demand for denser LiDAR data is growing, higher and higher pulse repetition frequency (PRF) LiDAR systems are introduced. pulse repetition frequency (PRF) is the number of pulses of a repeating signal in a specific time unit, normally measured in pulses per second. The state-of-the-art LiDAR systems are capable of providing as high as 500 kHz or even 1000 kHz pulse repetition rate.

The maximum PRF is an important parameter of a LiDAR system since this is the primary factor to determine the point density. An increase in point density has obvious advantages; the denser the data, the better the potential to extract more detailed geospatial information.

The other important parameter is the Time-of-Flight principle (ToF). It is a method for measuring the distance between a sensor and an object, based on the time difference between the emission of a signal and its return to the sensor, after being reflected by an object.

**Problem**

Lidar uses active sensors that supply their own illumination source. The energy source hits objects and the reflected energy is detected and measured by sensors. Distance to the object is determined by recording the time between transmitted and backscattered pulses and by using the speed of light to calculate the distance traveled.

these devices typically capture data at a rate between hundreds and thousands of points per second. These devices are the clear favorite when the ranges to be measured are long; however, in high ranges with high PRF we face matching problem. This is because of the limitation caused by speed of light. To make it clearer assume that a signal is making a 2-kilometer round trip, the speed of light would limit the signal rate to 150,000 pulses per second. In addition, based on this simple equation ‘x= vt’ we can calculate the range using time of flight (1 micro second for Galaxy Prime sensor) as bellow:

As is the distance to go and return, the range will be half of that so:

So, in distances higher than 150 m we are not capable of distinguish which received pulse is related to which transmitted one.

The reason is that each laser pulse must have time to return to the detector before the next one is sent out otherwise matching every back scattered light to their corresponding emitted pulse will be challenging.

However, Galaxy prime aerial laser scanner has PRF = 1 MHz it means that it shoots 1 million pulses per second or every 1 micros second it shoots a transmit laser beam to the earth. the travel time (time of flight) depends on the range (the distance from sensor to the earth). For this sensor after ranges higher than 150 meters matching return pulses to transmitted ones might face difficulties.

Moreover, since there are particles suspended in the air and very close to the sensor, laser beam will hit them and return. This return light can be considered as a back scattering from a true terrestrial point and will be mixed with the ground points in longer ranges.

To address this issue, we wish to divide the maximum range into various ranges based on PRF and consider them as radius of circles which the center of the circle is the center of the sensor. As Galaxy Lidar sensors mounted on airplanes, the absolute position and orientation of the sensor will be determined using a [Global Positioning System](https://en.wikipedia.org/wiki/Global_Positioning_System) receiver and an [Inertial Measurement Unit](https://en.wikipedia.org/wiki/Inertial_Measurement_Unit) (IMU) which are installed on the aircraft. Hence, to define the circles the 3D coordinate of circle center is known and only the radius or range related to PRF should be calculated.

As there is an uncertainty of 100 meters, with the distance of 100 m around every circle the probability of noise existence will be zero. So, based on the distance of every point respect to each circle, the probability will be calculated. Finally, the probability vector will be added to the features to train the SVM model. Then the raw data will be classified to noise and non-noise using support vector machine.

