Classifying Single Photon LiDAR data using Machine Learning and Open Source tools

Workshop on Single Photon
and Geiger-Mode LiDAR

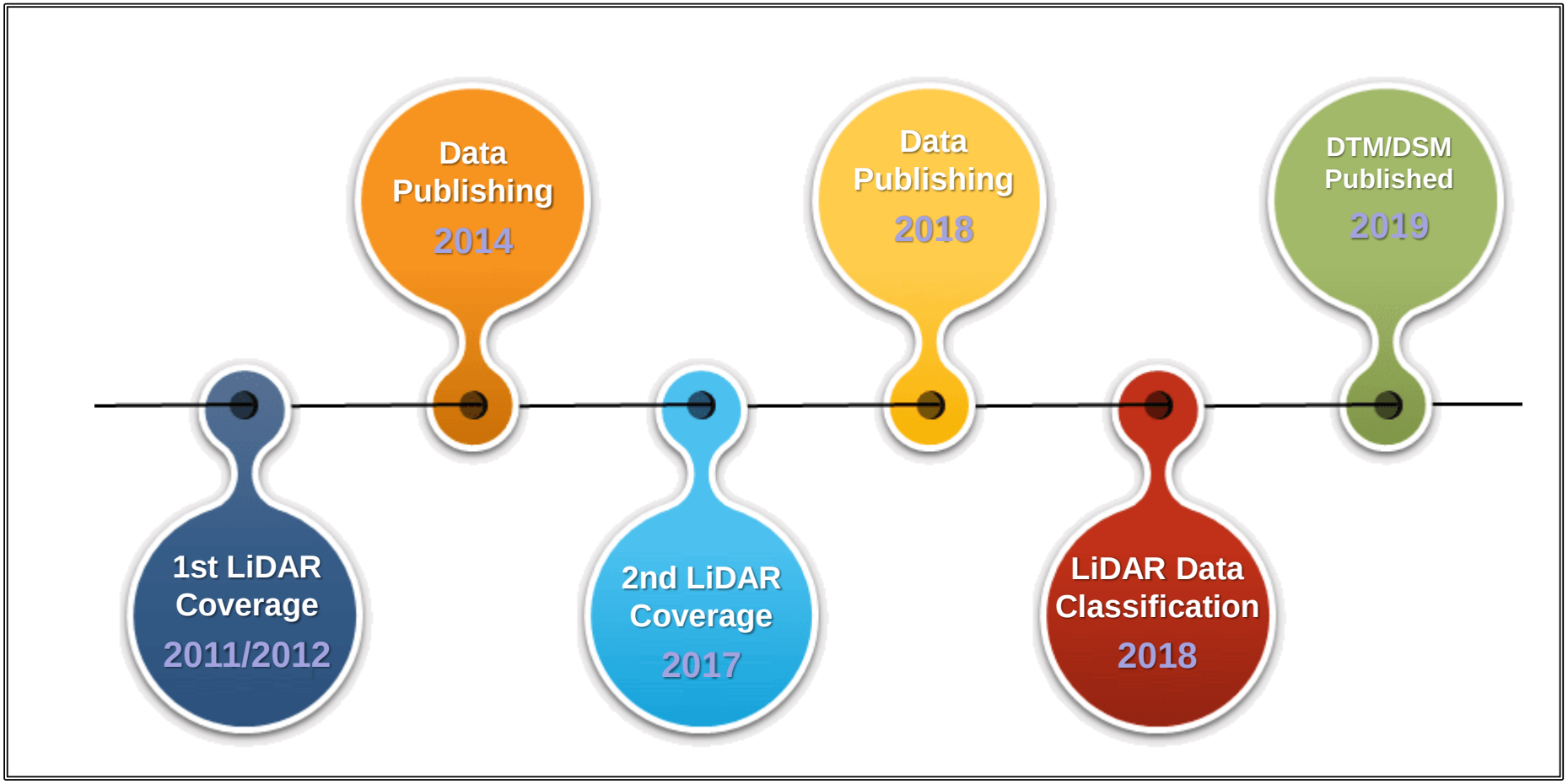
Víctor J. García Morales - Tracasa
vgarcia@tracasa.es

6th March 2019

EuroSDR

tracasa

Timeline



Nafarroako Gobernua Gobierno de Navarra

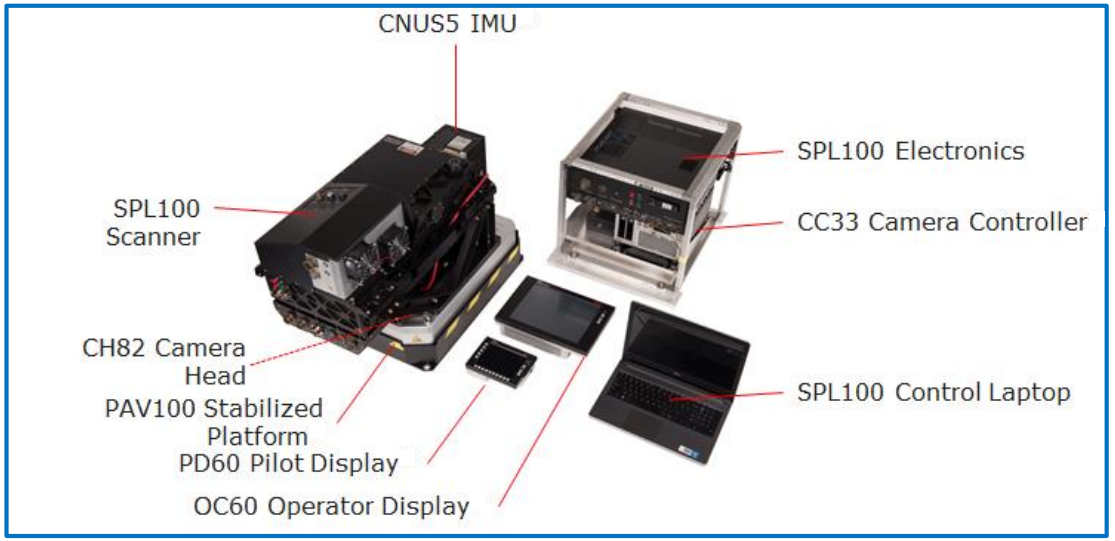
GOBIERNO DE ESPAÑA MINISTERIO DE FOMENTO INSTITUTO GEOGRÁFICO NACIONAL

EuroSDR

tracasa

Data Capture Equipment

Beechcraft B200 King Air

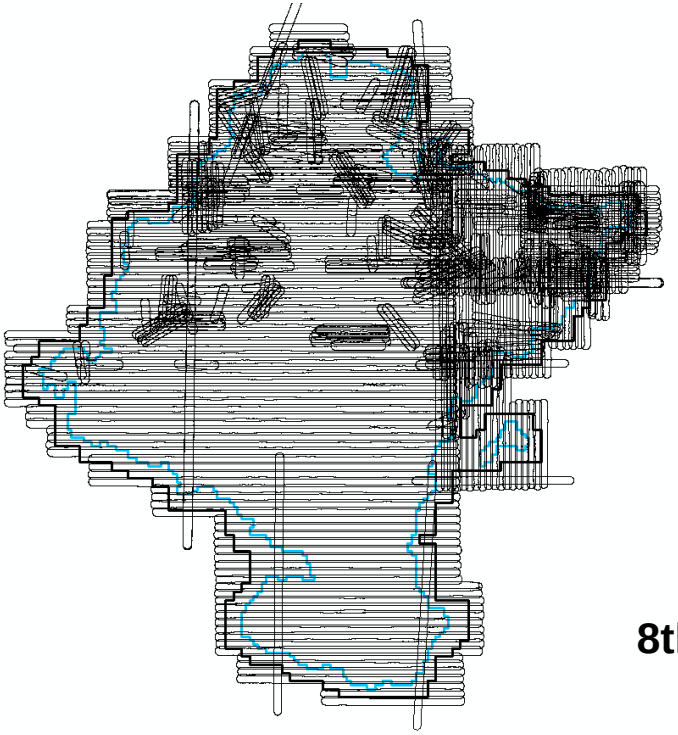


Sensor LiDAR SPL100 – RCD30 Medium Format Camera – Stabilized Platform

EuroSDR

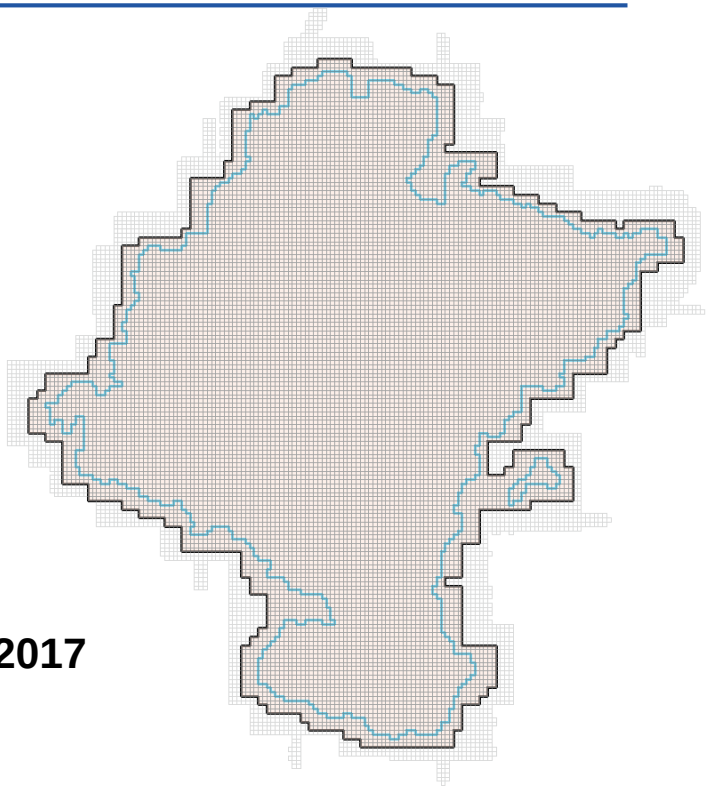
tracasa

LiDAR Data SPL100 - Navarra



Technical Specification

Area: 10.391 Km^2
Point Density: 10 pts/m^2
Overlap: 15%
Accuracy: XY: 20 cm Z: 15 cm



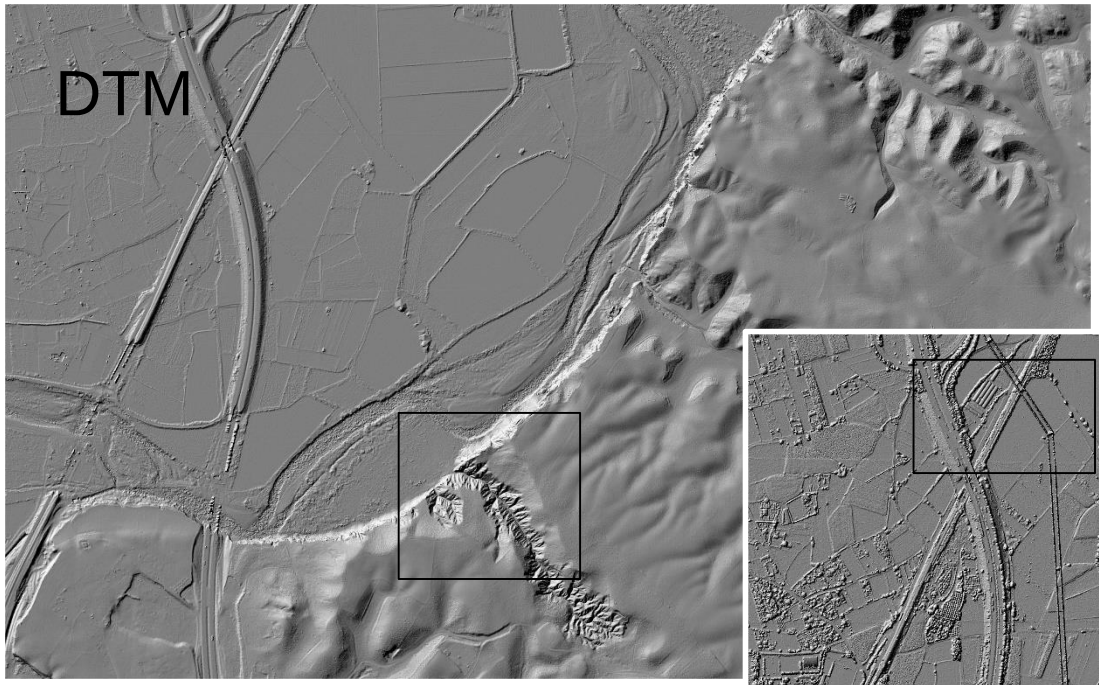
8th September – 16th November 2017

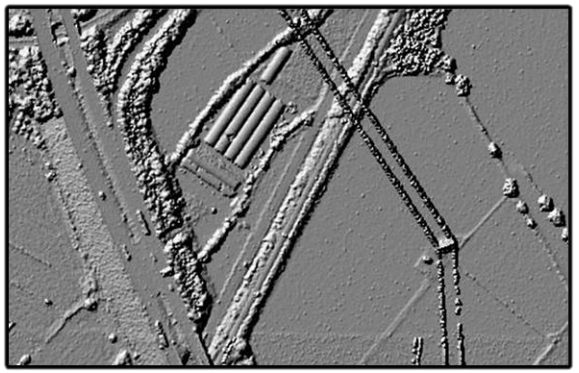
SPL100	Project
Flightlines	482
Total lenght	16.000 km
Height (a.s.l.)	3.900 – 6.300 m
Swath (average)	2300 m
Max. lenght	110 Km
Density per fligtline	14 ppsm
Speed (knots)	200
Sessions/Days	40/24

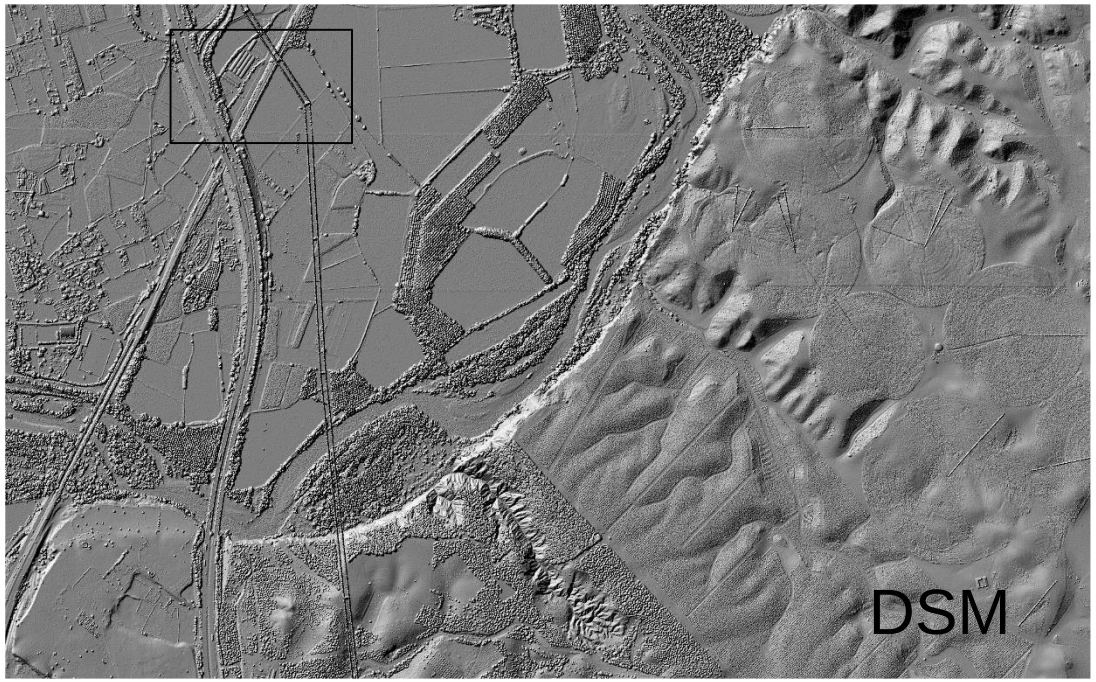
Information	
Total points	***580.696.951.479***
Points Sensor Noise	***98.114.268.808***
% Points Sensor Noise	***16,84%***
Valid Points	***482.582.682.671***
1x1 km Blocks	***16.202***
Color	***RGBNIr***
Classification	***Automatic Classification***
Classes	***Ground, Vegetation (Low. Medium, High), Building, Noise, Sensor Noise***

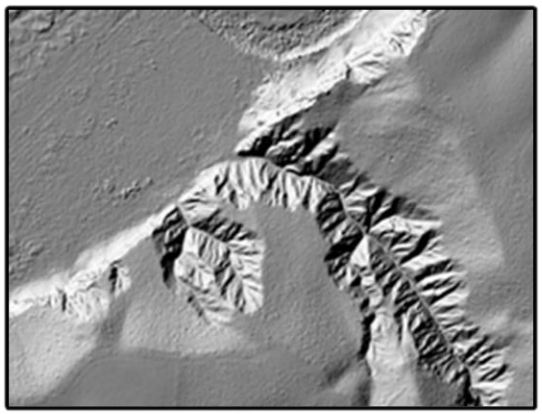
tracasa

Final product: **DTM/DSM**









25cm / **50 cm** / 1m / **2m**

tracasa

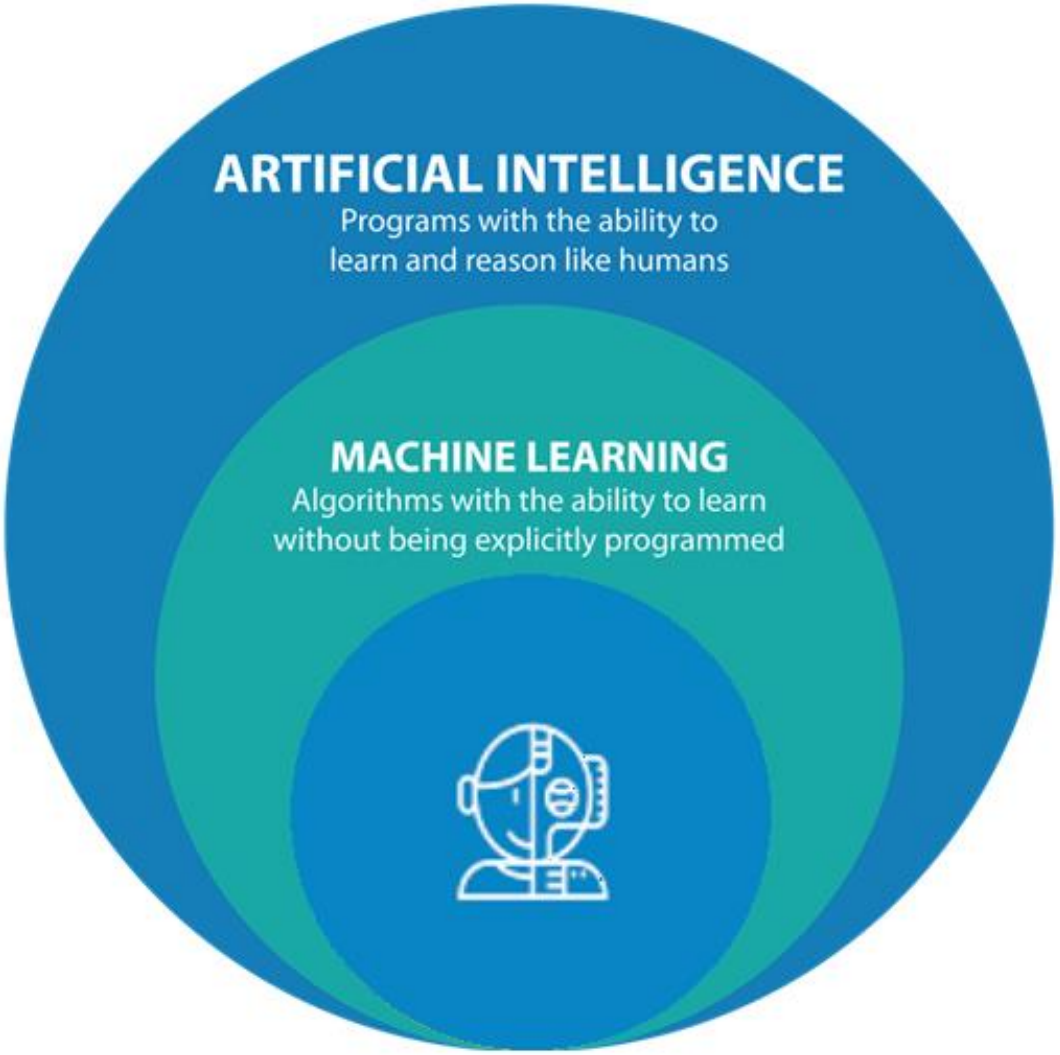
Analysis of the problem

- Final product requires very accurate data classification
- Tested Commercial/Open Source SW - Critical: Licenses, computing time, result
- Heterogeneous project Area: Need homogeneous solution
- Same decision in classification

EuroSDR

tracasa

Solution



EuroSDR

tracasa

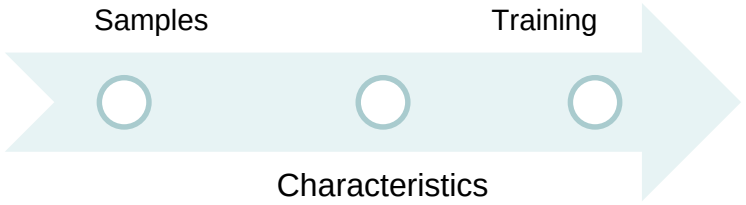
Expectative…

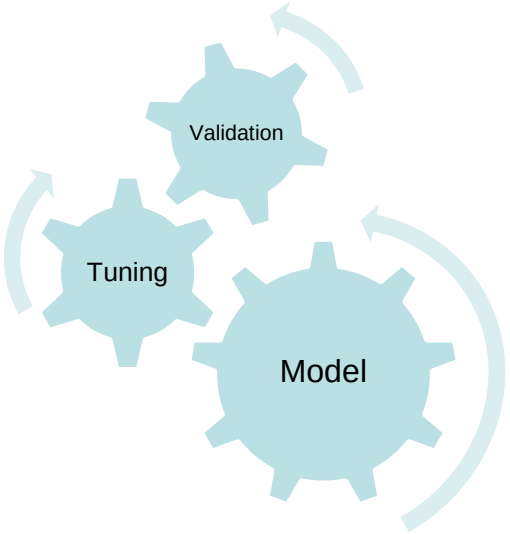
Reallity??

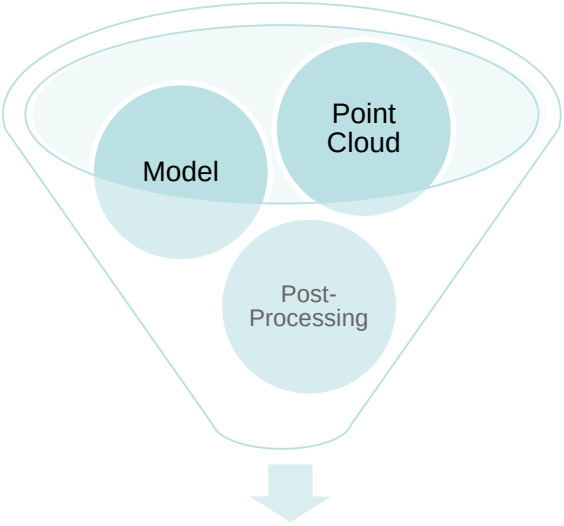
EuroSDR

tracasa

Workflow







Data Classified

EuroSDR

tracasa

Infraestructure

Training

Training Cluster - UPNA

15 computacional nodes – 13 CPU / 2 GPU
960 Gb RAM – 12 Tb Storage

Classification

High Performance Computer (HPC) – Nasertic

38 nodes x 20 cores = 760

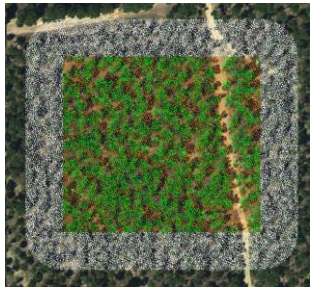
Tracasa Head Quarters - HTCondor

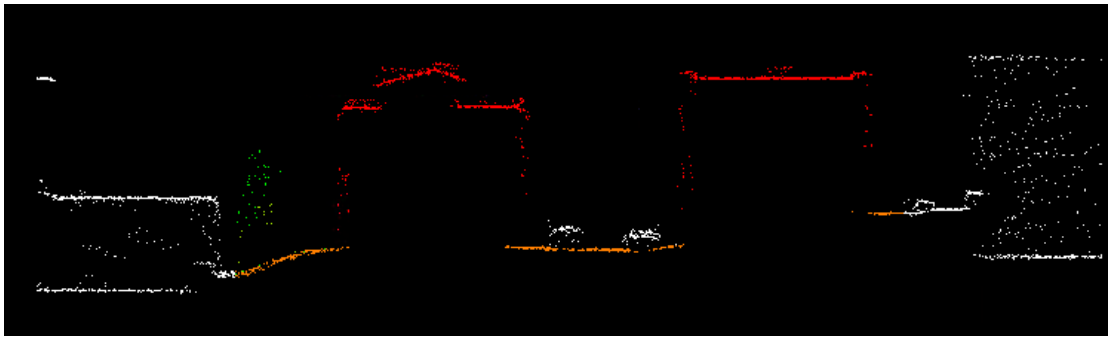
125 Computers working in parallel, distributed processing with different performance (1-2 cores)

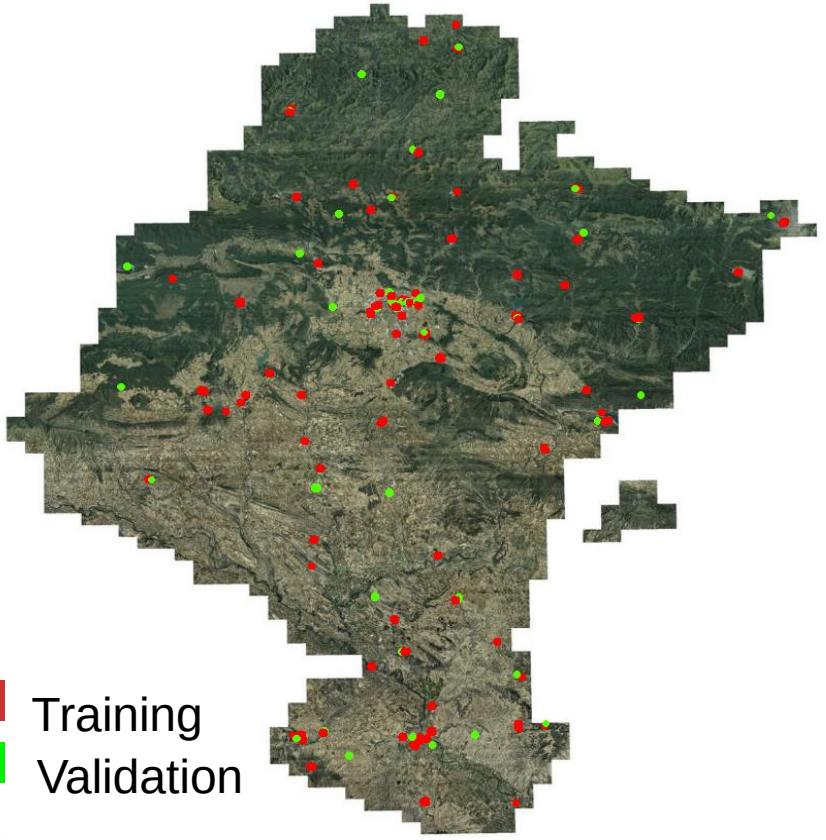
EuroSDR

tracasa

Machine Learning: Supervised Method - Samples







- 160 samples manually classified (60 Million points)
- 75% Training – 25% Validation.
- Every sample 50 m neighbourhood
- Classes: Ground, Low/medium/High Vegetation, Building, Low points.
- Need of very good classification.
- 0.012% Data clasified

EuroSDR

tracasa

Feature Extraction: >100 Characteristics

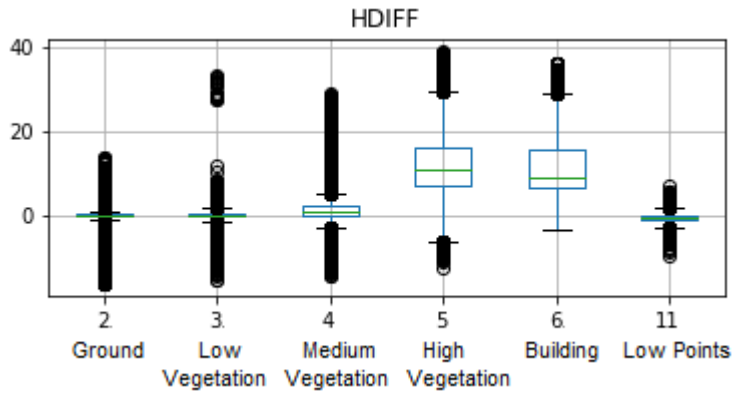
Variable	Created by
HDIFF	Tracasa
HeightAboveGround	PDAL
Red	LiDAR
Green	LiDAR
Blue	LiDAR
Infrared	LiDAR
NDVI	Tracasa
Intensity	LiDAR
CartoC2	Tracasa
CartoC6	Tracasa
Eigenvalue0 0-7	CGAL
Eigenvalue1 0-7	CGAL
Eigenvalue2 0-7	CGAL
Distance to plane 0-7	CGAL
Elevation 0-7	CGAL
Verticality 0-7	CGAL
Echo scatter 0-7	CGAL
Vertical dispersion 0-7	CGAL
Linearity 0-7	Tracasa
Planarity 0-7	Tracasa
Sphericity 0-7	Tracasa
Omnivariance 0-7	Tracasa
Anisotropy 0-7	Tracasa
Eigenentropy 0-7	Tracasa
ChangeOfCurvature 0-7	Tracasa

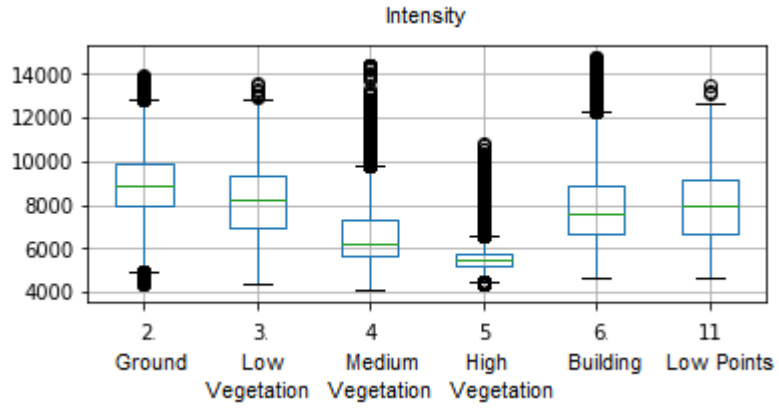
CGAL Characteristics - Neighbourhood

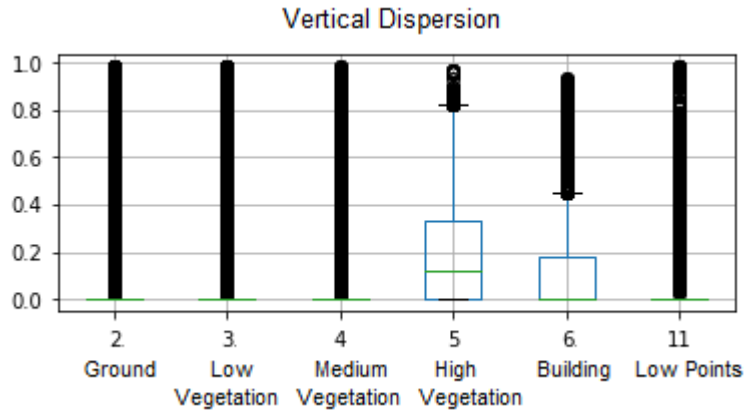
Minimum Voxel Size 25 cm

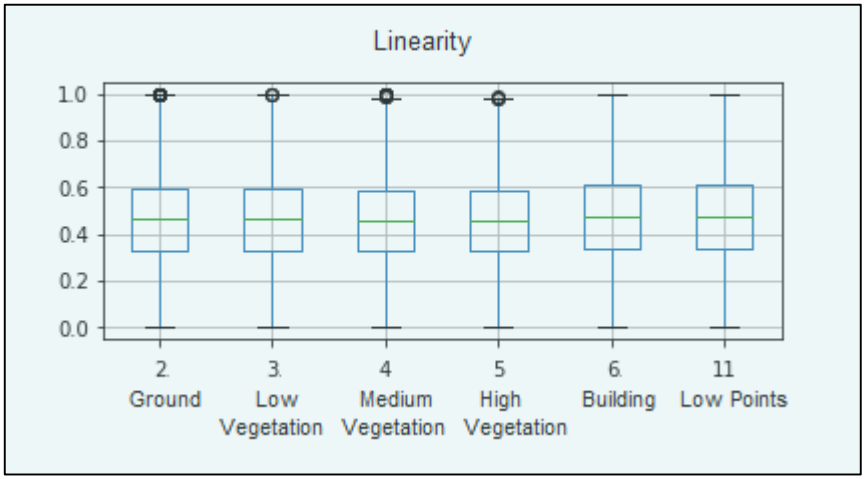
8 Scales: 50 cm, 1m, 2m, 4m, 8m, 16 m, 32m.

Box Plot Analysis





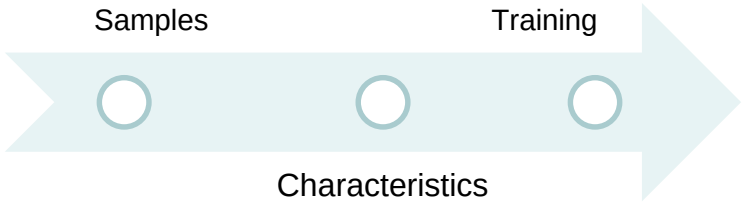


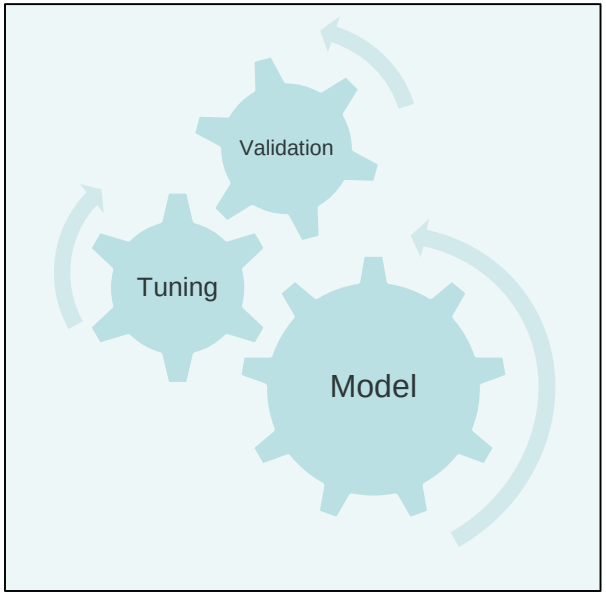


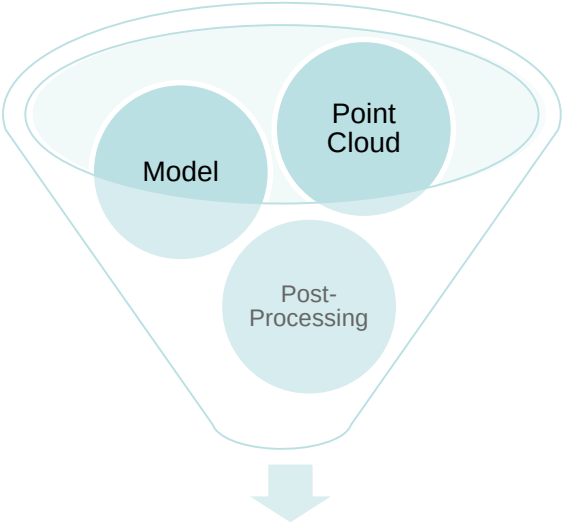
Combined analysis of the characteristics

Selection of characteristics

Workflow







Data Classified

EuroSDR

tracasa

Training Algorithm

Different algorithm were tried:

- KNN
- SVM
- Decision tres
- Random Forest
- NN
- AdaBoost
- Naive Bayes
- Logistic Regression
- Extra Trees
- XG Boost

Best results: Random Forest and **XG Boost**

XGBoost more efficient and accurate.

Performance measures

$$Recall = \frac{TP}{TP + FN} \qquad Precision = \frac{TP}{TP + FP}$$

$$F - measure = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

	Predicted Class		
Real Class		Positive	Negative
	Positive	True Positive	False Negative
	Negative	False Positive	True Negative

Cost Matrix

		Predicted Class					
		Ground	Low Vegetation	Medium Vegetation	High Vegetation	Building	Noise
Real Class	Ground (2)	0	25	40	80	90	60
	Low Vegetation (3)	10	0	0	40	80	65
	Medium Vegetation (4)	50	0	0	0	70	70
	High Vegetation (5)	100	60	0	0	60	75
	Building (6)	100	80	70	60	0	80
	Noise (11)	80	85	90	95	100	0

The importance of a point misclassified

Exploratory Data Analysis

Only 10% Samples where used for the training

Training Cluster: 100% Points of the Samples 24h
10% Points of the Samples 1h

	f1-measure (por clase) – en test					
id	**f1 test 2**	**f1 test 3**	**f1 test 4**	**f1 test 5**	**f1 test 6**	**f1 test 11**
10%	0,8592	0,0489	0,8223	0,9727	0,9404	0,4436
30%	0,8596	0,0511	0,8204	0,9722	0,9374	0,4419

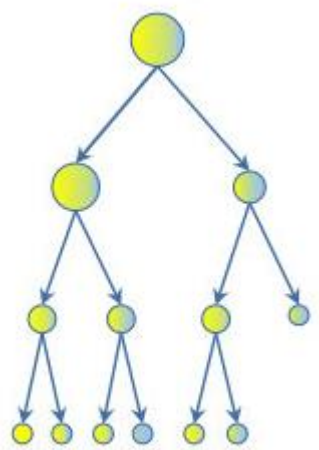
Values calculated with the Validation samples

EuroSDR

tracasa

Tuning

- n estimators (number of trees) = 100
- learning rate
- max depth = 6 Levels
- min child weight
- gamma
- subsample
- colsample bytree
- reg alpha
- use class weights



Results

f1-measure (per class)					
f1 test 2	**f1 test 3**	**f1 test 4**	**f1 test 5**	**f1 test 6**	**f1 test 11**
0,85735	**0,11575**	0,81802	0,97464	0,94309	0,42676

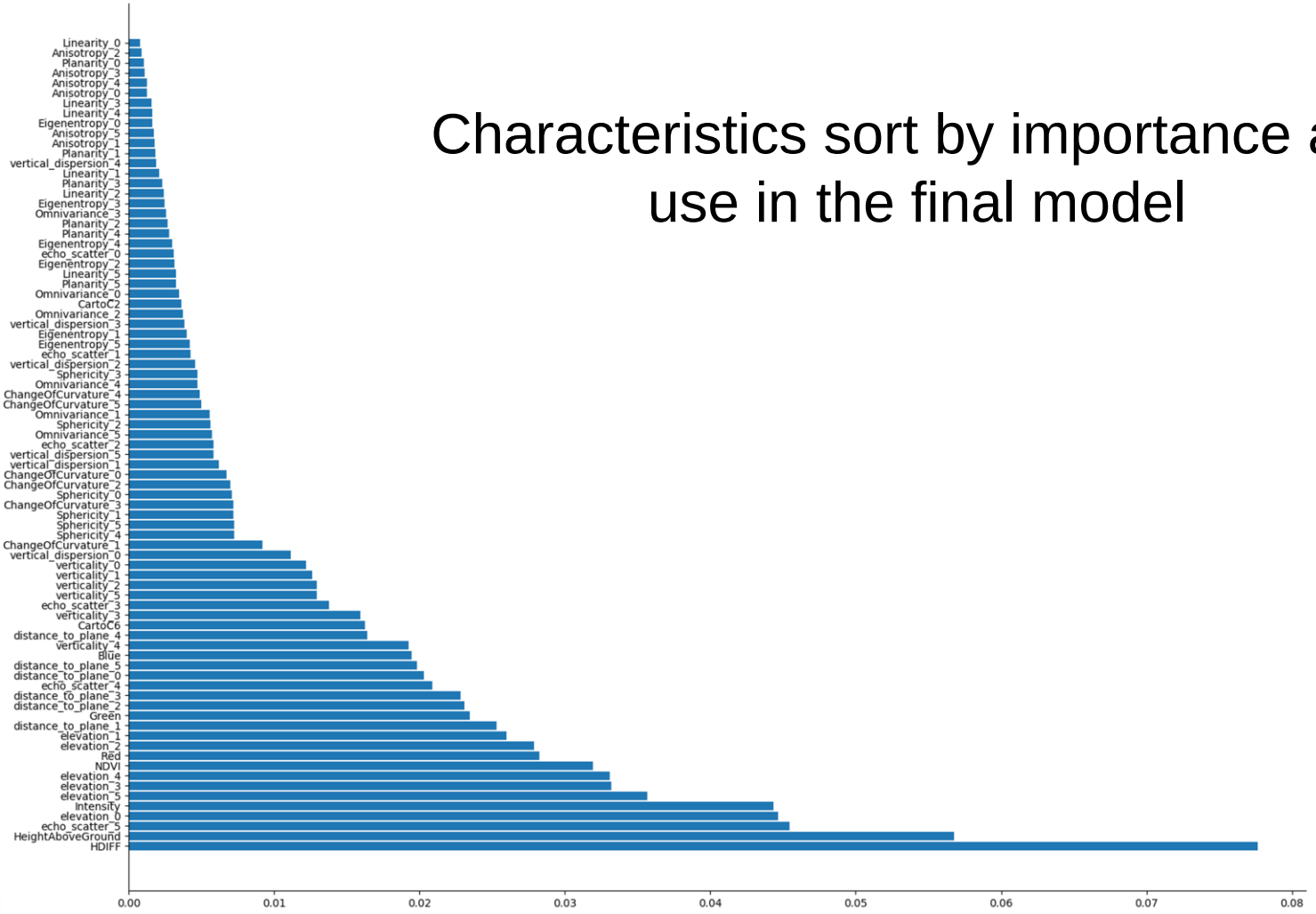
Confusion Matrix

	Predicted Class						
Real Class		2	3	4	5	6	11
	2	4598450	74670	216592	1652	60597	18462
	3	830416	79434	239919	308	125	1893
	4	287355	59351	2236707	96090	5331	2248
	5	511	75	64305	3410700	4169	42
	6	11823	3444	13920	10317	915228	711
	11	28134	3383	10038	9	25	24175

EuroSDR

tracasa

Final Model

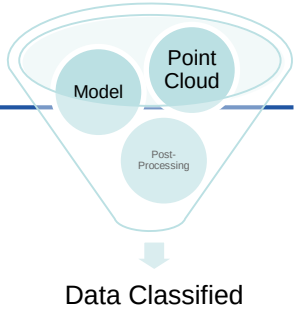


Characteristics sort by importance and use in the final model

EuroSDR

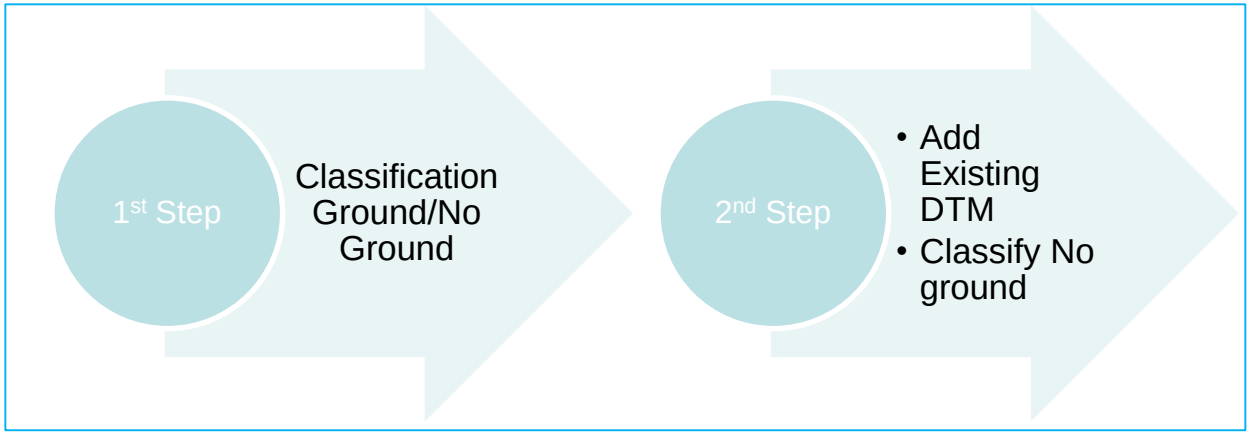
tracasa

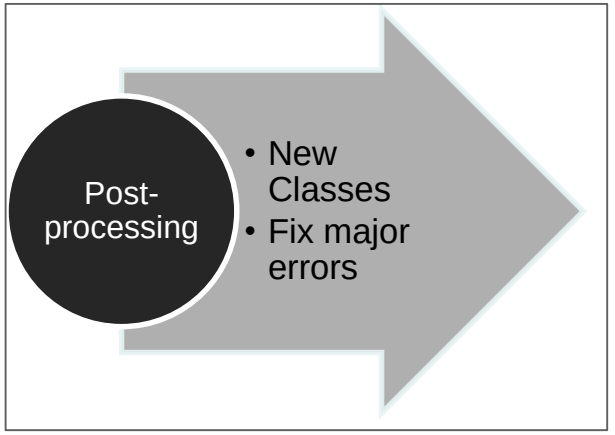
Classification Workflow



Massive data Classification

High Performance Computer (HPC) – Nasertic	Tracasa Head Quarters – HTCondor	Tracasa Head Quarters - HTCondor
100h	150h	100h





Cars, water, bridges, Noise,...

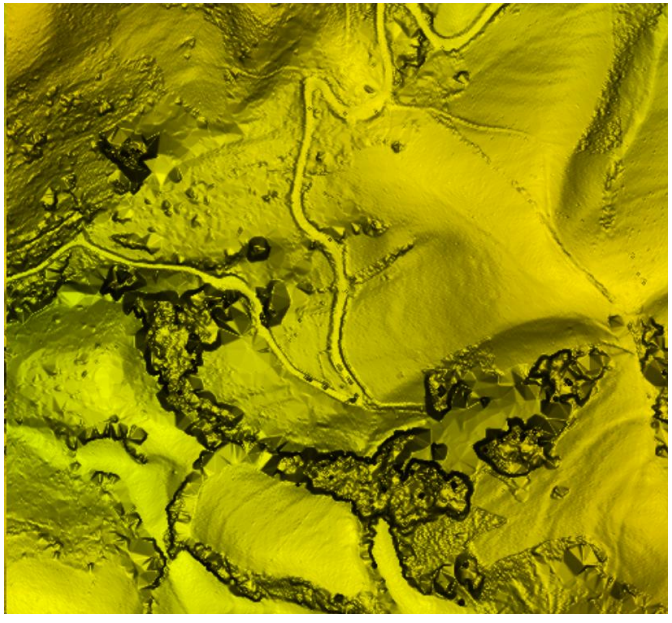
PDAL + plugins (available/ new)

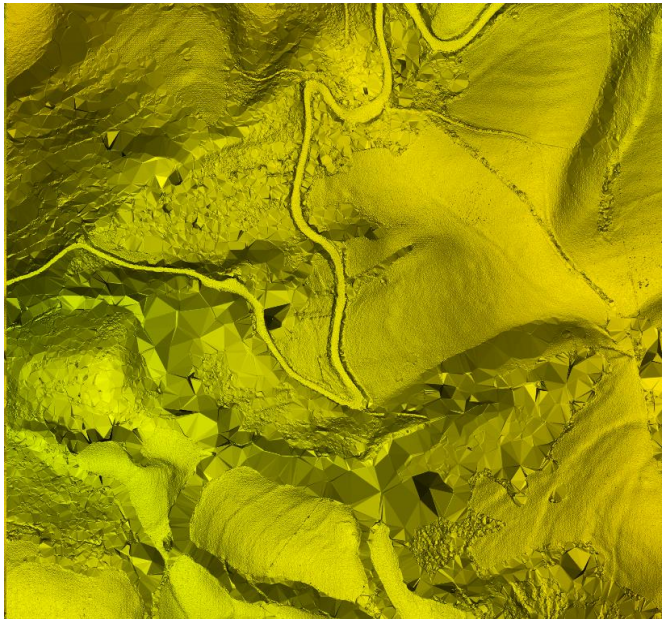
Characteristics extraction // Classify // Post- processing

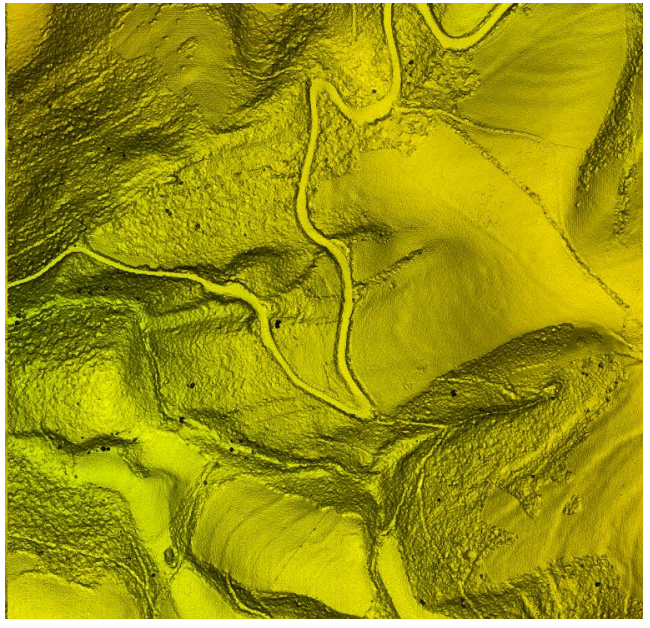
EuroSDR

tracasa

Data Classified



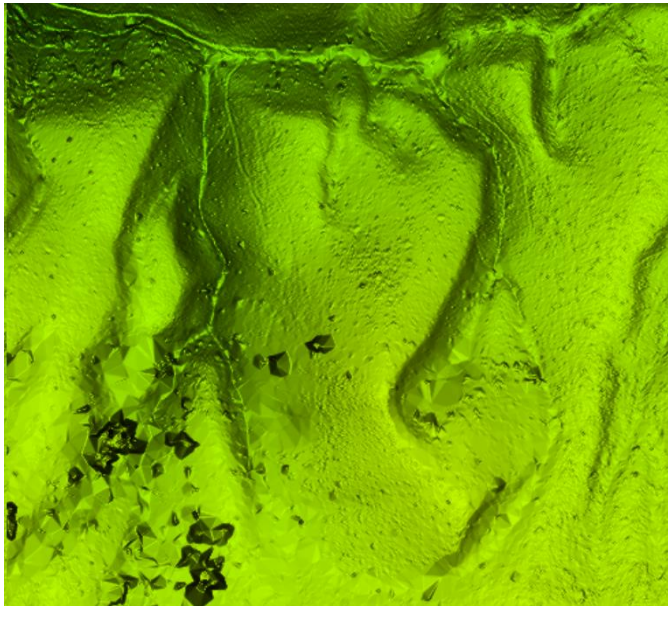


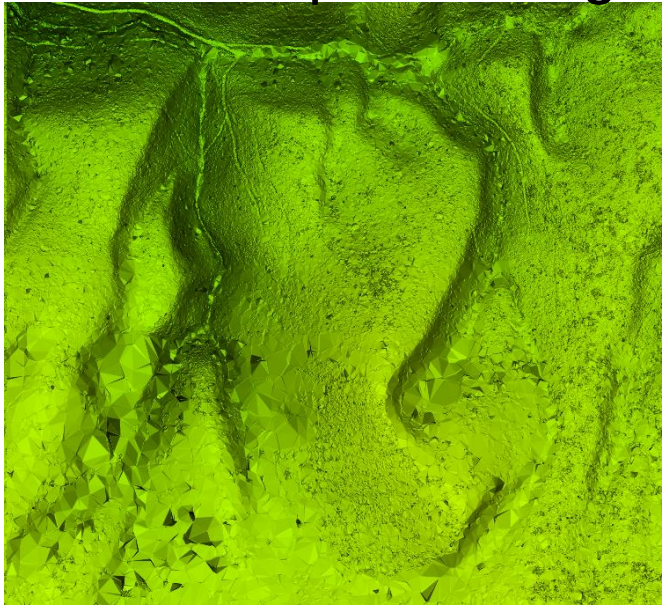


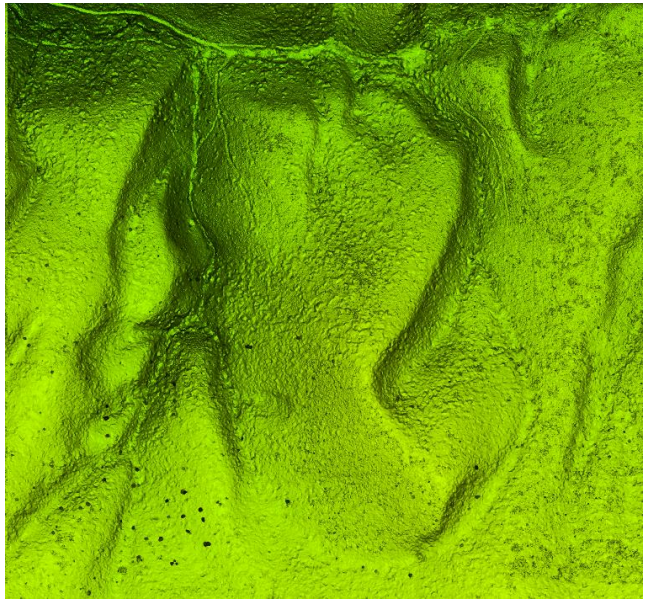
AI Classification

Post- processing

DTM

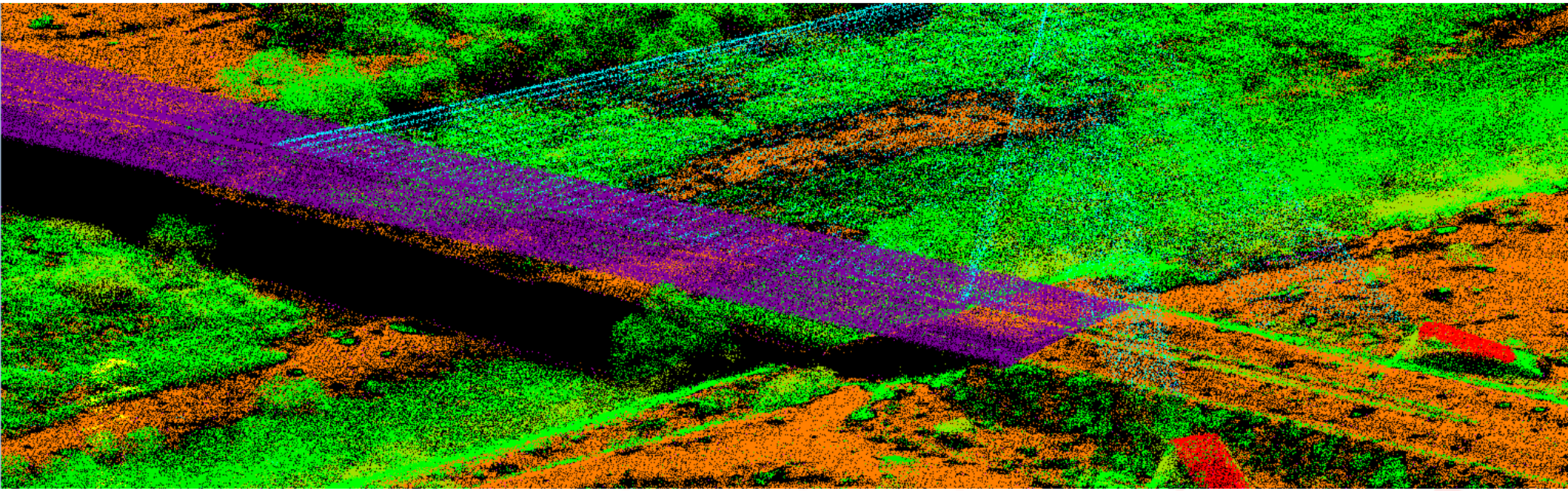






Final Remarks

- Most Important: Samples (Training – Validation)
- Objective methodology to asses the Model
- Manual Classification vs Automatic Classification
- Possibility of using this technique with other Point Cloud (data/sensor/technology)
- Final product not perfect : Train model to avoid the post-processing
- Use of Open Source libraries and distributed processing



Classifying Single Photon LiDAR data using Machine Learning and Open Source tools

Víctor J. García Morales - Tracasa
vgarcia@tracasa.es

EuroSDR

6th March 2019

tracasa