

Clasificación de la Cobertura LiDAR 2017 de Navarra con Inteligencia Artificial y herramientas open source

JORNADAS SIG LIBRE 2019

Alvaro Huarte
Sistemas de Información Territorial
Tracasa (www.tracasa.es)



**Jornadas
SIG libre**

29 - 30 de mayo 2019
Girona



Contenido

- ¿Qué es la tecnología LiDAR?
- Marco del proyecto
- Procesamiento de los datos
 - Preprocesos
 - Clasificación basada en algoritmos de Machine Learning supervisado
 - Postproceso de mejora de resultados
 - Reasignación de nuevas clases
- Productos
 - Bloques 1x1 km reclasificados
 - MDT y MDS en Hojas 1:10.000
 - Hillshade coloreado 1m
 - Conversor de ficheros LAS/LAZ



¿Qué es la tecnología LiDAR?

El LiDAR (de Light Detection And Ranging) es una técnica de teledetección óptica que utiliza la luz de láser para obtener una muestra de puntos de la superficie de la tierra produciendo mediciones exactas de la posición de los objetos.

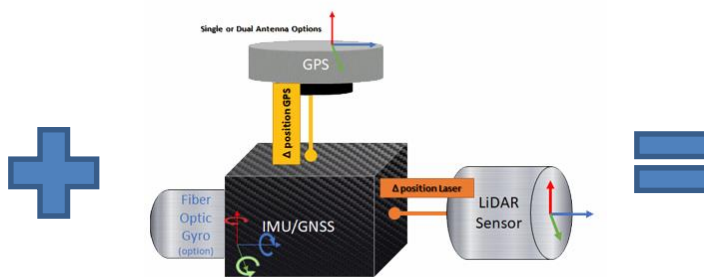
La documentación de ESRI la describe de forma más precisa:

“LIDAR es un sensor óptico activo que transmite rayos láser hacia un objetivo mientras se mueve a través de rutas de topografía específicas. El reflejo del láser del objetivo lo detectan y analizan los receptores en el sensor LIDAR. Estos receptores registran el tiempo preciso desde que el pulso láser dejó el sistema hasta cuando regresó para calcular la distancia límite entre el sensor y el objetivo. Combinado con la información posicional (GPS e INS), estas medidas de distancia se transforman en medidas de puntos tridimensionales”.

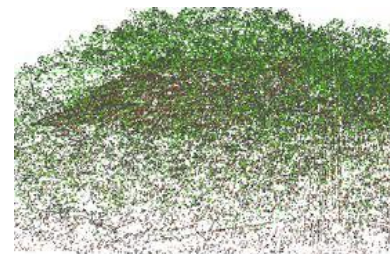
Sensor aerotransportado



Geoposicionamiento



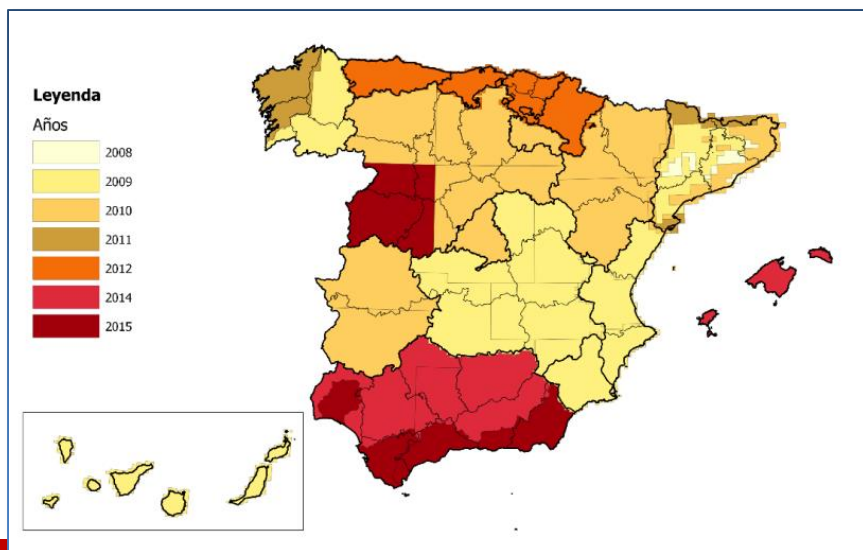
Nube puntos



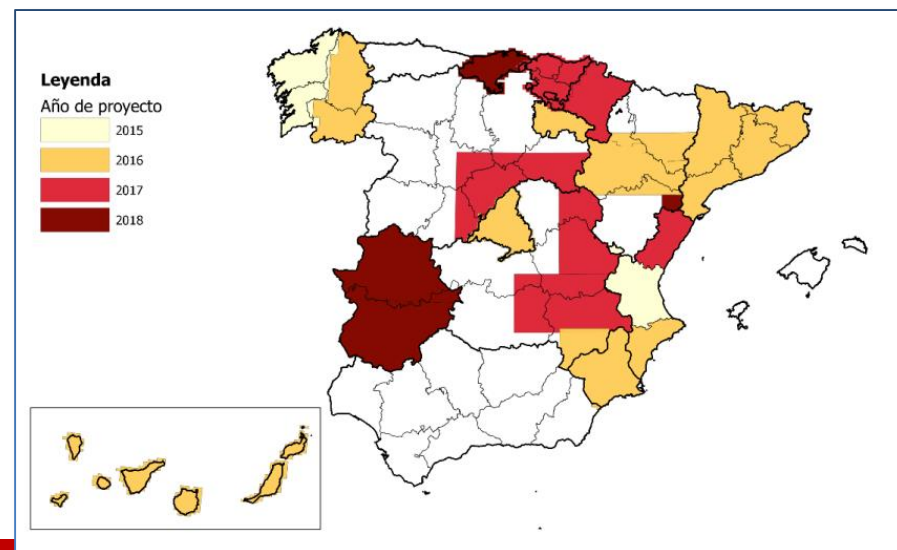
Marco del proyecto – PNOA (Plan Nacional de Ortofotografía Aérea)

- El objetivo es cubrir todo el territorio de España mediante nubes de puntos obtenidas mediante sensores LiDAR con una frecuencia de aprox. 6 años.
- Plan cofinanciado entre el IGN/CNIG, varios Ministerios y las Comunidades Autónomas.
- Precisión altimétrica de 20 cm RMSE Z (Anteriormente ~1-2m).

1º Cobertura 2008-2015



2º Cobertura 2015-...



Marco del proyecto – Coberturas LiDAR en la Comunidad Foral de Navarra

Encargo del Gobierno de Navarra a través de la Dirección General de Obras Públicas de Navarra.



1º Cobertura 2011-2012



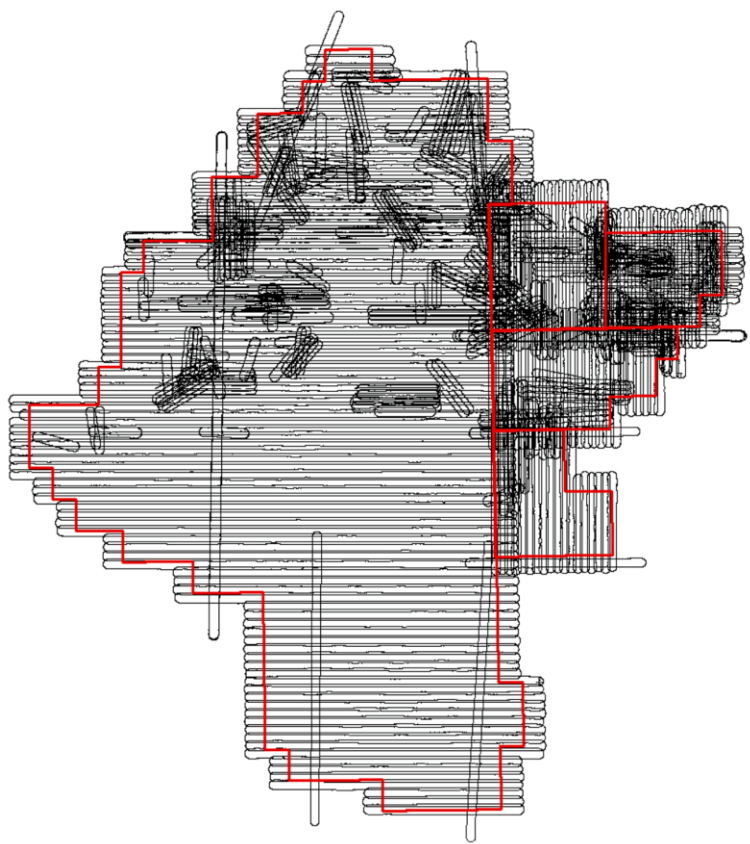
2º Cobertura 2017



ftp://ftp.cartografia.navarra.es/5_LIDAR/

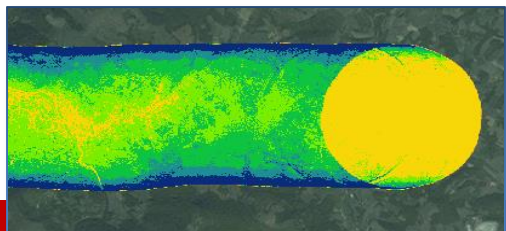
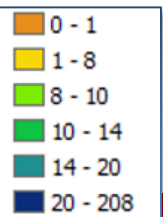
Marco del proyecto – Cobertura LiDAR 2017 de Navarra

Vuelo ejecutado con el sensor SPL100 (Single Photon LiDAR) + Cámara de medio formato RCD 30 montados en un avión Beechcraft B200 King Air.



SPL100 Ejecutado	
Nº de Puntos	580.696.951.479
Líneas de Vuelo	482
Longitud total	16.000 km
Altura de vuelo (m s. n. m.)	3.900 – 6.300 m
Ancho pasada (m)	2300 m
Bloques	16.203
Longitud máxima	110 Km
Densidad	14 puntos/m ²
Velocidad	200 Nudos
Pasadas Transversales	10

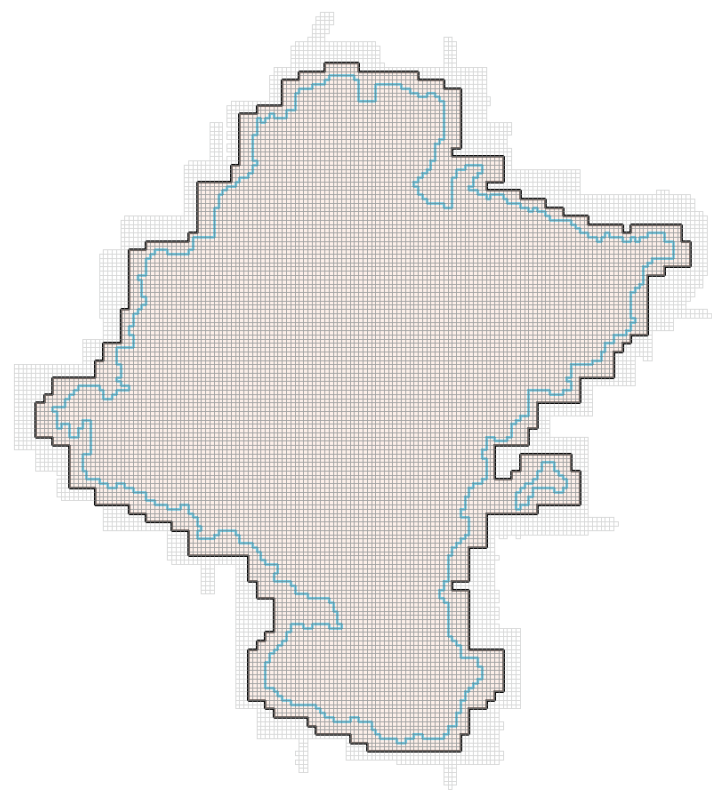
Fecha inicio vuelo: 8 Septiembre 2017
Fecha finalización vuelo: 16 Noviembre 2017



Argggh!

Marco del proyecto – Cobertura LiDAR 2017 de Navarra

Especificaciones del vuelo ejecutado.



Ficheros LiDAR LAZ procesados	
SRID	EPSG:25830
Cota	Elipsoidal GRS80
Bloques 1 x 1 km	16.203
Versión fichero	LAZ v1.4
Formato punto	8 (RGB + Nir)
Tamaño total datos	~6Tb
Nº total puntos	580.696.951.479
Densidad	14 pts/m²
% puntos de ruido	16,84%
Puntos preclasificados en varias clases base	Ground, Vegetation, Building, Noise, Sensor Noise, Overlap.

Procesamiento de los datos

Dos objetivos: Clasificar la nube de puntos LiDAR y generar los productos MDS y MDT por hojas 1:10.000.



Ficheros de entrada

- 16.203 archivos LAZ de 1km x 1km.
- ~6TB de espacio en disco.
- LAZ version: 1.4 .
- Point format: 8 (RGB+NIR).

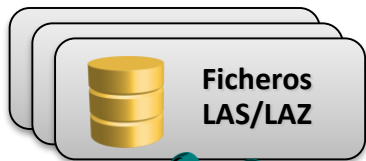
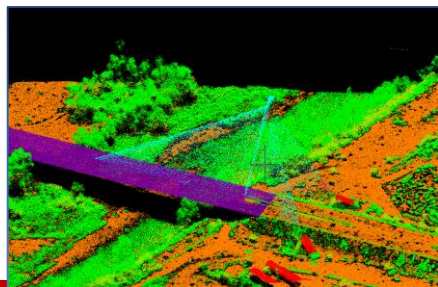


Proceso de clasificación

- Reasignación y recuperación de ruido.
- Asignación de datos cartográficos.
- **Clasificación.**
- Postproceso de mejora de resultados.
- Asignación de clases extra (Puentes, Agua, elementos dispersos en viales: coches...).

1º Producto:

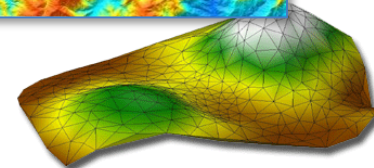
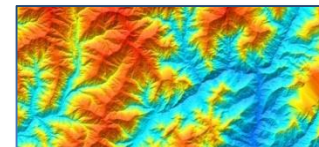
Ficheros reclasificados



2º Producto:

Hojas 1:10.000 MDS/MDT

- Extracción de clases por hojas 1:10.000.
- Generación de TIN temporales para evitar huecos.
- Exportación archivos MDT/MDS en Gtiff, Ascii Grid y XYZ a 25cm, 50cm, 1m y 2m.



Clasificación – ¿Cómo?

Teniendo en cuenta el volumen y densidad de los datos ¿Cómo clasificamos?

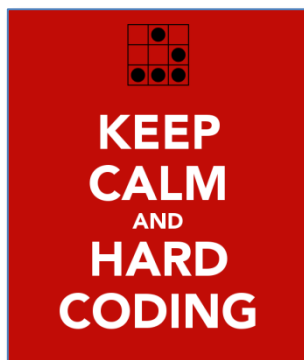


Ficheros de entrada

- 16.203 archivos LAZ de 1km x 1km.
- ~6TB de espacio en disco.
- LAZ version: 1.4 .
- Point format: 8 (RGB+NIR).

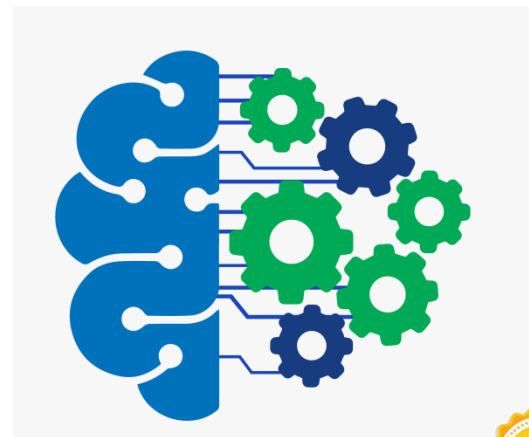


¿?



FME, LasTools, MCC LIDAR,
Orfeo Toolbox, PCL, PDAL, ...

¿?



IA - Machine Learning



Clasificación – Machine Learning supervisado

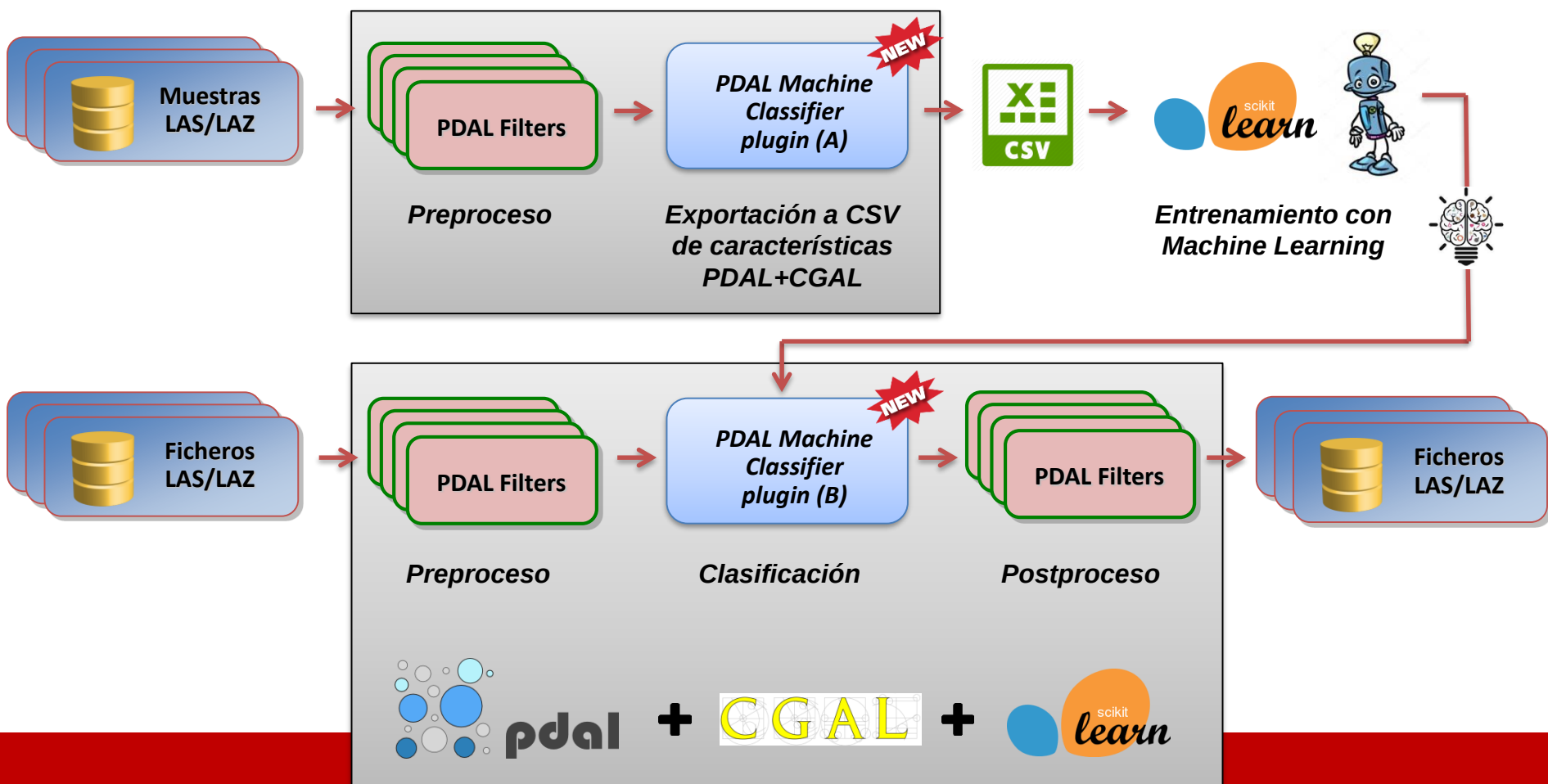
Los puntos LiDAR se clasifican aplicando un algoritmo de Machine Learning basado en aprendizaje supervisado.

... requiere dos pasos principales:

- Entrenamiento con muestras “verdad”
 1. Se clasifican manualmente un conjunto heterogéneo de puntos que representen las posibles casuísticas del terreno y de los objetos sobre él.
 2. Estudio estadístico para la selección óptima de características de los puntos.
 3. Entrenamiento y evaluación de diferentes modelos supervisados.
 4. Con el modelo con mejores resultados, se aplica una búsqueda óptima de sus parámetros de configuración.
- Clasificación final de los puntos con el modelo ya entrenado

Clasificación – Flujo y esquema de aplicaciones

Utilizamos **PDAL** como plataforma base del proceso, **CGAL** añade características geométricas a los puntos, y entrenamos y clasificamos con **scikit-learn** mediante técnicas de Machine Learning supervisado.



Clasificación – Machine Learning supervisado

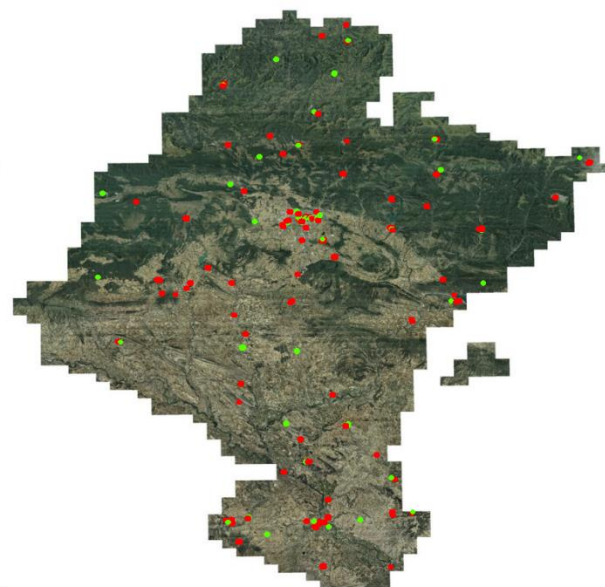
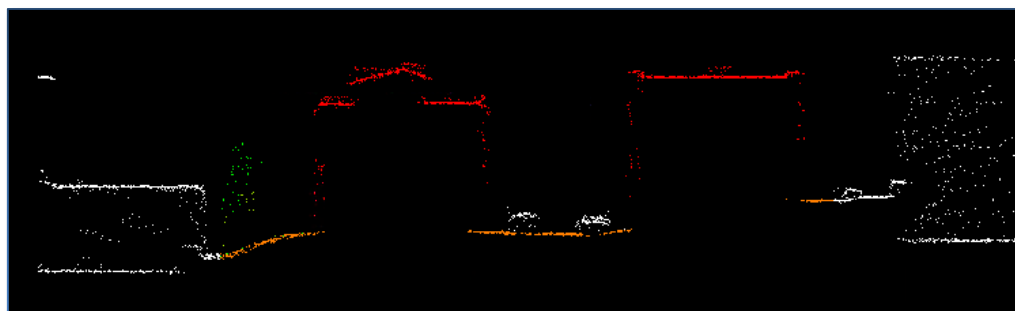
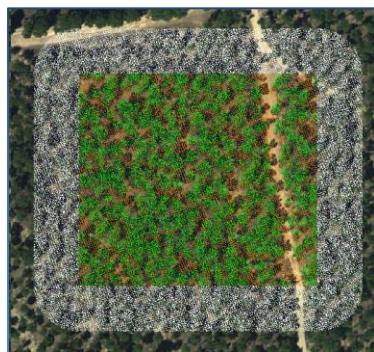
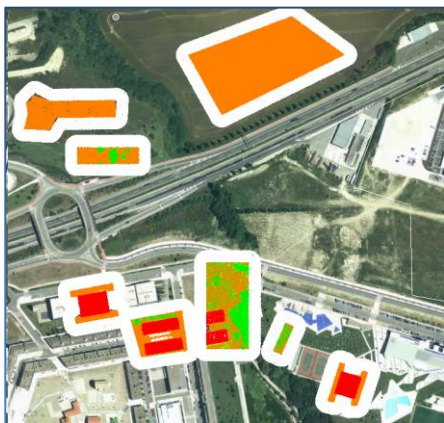
Los puntos LiDAR se clasifican aplicando un algoritmo de Machine Learning basado en aprendizaje supervisado.

... requiere dos pasos principales:

- Entrenamiento con muestras “verdad”
 1. Se clasifican manualmente un conjunto heterogéneo de puntos que representen las posibles casuísticas del terreno y de los objetos sobre él.
 2. Estudio estadístico para la selección óptima de características de los puntos.
 3. Entrenamiento y evaluación de diferentes modelos supervisados.
 4. Con el modelo con mejores resultados, se aplica una búsqueda óptima de sus parámetros de configuración.
- Clasificación final de los puntos con el modelo ya entrenado

Clasificación – Machine Learning supervisado

1º Paso: Clasificación manual de puntos muestra “verdad” para el entrenamiento supervisado del algoritmo.



■ Entrenamiento
■ Validación

Clasificación – Machine Learning supervisado

Los puntos LiDAR se clasifican aplicando un algoritmo de Machine Learning basado en aprendizaje supervisado.

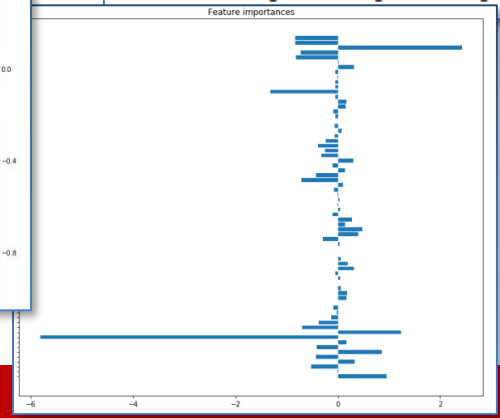
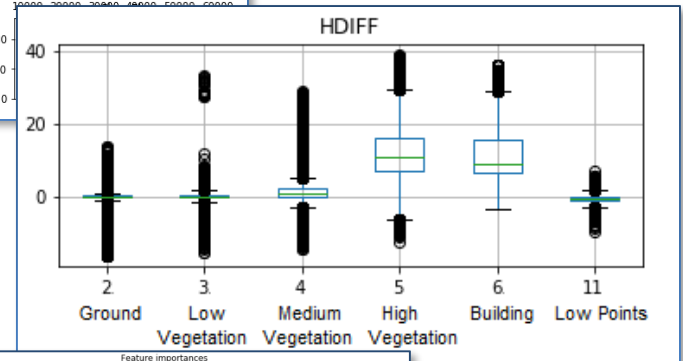
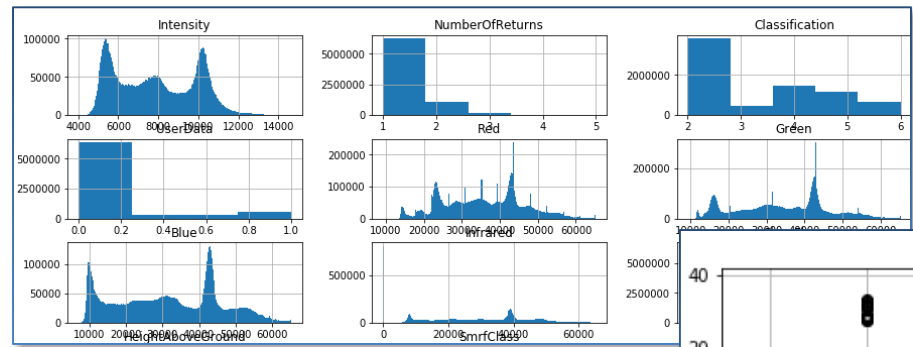
... requiere dos pasos principales:

- Entrenamiento con muestras “verdad”
 1. Se clasifican manualmente un conjunto heterogéneo de puntos que representen las posibles casuísticas del terreno y de los objetos sobre él.
 2. Estudio estadístico para la selección óptima de características de los puntos.
 3. Entrenamiento y evaluación de diferentes modelos supervisados.
 4. Con el modelo con mejores resultados, se aplica una búsqueda óptima de sus parámetros de configuración.
- Clasificación final de los puntos con el modelo ya entrenado

Clasificación – Machine Learning supervisado

2º Paso: Estudio estadístico de las características de los puntos con **scikit-learn**.
Nos permite seleccionar aquellos atributos más importantes para los algoritmos.

Variable	
LAS	Intensity
	Infrared
	Red
	Green
	Blue
	ReturnNumber
PDAL	NumberOfReturns
	SmrfClass (SMRF Filter)
	HeightAboveGround (HAG)
	HDIFF, NDVI, SAVI,...
CGAL	Eigenvalue0 0-6
	Eigenvalue1 0-6
	Eigenvalue2 0-6
	Distance to plane 0-6
	Elevation 0-6
	Verticality 0-6
	Echo scatter 0-6
	Vertical dispersion 0-6
	Linearity 0-6
	Planarity 0-6
	Sphericity 0-6
	Omnivariance 0-6
	Anisotropy 0-6
	Eigenentropy 0-6
	ChangeOfCurvature 0-6



Clasificación – Machine Learning supervisado

Los puntos LiDAR se clasifican aplicando un algoritmo de Machine Learning basado en aprendizaje supervisado.

... requiere dos pasos principales:

- Entrenamiento con muestras “verdad”
 1. Se clasifican manualmente un conjunto heterogéneo de puntos que representen las posibles casuísticas del terreno y de los objetos sobre él.
 2. Estudio estadístico para la selección óptima de características de los puntos.
 3. Entrenamiento y evaluación de diferentes modelos supervisados.
 4. Con el modelo con mejores resultados, se aplica una búsqueda óptima de sus parámetros de configuración.
- Clasificación final de los puntos con el modelo ya entrenado

Clasificación – Machine Learning supervisado

3º Paso: Entrenar y evaluar los puntos “verdad” dividiendo la muestra total de puntos en dos grupos “training” (65%) y “test” (35%).

El rendimiento del modelo se evalúa con los típicos ratios de la matriz de confusión:

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

- **VP** es la cantidad de *positivos* que fueron *clasificados correctamente* como positivos por el modelo.
- **VN** es la cantidad de *negativos* que fueron *clasificados correctamente* como negativos por el modelo.
- **FN** es la cantidad de *positivos* que fueron *clasificados incorrectamente* como negativos.
- **FP** es la cantidad de *negativos* que fueron *clasificados incorrectamente* como positivos.

Accuracy

$$Acc = \frac{VP + VN}{VP + VN + FP + FN}$$

$$TPR = Recall = \frac{VP}{VP + FN}$$

$$Precision = \frac{VP}{VP + FP}$$

$$TNR = Especificidad = \frac{VN}{VN + FP}$$

Clasificación – Machine Learning supervisado

Los puntos LiDAR se clasifican aplicando un algoritmo de Machine Learning basado en aprendizaje supervisado.

... requiere dos pasos principales:

- Entrenamiento con muestras “verdad”
 1. Se clasifican manualmente un conjunto heterogéneo de puntos que representen las posibles casuísticas del terreno y de los objetos sobre él.
 2. Estudio estadístico para la selección óptima de características de los puntos.
 3. Entrenamiento y evaluación de diferentes modelos supervisados.
 4. Con el modelo con mejores resultados, se aplica una búsqueda óptima de sus parámetros de configuración.
- Clasificación final de los puntos con el modelo ya entrenado

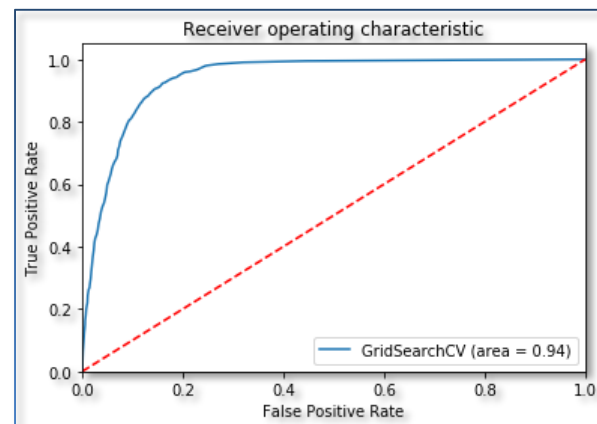
Clasificación – Machine Learning supervisado

4º Paso: Evaluados varios algoritmos supervisados, el modelo con mejores resultados fue uno de tipo “Ensemble Decision Tree”:



XGBoost

*“Parameter Tuning”: Escogido el algoritmo con mejor rendimiento, con el componente **GridSearchCV** se realiza una búsqueda automática de la óptima configuración de sus parámetros de ejecución:*

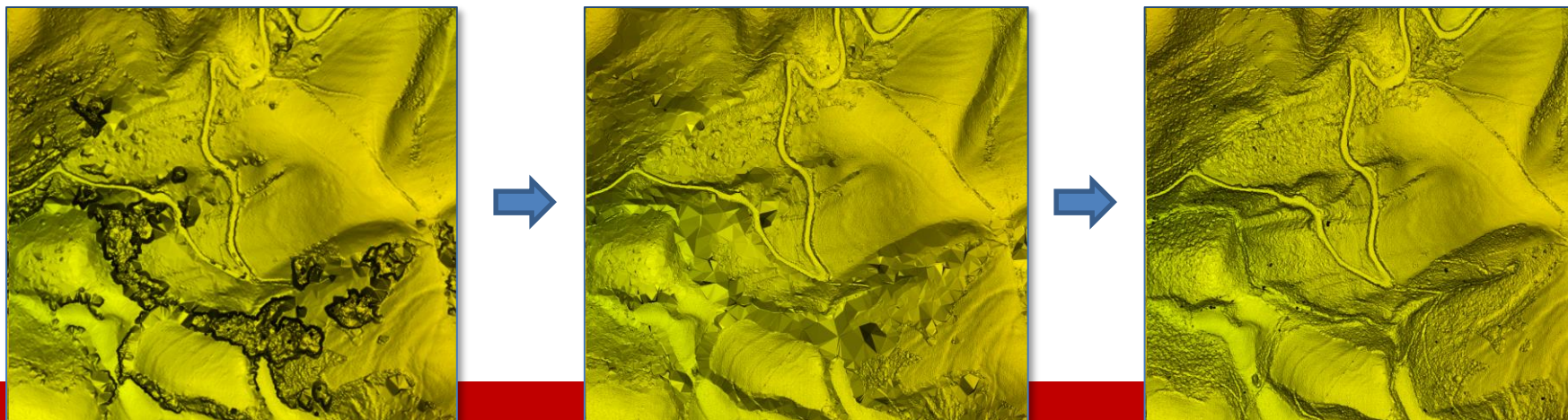


Clasificación – Rendimiento y resultados

Métricas de rendimiento por clases:

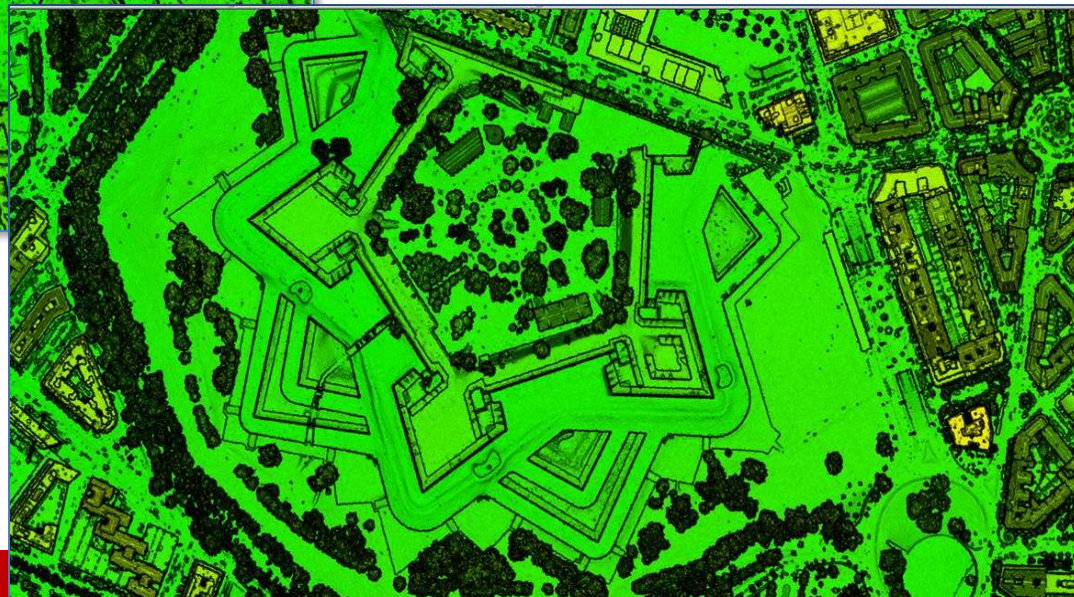
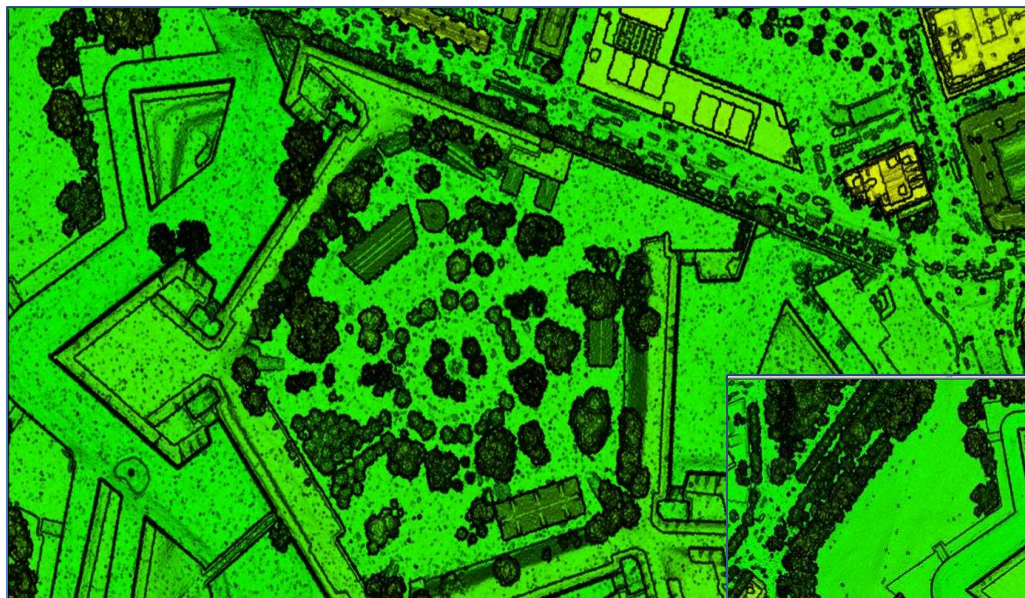
f1-measure (per class)					
f1 test 2	f1 test 3	f1 test 4	f1 test 5	f1 test 6	f1 test 11
0,85735	0,11575	0,81802	0,97464	0,94309	0,42676

Necesidad de **postproceso**, el producto **MDT** (Modelo Digital del Terreno), no tanto como el **MDS** (Modelo Digital de Superficie), es **muy sensible** a errores de clasificación entre tipos “suelo” y “vegetación”. Pocos puntos mal etiquetados malogran el resultado final:



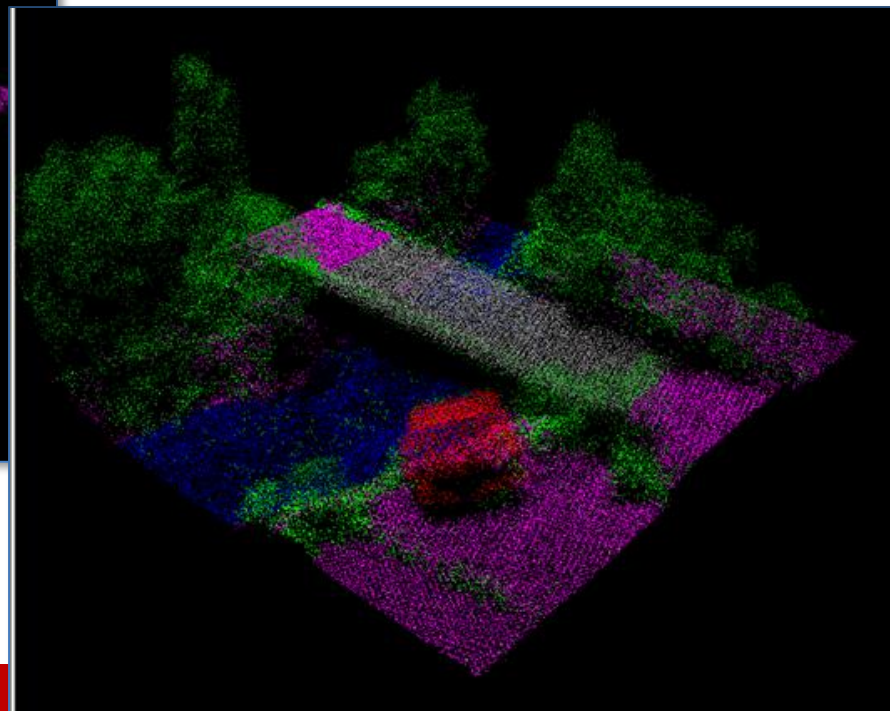
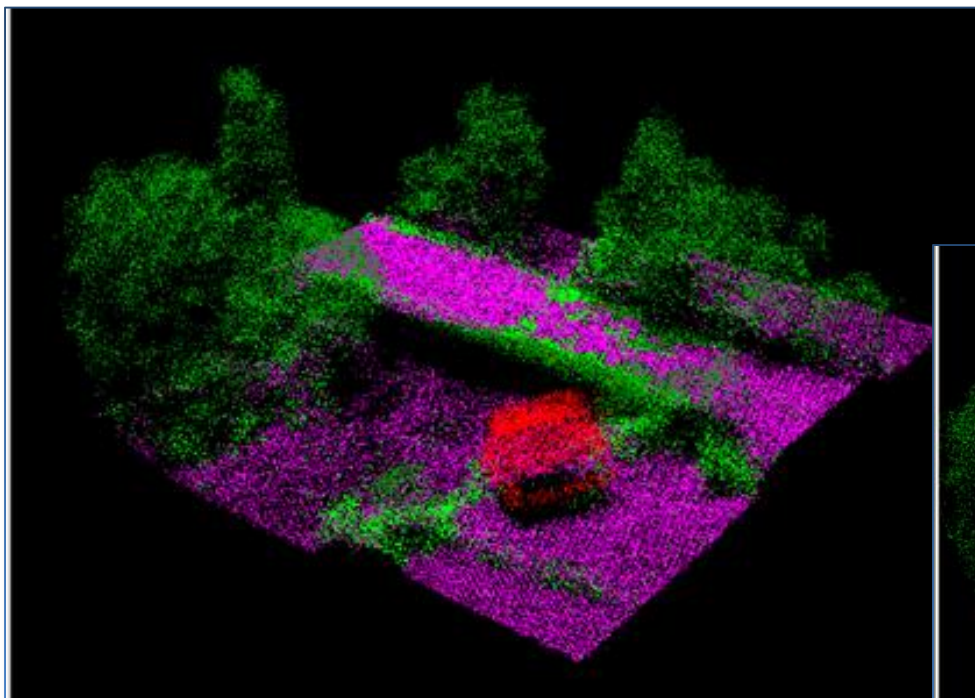
Postproceso de la Clasificación

Los datos clasificados se procesan con un nuevo plugin “ad-hoc” en **PDAL** para corregir errores y para reclasificar nuevas clases “Agua”, “Puentes” y “artefactos en viales”.



Postproceso de la Clasificación

Los datos clasificados se procesan con un nuevo plugin “ad-hoc” en **PDAL** para corregir errores y para reclasificar nuevas clases “Agua”, “Puentes” y “artefactos en viales”.



Procesamiento de los datos – ¿Dónde?

Teniendo en cuenta el volumen y densidad de los datos ¿Dónde procesamos?

Los procesos a ejecutar son pesados pero pueden ser paralelizados. En la mayoría de los casos cada bloque de 1x1 km establece la unidad mínima de ejecución y pueden ser procesados de forma independiente.

Los trabajos se lanzan en un par de clusters de alta computación en Navarra (NASERTIC y UPNA) y en el entorno de hardware distribuido de la propia Tracasa.



HPC - NASERTIC

38 nodos proceso x 20 cores
128 Gb RAM por nodo

Preprocesos



Cluster – UPNA

15 nodos proceso (13 CPU / 2 GPU)
960 Gb RAM, 12 Tb Storage

*Entrenamiento
de algoritmos*



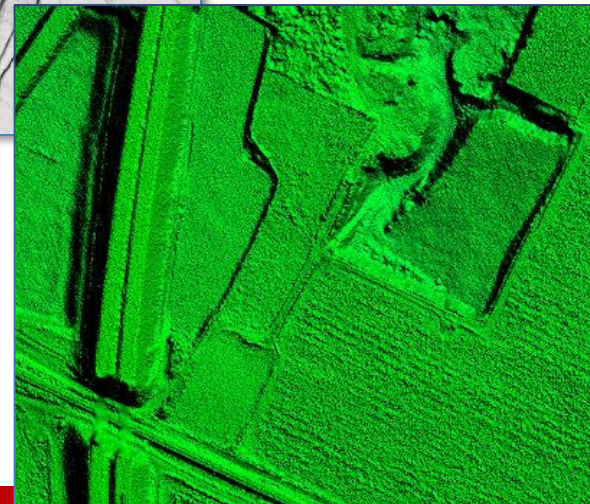
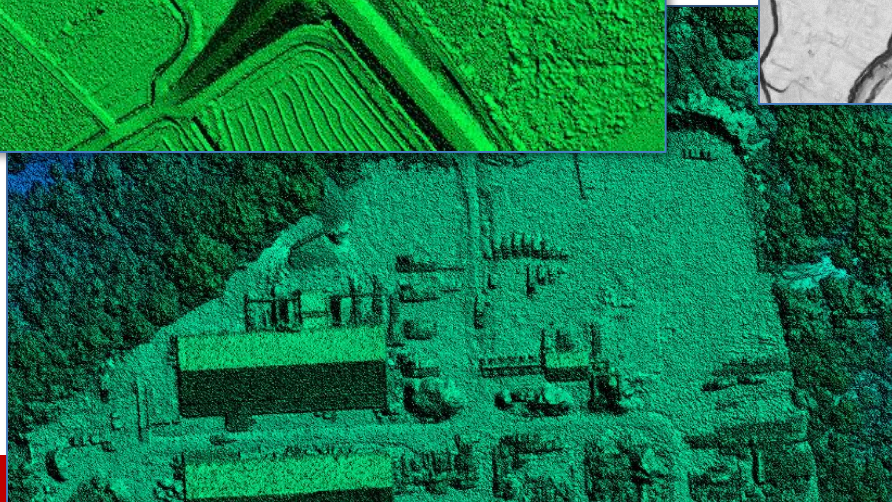
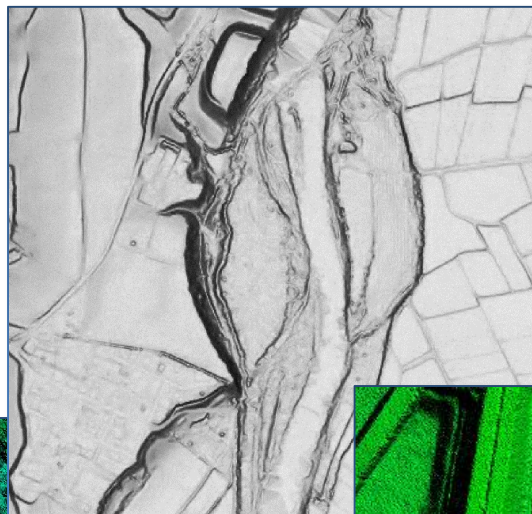
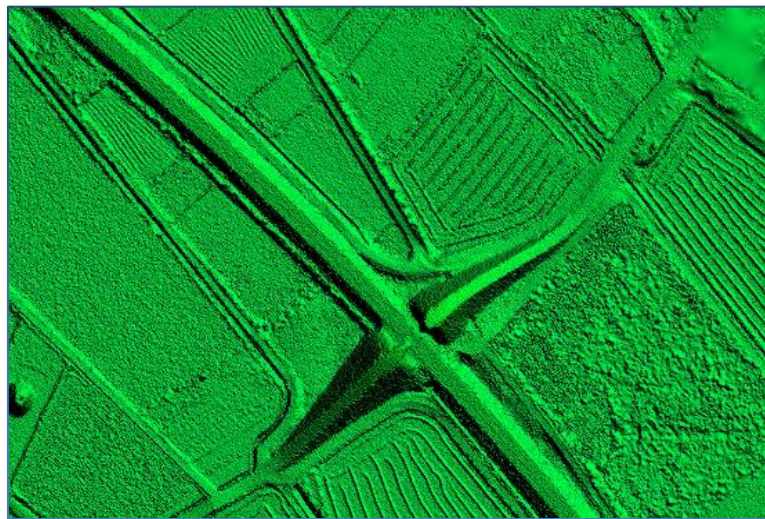
HTCondor - TRACASA

125 PCs en procesos distribuidos
16 Gb RAM por CPU

*Clasificación y
postprocesos*

Productos – MDT y MDS

Se generan para 401 hojas cartográficas 1:10.000 el Modelo Digital del Terreno (Clases “suelo”) y el Modelo Digital de Superficie (Clases “suelo”, vegetación,...). Los ficheros se exportan a formato Gtiff, Ascii Grid y XYZ a resoluciones de 25cm, 50cm, 1m y 2m en cota Ortométrica.



Productos – “hillshade” coloreado

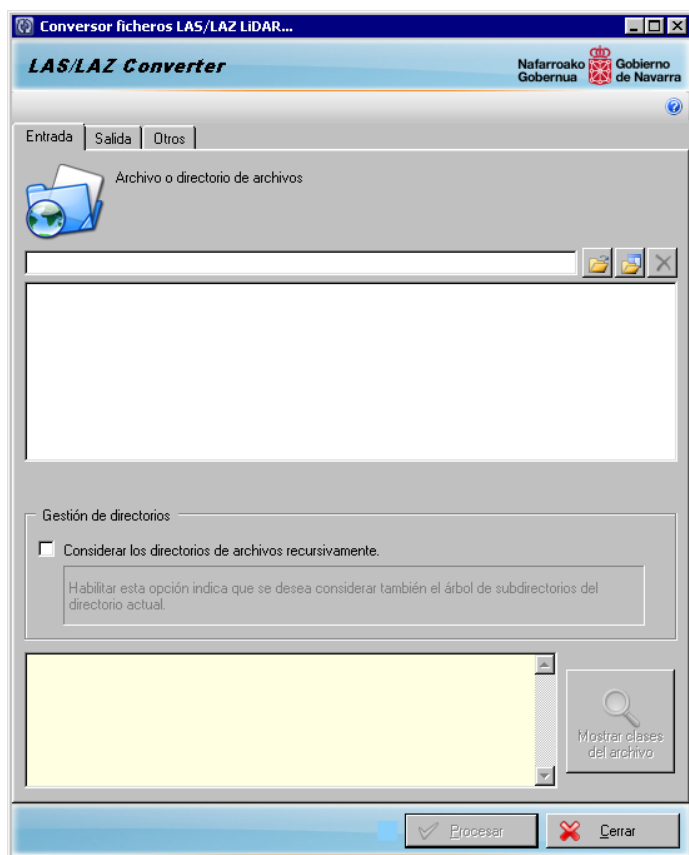
Se genera también una imagen “hillshade” coloreada por clases de cada uno de los bloques reclasificados (1x1 km). Permite la rápida visualización del contenido de los ficheros LiDAR.

Utilizamos **GDAL** en combinación con herramientas propias.



Productos – Herramienta “LiDAR Converter”

La Dirección General de Obras Públicas de Navarra ofrece gratuitamente una herramienta desktop para la conversión y extracción de información de ficheros LAS/LAZ de forma sencilla y rápida.



Funcionalidades:

- Comprimir/descomprimir ficheros LAS/LAZ.
- Cambio de sistema de referencia.
- Cambio cota (elipsoidal <> ortométrica).
- Cambio de la versión y formato de punto.
- Filtrado de clases.
- Inspeccionar atributos de un fichero.
- Filtrar por rango espacial.

Trabaja sobre fichero/s o directorios.
Sólo para Microsoft Windows (x64).
Basada en PDAL.

ftp://ftp.cartografia.navarra.es/5_LIDAR/

¿ Futuro ?



- Implementación de una nueva función de coste para la validación de los algoritmos que penalice en mayor grado errores más graves en el etiquetado de clases.
- Evaluación de algoritmos basados en redes neuronales.
- Implementar la ejecución de los procesos basados en GPU.

¿ Preguntas ?



Muchas gracias

Alvaro Huarte
ahuarte@tracasa.es