Lightweight model design for automatic bare earth point detection in 3D bathymetric LiDAR point clouds

X. Pellerin Le Bas, C. Delblond, E. Marrec, L. Froideval, C. Monpert

HYDRO 2025

28th – 30th October, Liverpool

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

SHOM
L'océan en référence

SCIENTEAMA
EXPLORE VOS PRODUITS DE DEMAIN

cnrs

M2C

Xavier PELLERIN LE BAS - SCIENTEAMA

xavier.pellerin@scienteama.fr

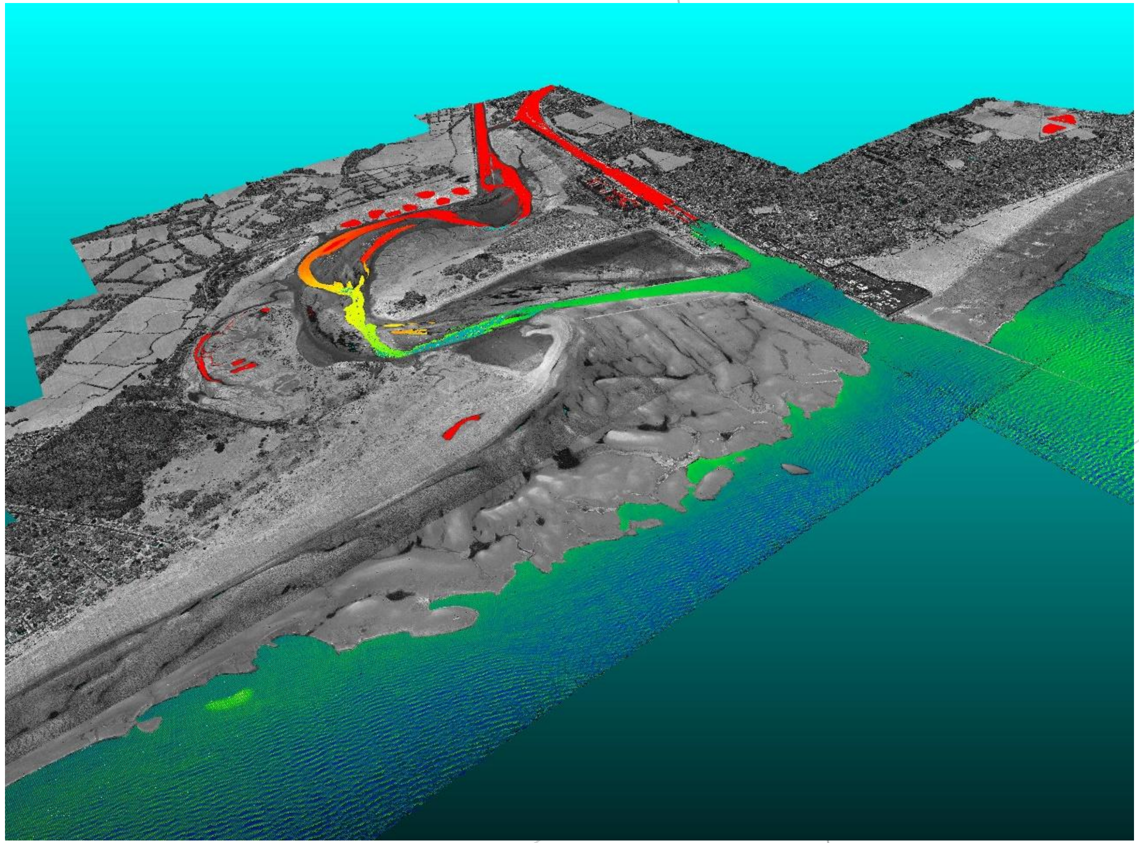
1 – Partnership – SHOM

- SHOM, France's national hydrographic service
  - Official operator for maritime and coastal geographic information
- Missions:
  - Understand and describe the physical marine environment (relations with atmosphere, seabed, coastal areas)
  - Forecast its evolution
  - Disseminate corresponding information
- **Litto3D** – SHOM and IGN
  - French coastal mapping program
  - Provide **high-resolution**, **accurate** and **continuous land-sea** elevation model for **entire coastline**

Litto3D – Topo-Bathy LiDAR survey of Saint-Pierre and Miquelon

1 – Partnership – CNRS: M2C Lab

- LiDAR team of M2C Laboratory
  - LiDAR operator of National Centre for Scientific Research (CNRS)
  - 15 years of coastal airborne LiDAR survey
  - Point cloud experts in acquisition and processing
  - SfM – Photogrammetry for coastal protection
  - Coastal and ocean altimetry (SWOT)
- Common platform with LiDAR bathymetry
  - French topo-bathymetric platform
  - Nantes, Rennes and Caen

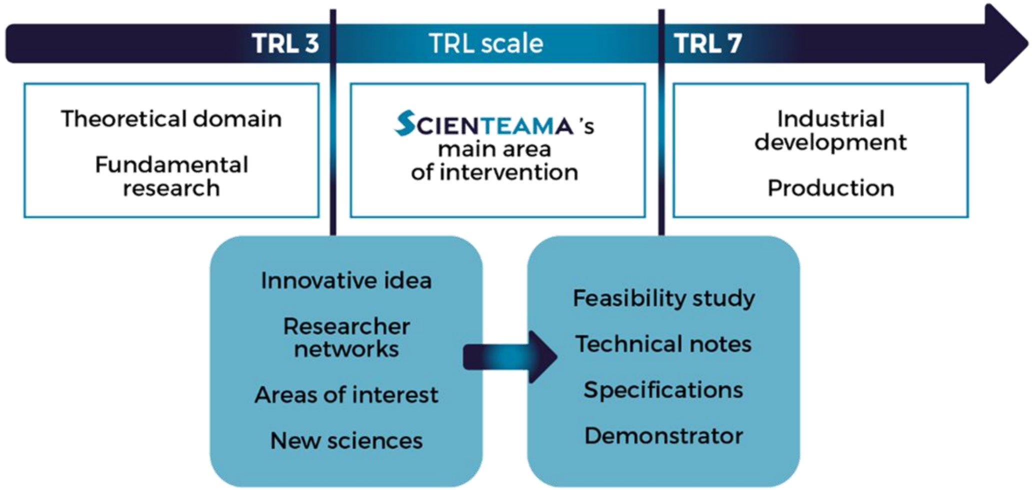


Surface water classification of Topo LiDAR survey of Orne estuary (Normandy, France)

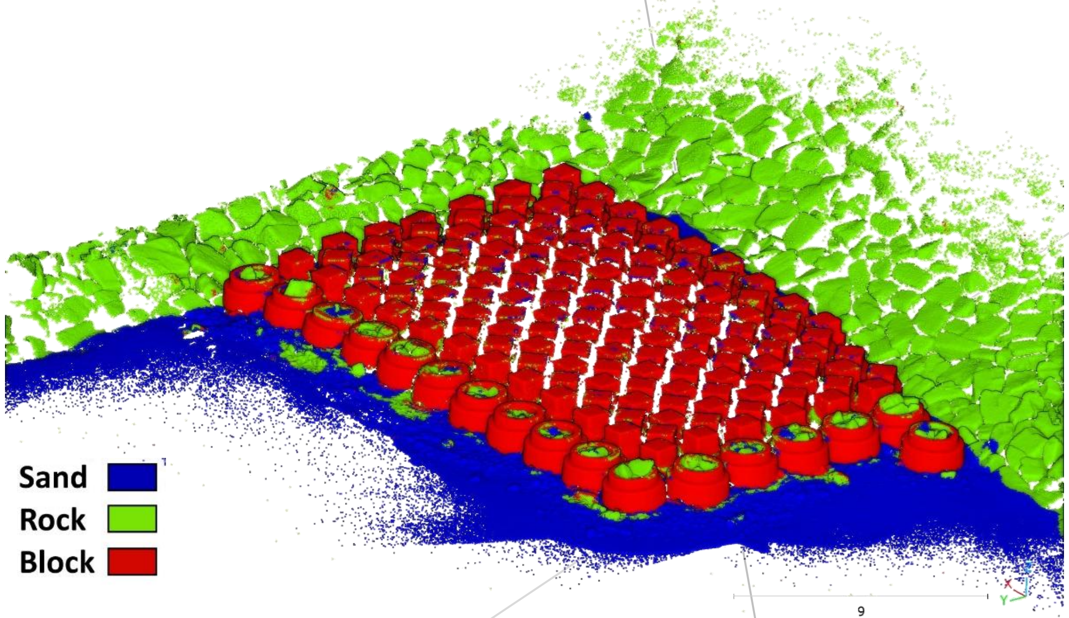
1 – Partnership – Scienteama

- SCIENTEAMA, research & development company
  - Signal processing: ultra-low and noisy signal
  - Instrumentation design: own RADAR architecture
  - Energy harvesting, Autonomous sensor and IoT
  - **Machine Learning:** Active software developpment, Automatic classification of 3D point clouds

SCIENTEAMA
EXPLORE VOS PRODUITS DE DEMAIN



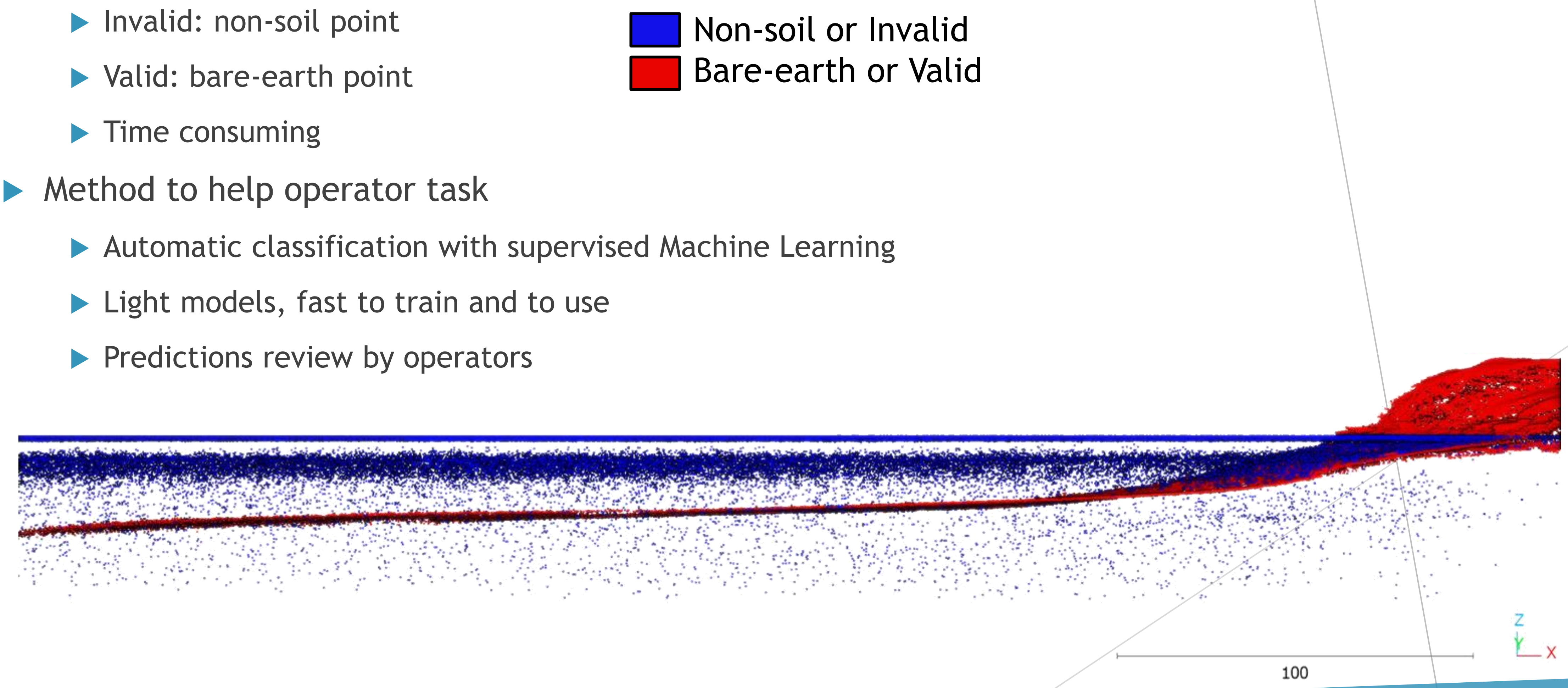
Automatic classification with Machine Learning



Classification of SfM model as sand, rock and concrete block (Normandy, France)

2 – Classification challenge

- Coastal mapping
  - Segment Topo-Bathy LiDAR point clouds as:
    - Invalid: non-soil point
    - Valid: bare-earth point
    - Time consuming
  - Method to help operator task
    - Automatic classification with supervised Machine Learning
    - Light models, fast to train and to use
    - Predictions review by operators



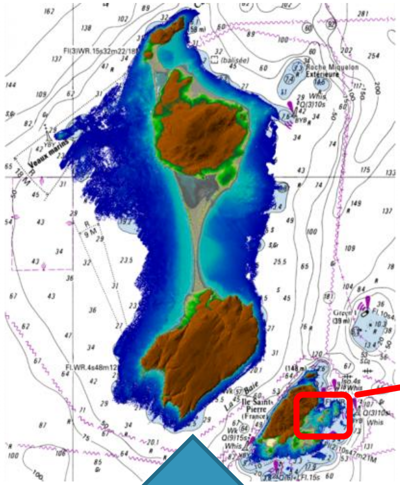
2 – Dataset

- Saint-Pierre and Miquelon survey
  - Acquisition: Hexagon company
  - Lidar Hawkeye 5, Leica
  - Topo-Bathy (up to 20m to Chart Datum)
  - Topo, shallow and deep LiDAR
  - Total area: 715 km^2
  - Imagery capture
- Used dataset
  - Part of Saint-Pierre
  - 12 km^2: 12 x Tiles (1x1km)
  - 500 million points

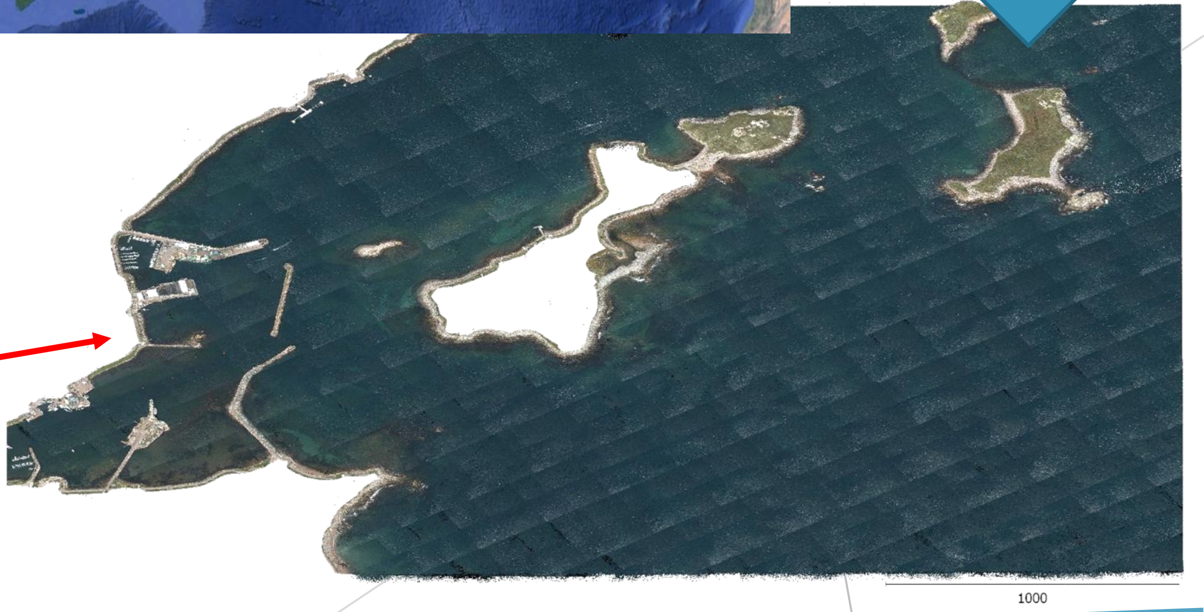
Liverpool

Saint-Pierre and Miquelon

Used dataset (500 Mpts)



SPM survey



3 – cLASpy_T

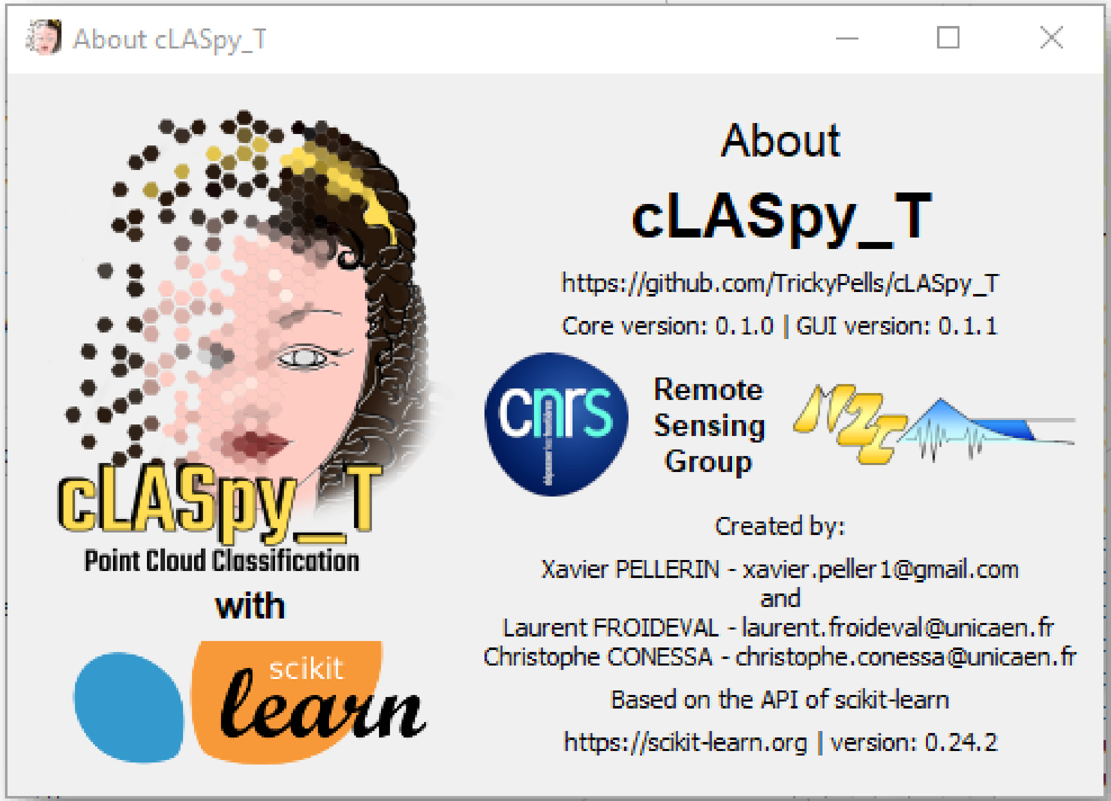
- **c**lassification **LAS** **py**thon **T**ools

Open software → CeCILL License[1]

Work in progress !

Design ML models to classify 3D point clouds

- Linux and Windows OS
- Based on **scikit-learn API**
- **Supervised algorithms:**
  - *Random Forest*
  - *Gradient Boosting*
  - *Multi-Layer Perceptron* → Neural network
- **Non-supervised algorithm:**
  - *KMeans* → Clustering



1: http://www.cecill.info/index.en.html

3 – cLASpy_T

- classification LAS python Tools

Open software → CeCILL License[1]

Work in progress !

Design ML models to classify 3D point clouds

- **Github:** https://github.com/TrickyPells/cLASpy_T
- **Documentation:** https://claspy-t-project.readthedocs.io/en/latest/
- **Article:** Pellerin Le Bas X., Froideval L., Mouko A., Conessa C., Benoît, L. and Perez L., **2024** *A new open-source software to help the design of models for automatic 3D point cloud classification in coastal studies*

DOI: https://doi.org/10.3390/rs16162891

remote sensing

MDPI

Technical Note

A New Open-Source Software to Help Design Models for Automatic 3D Point Cloud Classification in Coastal Studies

Xavier Pellerin Le Bas [1,*], Laurent Froideval [2], Adan Mouko [2], Christophe Conessa [2], Laurent Benoit [2] and Laurent Perez [2]

[1] Scienteama, 4 Avenue de Cambridge, 14200 Hérouville-Saint-Clair, France
[2] Normandie Univ, UNICAEN, UNIROUEN, CNRS, M2C, 14000 Caen, France; laurent.froideval@unicaen.fr (L.F.); adan-wealth-sy.mouko.auditeur@lecnam.net (A.M.)
* Correspondence: xavier.pellerin@scienteama.fr

Abstract: This study introduces a new software, cLASpy_T, that helps design models for the automatic 3D point cloud classification of coastal environments. This software is based on machine learning algorithms from the scikit-learn library and can classify point clouds derived from LiDAR or photogrammetry. Input data can be imported via CSV or LAS files, providing a 3D point cloud, enhanced with geometric features or spectral information, such as colors from orthophotos or hyperspectral data. cLASpy_T lets the user run three supervised machine learning algorithms from the scikit-learn API to build automatic classification models: RandomForestClassifier, GradientBoostingClassifier and MLPClassifier. This work presents the general method for classification model design using cLASpy_T and the software's complete workflow with an example of photogrammetry point cloud classification. Four photogrammetric models of a coastal dike were acquired on four different dates, in 2021. The aim is to classify each point according to whether it belongs to the 'sand' class of the beach, the 'rock' class of the riprap, or the 'block' class of the concrete blocks. This case study highlights the importance of adjusting algorithm parameters, selecting features, and the large number of tests necessary to design a classification model that can be generalized and used in production.

check for updates

Citation: Pellerin Le Bas, X.; Froideval, L.; Mouko, A.; Conessa, C.; Benoit, L.; Perez, L. A New Open-Source Software to Help Design Models for Automatic 3D Point Cloud Classification in Coastal Studies. *Remote Sens.* **2024**, *16*, 2891. https://doi.org/10.3390/rs16162891

Academic Editors: Scot Smith and Kamal Darwish

Received: 31 May 2024
Revised: 22 July 2024

Keywords: machine learning; classification; coastal environment; point cloud processing; photogrammetry

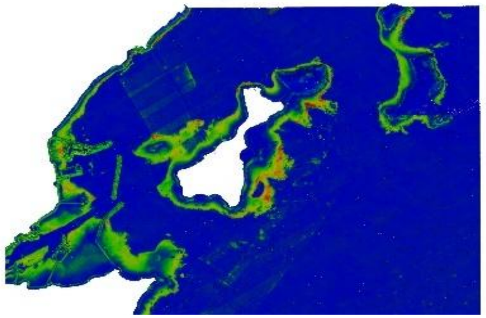
1. Introduction

3D point clouds have taken on an increasingly important role in the world of geospatial data, mostly due to the growing use of laser scanning methods, Structure from Motion (SfM) photogrammetry, and Red, Green and Blue-Depth (RGB-D) sensors used in robotics [1]. Specifically, point clouds are often used for environmental studies such as coastal geomorphology monitoring [2–4], coastal infrastructure monitoring [5–7], and cliff monitoring [8–10]. Point clouds are generally used as a basis for generating products that are very useful in geosciences: rasterized products such as digital terrain models

4 – Method

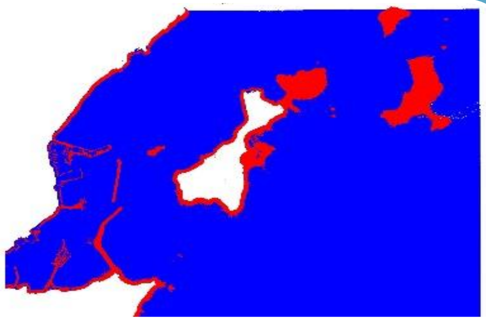
Dataset

- 12 tiles (500 Mpts)



Features

- Compute geometric features

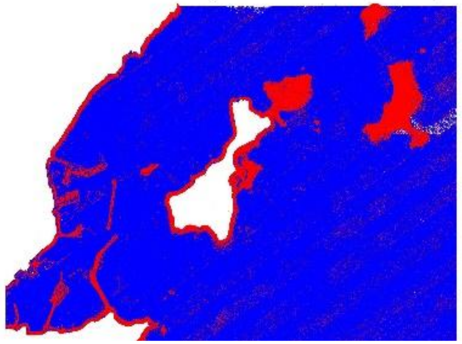


Labels

- Non-soil: 0
- Bare-earth: 1

Training

- Zone selection
- Training models



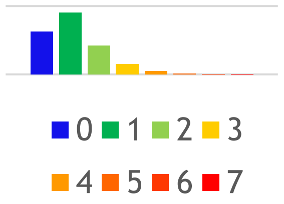
Predictions

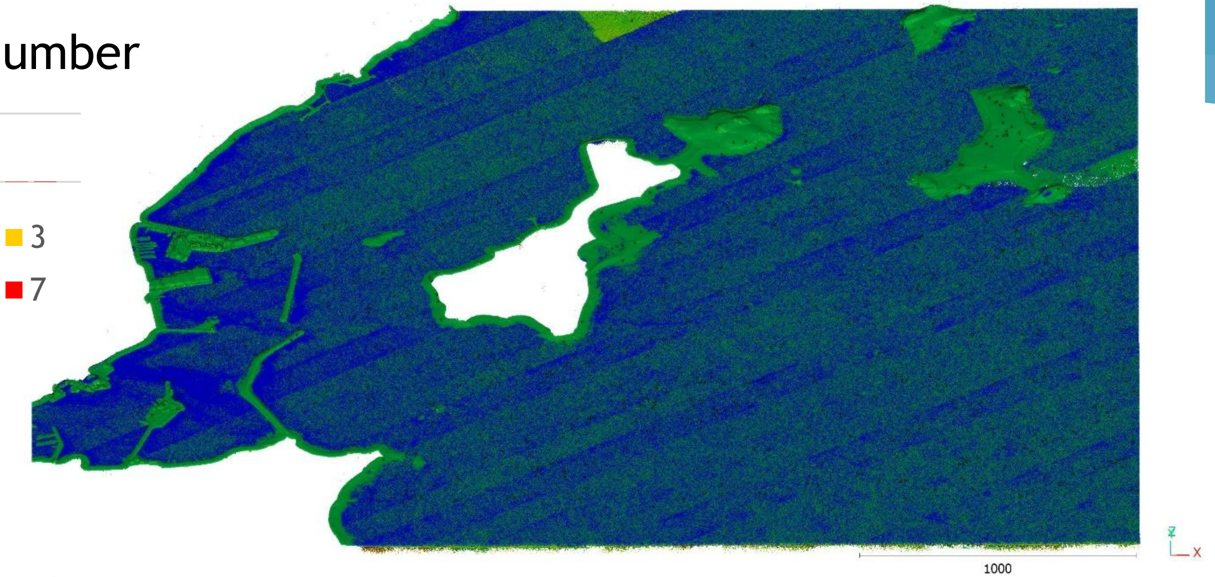
- Test – 500 kpts
- Production – 44.5 Mpts

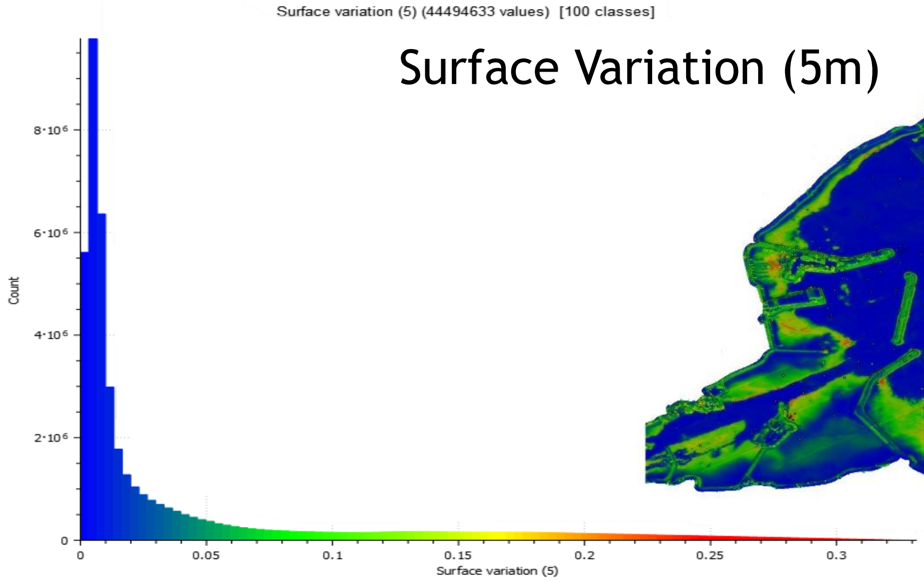
5 – Dataset and Features

- **Standard LAS features**
  - Intensity
  - Return number
  - Number of returns
  - User_data: topo, shallow and deep LiDAR
- **Compute Geometric features (1 and 5 m)**
  - Anisotropy
  - Mean curvature
  - Omnivariance
  - Sphericity
  - Surface variation
  - Verticality

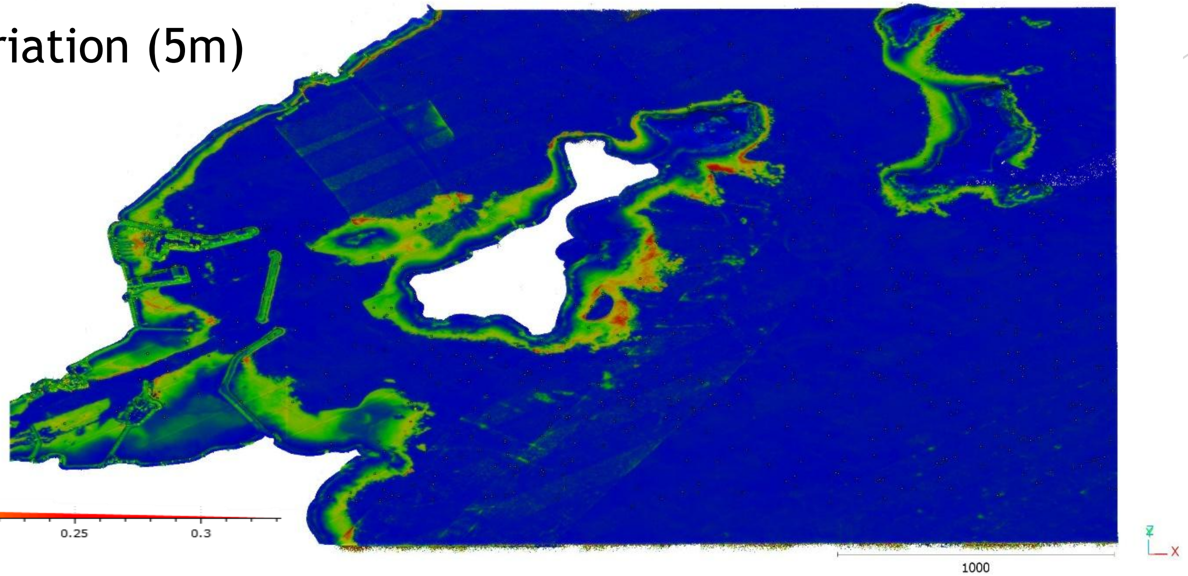
Return number





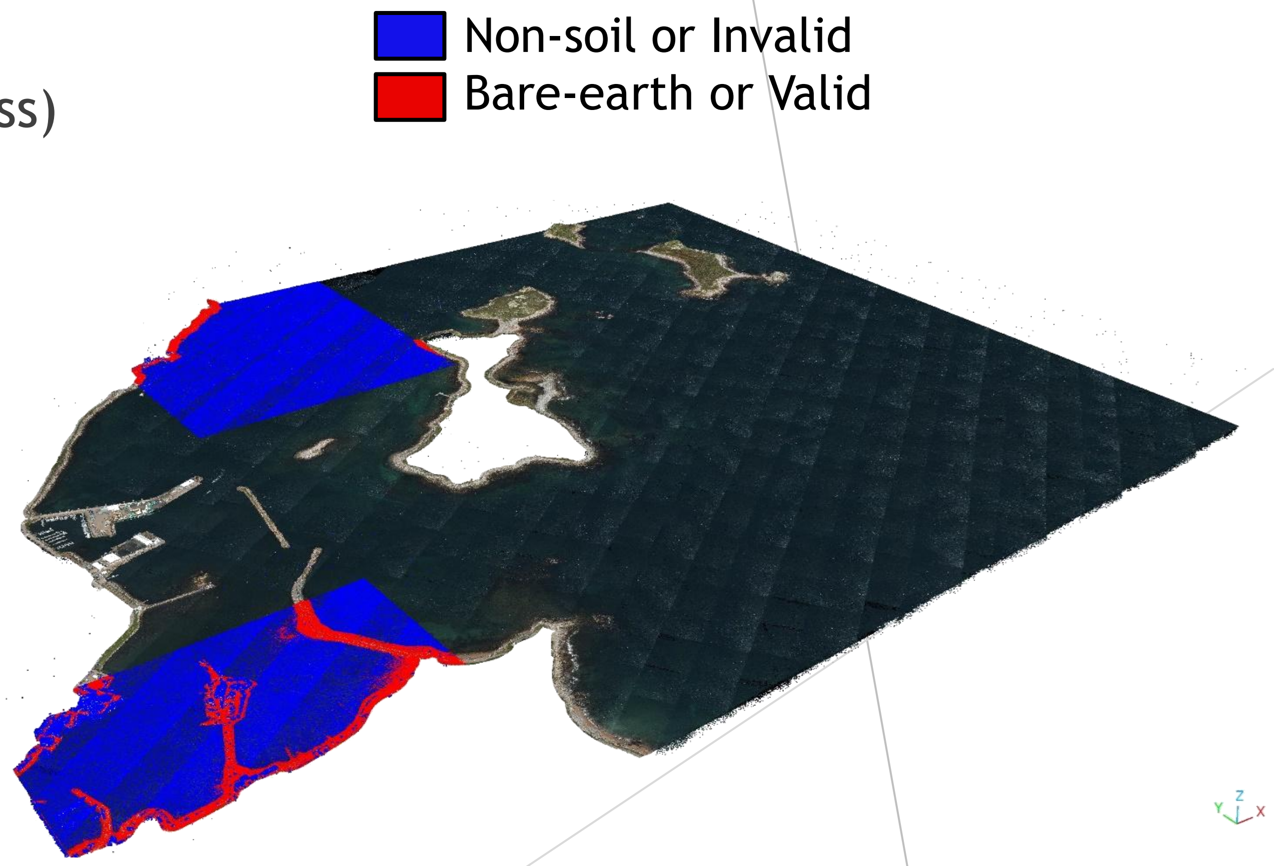


Surface Variation (5m)



6 – Bare-earth model

- **Trainset - 2 million points**
  - 2 zones: subsample 2 tiles (1Mpts/zone)
  - 2 classes: Non-soil & Bare-earth (1Mpts/class)
- **Training**
  - Decision Trees: Gradient Boosting Classifier
  - Number of trees: 50
  - Maximum depth: 8
  - Minimum sample split: 500
  - Model trained:
    - 500 kpts
    - In 5 min 38 sec
    - All features (previous slide)

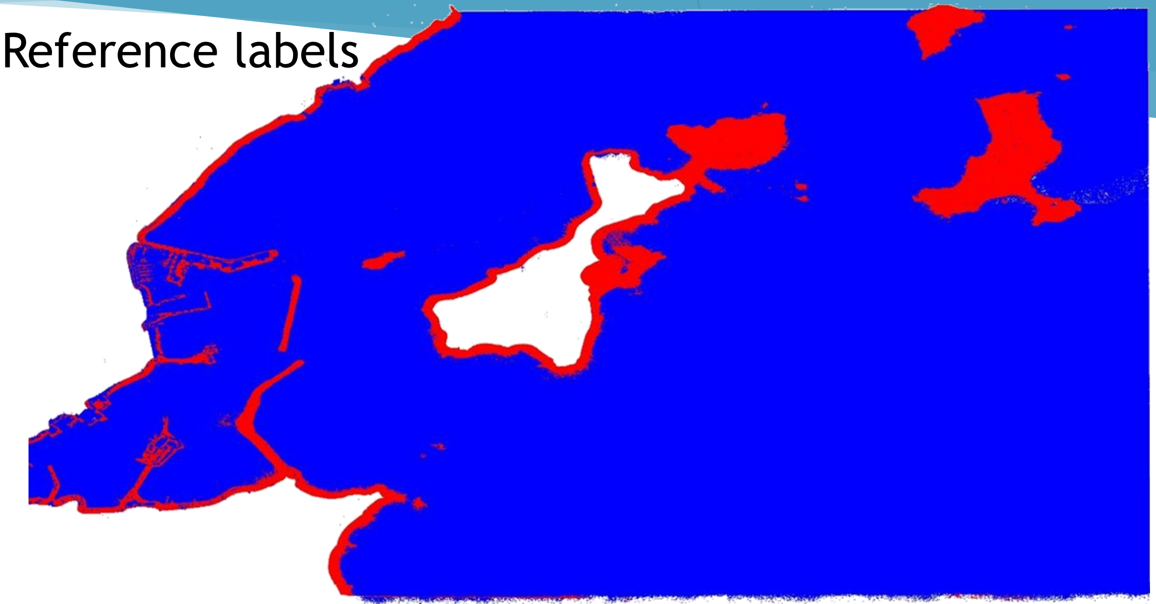


6 – Bare-earth model

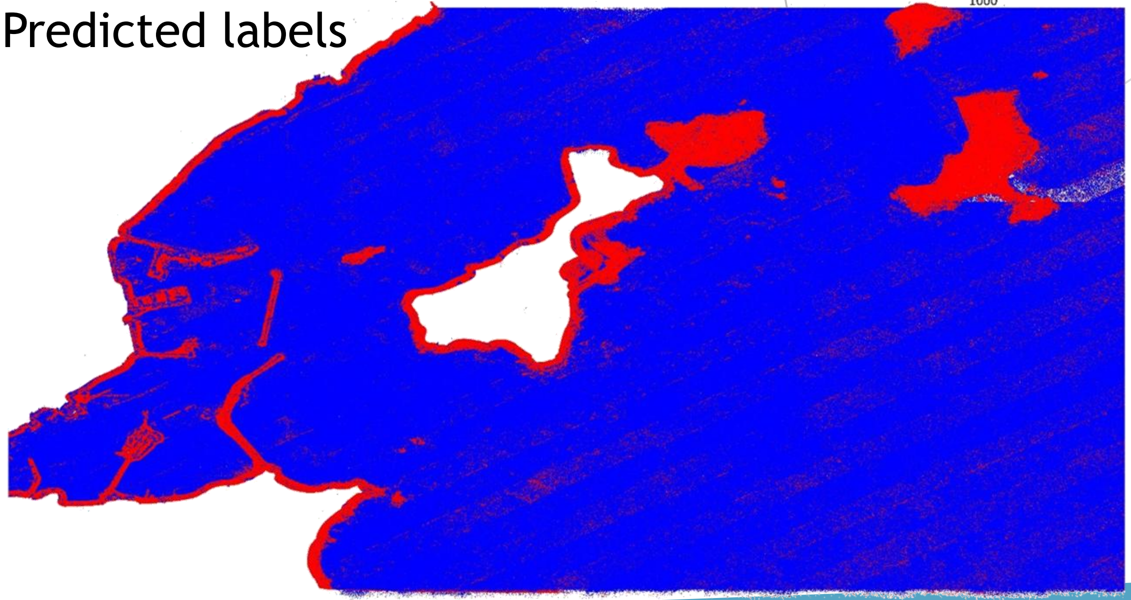
- **Predictions**
  - 12 tiles subsample: 44,5 Mpts
  - Time prediction: 7 min 50
  - Global accuracy: 95%
- **Scores:**

Class	Metric	Test 500 kpts	12 Tiles 44.5 Mpts
Non-soil	Precision	96%	99%
	Recall	94%	95%
Bare-earth	Precision	94%	86%
	Recall	96%	96%
Global Accuracy		95%	95%

Reference labels



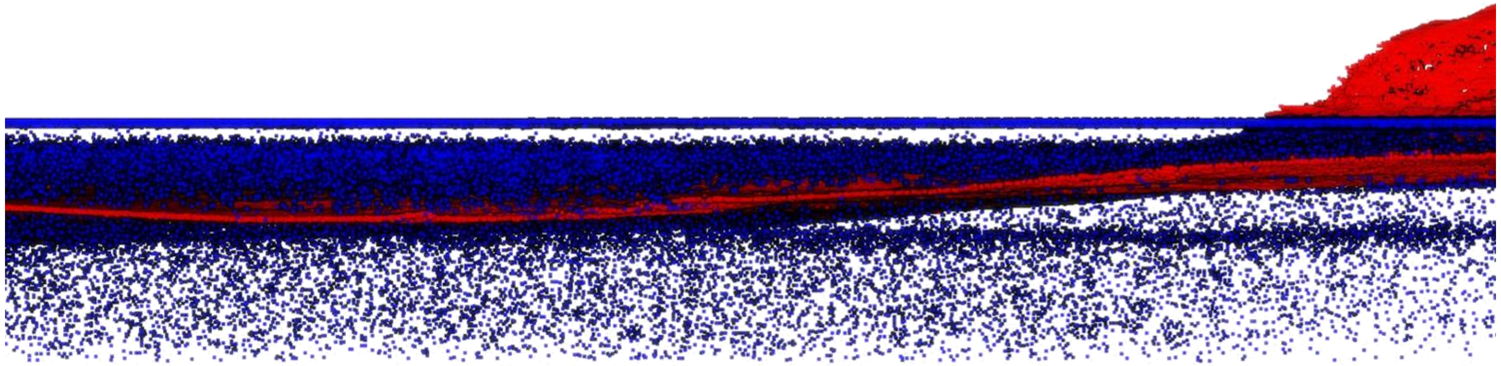
Predicted labels

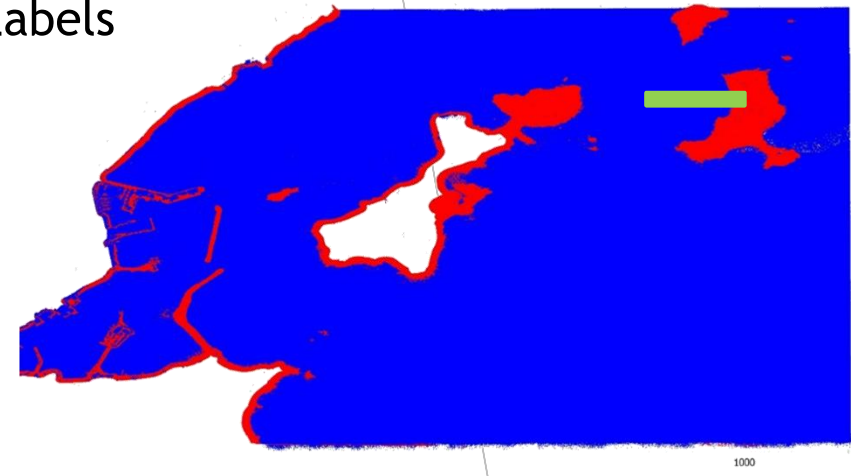


6 – Bare-earth model

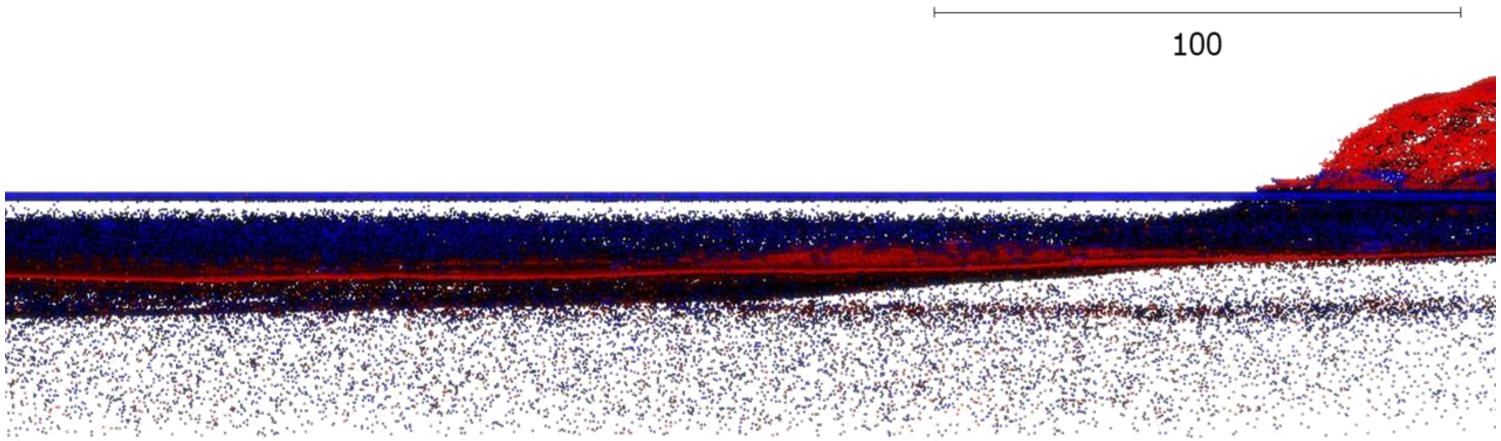
- **Zoom**

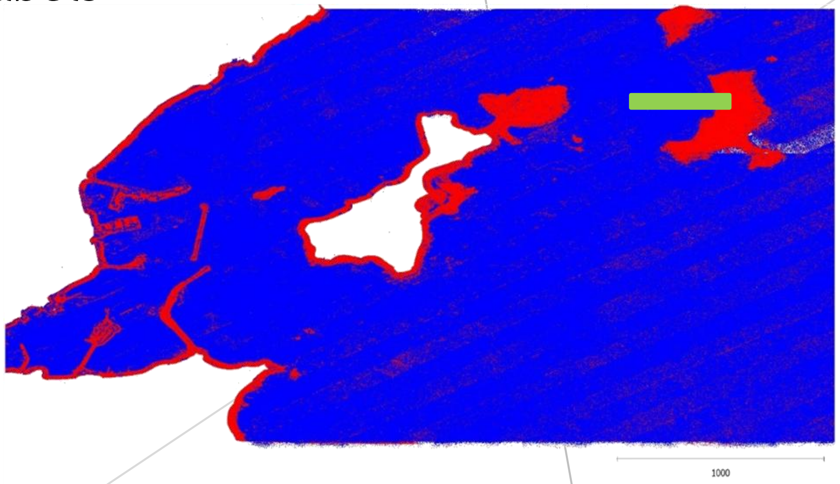
Reference labels





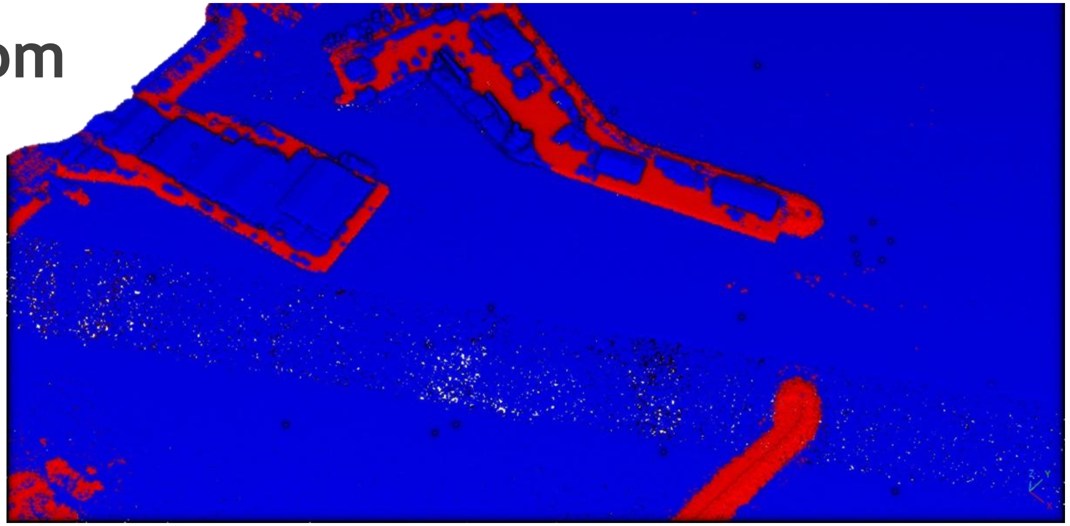
Predicted labels

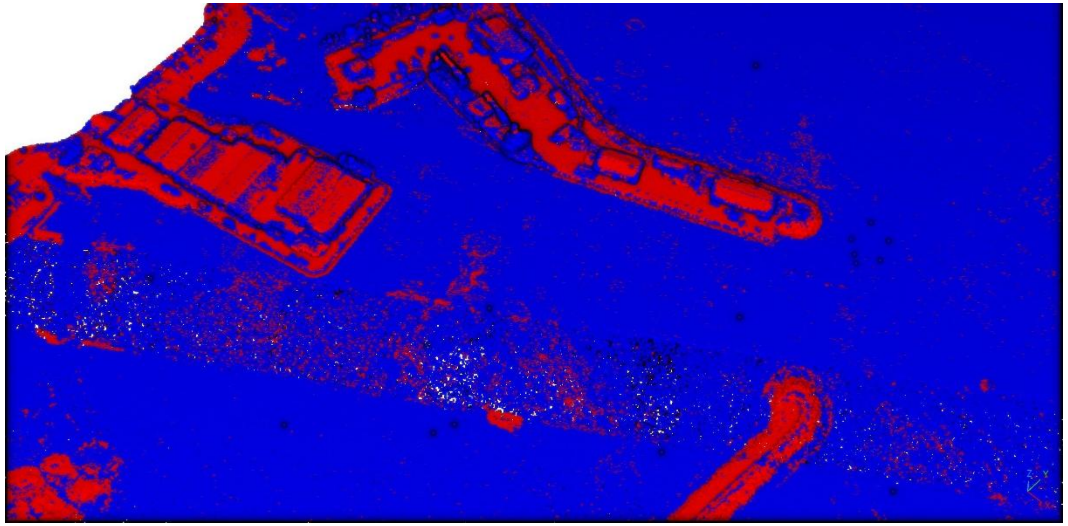




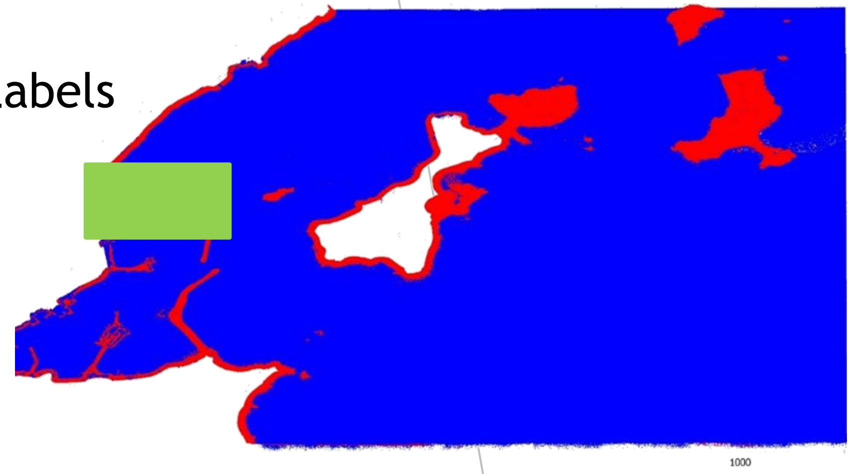
6 – Bare-earth model

- **Zoom**





Reference labels

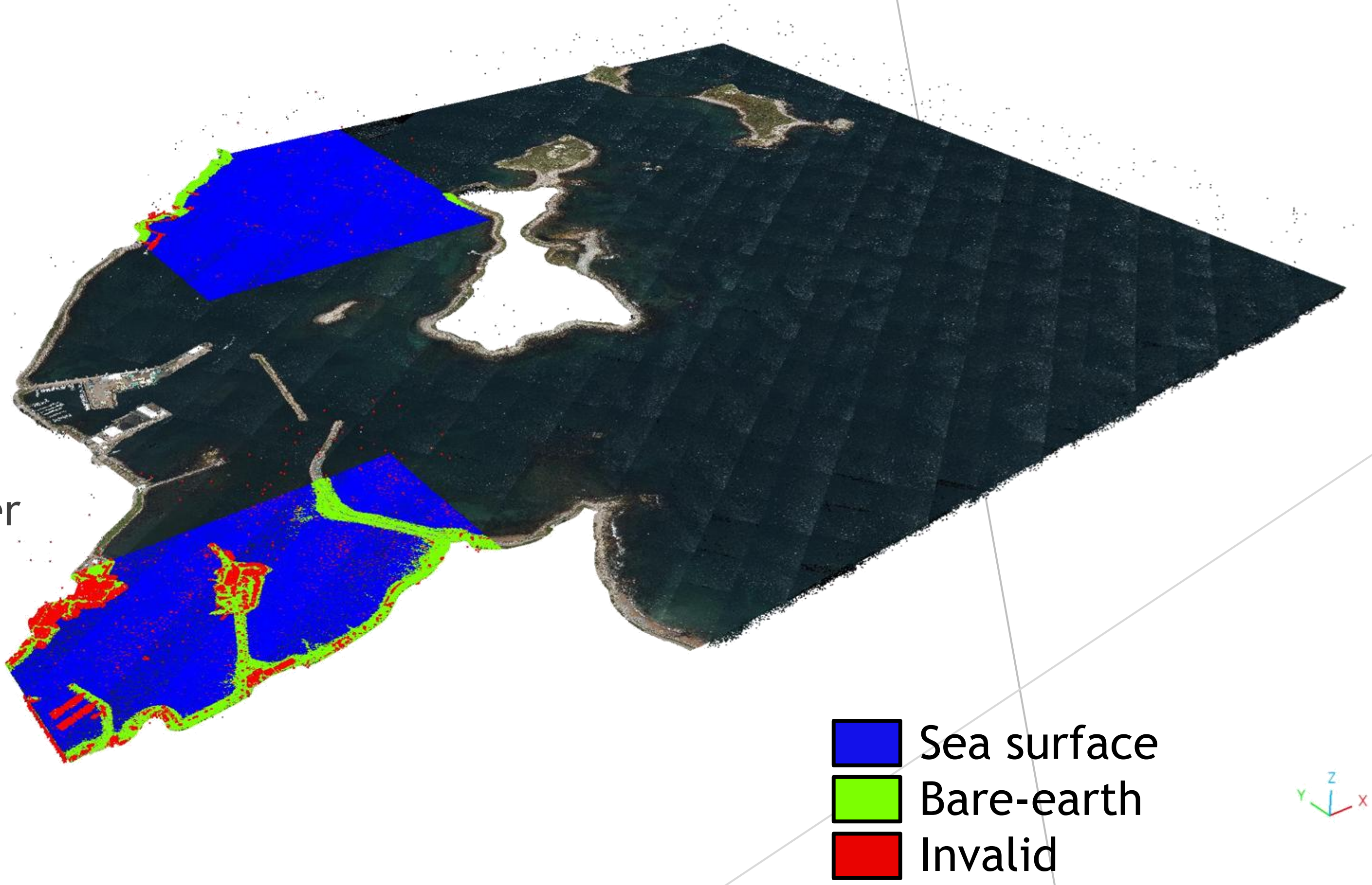


Predicted labels

7 – 3 classes model

- **Trainset - 2 million points**
  - 2 zones: subsample 2 tiles (1Mpts/zone)
  - 3 classes 666 666 pts/class
    - Sea surface
    - Bare-earth
    - Invalid

- **Training**
  - Decision Trees: Gradient Boosting Classifier
  - Number of trees: 50
  - Maximum depth: 8
  - Minimum sample split: 500
  - Model trained:
    - **500 kpts**
    - In **15 min 38 sec**
    - Same features (previous trainset)

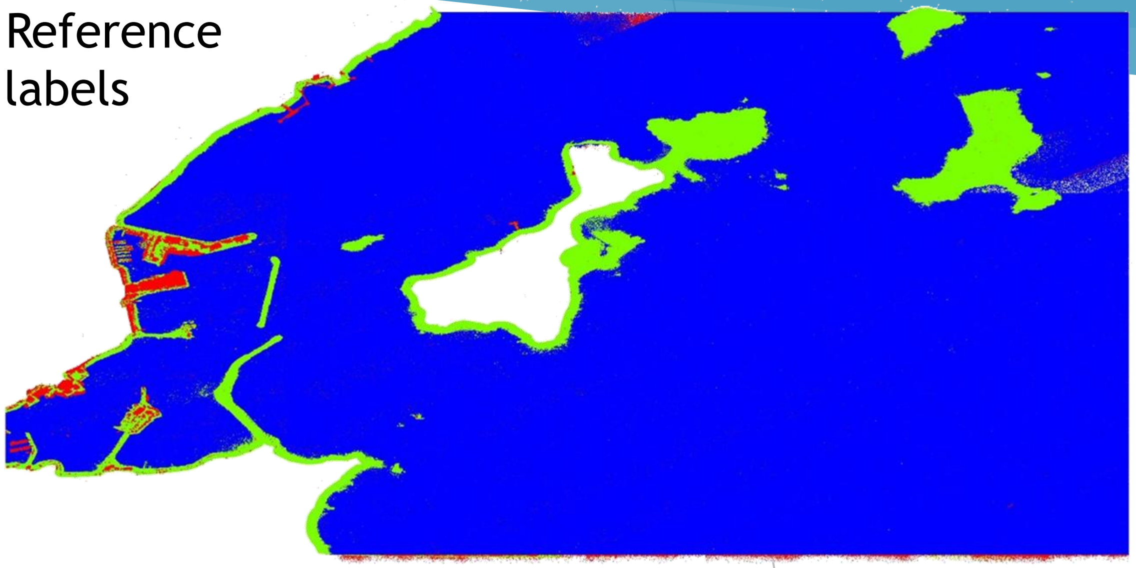


7 – 3 classes model

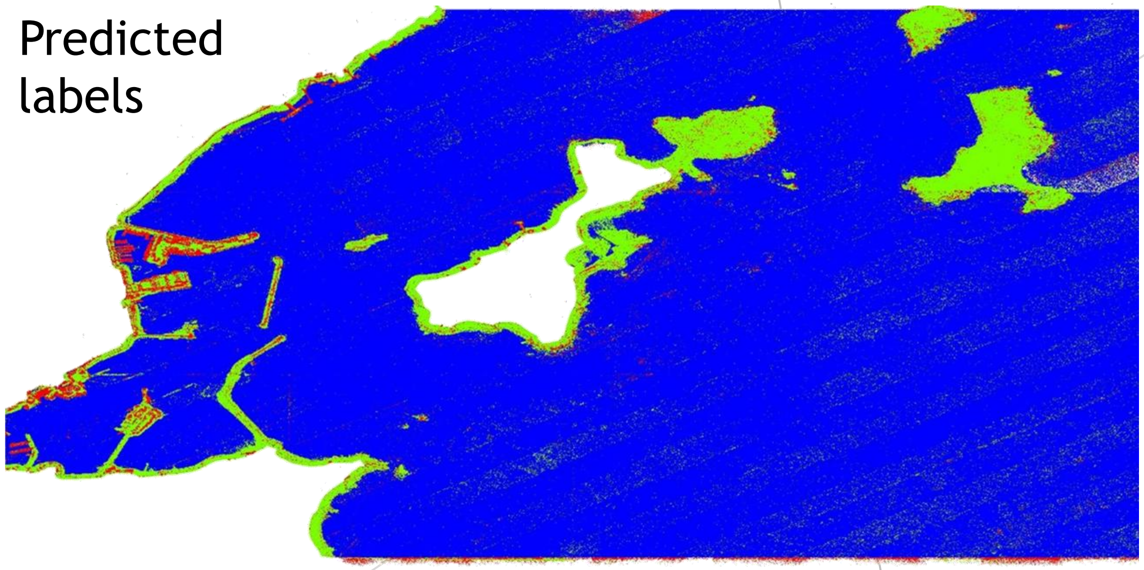
- **Predictions**
  - 12 tiles subsample: 44,5 Mpts
  - Time prediction: 8 min 35
  - Global accuracy: 95%
- **Scores:**

Class	Metric	Test 500 kpts	12 Tiles 44.5 Mpts
Sea surface	Precision	96%	95%
	Recall	97%	96%
Bare-earth	Precision	93%	90%
	Recall	94%	94%
Invalid	Precision	95%	86%
	Recall	92%	76%
Global Accuracy		94%	93%

Reference labels

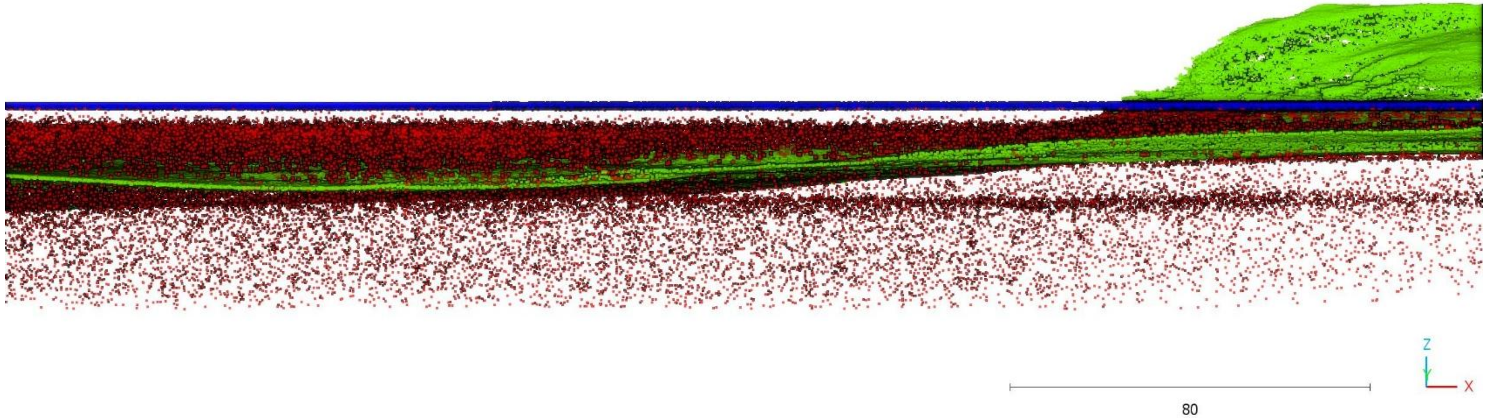


Predicted labels

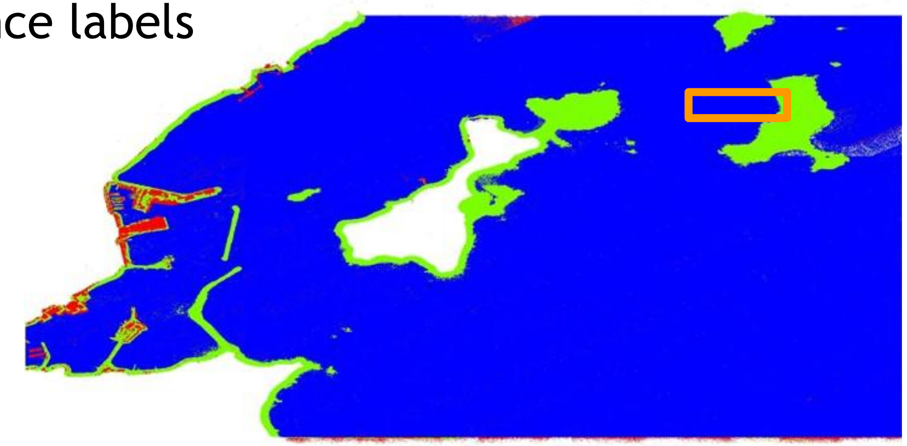


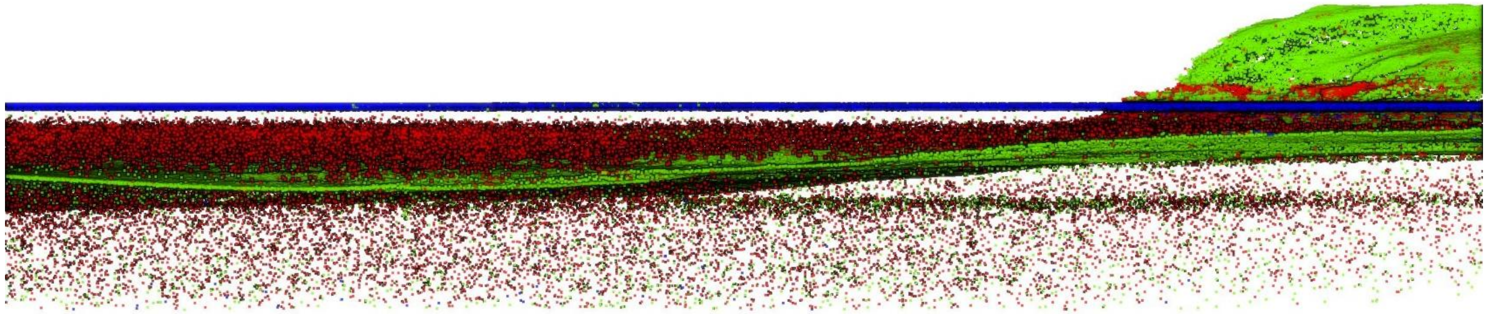
7 – 3 classes model

- Zoom

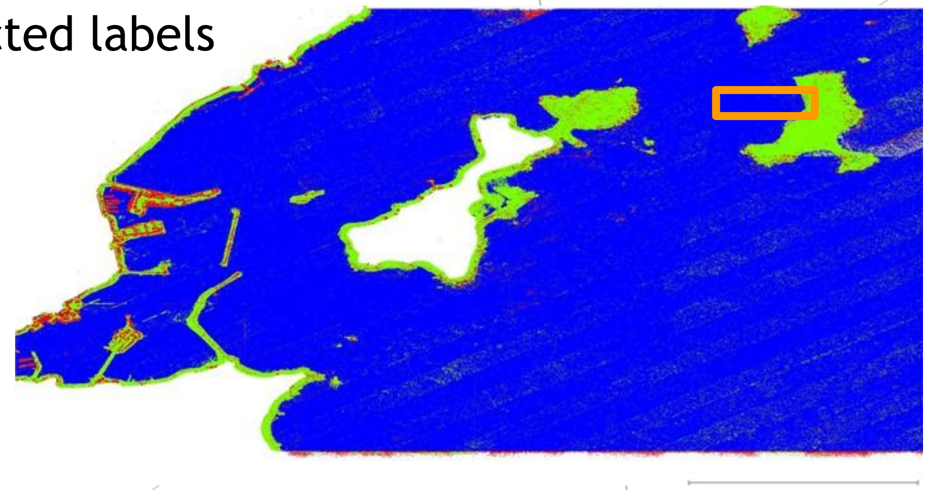


Reference labels



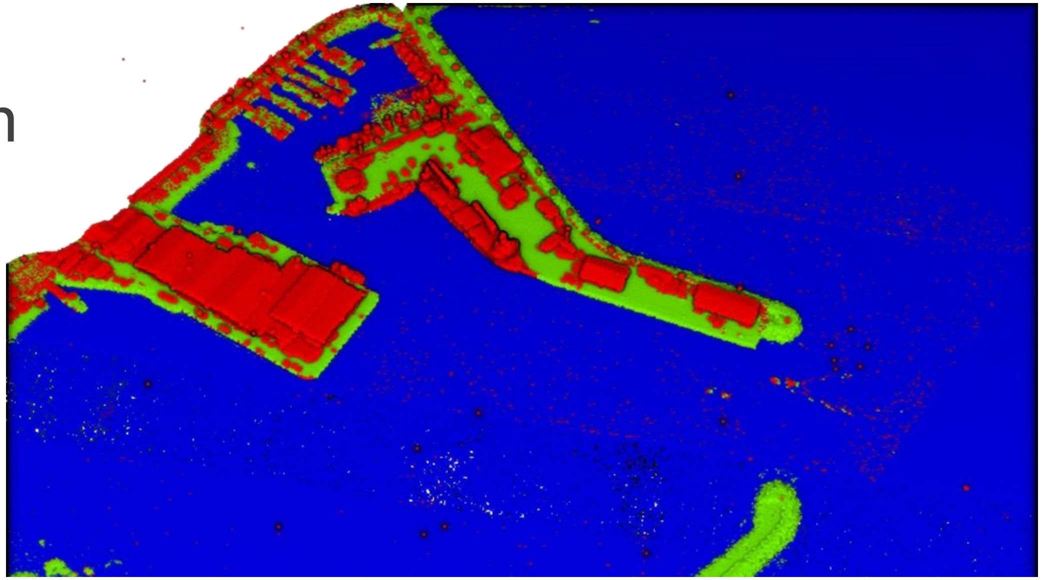


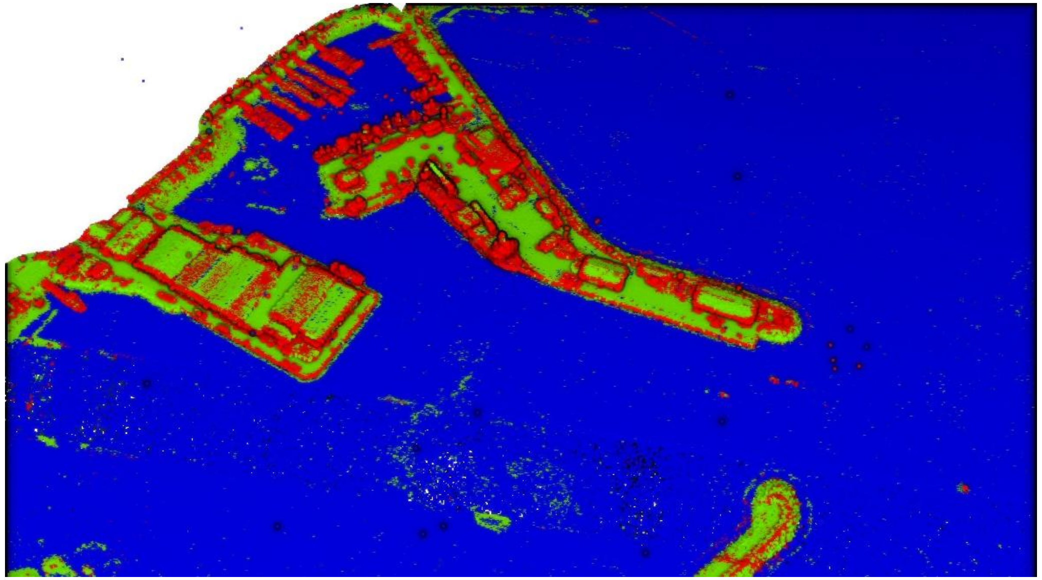
Predicted labels



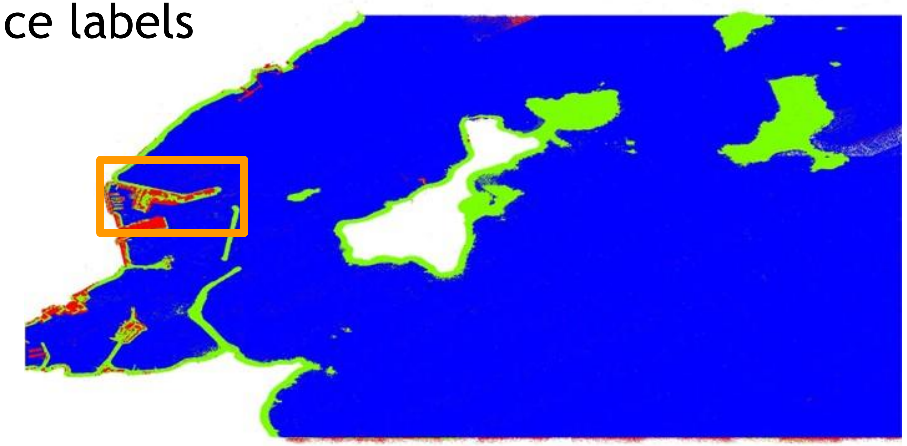
7 – 3 classes model

- Zoom

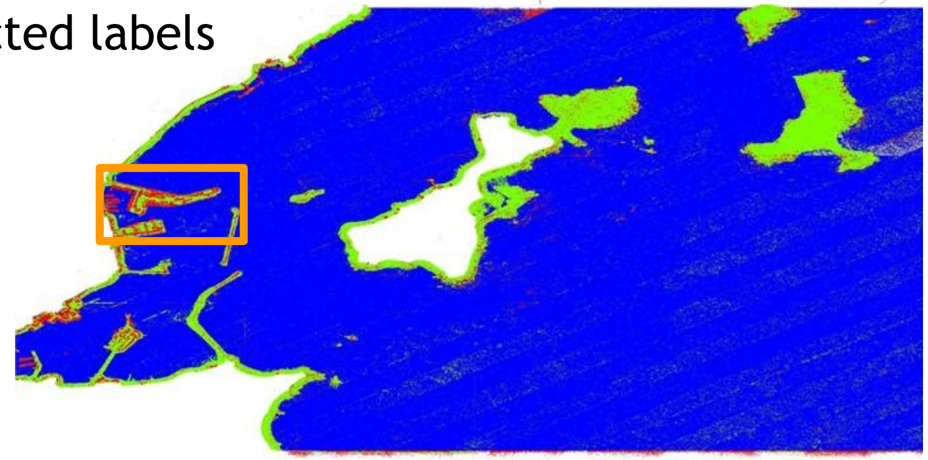




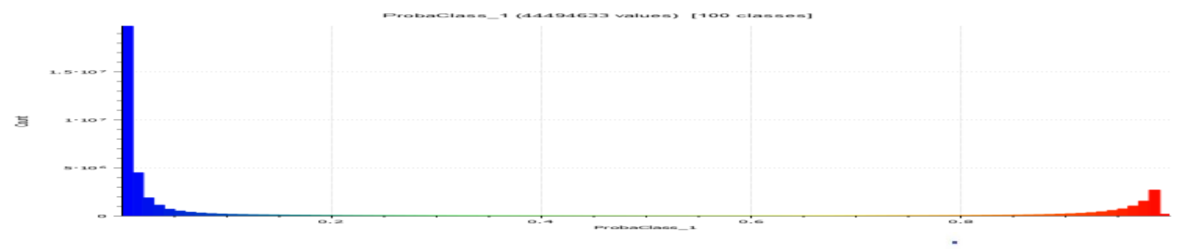
Reference labels



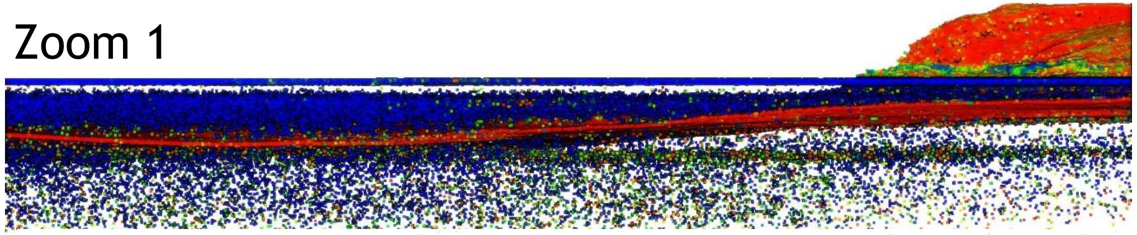
Predicted labels



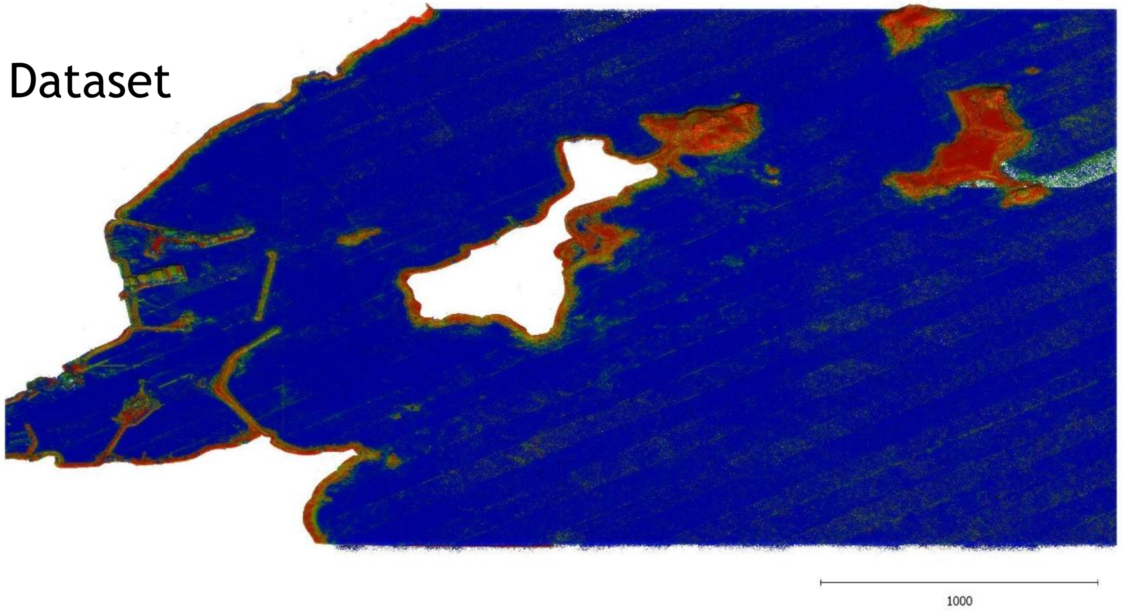
8 – Bare-earth likelihood



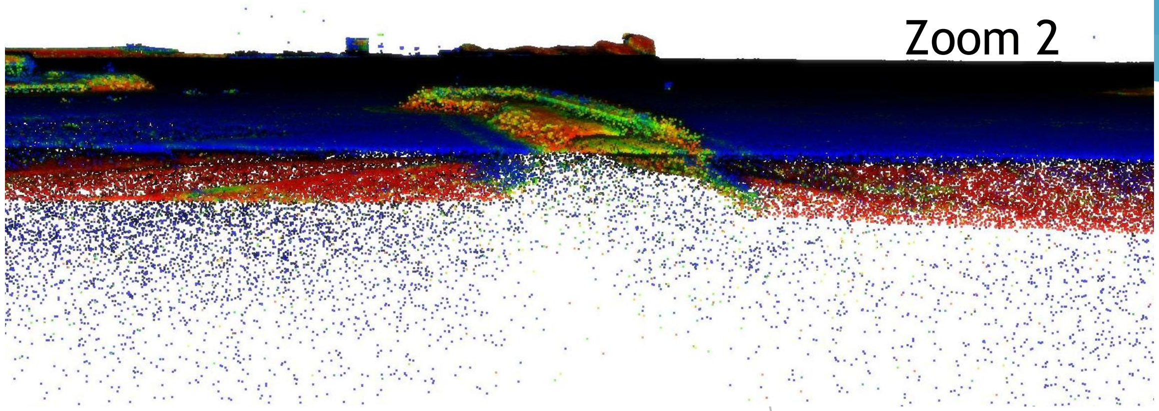
Zoom 1



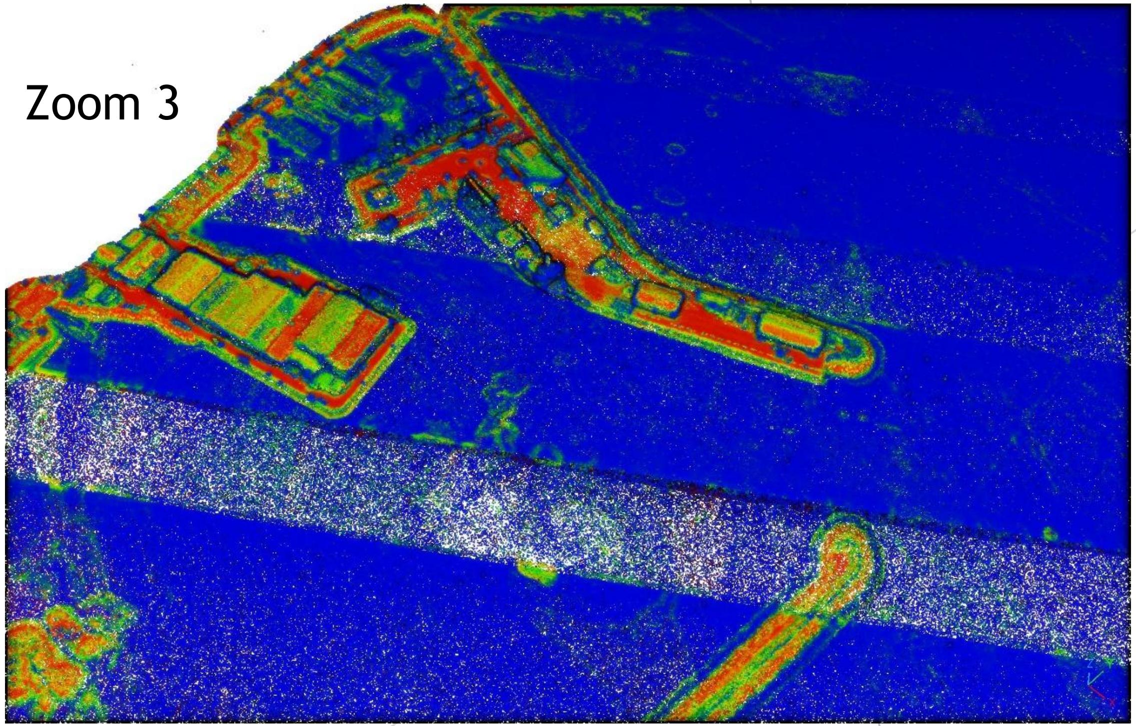
Dataset



Zoom 2



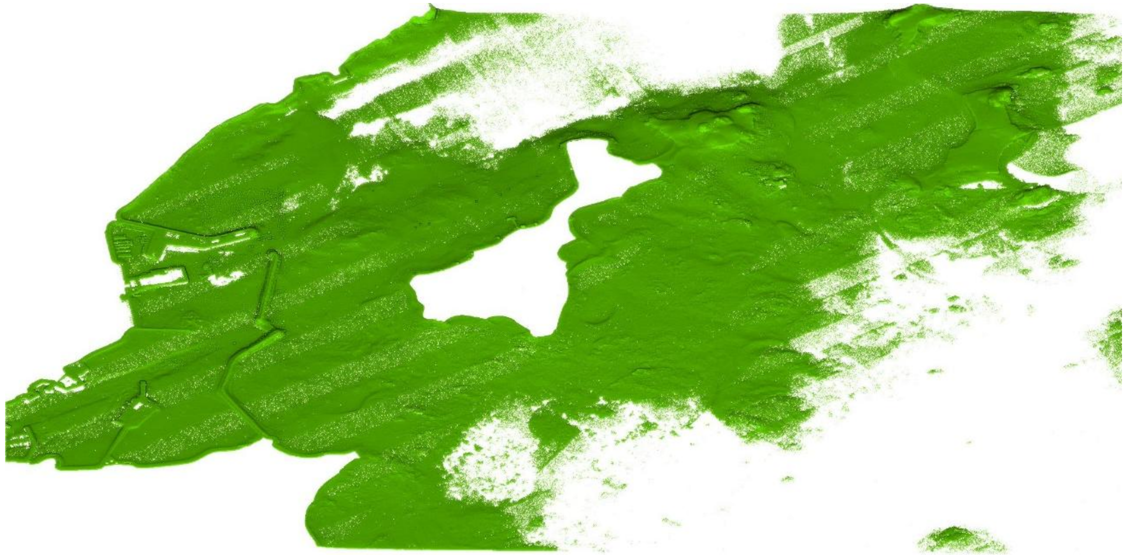
Zoom 3



7 – Bare-earth likelihood

Bare-earth reference

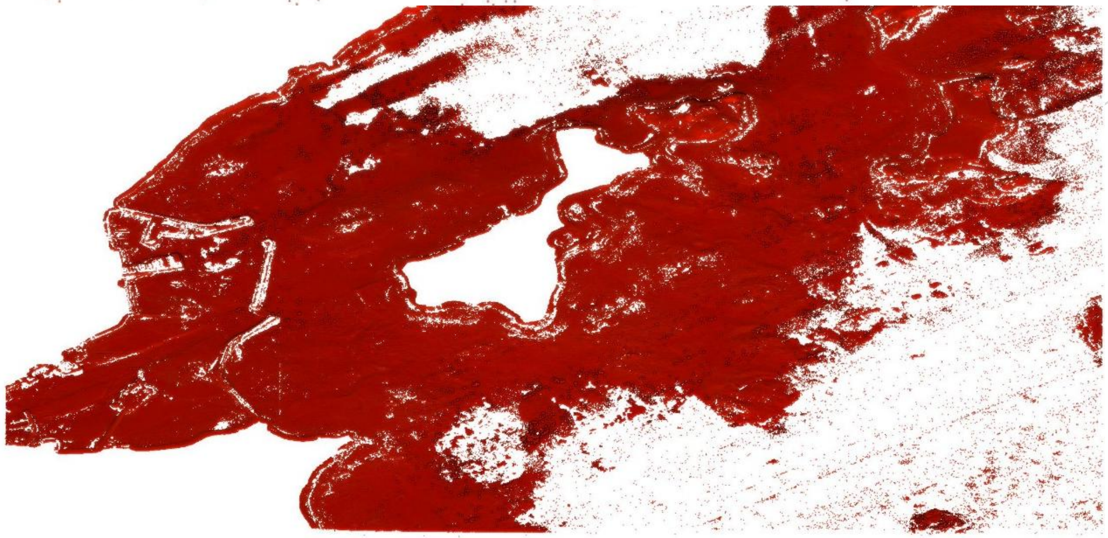
Zoom 1



Bare-earth reference

Bare-earth predicted Likelihood > 95%

Zoom 1 – 95% likelihood



Dataset – 95% likelihood

8 – Conclusion

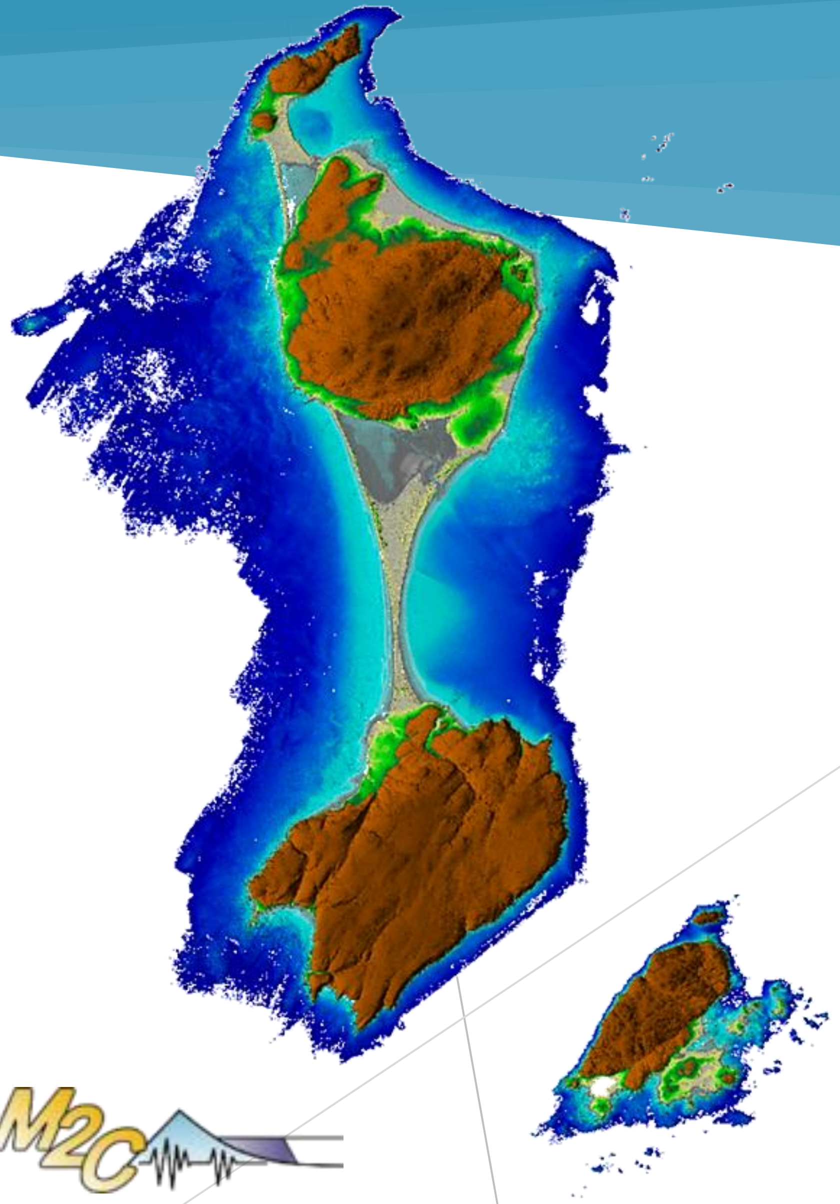
- **Model 2 classes**
  - **Non-soil:** Precision 99% and Recall 95% over 12 tiles (500 million points)
  - **Very promising to remove non-soil points**
  - **Bare-earth:** Bad generalization, Precision from 94% to 86%
  - **Increase** bare-earth point **variety**, add other tiles in trainset
- **Model 3 classes**
  - **Sea surface:** Precision 95% and Recall 96% over 12 tiles
  - **Very promising to remove only sea surface**
  - Bare-earth: Precision 90% and Recall 94%
  - Bare-earth likelihood
    - >95% : 60% of total bare-earth points
    - >90% : 80% of total bare-earth points

Measurement of uncertainty
Customizable threshold

8 – Conclusion

- **Results**
  - Non-soil and Sea surface are **well modelled**
  - **Understandable and analyzable models**, not a black box
  - **Already scalable:** from 500 kpts to 500 Mpts
    - Predictions:
      - Subsample: 44.5 Mpts → 10 min
      - All dataset: 500 Mpts → less than 1h (12 tiles – Multithreading)
- **Improvements**
  - Combine 2 and 3 classes models
  - Train more complex models, to fit bare-earth points (⚠Overfitting)
  - Train with more zones, **to increase variety**
  - Add 10m geometric features

Thank you for your attention !



HYDRO 2025

28^{th} – 30^{th} October, Liverpool

Xavier PELLERIN LE BAS - SCIENTEAMA

xavier.pellerin@scienteama.fr

RÉPUBLIQUE FRANÇAISE
Liberté
Égalité
Fraternité

SHOM
L'océan en référence

SCIENTEAMA
EXPLORE VOS PRODUITS DE DEMAIN

cnrs

M2C