
3D Spectral Classification of UAV-SfM Point Clouds: Mapping Sediment in Alpine Rivers



© 2019 Bob de Graffenried

Student:

Fulvia Malvido Montandon

Supervisors CREALP:

Joshua Theule, Ferran Galán

Supervisors EPFL:

Prof. Christophe Ancey, Bob de Graffenried

September 2026

Contents

1	Introduction	3
2	State-of-the-art / Literature review	5
2.1	High-resolution topographic surveys for river monitoring	5
2.2	ML-based classification of 3D point clouds	5
2.3	Quantifying morphological changes: erosion and deposition	7
3	Methodology	9
3.1	Study area and data	9
3.1.1	Data Acquisition and Processing	10
3.1.2	Class Selection and Geomorphological Units	10
3.2	Building Blocks for the Classification Framework	11
3.2.1	Multi-Scale Geometric and Spectral Descriptors	11
3.2.2	Machine Learning Classifiers	13
3.2.3	Evaluation Metrics	13
3.3	The Proposed Classification Framework	14
3.3.1	Dataset Labeling and Training Split	17
3.3.2	Hyperparameter Tuning	17
3.3.3	Feature Selection and Optimization	17
3.3.4	SHAP for Feature Interpretation	18
3.4	Application of the Developed Pipeline for Multi-Temporal Change Detection	18
3.4.1	Cloud Co-registration and M3C2 Distance Computation	19
3.4.2	Sediment Budgets and Spatial Analysis	19
4	Results	20
4.1	Preliminary results	20
4.2	Model evaluation on the Navisence datasets	21
4.2.1	Feature and Scale Selection	23
4.2.2	Model Interpretability via SHAP Analysis	24
4.3	Classification results	28
4.4	Multi-temporal Analysis: Summer 2024	29
4.4.1	LOD calculation and M3C2 results	30
4.4.2	Geomorphic change analysis	31
5	Discussion	33
5.1	Rolling Protocol and Classification Performance	33
5.2	Feature Importance and Model Interpretability	34
5.3	Challenges and Limitations of RGB-SfM Classification	35
5.4	Framework Contributions and Limitations	36
5.5	Combining Classification and 3D Change Detection for Geomorphic Monitoring	36
6	Future perspectives	37
7	Conclusion	38
8	AI statement	38

A	Appendix	43
A.1	Reproducibility tests -Installation check	43
A.2	Intermediate results: scales and Large Woody Debris	43
A.3	Hyperparameters for each model	44
A.4	RGB Point Clouds ROI	45

Acknowledgments

I would like to thank Joshua, who helped me a lot throughout the project. He introduced me to his world of remote sensing and took me on a flight over the mountains. He shared valuable knowledge and advice with me, talking about both work and life.

I would also like to thank Ferran, who gave me relevant feedback at just the right times, directing and redirecting my work in the best ways. More generally, I would like to thank the whole CREALP team, for all the small interactions and big discussions.

I would like to thank Christophe, who allowed me to do this thesis, which I felt was the right one for me from the very beginning. I also want to thank Bob and Axel, for the time spent out in the field, and for one very good raclette evening.

I would like to thank my family, who supported me throughout my studies, which are now coming to an end with this master's thesis. A special thought for my sister, who is my greatest supporter and who I can trust with all my problems, big and small.

Finally, I would like to thank Côme, all my friends, and my flatmates, who made my studies a special chapter of my life. A special thanks to Point Vélo, without which I probably wouldn't be finishing my master's.

Abstract

Mountain rivers are dynamic environments where sediment transport and morphological evolution are strongly driven by hydrological regimes. Quantifying these changes requires high-resolution topographic monitoring, yet traditional field methods and LiDAR surveys are often limited by cost and spatial or temporal coverage. UAVs combined with Structure-from-Motion (SfM) photogrammetry offer an accessible alternative, but automating the classification of complex 3D point clouds remains challenging when only RGB data are available.

This thesis adapts the 3DMASC tool to classify UAV-SfM point clouds using RGB and topographic information. Multi-scale geometric descriptors and illumination-invariant spectral indices were integrated into a machine learning workflow. To assess generalizability, a rolling protocol with cumulative training across repeated surveys was evaluated on five datasets (2024–2026) from the Navisence River (Valais, Switzerland). XGBoost achieved the highest performance, reaching an Overall Accuracy of 85.5% and a macro F1 score of 85.2% on the final test dataset. SHAP analysis showed that small (0.5–1.0 m) and large (6.5 m) scales were most important, with spectral features mainly contributing to vegetation and water classification, while geometric features helped characterize 3D sediment structures.

Combining the classified point clouds with M3C2 further enabled the analysis of geomorphic change over a 270 m reach following the 2024 floods, revealing a net deposition of 3,803 m³. Fine sediments and big cobbles and boulders were mainly associated with erosion, while small cobbles showed more deposition. Overall, the results show that RGB UAV-SfM point clouds combined with interpretable multi-scale classification can provide an accessible framework for multi-temporal geomorphic monitoring of Alpine rivers.

Résumé

Les rivières de montagne sont des environnements dynamiques où le transport des sédiments et l'évolution morphologique sont fortement contrôlés par les régimes hydrologiques. Quantifier ces changements nécessite un suivi topographique à haute résolution, mais les méthodes de terrain traditionnelles et les relevés LiDAR sont souvent limités par leur coût ainsi que par leur couverture spatiale ou temporelle. Les UAV combinés à la photogrammétrie Structure-from-Motion (SfM) offrent une alternative plus accessible, mais l'automatisation de la classification de nuages de points reste difficile lorsque seules des données RGB sont disponibles.

Cette thèse adapte l'outil 3DMASC pour classer des nuages de points UAV-SfM à partir d'informations RGB et topographiques. Des descripteurs géométriques multi-échelles et des indices spectraux invariants à l'illumination ont été intégrés dans un workflow de machine learning. Afin d'évaluer la généralisabilité, un protocole roulant avec entraînement cumulatif sur des relevés répétés a été appliqué à cinq jeux de données (2024–2026) de la Navisence (Valais, Suisse). Parmi les classificateurs testés, XGBoost a obtenu les meilleures performances, avec une Overall Accuracy de 85.5% et un macro F1 de 85.2% sur le jeu de test final. L'analyse SHAP a montré que les petites échelles (0,5–1,0 m) et la grande échelle (6,5 m) étaient les plus importantes, les caractéristiques spectrales contribuant principalement à la classification de la végétation et de l'eau, tandis que les caractéristiques géométriques permettaient de caractériser les structures 3D des sédiments.

La combinaison des nuages de points classifiés avec M3C2 a également permis d'analyser les changements géomorphologiques sur un tronçon de 270 m après les crues de juin 2024, révélant une déposition nette de 3,803m³. Les sédiments fins ainsi que les gros galets et blocs étaient principalement associés à l'érosion, tandis que les petits galets présentaient davantage de déposition. Dans l'ensemble, ces résultats montrent que les nuages de points RGB issus d'UAV-SfM, combinés à une classification multi-échelle interprétable, peuvent constituer une approche accessible pour le suivi géomorphologique multi-temporel des rivières alpines.

Key-words: river geomorphology, UAV-SfM photogrammetry, RGB point clouds, river monitoring, machine learning, multi-scale classification, change detection

1 Introduction

Mountain rivers are among the most dynamic landscapes on Earth: discharge, sediment transport, and channel morphology are all strongly controlled by snowmelt, glacier melt, and flood events [1, 2]. These events vary significantly depending on the region, climate, and catchment characteristics, making the geomorphic evolution of these rivers hard to predict [3]. However, sediment transport in mountain rivers is difficult to quantify, particularly in proglacial streams, which are generally characterized by high flow velocities, high sediment concentrations, and fast changes [4, 5].

Literature has shown that even within a single reach, the geomorphic response to a flood is very scale-dependent: net morphological change can be up to an order of magnitude smaller than the actual erosion and deposition volumes, as the local gains and losses partly cancel out at coarser scales [6, 7]. Survey frequency matters too, going from weekly to daily surveys can double the amount of morphological change detected over the same period, which shows that even small flow pulses can cause changes that low-frequency surveys could miss. At the process level, bank erosion can contribute as much to sediment storage change as the more classic braiding mechanisms like bar development or meander cutoffs [8]. This complexity is amplified in glacierized Alpine catchments, where hydropower regulation adds another layer on top of natural sediment supply: in the Borgne d’Arolla (Valais), for instance, water abstraction at upstream intakes combined with occasional sediment-laden flushing events has caused aggradation downstream, with sediment loads further increased by faster glacier melt since the 1980s, leaving the channel highly unstable and largely free of vegetation [9]. All of these dynamics greatly affect flood risk, infrastructure safety, and river ecosystems, which is why accessible monitoring tools for Alpine rivers are needed [10].

Traditionally, monitoring these changes has relied on ground-based methods like Wolman pebble counts and terrestrial laser scanning (TLS), or airborne LiDAR [11, 12]. These methods have several disadvantages: cost of the instruments, time in the field, limited spatial range, or inconvenience for remote or fast-changing mountain rivers. Over the last decade, Unmanned Aerial Vehicles (UAVs), combined with Structure-from-Motion (SfM) photogrammetry, have become a much more accessible alternative, allowing the generation of high-resolution 3D point clouds of river reaches [13, 14]. Nowadays, a growing number of tools and softwares make point cloud processing fast and easy. Moreover, the development of image-based techniques applied to drone videos, such as Large-Scale Particle Image Velocimetry (LSPIV), have made it possible to estimate surface flow velocities without any instruments in the water [15, 16]. Together, this means channel morphology and flow could both be characterized just from drone footage.

In parallel to these technological advances, computational capacities and machine learning algorithms have also developed rapidly, as well as their application to environmental remote sensing data. For example, semantic segmentation has been successfully applied to high-resolution natural scene images [17]. Most existing classification approaches deal with the heavy point clouds by rasterizing them into 2D grids first, then applying classifiers like maximum likelihood, SVM, or decision trees [18]. But using spectral data is challenging in fluvial environments, for instance water surfaces scatter light, weather and lighting conditions can affect the colors, vegetation often overhangs the water and makes the land-water boundary hard to detect [19, 20]. For example, riparian vegetation and water can present similar colors, making them very difficult to distinguish on a purely spectral basis in the visible RGB range. Shadows casted by vegetation and boulders add another complication by altering colors and blurring class boundaries. Moreover, rasterizing point clouds smooths out fine-scale geometry and loses the spatial point pattern information that is particularly informative in complex natural terrain [21].

To solve this, recent frameworks classify directly in 3D [22]. Building on earlier tools that use spatial data for segmentation, like CANUPO [23], the 3DMASC tool combines spectral and geometric features at multiple scales with an interpretable classifier, and has shown strong performances for this task [24]. However, 3DMASC was originally built for bi-spectral topo-bathymetric LiDAR data (green and near-infrared wavelengths), which is expensive and operationally demanding to acquire compared to UAV-SfM photogrammetry.

This leaves two problems for using this workflow more broadly in river monitoring. First, 3DMASC has mostly been tested on bi-spectral LiDAR rather than the simple RGB and topographic data produced by UAV-SfM surveys. Second, many machine learning approaches for point cloud classification, especially deep learning, work as black boxes: it's hard to know which features actually drive the classification, which is a problem when results are meant to inform management decisions or scientific interpretation [25, 26].

To address this gap, this thesis investigates whether the geomorphological evolution of Alpine rivers can be accurately monitored and quantified using drone imagery. Specifically, I aim to develop a point cloud classification framework based on the 3DMASC tool, adapted to UAV-SfM data containing only RGB and topographic information. The development of this framework involves two main steps. First, geometric and spectral features are adapted to RGB-only data [27]. Second, different algorithms, including Random Forest, XGBoost, and LightGBM, are tested to find the most effective classifier [26]. Finally, the model resulting from the pipeline is applied to several datasets, in parallel with the M3C2 algorithm, to quantify multi-temporal geomorphological changes, coupling erosion and deposition volumes to sediment classification [28, 29].

While designed as a generalizable pipeline for mountain rivers, it is developed and tested on the Navisence river (Valais, Switzerland), a highly dynamic site that has been monitored for several years now by the Environmental Hydraulics Laboratory (LHE-EPFL)[30, 31, 32] in collaboration with CREALP [33, 34].

The rest of this thesis is structured as follows. First, the state-of-the-art reviews the relevant literature on high-resolution topographic monitoring of rivers, point cloud classification and geomorphological change quantification. The methodology then presents the data, details the building blocks needed for the development of the framework, and proposes an example of an analytical method. Lastly, the results and discussion present how well the different algorithms performed, show the sediment budgets estimated for the Navisence river, and discuss the practical limitations encountered along the way.

2 State-of-the-art / Literature review

2.1 High-resolution topographic surveys for river monitoring

Quantifying geomorphic change in rivers requires topographic data that is detailed enough to distinguish different sediment types. At the same time, surveys need to be easily repeatable over time to capture even the morphological changes produced by seasonal floods. Several methods exist for this purpose, each with strengths and limitations depending on the type of surface, the spatial extent, and the required temporal frequency.

Terrestrial laser scanning and airborne LiDAR. TLS was one of the first methods to provide the point density and precision needed for reliable change detection in river channels. It can detect small-scale changes such as local sediment deposition, bank erosion, and channel incision, with vertical precision as low as ± 0.02 m on dry surfaces [11]. However, it is less precise in submerged areas, and it is labor-intensive, limited in spatial coverage by the scanner range, and expensive to operate at reach scale. Airborne LiDAR solves these coverage issues and can also penetrate the vegetation canopy to generate bare-earth terrain models of the riverbed and floodplain [12]. Moreover, topo-bathymetric LiDAR systems using green wavelengths can map submerged topography in shallow, clear water [19]. However, studies in a braided river showed that LiDAR-derived sediment budgets are very sensitive to processing choices, with cumulative errors potentially masking small-scale changes [6]. And like TLS, airborne LiDAR remains expensive and difficult to sufficiently repeat for event-scale monitoring in Alpine rivers.

UAV-SfM photogrammetry. The development of affordable drones combined with Structure-from-Motion (SfM) photogrammetry has made high-resolution topographic surveys much more accessible [13, 14]. UAV-SfM can produce Digital Elevation Models at resolutions around 0.02 m over river reaches of 50–100 m, and with refraction correction, can even map shallow submerged areas down to about 0.7 m depth in clear water [35]. Unlike airborne LiDAR or TLS, UAV surveys can easily be repeated at shorter intervals, allowing the morphological impact of individual flood events to be captured [9], with results comparable to those obtained from TLS for quantifying erosion, deposition, and bank retreat [7, 8]. However, SfM accuracy depends on GCP configuration and flight geometry, and unlike LiDAR, it cannot penetrate vegetation canopy in riparian zones. In these areas, or where water is too deep or turbid for SfM, UAV LiDAR performs better. The two methods produce similar volumetric change estimates on exposed terrain (differences below 1% in some studies [14]), but LiDAR produces fewer artifacts over water and vegetation. For this reason, combining both methods has been proposed for braided Alpine rivers where exposed gravel bars, shallow channels, and riparian vegetation are all present in the same survey area [19, 10].

The UAV sensor and spectral range selection plays an important role in surface classification. Multispectral cameras with near-infrared (NIR) bands help distinguish water, vegetation, and bare sediment more effectively, as indices such as NDVI (Normalized Difference Vegetation Index) capture contrasts that are not visible in the RGB range. However, multispectral sensors are more expensive and less commonly available than standard RGB cameras. Comparisons between the two have shown that multispectral data outperforms RGB for multi-class classification, though RGB-only data can still be sufficient for simpler classification tasks [20, 27]. Standard RGB cameras are the most widely used option in UAV-SfM surveys, and while they work well on exposed terrain, they struggle with spectrally similar surfaces like wet gravel, shallow water, and shadowed sediment, which is one of the challenges this thesis aims to address.

2.2 ML-based classification of 3D point clouds

Classifying surface types in river point clouds is a complex task for which several machine learning approaches have been developed, ranging from raster-based methods to direct 3D classification frameworks.

Early classification methods involved rasterizing point clouds into 2D images or Digital Elevation Models (DEMs) and applying pixel-wise classifiers such as maximum likelihood, Support Vector Machines (SVM), or decision trees [18]. The problem with these approaches is that rasterization smooths out fine-scale geometry

and discards spatial point pattern information, producing mixed pixels at class boundaries and losing the 3D structure of objects such as vegetation canopies, boulder fields, or undercut banks [21]. A review of point cloud segmentation algorithms presents the diversity of the existing approaches (edge-based, region-growing, model-fitting, attribute-based) but also shows that most were designed for structured environments like urban scenes and do not transfer well to heterogeneous and complex natural landscapes [22]. Classifying directly on the raw 3D point cloud preserves the full spatial information, but requires features that are explicitly defined in 3D.

Deep learning methods, such as PointNet, Point-CNN, and graph-based neural networks [25], learn spatial representations directly from raw point coordinates and attributes without manual feature engineering [25, 36]. While these architectures have achieved high accuracy on urban or indoor benchmark datasets [22], their application to fluvial remote sensing remains limited. A primary obstacle is data availability: deep learning models require large volumes of annotated data to train effectively, but large labeled collections of river point clouds are rare and time-consuming to produce. Furthermore, models trained on a specific channel reach frequently generalize poorly when transferred to other fluvial environments with different geomorphic characteristics. Furthermore, deep learning models act as "black boxes" that hide the physical or geometric features driving their decisions. This interpretability gap is problematic in fluvial geomorphology, where the underlying physical mechanisms are important to ensure reliable scientific and engineering conclusions.[26].

Geometric features and neighborhood selection. Feature-engineered machine learning pipelines compute explicit geometric and spectral descriptors from the raw point cloud, which are then used to train a classifier. Unlike deep learning, this approach produces models where each decision can be linked to physically meaningful features. For each point, a neighborhood is defined as the set of surrounding points within a given search radius, and features are computed from this local subset. The choice of radius controls what scale of information is captured: small radii encode fine surface texture, while large radii capture broader structural patterns. Eigenvalue-based dimensionality features such as linearity, planarity, and sphericity have been shown to effectively characterize local 3D structure [21]. Since objects in river scenes range from small cobbles to large trees, computing features at multiple scales is more effective than using a single fixed scale [23]. Moreover, keeping only the most informative features has been proven to reduce computational cost without significantly affecting accuracy [21].

CANUPO was one of the first frameworks to apply multi-scale dimensionality analysis directly on raw 3D point clouds of complex natural scenes [23]. It works by computing a multi-scale feature set for each point and then using these signatures to train a classifier. In fluvial and cliff environments, CANUPO achieved classification accuracies above 98% for separating riparian vegetation from ground and classifying mountain stream surfaces into vegetation, rock, gravel, and water. Its main limitation is that it only uses geometric information and omits spectral data, which makes it hard to distinguish classes that are geometrically similar but differ in color, such as wet gravel and shallow water.

3DMASC was developed as an extension of CANUPO, adding spectral features and a more complete feature engineering pipeline within the CloudCompare software [24]. It also classifies directly on 3D point clouds and computes features at multiple spherical neighborhood scales. These features include dimensionality descriptors (linearity, planarity, sphericity), geometry descriptors (slope, roughness, curvature, anisotropy, point count, first-order moment), and height descriptors, which can also be expressed as statistical summaries (mean, standard deviation, range, skewness) within each neighborhood. This combination of geometric and spectral information at multiple scales allows 3DMASC to distinguish surface types more effectively than single-scale or geometry-only approaches.

The 3DMASC framework was originally developed for bi-spectral topo-bathymetric LiDAR (green and near-infrared), where the two clouds are compared to extract additional spectral and geometric information (see Figure 1). In coastal and fluvial environments, it reached classification accuracies of at least 94% using models trained on 2000 samples per class. A feature and scale selection step reduces redundancy and helps identify the most informative descriptors; results show that neighborhood diameters below 7 m are generally sufficient.

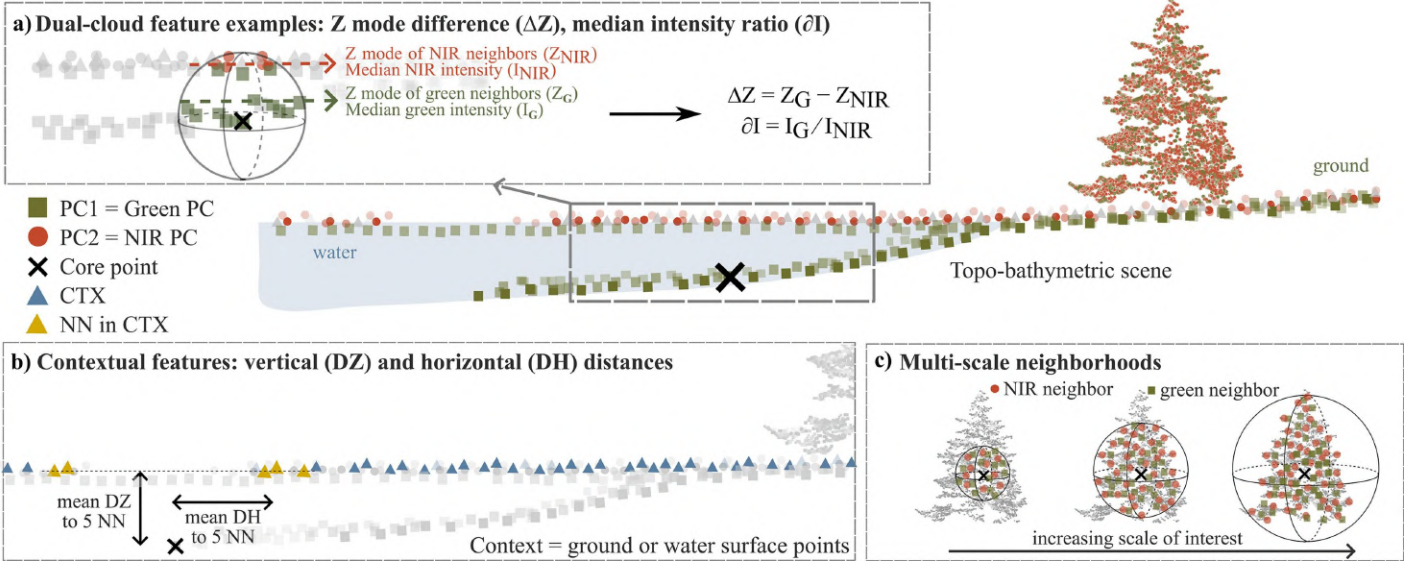


Figure 1: Illustration of the main characteristics of 3DMASC : a) examples of new dual-cloud features providing a better description of the differences between clouds, b) the generalized contextual attributes placing each point in its spatial setting and c) the multi-scale spherical neighborhoods used to describe the many aspects of 3D objects. Figure from the 3DMASC paper [24]

For classification, 3DMASC uses a Random Forest (RF) implemented via OpenCV, in a supervised learning approach. RF builds an ensemble of decision trees, each trained on a bootstrapped data subset with a random feature subset at each split, and combines predictions by majority vote. This reduces overfitting, handles high-dimensional feature spaces, and produces feature importance scores that help interpret what drives the classification [26]. The proportion of trees voting for the assigned class also serves as a per-point confidence indicator. Unlike deep learning approaches, 3DMASC is designed to be interpretable: in addition to feature importance scores, SHAP values can explain how each feature influences individual predictions.

Once point clouds are classified into geomorphological units, the next step is to measure how the river changes over time. The following section reviews the methods used to quantify these changes.

2.3 Quantifying morphological changes: erosion and deposition

Once repeat surveys are available, the next step is to quantify how the surface has changed between dates. Several methods have been developed for this, each with different trade-offs between accuracy, complexity, and data requirements.

One of the most widely used approaches for volumetric change estimation is the **DEM of Difference (DoD)**, which subtracts one DEM from another to produce a map of erosion and deposition [7]. DoD is used to compute sediment budgets and morphological activity across a wide range of river types and scales [8]. The main limitation is that the DoD collapses 3D point clouds onto a horizontal grid, losing vertical and overhanging geometry, and can misrepresent change on rough or undercut bank faces. On steep terrain, even small lateral misalignments between surveys can produce large apparent vertical differences that appear as real change [6]. Furthermore, DoD maps confuse survey noise from real change, unless an explicit uncertainty analysis is applied [7]. One way to avoid these errors is to remove values below the minimum Level of Detection (LoD). However, measurement errors vary spatially with slope, roughness, point density, and the presence of water or vegetation. Using spatially variable uncertainty estimates produces more physically meaningful results, keeping more valid observations while removing false positives [6]. Moreover, error does not come from a single source: filtering, interpolation, and differencing each contribute to the total uncertainty, as shown by a step-by-step error assess-

ment in a braided river context. Tracking and correcting these errors at every stage is therefore important for reliable sediment budget estimates.

Cloud-to-cloud (C2C) and cloud-to-mesh methods (C2M). To avoid the information loss from DEM gridding, original point clouds can be compared directly in 3D space without rasterization. C2C and C2M methods were early implementations of this approach in the CloudCompare software, but both have biases: C2C uses the nearest-point distance regardless of surface orientation, while C2M requires one of the clouds to be meshed, which is problematic for rough natural surfaces [6]. Multi-temporal TLS monitoring of boulder movement in a mountain riverbed confirmed that point cloud differencing outperforms DEM-based methods [4].

The **M3C2 (Multiscale Model-to-Model Cloud Comparison)** algorithm, also implemented in CloudCompare, compares two point clouds directly along locally estimated surface normals, avoiding the need for DEM gridding [37]. As illustrated in Figure 2, it measures distances along a cylinder projected in the normal direction and computes a local confidence interval, producing a spatially distributed LoD that identifies statistically significant change. Applied to TLS monitoring of a rapidly eroding bedrock canyon, it achieved LoDs down to 6 mm at 95% confidence, outperforming DoD on complex surfaces. M3C2-EP, an extended version, propagates per-point measurement uncertainty into the statistical test rather than assuming a uniform registration error [28]: over a three-week rock glacier monitoring period, it detected significant change at 12.5% fewer locations while recovering 25.2% more change volume. A comparison of DoD, C2C, C2M, and M3C2 for TLS-derived channel change found M3C2 to be the most accurate, with propagated errors of 0.016 m vs. 0.018 m for DEM-based methods [29]. For repeat UAV-SfM surveys, combining M3C2 with spatially variable LoD and consistent GCP configuration represents current best practice.

Building on these established methods for UAV-SfM acquisition, 3D classification, and cloud-to-cloud differencing, the next chapter describes how they are implemented in the workflow developed for the Navisence river.

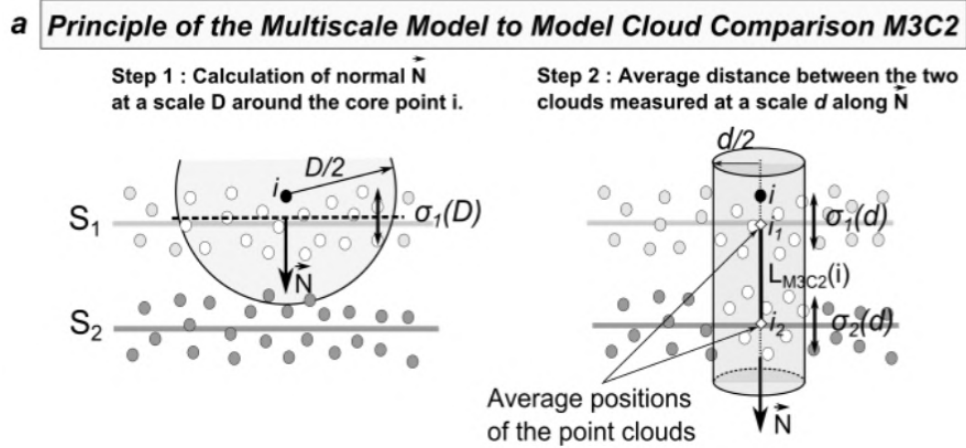


Figure 2: Description of the M3C2 algorithm and the two user-defined parameters D (normal scale) and d (projection scale). Step 1 : the normal is estimated from cloud 1. Step 2 : 2 sub-clouds are defined by the intersection of the reference and compared clouds with a cylinder of diameter d and axis (i, N) . Each sub-cloud is projected on the cylinder axis which gives a distribution of distances along the normal direction. These are used to define the mean position of each cloud i_1 and i_2 . The local point cloud roughness $\sigma_1(d)$ and $\sigma_2(d)$ and size n_1 and n_2 of the 2 sub-clouds are subsequently used to estimate a local confidence interval. Figure from the M3C2 paper [37]

3 Methodology

3.1 Study area and data

This study focuses on the upper part of the Navisence river, located in Val d’Anniviers, Valais, Switzerland (see Figure 3). The Navisence originates at an elevation of 2,300 m and meets the Rhône at an elevation of approximately 525 m asl, with a watershed area of 257 km². This 23km-long river flows down through a landscape that alternates between morainic alluvial deposits and rocky gorges. The upper river is primarily fed by the Zinal Glacier and partly by the Moming Glacier. The study site is located upstream of the village of Zinal, where the Navisence flows through the Plats de la Lé alluvial plain at approximately 1,700 m elevation. At this location, the riverbed slope drops to around 3%, much lower than the average 10% upstream. This reduction in flow velocity causes a decrease in transport capacity, leading to sediment deposition and the formation of a partially braided gravel-bed channel.

This region is characterized by an alpine climate, with cold winters and moderate summers. The catchment has an alpine nivo-glacial discharge regime, characterized by strong seasonality: peak flows occur in mid-to-late summer driven by snowmelt and glacier melt, and severe low flows occur during the freezing conditions of winter months. Summer storms can produce large flood events capable of significant morphological change and damage to downstream infrastructure.

The Navisence has been used for several years as an open-air laboratory by the Environmental Hydraulics Laboratory (LHE-EPFL) for experiments on mountain stream processes, including sediment photosieving, automated grain size detection, and sediment transport monitoring [30, 31, 32]. CREALP has additionally instrumented the site with geophones in the riverbed to study bedload transport [33, 34]. A measurement station located at the bridge just downstream of the study area provides hourly discharge data available through the cantonal monitoring platform Guardaval.

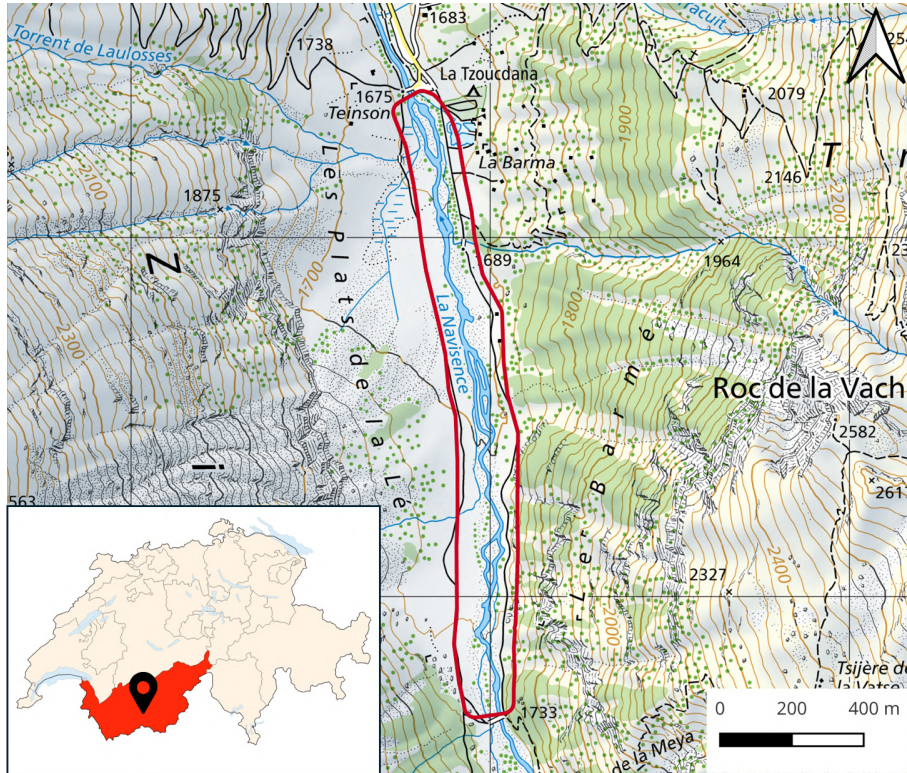


Figure 3: Study area of Plats de la Lé, datasets outline is indicated by the red shape, delineating the studied reach (1.7km long). At the bottom left, the Switzerland map with canton Valais highlighted in red. @swisstopo

Additional Navisence point clouds from a reach upstream of Plats de la Lé, close to the Zinal glacier, were used for preliminary model and feature-set comparisons. These datasets were not part of the performance analysis of the rolling protocol.

3.1.1 Data Acquisition and Processing

Data was collected during multiple summer fieldwork campaigns in Plats de la Lé between 2024 and 2026 using the DJI Mavic 3E and Mavic 3M UAVs by the LHE group. I participated in the field campaigns of June and July 2026 by taking drone footage. The automated flight missions were managed via Map Pilot Pro using terrain-following at an altitude between 40 and 50 m Above Ground Level (AGL).

Although the Mavic 3M can collect multispectral data, we only acquired the RGB imagery for the 3D reconstruction due to its higher resolution. The absolute spatial georeferencing was achieved using Ground Control Points (GCPs) measured via a Leica Zeno 20 collector paired with a Leica Zeno GG04 plus smart antenna. The LHE group processed the raw data with the Pix4Dmapper software to obtain the dense point clouds that I have used throughout the project. A summary of the acquisition and processing parameters is provided in Table 1.

Table 1: UAV hardware specifications and photogrammetric processing parameters

Parameter	Specification
UAV Platforms	DJI Mavic 3E and Mavic 3M
RGB Sensor	20 MP, 4/3 CMOS
Shutter Type	Mechanical
Focal Length	12.2 mm (24 mm equivalent)
Flight Altitude	40-50 m AGL
Spectral Bands	RGB
UAV Positioning	Onboard GNSS
GCP Surveying	Leica Zeno 20 + GG04 plus antenna
SfM Software	Pix4Dmapper
Analysis Software	CloudCompare
Point Cloud Density	$\approx 750 \text{ points/m}^2$

3.1.2 Class Selection and Geomorphological Units

I selected the classes inspired by the 3DMASC framework, with grain size categories adapted to the Wentworth scale as it links classes to flow regimes and sediment transport thresholds via the Hjulström curve. Given the spatial resolution of the data and the extent of the study area, the sediment classes were up-scaled relative to the original Wentworth categories. I defined class boundaries based on visually distinguishable differences in the RGB point cloud (see Figure 4). The eight classes are:

1. **Fine sediments:** silt and sand. Mostly uniform and homogeneous surface texture.
2. **Coarse sediments:** Small gravels that generate some rugosity. Sizes are smaller than 10 cm.
3. **Small cobbles:** we can distinguish the individual elements, often grouped. Sizes from 10 cm to 30 cm.
4. **Big cobbles and boulders:** bigger rocks, grouped and individual. Sizes larger than 30cm.
5. **Large woody debris:** fallen trees and logs that are present in the stream active area .
6. **Water.** The main river network as well as shallow lateral channels.
7. **Low vegetation:** grass and sparse bushes.

8. **High vegetation:** trees of all sizes, often creating some sort of canopy.

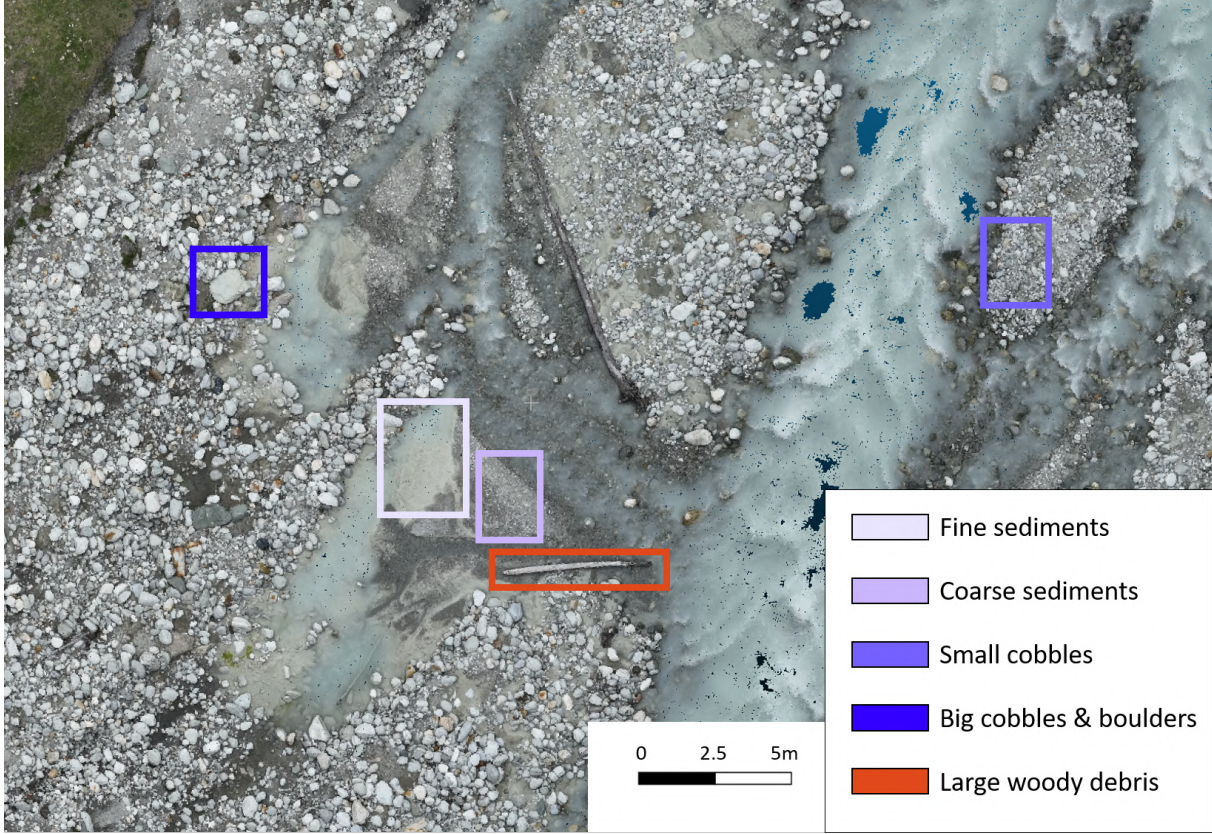


Figure 4: Image illustrating the four sediment classes and the large woody debris class. The image is taken from the RGB point cloud of 2 June 2026 point cloud.

3.2 Building Blocks for the Classification Framework

Before adapting the 3DMASC tool to the Classification Framework, I first test it using the data provided by the developers of 3DMASC. The code is available in their GitHub repository, while the dataset for the Ain river can be found in their webpage. I then validate it by comparing my findings to their published results using the evaluation metrics described in Section 3.2.3. This reproducibility step ensures that the tool is correctly installed and configured (results in Table 8 in the Appendix).

In the following sections, I first describe the geometric and spectral features used to characterize the point clouds. Then, I explore the different machine learning classifiers that can be used to train the model. Finally, I present the evaluation metrics that allow to evaluate the performance of the classification. These building blocks establish the basis that allow to construct the Classification Framework.

3.2.1 Multi-Scale Geometric and Spectral Descriptors

3DMASC characterizes the local environment around each point by computing features within a spherical neighborhood of a given radius. The size of this neighborhood, the scale, controls what level of detail is captured: small scales encode fine surface texture, while large scales capture broader structural patterns. Features are computed at multiple scales simultaneously to capture information across the full range of object sizes present in the scene. In practice, scales are defined in a parameter file provided as input before the feature computation step. I initially use scales ranging from 0.5 m to 9.5 m in steps of 0.5 m, chosen to cover the range from individual cobbles to high vegetation, and later reduce this range based on feature selection results (see subsection 4.2).

This section describes the feature set characterizing the Navisence RGB datasets.

I use the geometric features proposed in the 3DMASC workflow [38, 39]. Height-based features characterize the vertical structure (in meters) of the local neighborhood: Z_{min} ($z - Z_{min}$), Z_{max} ($Z_{max} - z$) and Z_{range} ($Z_{max} - Z_{min}$), where Z_{min} and Z_{max} are the minimum and maximum heights in the neighbourhood, and z corresponds to the elevation of the point being evaluated. Dimensionality and geometry-based features describe the local shape and aspect of the point cloud: dip angle, roughness, PCA components constructed with the structure tensor eigenvalues, anisotropy, sphericity, linearity, planarity, and curvature [40, 41, 42]. Additionally, I add a residual topography feature, computed as the root mean square height [43, 44, 45], providing a slope-invariant roughness descriptor that is more robust across reaches with varying gradients. After some preliminary tests, I remove features with irrelevant or misleading information from the feature set, such as point density, which depends on survey quality rather than surface properties.

To adapt the framework to RGB data, I compute several spectral indices and add them to the dataset as additional scalar fields (attributes in CloudCompare). Since near-infrared information is not available, standard indices such as the Normalized Difference Vegetation Index (NDVI) cannot be computed, and instead, I use RGB-based alternatives. The R, G and B values can be obtained directly from the spectral data of the point clouds (16-bit digital numbers DN).

The **Excess Green index (ExG)** amplifies the green channel relative to red and blue, and is widely used to separate vegetation from bare ground [46]:

$$ExG = 2 \cdot G - R - B \quad [16\text{-bit DN}] \quad (1)$$

where R, G, and B represent the red, green, and blue channel values.

The **Normalized Difference Index** between the red and green wavelengths (NDI_RG) provides a vegetation proxy analogous to NDVI but using only visible bands:

$$NDI_RG = \frac{G - R}{G + R + \epsilon} \quad [-] \quad (2)$$

with R and G as defined above, and ϵ denoting a very small positive quantity used to avoid division by 0.

To reduce the influence of varying lighting conditions and camera settings across acquisition dates, I replace the raw RGB values by **chromaticity** ratios, which normalize out overall brightness:

$$R_{norm} = \frac{R}{R + G + B}, \quad G_{norm} = \frac{G}{R + G + B}, \quad B_{norm} = \frac{B}{R + G + B} \quad [-] \quad (3)$$

Additionally, I derive three features in the HSV (Hue, Saturation, Value) color space, which separates chromatic information from luminance and provides more illumination-invariant descriptors than raw RGB [46]. Hue (H) captures color information that is relatively insensitive to lighting changes, and Saturation (S) distinguishes vegetation from spectrally similar desaturated backgrounds. I excluded the Value (V) feature from the set as it directly reflects brightness and is therefore sensitive to illumination differences between surveys.

Finally, I include color variance as a feature to detect sharp color contrasts and edges between surface types. It is expressed as the coefficient of variation of R, G, B values within the local neighborhood, which is more stable than absolute variance across varying exposure conditions:

$$ColorVariance = \frac{\text{StandardDeviation}(R, G, B)}{\text{Mean}(R, G, B) + \epsilon} \quad [-] \quad (4)$$

3.2.2 Machine Learning Classifiers

The baseline classifier used in 3DMASC is a **Random Forest (RF)**, implemented via the OpenCV library. RF builds an ensemble of decision trees, each trained on a bootstrapped subset of the data with a random feature subset at each split, and combines predictions by majority vote. The Scikit-learn RF implementation is also proposed in the 3DMASC framework as an alternative to the OpenCV version.

I evaluate two **Gradient Boosting** alternatives : XGBoost and LightGBM. Unlike RF, gradient boosting builds trees sequentially, with each new tree correcting the residual errors of the ensemble, and minimizes a loss function using gradient descent. This sequential approach allows gradient boosting methods to learn complex relationships in the data but makes it more prone to overfitting, which must be controlled through techniques such as regularization.

XGBoost has built-in L1/L2 regularization that prevents overfitting, and executes parallel computation, allowing it to handle big datasets [47]. **LightGBM** further improves training speed by using a histogram-based learning strategy that bins continuous features into discrete intervals, reducing the complexity of the split-finding step, and grows trees leaf-wise rather than level-wise, producing deeper and more accurate trees with fewer splits [48]. Both methods are well-suited to the high-dimensional feature matrices, and have shown great performance in comparison to RF in geospatial classification tasks [49].

Because classifier performance can vary depending on the chosen hyperparameters, machine learning workflows often include a hyperparameter tuning step. This process identifies the combination of hyperparameters that provides the best performance for the specific machine learning task.

3.2.3 Evaluation Metrics

Evaluation metrics are used to quantitatively assess how well the machine learning models perform at the classification task.

A **confusion matrix** compares predicted and ground-truth class labels for each class, providing a detailed view of which classes are correctly identified and which are most frequently confused with one another (Figure 5). The diagonal corresponds to correct predictions, while off-diagonal values indicate misclassifications.

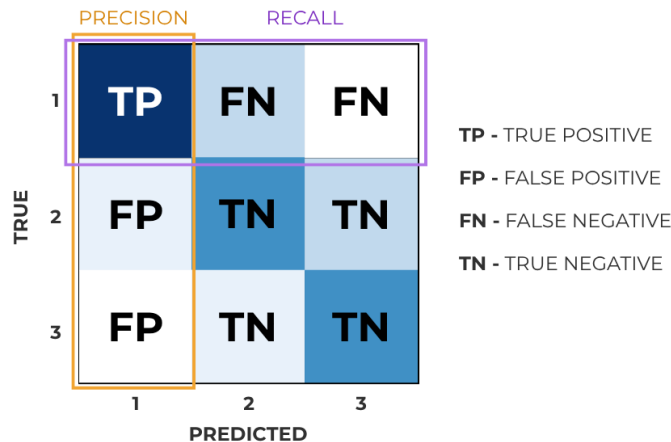


Figure 5: Illustration of a multi-class confusion matrix. For class 1, precision is computed along the predicted column (orange) and recall along the true row (purple).

Overall accuracy (OA) gives the proportion of correctly classified points across all classes:

$$OA = \frac{Trace(C)}{N} \quad [\%] \quad (5)$$

where the trace of the confusion matrix (C) is the sum of the diagonal elements and N is the total number of points. While informative, OA can be misleading for imbalanced class distributions and needs to be complemented by per-class metrics.

Precision measures how many of the points predicted as belonging to a given class are correctly classified (measuring exactness):

$$Precision = \frac{TP}{TP + FP} \quad [\%] \quad (6)$$

Recall (or sensitivity) measures how many of the actual points of a given class are correctly retrieved (measuring completeness):

$$Recall = \frac{TP}{TP + FN} \quad [\%] \quad (7)$$

The **F1 score** is the harmonic mean of precision and recall, providing a single balanced metric:

$$F1score = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall} \quad [\%] \quad (8)$$

A high F1 score means the model performs well on both metrics. The harmonic mean penalizes large imbalances between precision and recall, meaning a classifier must perform well on both metrics to achieve a high F1 score. For a model overall F1 score, the **macro F1 score** is calculated as the unweighted arithmetic average of the individual per-class F1 scores.

Finally, the RF classifiers output, for each point, the proportion of decision trees voting for the assigned class, which corresponds to a per-point prediction probability. The GB models include a prediction probability implementation. The average of the prediction probabilities gives a mean confidence, which allows to detect uncertain predictions and could be used to remove them from the final classification map.

3.3 The Proposed Classification Framework

This section presents the classification framework I have designed with the aim to expand the 3DMASC applicability to RGB-SfM point clouds. Moreover, I implement a rolling protocol that allows models to learn from several datasets. This aims to explore the capacity of the framework to classify data acquired on different dates or even other catchments. The proposed framework combines a labeling step in CloudCompare with RGB feature computation and machine learning training steps implemented in Python (see Figure 6). The machine learning process is performed in three steps: training, validation, and testing.

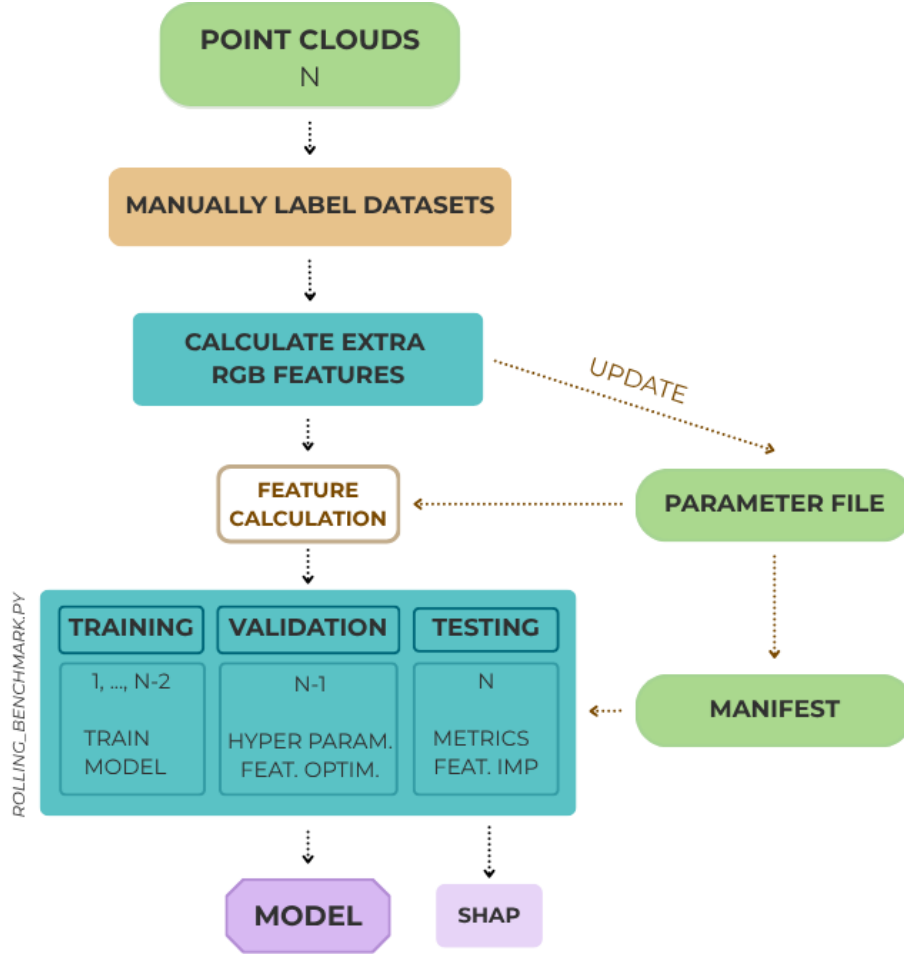


Figure 6: Architecture of the proposed classification framework adapted for N RGB-SfM datasets. Green represents the inputs, orange indicates processes executed within CloudCompare, and blue represents Python-based processing and machine learning steps. The final outcome of the framework is the classifier model, represented in purple. An additional step allows to perform the SHAP analysis.

The framework requires a minimum of three labeled datasets for a given area to achieve independent training, validation and testing steps (see Figure 7). If fewer than three labeled datasets are available, a single dataset can be spatially segmented into three independent areas that will fill each classification role. It is important that the three categories are spatially independent to avoid overly optimistic performance estimates.

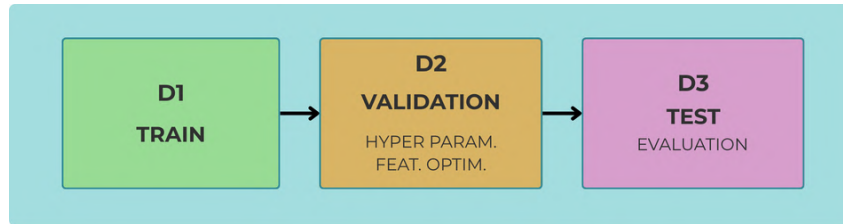


Figure 7: Simple scheme. The three datasets (D) are assigned to training, validation and testing roles.

For N labeled point clouds, datasets 1 through $N - 2$ are allocated to model training, dataset $N - 1$ is used for validation, and the last one is used for testing. When more than three labeled datasets are available, the framework implements the rolling protocol (Figure 8), based on cumulative training: the training set grows at each round by incorporating a new labeled dataset.

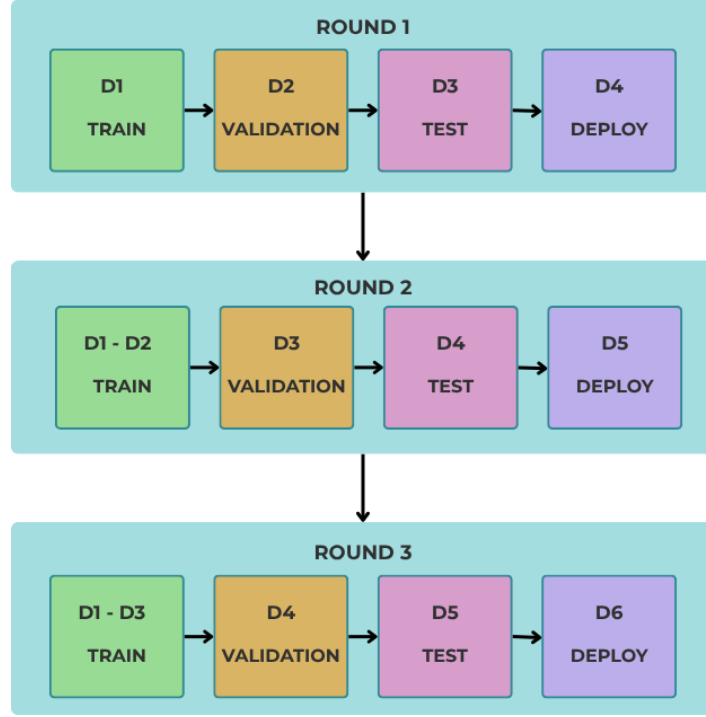


Figure 8: Rolling protocol. After testing on dataset N , the model can be deployed on a new, unseen dataset. When new labeled datasets become available, the window rolls forward one step and the procedure repeats, adding a round where an additional dataset is used for the training.

For each dataset, features are calculated through a function that calls CloudCompare, included in the main Python training script *rolling_benchmark.py*. To adapt to RGB-SfM data, RGB-based spectral indices are added to the feature set using an additional script *add_scalar_fields_invariant.py*. The configuration parameters are stored in a manifest file, which includes the class definitions and the paths to the original and labeled point clouds. Moreover, the parameter file contains the list of features and the clouds where they are computed. An example of a parameter file, as well as a detailed manual for the framework application, which includes the complete labeling protocol, are provided in the Appendix. The framework is called from the command line, which allows to select the model's inputs (see Table 2), for example:

```
python rolling_benchmark.py
--manifest "../manifests/manifest_Zinal.yaml"
--category plats --models xgboost,lightgbm
--params-key RGB_invariant
--prefer-cached --save-models
```

Table 2: Command-line arguments for `rolling_benchmark.py`

Argument	Type	Description
<code>-manifest</code>	string	Path to YAML manifest file containing dataset and configuration
<code>-category</code>	string	One category or comma-separated categories (default: <code>plats</code>) Example: <code>plats,glacier</code>
<code>-models</code>	string	Model selection (default: <code>xgboost</code>) Options: <code>xgboost</code> , <code>lightgbm</code> , <code>sklearn_rf</code> , <code>opencv_rtrees</code> Or: <code>all</code> (runs all four models)
<code>-params-key</code>	string	Parameter set key in manifest (default: <code>RGB_invariant</code>)
<code>-output-dir</code>	string	Output directory path Default: <code><paths.outputs>/classification</code> from manifest
<code>-save-models</code>	flag	Save final trained models from last round (XGBoost: <code>model.json</code> , LightGBM: <code>model.txt</code> , Random Forest: <code>model.joblib</code> , OpenCV: <code>model.yaml</code>)
<code>-prefer-cached</code>	flag	Reuse cached computed features instead of recomputing

3.3.1 Dataset Labeling and Training Split

The three machine learning steps are defined as follows. First, the training datasets are used to fit the model. Then, the validation dataset is used to improve the model through hyperparameter tuning and feature optimization. Finally, the test dataset contains unseen data on which model performance is assessed using the evaluation metrics. I performed manual annotation by visual interpretation of the RGB point clouds using the segmentation tool in CloudCompare. For each class, I used a subsample of the original dataset (≈ 2000 points) to limit computation time, with equal numbers per class to prevent the classifier from assigning disproportionate weight to more abundant classes.

3.3.2 Hyperparameter Tuning

The hyperparameter tuning step is performed using the validation dataset to identify the parameter combination that achieves the best performance, used for subsequent model training. A grid search is conducted over a predefined set of parameter combinations:

- **OpenCV RTrees:** Maximum tree depth, minimum sample count and number of trees.
- **Sklearn RF:** Number of estimators, maximum tree depth, minimum sample split thresholds and minimum samples per leaf.
- **XGBoost:** Learning rate (η), tree depth, subsample ratios, column subsample ratios, with fixed L_1 regularization ($\alpha = 0.0$) and L_2 regularization ($\lambda = 1.0$).
- **LightGBM:** Learning rate (η), tree depth, number of leaves, subsample ratios, column subsample ratios, number of estimators, with fixed L_2 regularization ($\lambda = 1.0$).

Table 10 in the Appendix contains the parameters for each model and all the possible values.

3.3.3 Feature Selection and Optimization

Not all computed features contain useful information for river classification. Furthermore, having redundant or irrelevant features can dilute the signal from more informative descriptors. Feature selection therefore serves two purposes: reducing computational cost and optimizing model performance by retaining only the relevant features. I decided to use Gini importance metrics generated by the Random Forest to rank features globally and per class, which allows to identify the most relevant descriptors for each surface type [42]. However, while Gini importance is useful for this initial dimensionality reduction, it only provides the magnitude of a feature’s importance, without indicating whether a feature positively or negatively correlates with a specific surface class.

3.3.4 SHAP for Feature Interpretation

To address the aforementioned limitation and improve the model's interpretability, I also compute SHAP (SHapley Additive exPlanations) values on the final predictions. Based on cooperative game theory, SHAP assigns each feature a value representing its contribution to each individual prediction. The main difference between the two approaches is that the Gini importance indicates which features matter globally, whereas SHAP describes how the features drive the predictions. This makes it possible to infer if the spectral and geometric descriptors used by the model align with the actual physical characteristics of each geomorphological class. For instance, it could demonstrate that a high Green value drives a prediction towards "vegetation".

3.4 Application of the Developed Pipeline for Multi-Temporal Change Detection

In this section, I propose an example of application of the Classification Framework for geomorphic analysis, presented in Figure 9. Once trained and validated, the classifier model is applied to point clouds from two separate acquisition dates within a defined Region of Interest (ROI). In parallel, the M3C2 algorithm [37] is used to compute the distance between the two surveys, yielding spatial patterns of erosion and deposition. Finally, the resulting datasets, comprising both classifications and the M3C2 distances, are imported into the QGIS software for spatial analysis and sediment budget quantification.

As a case study, I select a 270 m ROI from the June 11th and July 5th 2024 datasets to perform the analytical protocol. I select these specific dates because they capture the morphological impact of two major flood events, with discharges $Q = 74 \text{ m}^3/\text{s}$ and $Q = 63 \text{ m}^3/\text{s}$, roughly double a typical seasonal flood. These datasets were also part of the training data used to build the classifier, which introduces data leakage: the model has already seen these samples during training, and classification accuracy for these specific dates therefore may be overestimated. However, the primary objective of this section is to illustrate the integration of the classified outputs with M3C2 distance calculations to extract sediment budgets, rather than to benchmark predictive performance on unseen data.

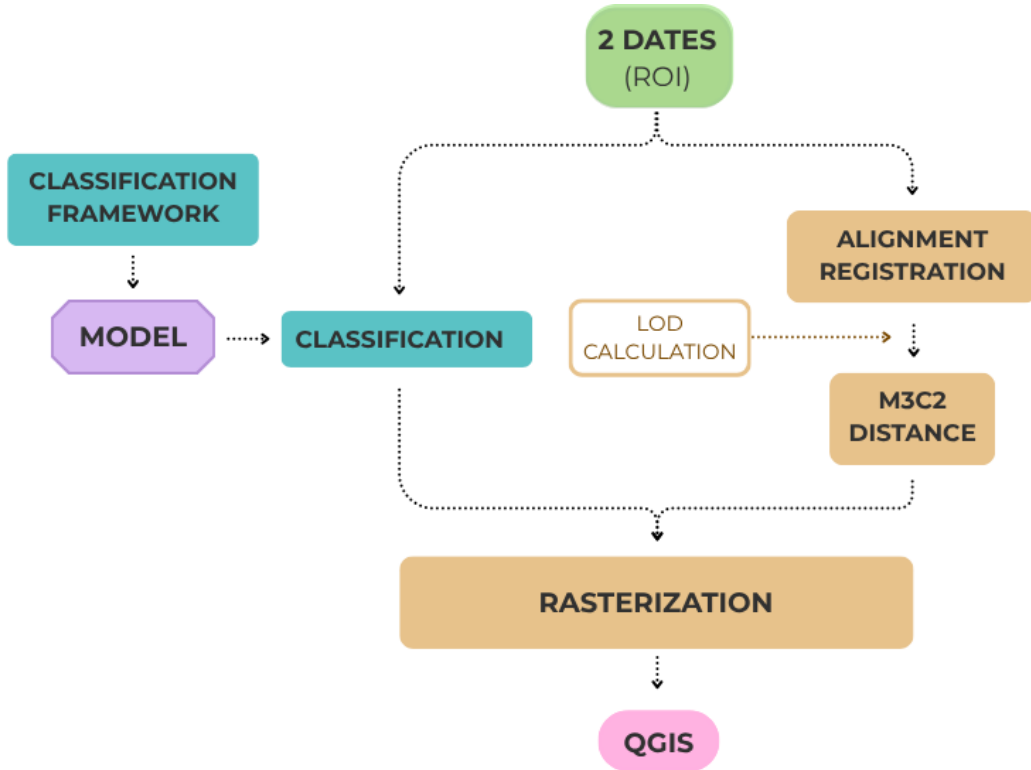


Figure 9: Analytical protocol for multi-temporal change detection and sediment budget calculation using the proposed Classification Framework and M3C2.

3.4.1 Cloud Co-registration and M3C2 Distance Computation

Before computing distances, the two clouds are aligned to minimize co-registration errors, which would propagate into the change estimates and lead to overestimation of the erosion and deposition volumes. This is done with the "Fine registration" tool in CloudCompare, which uses the Iterative Closest Point (ICP) method.

An additional step to remove noise is the calculation of the **Level of Detection (LOD)**. The LOD between two point clouds is the smallest measurable elevation change that exceeds local measurement errors and noise. One way to obtain the global LOD is to measure the apparent change on stable surfaces that haven't moved between the surveys:

$$LoD_{95\%} = \mu \pm 1.96 \cdot \sigma [m] \quad (9)$$

where μ is the mean elevation change and σ is the standard deviation in the stable areas. I delimited four stable areas and a Gaussian curve was fitted to obtain the mean and the standard deviation. Values below the LOD thresholds are not considered as significant change.

The M3C2 algorithm includes the calculation of a local LOD:

$$LoD_{95\%} = \pm 1.96 \left(\sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} + reg \right) [m] \quad (10)$$

where σ_1 and σ_2 represent local surface roughness, N_1 and N_2 are local point counts, and *reg* is the isotropic co-registration error derived from the ICP alignment. M3C2 uses this LOD to discriminate significant change from bias and noise.

3.4.2 Sediment Budgets and Spatial Analysis

The classified and M3C2-derived point clouds are rasterized and imported into QGIS for spatial analysis. The point clouds are subsampled fitting a grid of 20 cm to reduce the file size of the resulting rasters. The three rasters are spatially joined, allowing for the calculation of volumetric sediment budgets and tracking of sediment changes over time by analyzing the changing proportions of the classes. For this analysis, I chose the stable areas LOD method, as it results in a tangible value that can be physically interpreted, even if the method is less precise than the M3C2 LOD calculation that takes into account local roughness and errors. I also used Excel to analyze the data and produce statistics and plots.

4 Results

This chapter presents the results of the work in three parts. First, I present some preliminary results that allow to select the best model parameters/configurations. Then, I evaluate the performance of this model on the Navisence datasets and I apply it to unseen data. Finally, I apply the best model obtained through the classification framework to a multi-temporal geomorphic analysis.

4.1 Preliminary results

I trained several models using a simple protocol on Navisence datasets to compare their performances. The datasets were acquired on a reach upstream of Plats de la Lé, close to the Zinal glacier. These preliminary results use a different protocol from the rolling evaluation presented below and are therefore mainly used to compare the different models and feature sets. In addition to the different models, I explored the impact of the spectral indices on the classification performance (results in Table 3). This was done by examining the effect of three feature sets. The first one does not contain any spectral indices (only geometric features), the second one includes raw RGB-based features, and the last one uses light-invariant RGB indices (ExG, NDI_RG, normalized RGB, HSV Hue and Saturation).

Regardless of the feature set, the model achieving best OA and F1 score was XGBoost, followed closely by LightGBM. The two Random Forest models presented lower scores, however the spread between classifier performances narrowed from 10% to 1% when using the light-invariant RGB features. Moreover, the Gradient Boosting models consistently presented a higher mean confidence than the Random Forest implementations, with differences of up to 20%. In terms of computation time, OpenCV was the slowest classifier and Sklearn was the fastest. However, feature computation takes around 15–20 minutes for a Region of Interest composed of 227,000 points and ≈ 450 features, thus classification time can be neglected.

Models trained only on geometric features achieved Overall Accuracies (OA) ranging from 67.6% to 79.1% depending on the classifier. Adding raw RGB features improved performance slightly, with OA between 78.9% and 80.9%. The best results were obtained with the light-invariant RGB indices, where OA reached around 85% for all models. This indicates that adding spectral indices results in an improved classification process. Moreover, removing RGB indices that vary with light settings suggest an improved generalizability by reducing over-fitting to dataset-specific conditions.

Table 3: Comparison of models and spectral feature sets. The features are computed for scales 0.5 to 9.5m with a 0.5m-step, resulting in ≈ 450 features.

Feature set	Model	OA (%)	Macro F1 score (%)	Mean confidence (%)	Time (min)
No RGB indices	OpenCV	67.6	65.3	67.9	1.7
	Sklearn	68.6	67.1	68.1	0.2
	XGBoost	79.1	78.5	89.4	1.2
	LightGBM	78.4	77.4	91.3	0.4
RGB features	OpenCV	79.7	78.9	71.0	2.4
	Sklearn	78.9	78.1	71.2	0.2
	XGBoost	80.9	80.1	90.5	1.4
	LightGBM	79.8	78.9	92.4	0.6
Invariant RGB indices	OpenCV	85.2	85.2	72.5	1.4
	Sklearn	85.3	85.3	72.7	0.1
	XGBoost	86.5	86.5	90.5	0.8
	LightGBM	85.9	86.0	92.2	0.6

4.2 Model evaluation on the Navisence datasets

We have five datasets for the Plats de la Lé reach of the Navisence (≈ 1.7 km long), acquired during the summers of 2024-2026. The two dates in 2024, 11th June and 5 July, correspond to the before and after of an important flood event. The 2025 datasets were also acquired in June and July, and the final dataset corresponds to the 2 June 2026. I labeled the five datasets and applied the rolling protocol for three rounds (see Table 4).

Table 4: Rolling protocol: dataset role assignment at each round

Round	Training date(s)	Validation date	Testing date
1	11 June 2024	5 July 2024	30 June 2025
2	11 June 2024 5 July 2024	30 June 2025	23 July 2025
3	11 June 2024 5 July 2024 30 June 2025	23 July 2025	2 June 2026

I constructed an XGBoost model using a reduced feature set with scales ranging from 0.5 m to 6.5 m, and only light-invariant spectral features, resulting in a feature set composed of 299 features. Moreover, the large woody debris class was subsampled to 70% of the 2,000-point target to avoid over-representation, as this class is naturally scarce. The model and feature selection are based on the preliminary results in Table 3 and the intermediate tests shown in Table 9 in the Appendix. From now on, I will use this precise XGBoost model obtained through the rolling protocol and I will refer to it as **XGBoost-RP**. The evaluation metrics obtained for each round are presented in Table 5, and the evolution of the Overall Accuracy can be visualized in Figure 10. Both metrics present a similar evolution, presenting the highest performance in the second round, with OA = 87.8% and macro F1 score = 87.8%. The third round presents a decrease in performance, with an Overall Accuracy of 85.5%. Performance improves when moving from one to two training datasets, but decreases slightly when the third dataset is added. With only three rounds, no clear trend can be identified.

Table 5: Evaluation metrics of the XGBoost-RP

Round	Overall Accuracy	Macro F1 score
1	79.0%	77.8%
2	87.8%	87.8%
3	85.5%	85.2%

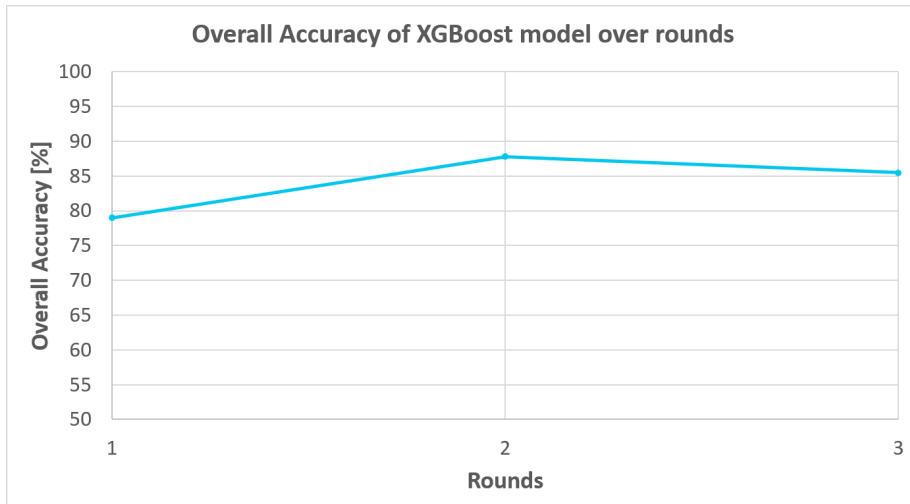


Figure 10: Visualization of the Overall Accuracy of the XGBoost-RP calculated in the validation step of the three rounds of the rolling protocol.

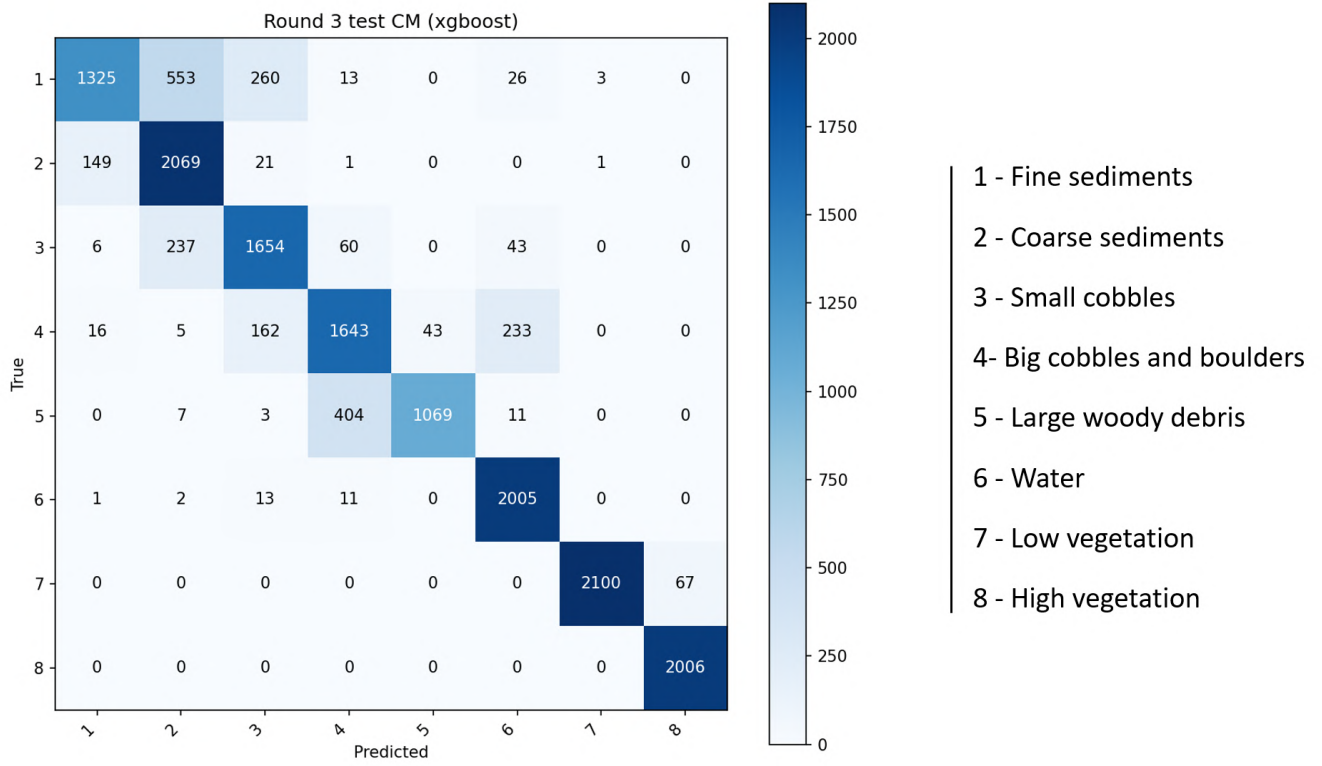


Figure 11: Confusion matrix for last round of the XGBoost-RP model. Predicted classes are shown on the x-axis and ground-truth classes on the y-axis. The color gradient from white to dark blue indicates an increasing number of points. The diagonal corresponds to the true positives (correctly classified).

While the overall accuracy and macro F1 score provide a general overview of model performance, the performance of individual classes can reveal patterns that are not apparent from these overall metrics alone. The confusion matrix, see Figure 11, compares the predicted classes to the ground-truth (labeled classes) and allows to detect which classes have been correctly classified (true positives) and which ones are mislabeled by the classifier. There is a dominant diagonal, which corresponds to the correctly classified points. The coarse sediments, water and both vegetation classes have the highest proportion of true positives. The first row of the matrix shows that fine sediments are frequently confused with other coarser sediment classes. Moreover, big cobbles and boulders are also confused with the smaller cobbles and water classes. Also, more than 25% of large woody debris is confused with big boulders. In general, misclassifications are concentrated in sediment classes, while water and vegetation are accurately classified.

Table 6: Per-class metrics of the last round of the XGBoost-RP model

	Fine sediments	Coarse sediments	Small cobbles	Big cobbles and boulders	Large woody debris	Water	Low vegetation	High vegetation
F1 score	72.1%	80.9%	80.4%	77.6%	82.0%	92.2%	98.3%	98.4%
Precision	88.5%	72.0%	78.3%	77.1%	96.1%	86.5%	99.8%	96.8%
Recall	60.8%	92.3%	82.7%	78.2%	71.6%	98.7%	96.9%	100.0%

The per-class metrics in Table 6 provide additional insights into these misclassifications. Precision indicates how reliable a prediction is when the model assigns a given class, while recall reflects the model’s ability to identify all actual points of that class. Fine sediments have the lowest F1 score (72.1%), the combination of a low recall of 60.8% and a high precision of 88.5%. This indicates that predictions assigned to fine sediments are reliable, but a large proportion of actual fine sediment points are classified as other classes. Conversely, coarse

sediments show a high recall (92.3%) but a lower precision (72.0%), indicating that most actual coarse sediment points are correctly identified, but many points predicted as coarse sediments belong to other classes. Large woody debris follows a similar pattern to fine sediments, with high precision (96.1%) but low recall (71.6%). This coincides with the confusion matrix results, which show it's confusion with big cobbles and boulders. Water and both vegetation classes achieve higher performances, with F1 scores of 92.2%, 98.3%, and 98.4%, respectively. High vegetation has a clear separation from other classes, as it has maximum recall.

4.2.1 Feature and Scale Selection

Exploring the most relevant features and scales is useful to further optimize the feature selection during the validation process in the rolling protocol. The top-20 most relevant features for classifier prediction are shown in Figure 12 ranked by importance. The two most important features are the PCA3 and the Excess Green Index (ExG). PCA3 is an eigenvalue-based geometric feature that represents the surface variation, distinguishing planar vs. volumetric structures. PCA3 appears several times at different scales across the full range, while ExG is most prominent at small scales. Normalized red and green values also frequently appear across varying scales. ZRANGE (Zmax - Zmin) appears at scales 1 and 6.5 m, capturing structure at both fine and coarse spatial levels. Sphericity at 1 m is also among the top-20 most informative features. Overall, 60% of the top-20 features are spectral, mostly based on Green and Red; whereas PCA3 and ZRANGE are the most prominent geometric features.

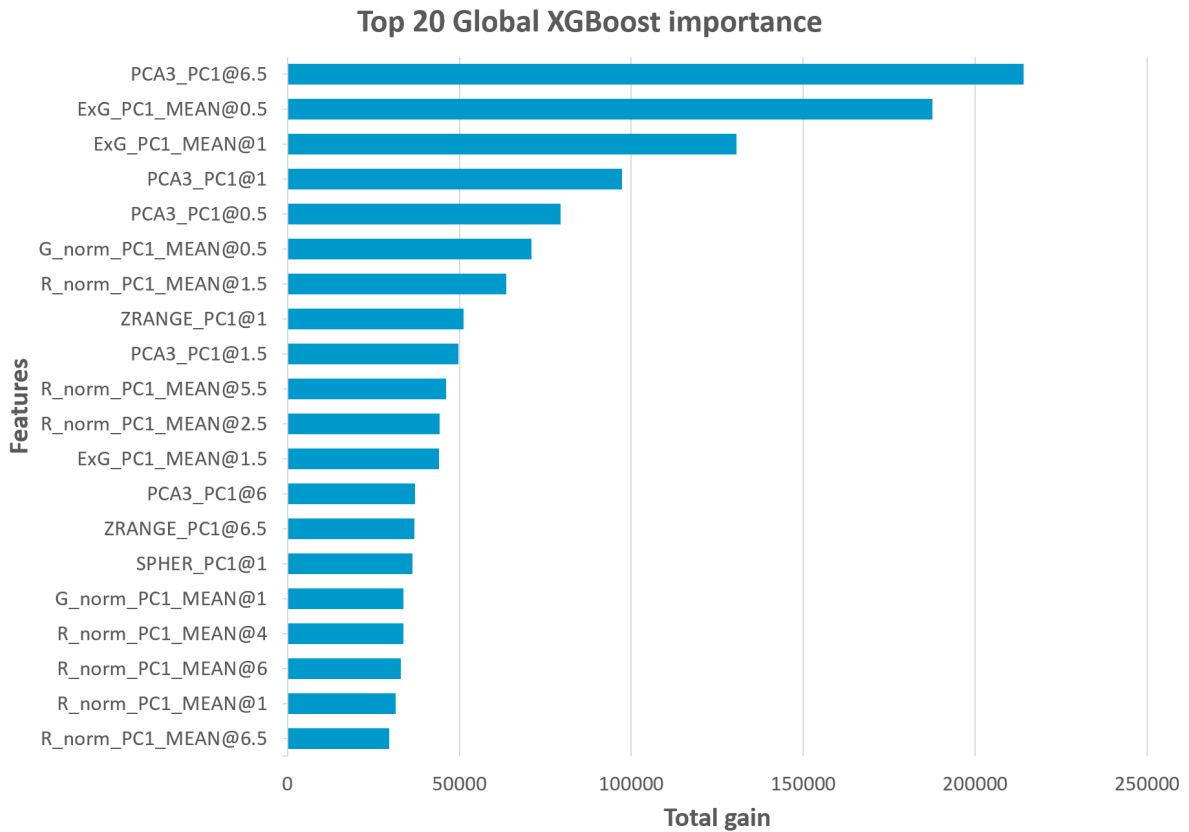


Figure 12: Top-20 ranking of global feature importance from the last round of XGBoost-RP classification (validation dataset). On the x-axis, the total gain (Gini index), and on the y-axis, the features ordered from higher importance to lower importance, in the form FEATURE_DATASET(_STATISTIC)@SCALE.

4.2.2 Model Interpretability via SHAP Analysis

I then performed a SHAP analysis on the final round of the XGBoost-RP model to understand how individual features drive each prediction. The cumulative SHAP importance curve (Figure 13) shows the combined contribution of the top features to the model's prediction. The curve rises quickly at first and then progressively flattens, indicating that approximately 50 features can explain 50% of the prediction, whereas ≈ 135 features are needed to reach 80% cumulative importance, and around 190 to reach 90%. These results indicate that the model's predictive information is not concentrated in only a handful of features, but rather a wide range of descriptors.

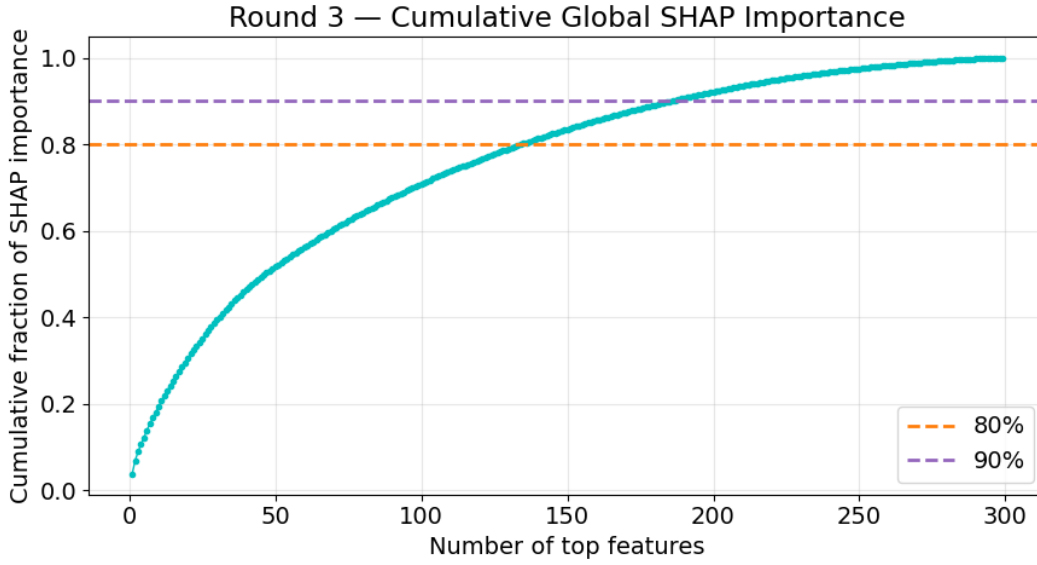


Figure 13: Cumulative global SHAP importance for the XGBoost-RP model. The blue curve shows the number of top features (x-axis) that correspond the cumulative fraction of SHAP importance (y-axis), where the thresholds 80% and 90% are displayed with the orange and purple dotted lines, respectively.

The top-10 features are ranked by mean absolute SHAP values, indicating the magnitude of impact of the most relevant descriptors contributing to the overall classification, see Figure 14. The difference with the top-20 ranking produced by the XGBoost model (see Figure 12) is the metric used to evaluate the features. The former one, uses SHAP values, while the latter uses the Gini importance. The three most influential features are all related to the color green: Excess Green (ExG) at scales 0.5 and 1 m, and normalized Green (G_norm) at 0.5 m. Importance then shifts toward PCA-derived features, normalized Green at larger scales, ZRANGE, and normalized Red (R_norm). This ranking shows the global impact of the features, as well as the magnitude of their contribution to each specific class. For instance, the top feature ExG at scale 0.5 m has a big effect on the fine sediments, large woody debris, and water classes, but not on the larger sediment classes nor the vegetation.

This differential impact on the classes is displayed more clearly on the SHAP heatmap presented in Figure 15. Ranked second, the normalized Green feature presents an opposed pattern to ExG at the same scale, highly influencing the vegetation classes. In fourth place, PCA-derived feature at scale 1 m is mainly discriminative for the coarse sediments class, despite being a globally important feature. The remaining features are distributed more evenly across classes. Overall, the heatmap shows that small sediment classes have distinct importance profiles, whereas larger sediment classes present a more uniform descriptor distribution.

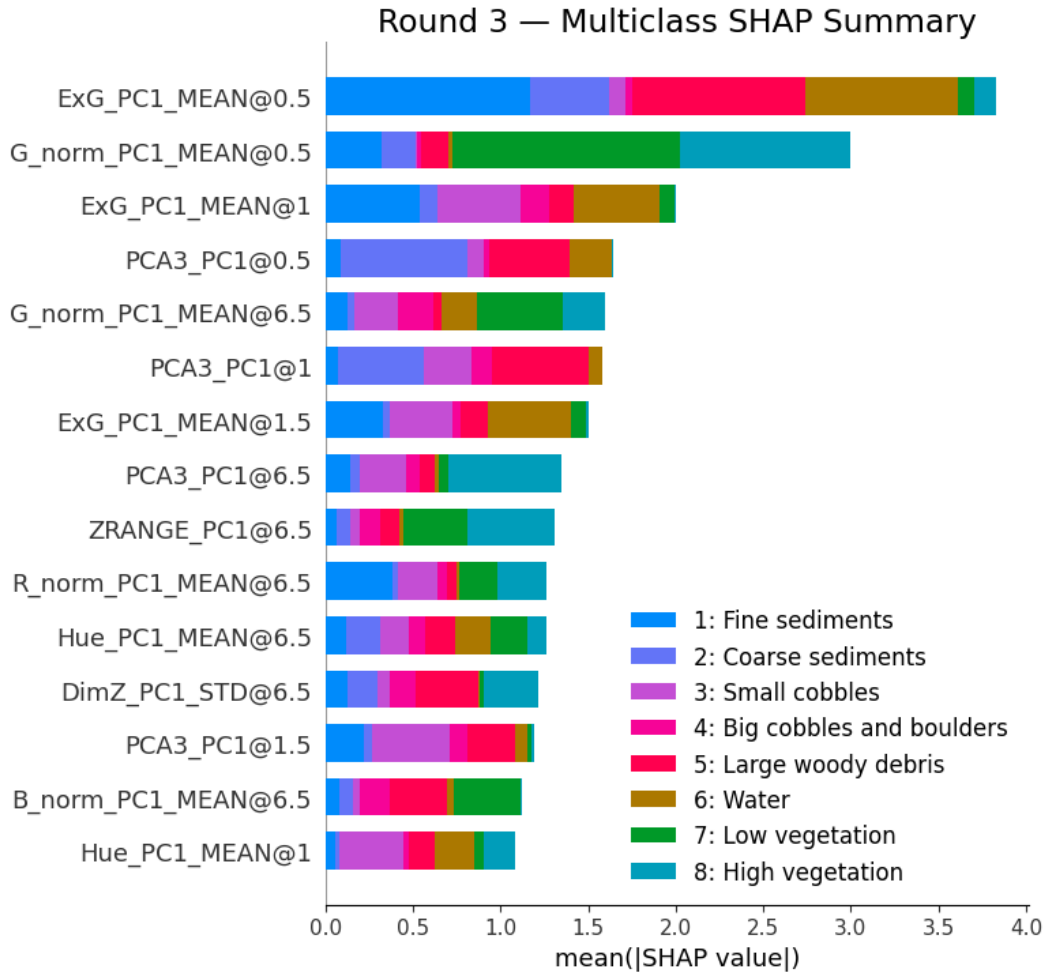


Figure 14: Multi-class SHAP summary of the XGBoost-RP model. Mean absolute SHAP value per feature (x-axis), ranked from highest to lowest impact (y-axis). Features are named in the form FEAT_DATASET(_STATISTIC)@SCALE. Each color in a feature's horizontal bar corresponds to a specific class, indicating how much that feature contributes to it's prediction.

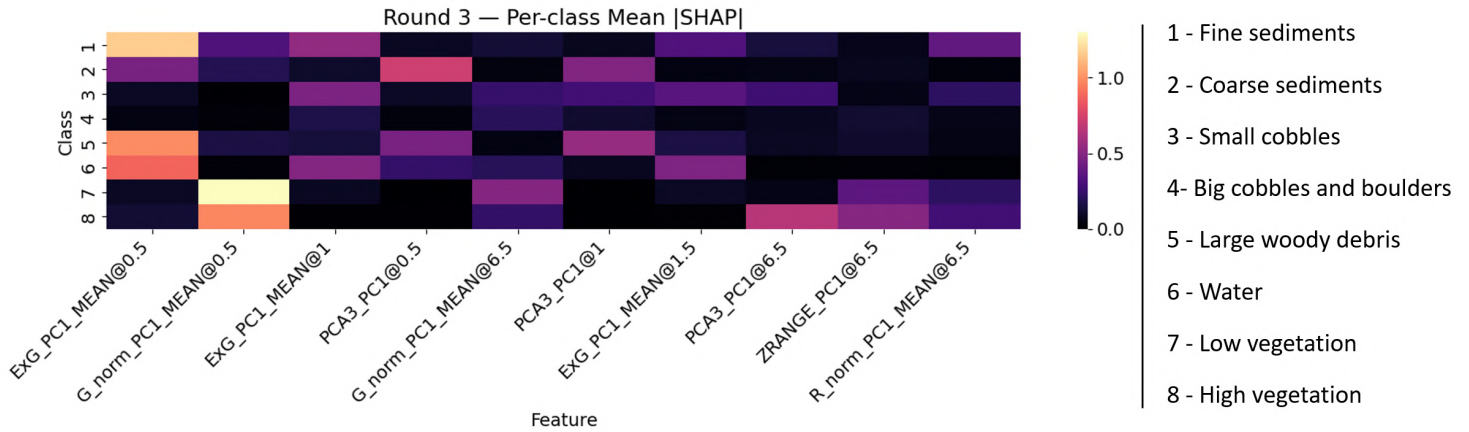


Figure 15: SHAP heatmap for the XGBoost-RP model, shows mean absolute SHAP values per feature per class. On the x-axis, features are ordered from highest to lowest global impact, in the form FEAT_DATASET(_STATISTIC)@SCALE. Colors range from black (no effect) to yellow (strong impact).

However, a global ranking does not indicate if the impact of the feature is positive or negative towards the prediction of the classes. The SHAP beeswarm plots complement the global feature impact by ranking class-specific top features and providing the magnitude and the direction of their effect.

The four sediment classes present very different feature profiles, provided in Figure 16. For the fine sediments class, the mean ExG at scales 0.5 m and 1 m is the strongest discriminator, with SHAP values ranging from -3 to 1.5. High values of PCA3 (associated with volumetric structures), of color variance and ZRANGE decrease the model's prediction towards fine sediments. This is coherent with the planarity, homogeneity and low height range characteristic of fine sediments. Coarse sediments, similar to fine sediments, also present a negative SHAP value (against the class prediction) with high values of PCA3, sphericity, color variance and dip (orientation of a planar surface). Small cobbles appear to have a distinctive color, illustrated by the high density of a small range of Hue values with a positive SHAP value. As for the smaller sediment classes, high PCA3 and sphericity values have a negative effect on the prediction of small cobbles. Finally, bigger sediments (4) are positively predicted by a high dip angle, high Zmin and high sphericity, which matches the description of big boulders. Moreover, a high color variance has a negative impact, as individual boulders do not present varying colors.

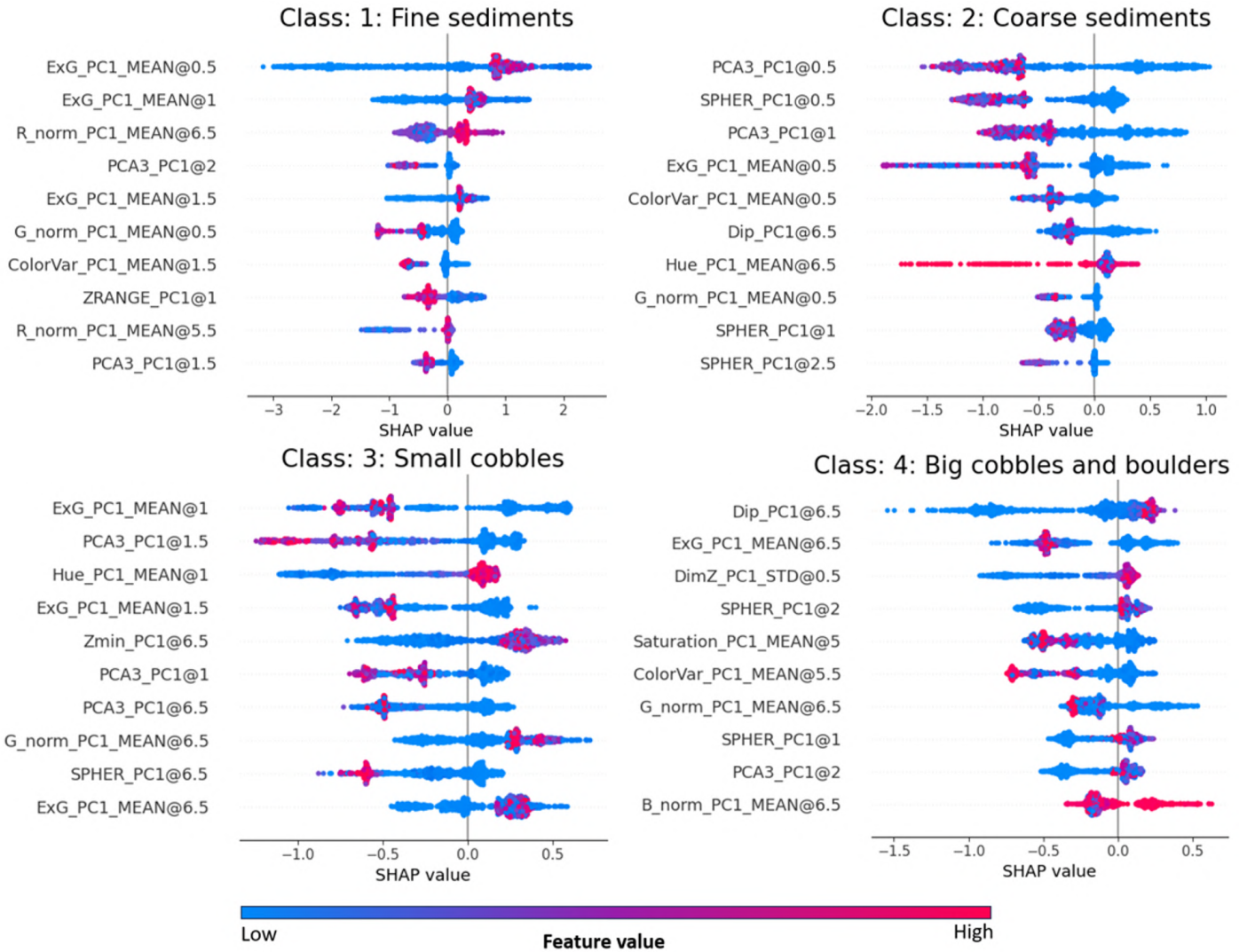


Figure 16: SHAP beeswarm plots for the XGBoost-RP model for the sediment classes: fine sediments (top left), coarse sediments (top right), small cobbles (bottom left) and big cobbles and boulders (bottom right). Each dot represents one sample, colored by feature value (blue = low, red = high), stacked vertically to show higher density. The x-axis shows the SHAP value and the y-axis ranks features by mean absolute impact on the class.

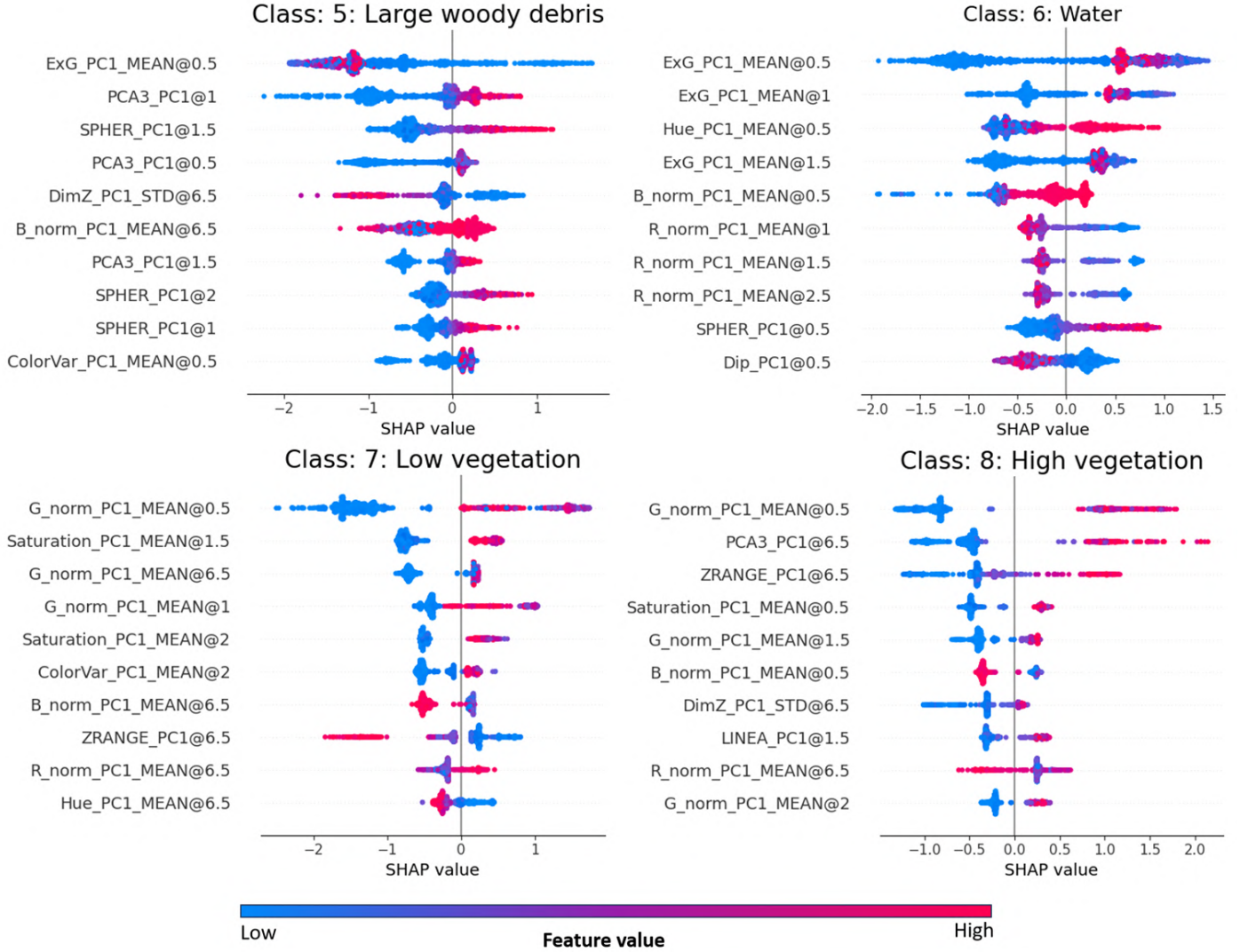


Figure 17: SHAP beeswarm plots for the XGBoost-RP model for the classes large woody debris (top left), water (top right), low vegetation (bottom left) and high vegetation (bottom right). Each dot represents one sample, colored by feature value (blue = low, red = high), stacked vertically to show higher density. The x-axis shows the SHAP value and the y-axis ranks features by mean absolute impact on the class.

The beeswarm plots for the remaining classes are provided in Figure 17. The prediction of the large woody debris class is positively impacted by high values of PCA3 and sphericity, which are characteristics of logs and fallen trees. The features that have a strong negative SHAP value are the ExG at scale 0.5 m and DimZ at scale 6.5 m. For water, on the contrary, high ExG values at small scales impact positively its prediction. Water has a distinctive color even with different light conditions, which is illustrated by the cluster of same high Hue values with a positive SHAP value. Moreover, the most relevant features for the water class are mainly spectral. Finally, both vegetation classes present features that are less distributed horizontally across the SHAP values range. This means that the prediction is either pushed towards or against the class, with almost no features having no impact on the prediction. However, the two vegetation classes have different features that distinguish them. On the one hand, high values of normalized Green, saturation have a positive impact on low vegetation; while high ZRANGE has a negative impact. On the other hand, the prediction of high vegetation is positively impacted by high values of normalized Green, PCA3 and ZRANGE, which are characteristics of trees.

4.3 Classification results

I applied the model from the last round of XGBoost-RP to a ROI from the point cloud acquired on 2 June 2026, an unseen dataset (see Figure 18). Water is detected accurately, including in shallower parts of the channel. Both vegetation classes also correspond well to the RGB patterns, though some areas of low vegetation are confused with fine sediments, for example in the lower-left corner of the ROI. Large woody debris is correctly identified in most cases, but the class is over-represented, with some cobbles being misclassified as woody debris. River banks are generally classified as big cobbles and boulders. Overall, the classification is able to correctly capture most of the landscape.

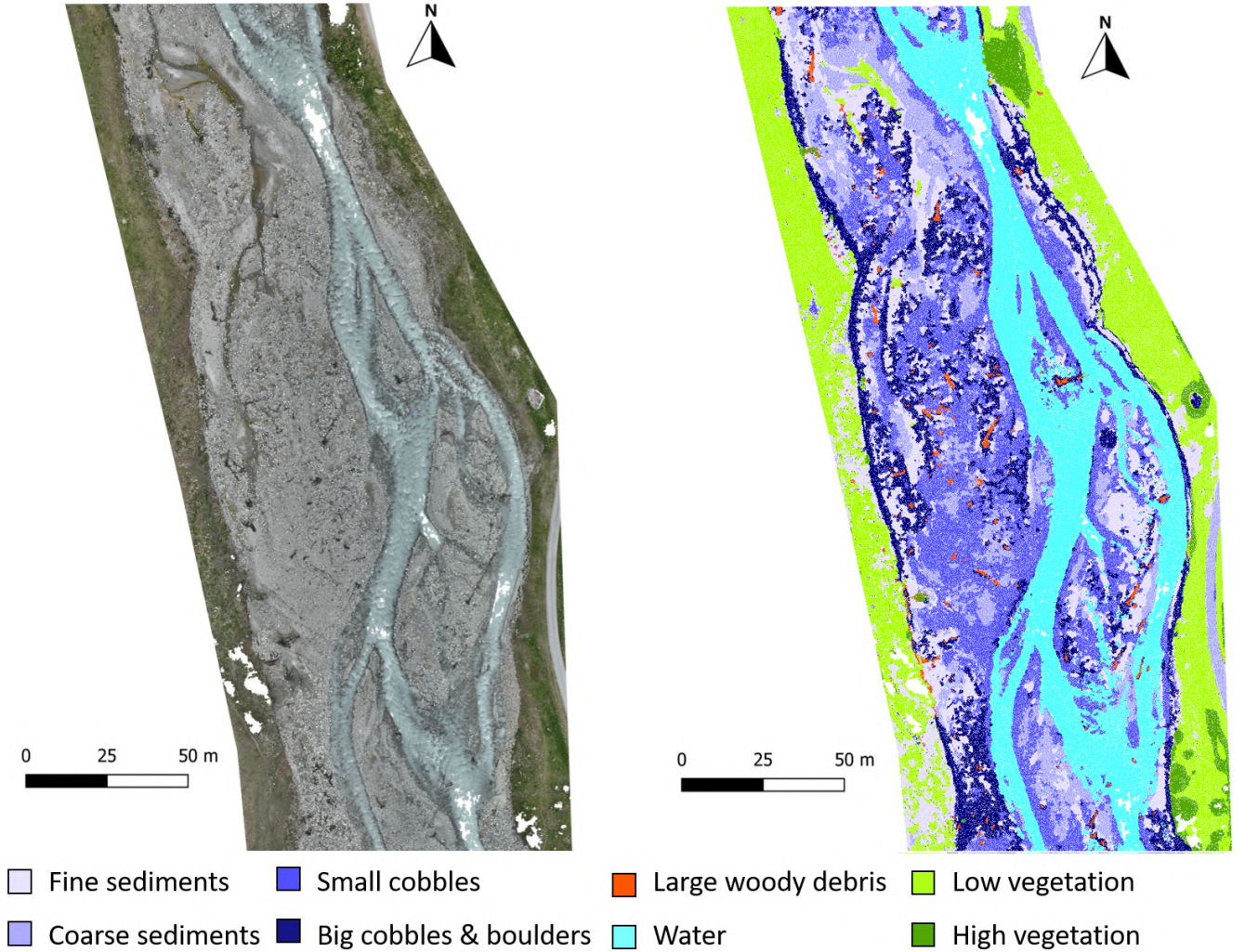


Figure 18: Classification result for a Region of Interest (ROI) segmented out from the dataset 2 June 2026. Left: RGB point cloud. Right: raster of the XGBoost-RP classified point cloud (resolution of 20 cm).

Table 7: Predicted class proportions for 2 June 2026, XGBoost-RP model

	Fine sediments	Coarse sediments	Small cobbles	Big cobbles and boulders	Large woody debris	Water	Low vegetation	High vegetation
June 2026	13.3%	9.1%	18.0%	12.6%	1.0%	25.3%	17.2%	3.5%

The class distributions of the classified ROI are summarized in Table 7. Water (25.3%), small cobbles (18%), and low vegetation (17.2%) represent the majority of the ROI surface area, while large woody debris (1%) and high vegetation (3.5%) are the least represented classes.

4.4 Multi-temporal Analysis: Summer 2024

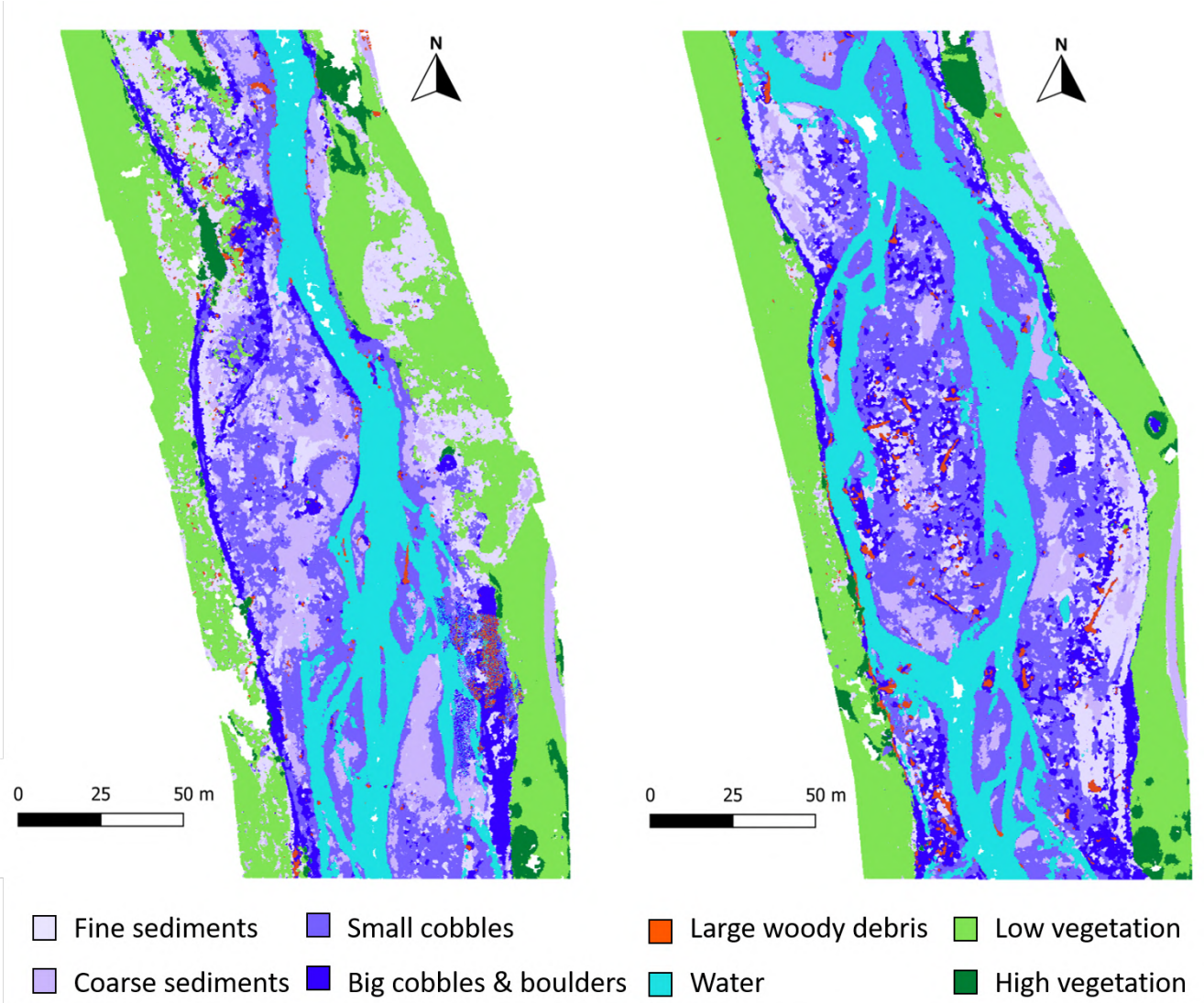


Figure 19: Classification result for June (left) and July (right) 2024, using the XGBoost-RP model. The classified point clouds are rasterized with resolution = 20 cm.

I classified a ROI extracted from the datasets acquired in June and July 2024 (see Figure 19) to analyze the geomorphological impact of the two major flood events that took place on June 2024. These two datasets are also part of the training of the XGBoost-RP model, which is used to classify them. This introduces data leakage and leads to a potentially over-estimated performance. However, they serve to illustrate the process rather than to show performance on unseen data. The original RGB point clouds can be found in Figures 24 and 25 in the Appendix for visual comparison. The classification outcomes are similar as for the 2026 dataset. Different river channels are visible through the water class, and river banks are identified as big cobbles and boulders. Low vegetation is occasionally misclassified as fine sediments or high vegetation. Moreover, fine sediments and large woody debris are over-represented. Also, there is an area in the June ROI that contains many misclassified points, resulting in a noisy region. Comparing the RGB point clouds and the classified raster allows to infer that the general landscape is well classified.

Furthermore, the classification of this region shows that in June, the river mainly flows through a single channel, whereas in July, the water is distributed across a more complex network of channels. The sediment classes also appear more spatially organized in June, while a greater mix of classes is visible in July. The classified maps therefore provide a visual overview of the spatial distribution of the different sediment classes and how it changes between the two dates.

The predicted class proportions change between the two dates (see Figure 20). Fine and coarse sediments increase slightly (by about 2% and 5% respectively), while small cobbles decrease by almost 10% and big cobbles and boulders by about 2%. Water coverage decreases by roughly 6%, and low vegetation increases by about 10%. Large woody debris stays similar across both dates. In June, small cobbles are strongly dominant among sediment classes, whereas the July distribution is more balanced.

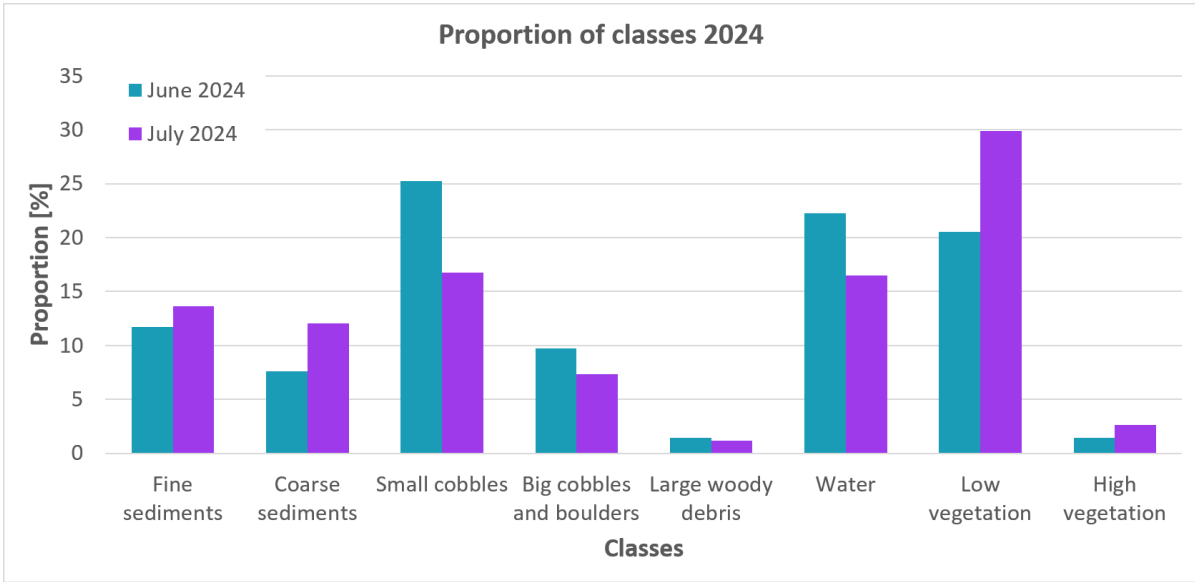


Figure 20: Bar plot with the predicted class proportions for June (blue) and July (purple) 2024.

4.4.1 LOD calculation and M3C2 results

To determine the level of detection (LoD), I measured elevation change in four stable areas outside the active channel. The mean and standard deviation of elevation change in these areas are $\mu = 0.02 \text{ m}$ and $\sigma = 0.06 \text{ m}$, giving:

$$\text{LoD}_{95\%} = \mu \pm 1.96 \cdot \sigma = 0.02 \pm 1.96 \cdot 0.06 = [-0.10, 0.14] \text{ m} \quad (11)$$

The LoD is asymmetric, reflecting a small upward shift in the second dataset relative to the first. The M3C2 distance map between June and July 2024 is shown before LoD filtering (see Figure 21). Two main erosion zones (red) are visible on the outer banks of the river, while deposition (blue) occurs primarily on the inside of the active channel. A thin erosion line is also visible on the left bank at the bottom, suggesting some minor bank erosion.

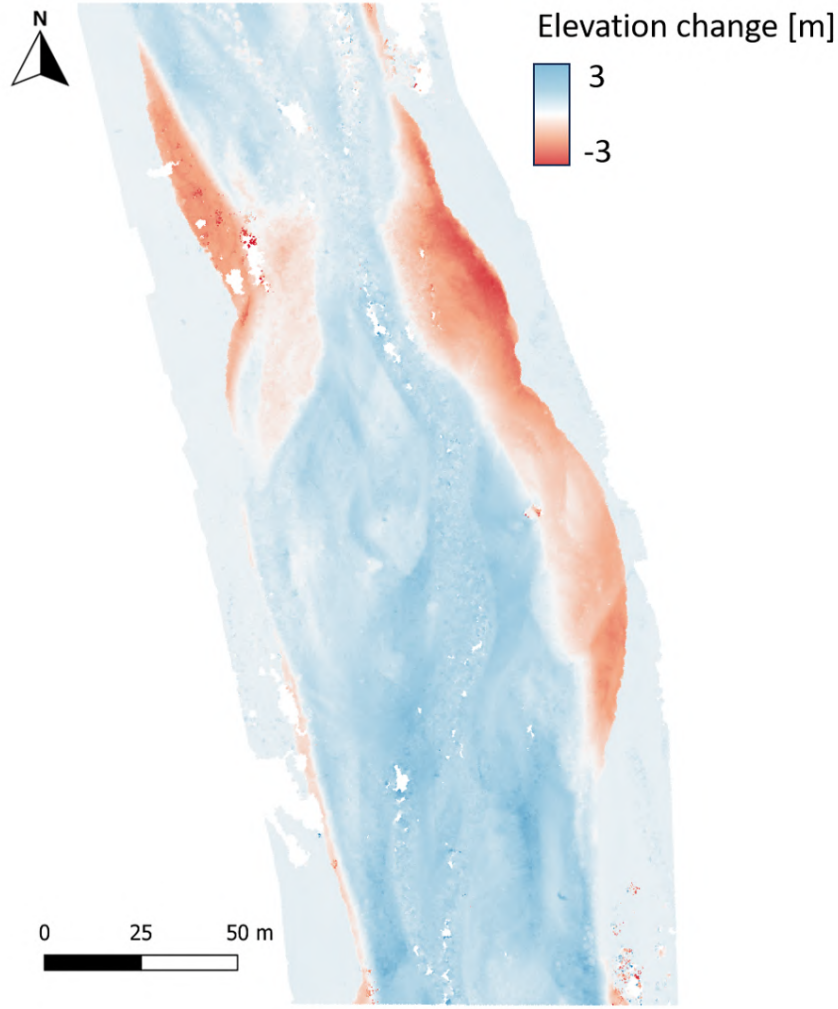


Figure 21: Raster of the raw M3C2 distances result for June–July 2024 (no LOD threshold is applied). Erosion is shown in red and deposition in blue. Lighter colors indicate small or no elevation change. The resolution of the raster is 20 cm).

4.4.2 Geomorphic change analysis

I calculated the erosion and deposition volumes by multiplying the area of each cell in the raster grid by the M3C2 distance:

$$Volume = 0.2 [m] \cdot 0.2 [m] \cdot M3C2 \text{ distance} [m] = 0.04 \cdot M3C2 \text{ distance} [m^3] \quad (12)$$

This allows to obtain the erosion and deposition volumes at the cell level. The addition of all eroded and deposited cells result in a total erosion $6,621 m^3$ and total deposition $10,424 m^3$. By adding these two quantities, the net change of the entire ROI results in $3,803 m^3$ (net deposition).

The combination of the classification results and the M3C2 volumes allows to link geomorphic changes with the surface classes at each date. For erosion, the classes assigned to the eroded surfaces correspond to the June classification, while for deposition, the classes correspond to the July classification. Comparing these class-associated volumes shows that the different classes do not contribute equally to the morphological change (see Figure 22). Fine sediments and big cobbles and boulders are mostly associated with erosion, while coarse sediments and small cobbles show more deposition than erosion. Large woody debris does not change much, although its proportion increases slightly.

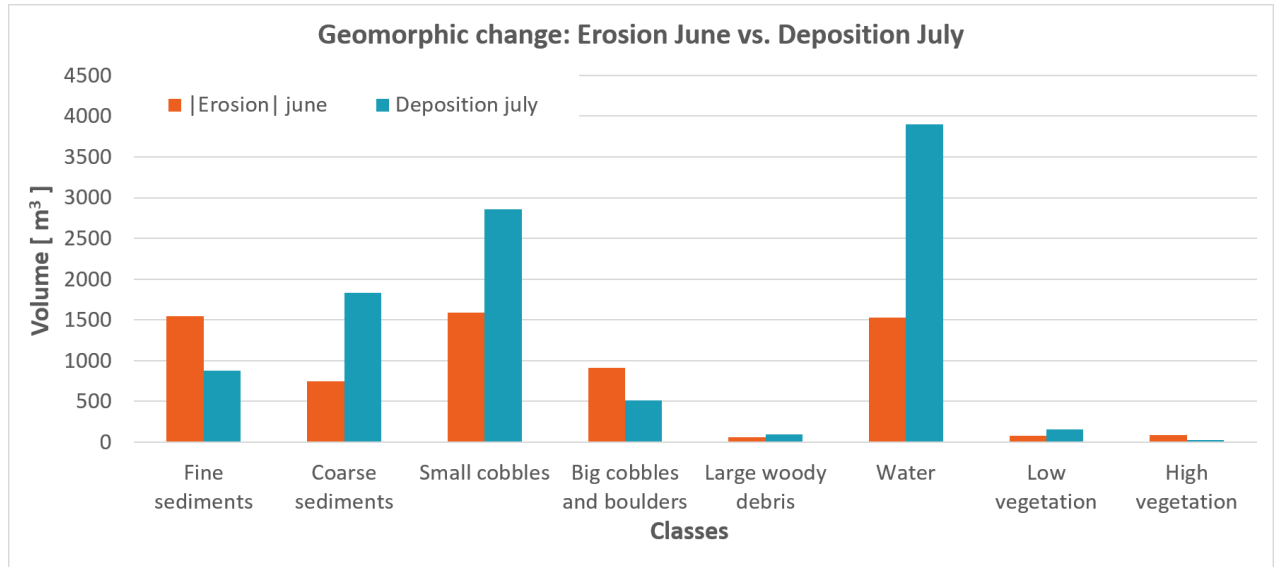


Figure 22: Per-class geomorphic change summary for June–July 2024. Absolute erosion in classes from the June dataset is shown in red, deposition in classes from July dataset in blue.

However, class proportions and geomorphic change volumes describe two different aspects of the river. Class proportions represent the surface at each date, while the M3C2 volumes describe the amount of elevation change between the two surveys. A class can cover a larger area in July and still be associated with erosion, as observed for fine sediments. Conversely, small cobbles show more deposition than erosion and also represent a larger proportion of the July surface. Class proportions cannot describe the magnitude or direction of geomorphic changes, but combining the classification with M3C2 can provide a more complete picture of the changes occurring during the flood period.

5 Discussion

The aim of this thesis was to develop and evaluate a classification framework for RGB point clouds derived from UAV-SfM surveys and to investigate how this tool could support geomorphic monitoring of Alpine rivers. The results show that the proposed approach can provide good and interpretable classifications using RGB and topographic information, while also highlighting several limitations related to sediment classes, illumination and model generalization. An application of this tool, combined with change detection methods, can support geomorphic change analysis.

5.1 Rolling Protocol and Classification Performance

The main contribution of this study is the evaluation of classification performance across successive UAV-SfM surveys using a rolling protocol. Unlike a conventional train-test split, this approach applies the model to datasets acquired at different dates, adding surveys to the training data in a monitoring context. This requires a more demanding evaluation of the model, as acquisition conditions change.

The classification framework developed in this thesis is based on the 3DMASC methodology. In the original study, 3DMASC was applied to bi-spectral topo-bathymetric LiDAR data [24] and achieved an overall accuracy (OAs) of 98% (see Table 8 in the Appendix). These high performances were obtained using training and testing areas extracted from the same dataset. Although the areas were spatially separated and independent, they shared the same acquisition conditions. The framework proposed in this thesis applies the methodology in a different context, using UAV-SfM point clouds containing only RGB and topographic information, and evaluates the model across repeated surveys.

The rolling protocol results show that the amount of training data can have a large effect on performance. Combining two datasets in the training process increases the OA from 79.0% to 87.8%, but performance decreases to 85.5% when a third training dataset is added (Table 5). Three rounds are not enough to identify a clear trend, performance could be reaching a plateau, or the differences could partly be related to the characteristics of the test dataset at each round. These results show that adding training datasets can greatly affect model performance and may help the model learn a wider range of conditions.

The final round is trained on the largest number of surveys and tests the XGBoost-RP model on the 2 June 2026 dataset. It produces an OA of 85.5% and a macro F1 score of 85.2%. Although its performance is lower than in the second round, the model has been trained with more acquisition conditions and therefore, it may be less affected by the specific characteristics of a survey. The lower performance could also indicate that the test dataset differs more from the previous surveys. This illustrates one of the advantages of the rolling protocol: performance is evaluated each time a new survey is introduced, enabling the identification of changes and problems in the data being classified.

The per-class metrics provide a more detailed view of the performance of the final round (see Figure 11 and Table 6). Water and both vegetation classes are particularly well classified, with F1 scores above 92%, while the sediment classes show lower performances. Fine sediments have the lowest F1 score, mainly due to their low recall (60.8%). This means that predictions classified as fine sediments are generally correct, but many actual fine sediment points are assigned to other classes. In contrast, coarse sediments have a high recall but a lower precision, showing that most actual coarse sediment points are detected, but that many points predicted as coarse sediments belong to other classes, such as large woody debris. These results show that the overall OA of 85.5% hides important differences in performance between classes, which can be revealed by more detailed per-class analyses.

The results also show that the choice of feature set has a larger effect on performance than the choice of classifier. With invariant RGB indices, the four tested models differed by only about 1% in OA, whereas the difference was greater than 10% when only geometric features were used. The light-invariant indices may

provide a more stable representation of the landscape across different surveys. XGBoost achieved the highest performance among the tested classifiers, which is consistent with previous comparisons of tree-based classifiers [49]. However, this difference was relatively small once the feature set was optimized.

5.2 Feature Importance and Model Interpretability

The SHAP analysis provides further insights into how the model distinguishes the different components of the fluvial landscape. The classes present distinct profiles, each with a different combination of spectral and geometric features. These profiles contribute to identify the classes, although the differences between them are not always clear.

The SHAP results illustrate more in detail the complementary role of spectral and geometric information. For instance, spectral features are particularly important for classes such as water and vegetation, while geometric features provide additional information about the 3D structure of objects. This is particularly visible for the two vegetation classes (see Figure 17). While normalized Green contributes to the prediction of both low and high vegetation, PCA3 and ZRANGE are mainly associated with high vegetation, reflecting the spatial configuration of trees.

The importance of the features also varies with the spatial scale at which they are calculated. The most important descriptors are mainly found at the smallest scales (0.5–1 m) and at the largest scale (6.5 m), while intermediate scales are not in the top rankings (see Figures 14 and 15). This suggests that both local surface properties and broader spatial structure contribute to the classification. Moreover, the lower importance of intermediate scales may also indicate that some descriptors at these scales are redundant with features calculated at smaller or larger scales [24]. However, the cumulative SHAP analysis (Figure 13) shows that these features still contribute to the overall prediction.

More generally, the SHAP results are coherent with the physical characteristics of the different classes (Figures 16 and 17). On the one hand, planar and homogeneous surfaces, characterized by low values of PCA3 and color variance, are associated with fine sediments. On the other hand, geometric descriptors of 3D structure (such as of PCA3, ZRANGE, dip and sphericity) contribute to the prediction of boulders, large woody debris and high vegetation. The interpretability is one of the advantages of the proposed approach, as the model predictions can be related to observable properties of the landscape. This is illustrated in Figure 23 for a few key features: spectral indices like ExG and normalized green are able to distinguish vegetation from the active channel, while PCA3 captures both small-scale roughness on planar surfaces (sediments, water, low vegetation) and large-scale 3D structure (trees and river banks).

Overall, the SHAP analysis shows that the classification relies on different combinations of spectral and geometric information depending on the class and the spatial scale. This provides a way to better understand the behavior of the model and the information used for classification.

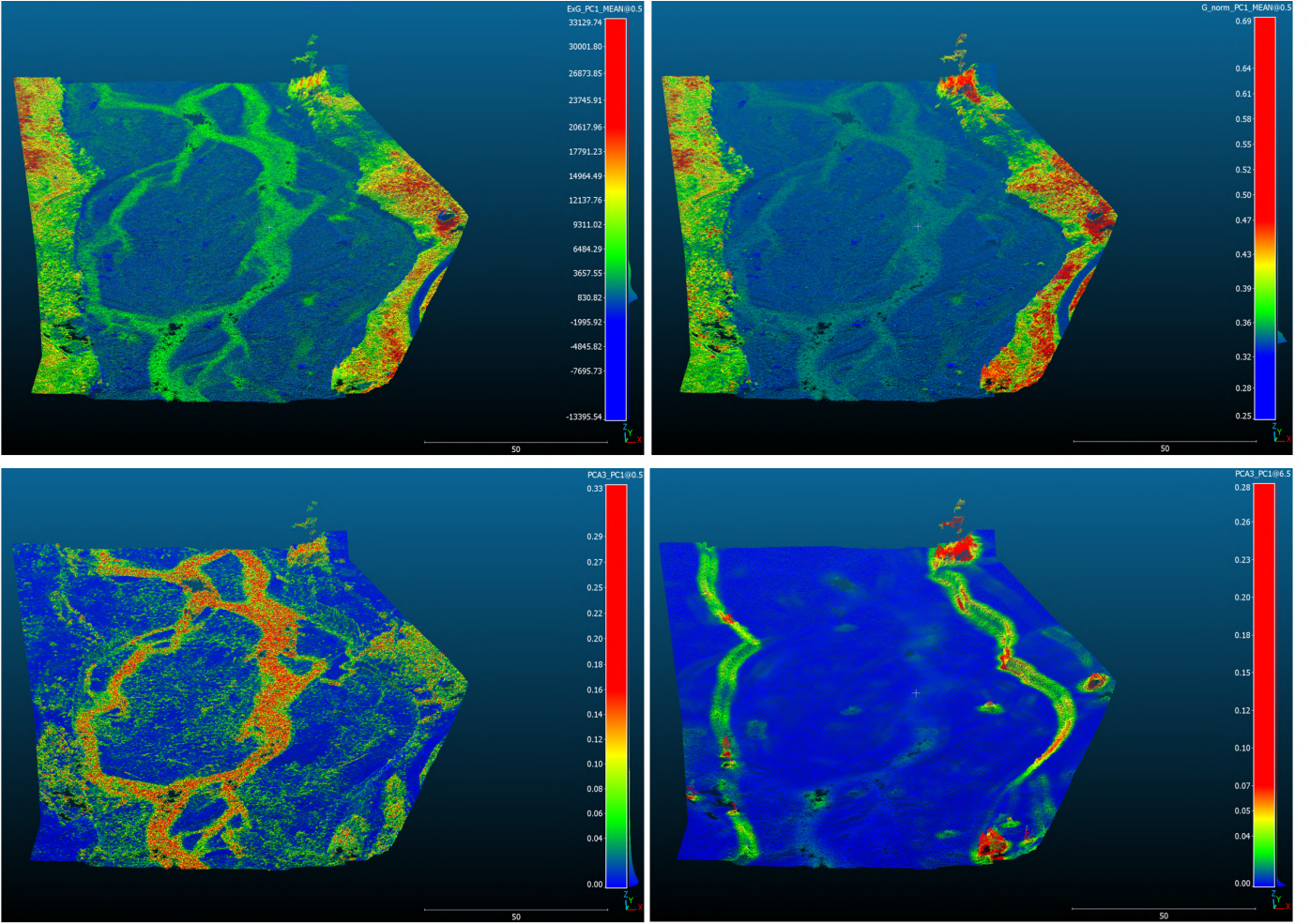


Figure 23: Selected features for the ROI acquired on June 2024. On the top, the most relevant spectral features ExG (left) and normalized Green at 0.5m (right). On the bottom, the PCA3 geometric feature at scales 0.5m (left) and 6.5m (right). High PCA3 values, in red, indicate volumetric structures, whereas low PCA3, in blue, indicates planar structures.

5.3 Challenges and Limitations of RGB-SfM Classification

The main classification difficulties are related to the sediment classes and to some visually similar surfaces. Sediment properties vary gradually with increasing sediment size, making them difficult to represent using discrete classes. Moreover, because the labeling process involved my own interpretation, some labeled areas may contain a mixture of sediment sizes. This reflects the complexity of natural landscapes and the challenge of delineating these classes. This may have contributed to the confusion between sediment classes observed in the final classification.

Large woody debris is also sometimes classified as big cobbles and boulders. Both classes can form isolated 3D structures above the surrounding surface and therefore present similar geometric characteristics. This is consistent with the SHAP analysis, where geometric descriptors such as PCA3 and sphericity are important for both classes. Some low vegetation is also confused with fine sediments in the classification results (see Figure 18). These surfaces can have similar colors and relatively low 3D structure, which makes it difficult to distinguish them.

Another challenge is the effect of lighting and shadowing. Variations in illumination can result from vegetation, boulders, and changes in sun position and terrain orientation. These variations affect RGB values and

can change the properties of the same surface between surveys. Light-invariant RGB indices improve the performance of the model, suggesting that their influence can be partly reduced (see Table 3). This remains an important limitation for multi-temporal applications.

Finally, the class distribution is not uniform across the landscape. For example, large woody debris represents only a small proportion of the surveyed area, which can make the classification of minority classes more difficult. In this study, the proportion of large woody debris was reduced in the training data to avoid its over-representation. However, class imbalance remains a challenge for naturally scarce classes and when class proportions vary between surveys.

5.4 Framework Contributions and Limitations

The results highlight several advantages of the proposed framework. First, it uses standard RGB-SfM data, making the approach more accessible than methods relying on LiDAR or multispectral data. Second, the rolling protocol provides a flexible way to incorporate new surveys, practical for repeated surveys and long-term monitoring. Third, the use of SHAP makes the classification easier to interpret by showing which features contribute to the prediction of each class.

However, the approach also implies several limitations. First, supervised machine learning requires labeled training data, which can be time-consuming to produce, especially for sediment classes where boundaries are not always clear. In addition, the framework was developed and evaluated on the Navisence. Its performance on other rivers may be affected by differences in sediment color, vegetation, illumination or acquisition settings. Further testing on other reaches and rivers could allow to assess its generalization ability.

5.5 Combining Classification and 3D Change Detection for Geomorphic Monitoring

The M3C2 analysis shows a net deposition of $3,803 \text{ m}^3$ between June and July 2024, with $6,621 \text{ m}^3$ of erosion and $10,424 \text{ m}^3$ of deposition. The spatial distribution of the changes (Figure 21) is consistent with the response of a braided river to large flood events, with mobilization of gravel bars, scouring along the outer river banks and deposition within the active channel. The results indicate a reorganization of the river morphology and channel widening during the flood period.

Combining these results with the classification provides additional information on the type of surfaces that were eroded and deposited (see Figure 22). Fine sediments and big cobbles and boulders were predominantly associated with erosion, while coarse sediments and small cobbles showed more deposition than erosion. This suggests a reorganization of the sediment composition during the large 2024 floods, which led to sediment coarsening. The studied ROI is a short test reach, but further analyses on a longer reach could reveal downstream sorting and other large-scale geomorphological patterns.

This method has limitations, as it only captures sediment class at the surface, while geomorphic change involves volumes. However, this method provides a practical way to estimate geomorphic change according to sediment sizes using accessible tools, without the need of in-situ measurements.

Overall, the two tools provide complementary information: M3C2 quantifies where and how much the surface changed, while point cloud classification indicates which surface classes were involved in these changes.

6 Future perspectives

Operational and methodological improvements

First, exploring other spectral bands and indices could help to further distinguish some classes. For instance, adding near-infrared (NIR), which can easily distinguish vegetation, water and sediment classes. Moreover, switching from ExG to normalized ExG could be more adapted to the monitoring context. However, adding multi-spectral data would come at the cost of using more specific and potentially less accessible sensors.

Another possible improvement would be to include a pre-processing step comparing the features of different surveys. This could help detect drifts in certain features between acquisition dates and potentially allow compensation measures to be included within the protocol. This could be particularly useful for multi-temporal applications where illumination and acquisition conditions vary a lot.

Feature selection could also be integrated directly into the rolling protocol. In this study, the feature importance and SHAP analyses were used to identify the most relevant descriptors, but feature selection was not integrated into the classification procedure itself. Reducing the number of features could simplify the model and reduce the computation time required for feature calculation. In this study, I have analyzed only small reaches, due to the limitation of feature computation to small datasets. However, solving this problem could help analyze longer river reaches at the kilometer scale.

Class imbalance could also be addressed further. A first round of general classification could help estimate class proportions, and adapt the number of points for each class used for training to each surveyed region.

Generalization to other environments

Another important step would be to test the framework on other river reaches and regions. This study focused on the Navisence, so testing the approach under different sediment, vegetation and illumination conditions would allow to assess its generalization. The rolling protocol could then be used to progressively include new labeled datasets from different catchments.

Future applications

The classification and change detection could be analyzed per-pixel to identify how sediments changed in each zone of erosion and deposition. This could help move from a general description of morphological change towards a more detailed analysis of the processes controlling sediment redistribution in the river. This information could also be useful for applications related to channel restoration and aquatic habitat, for example for fish and macroinvertebrate habitats.

Finally, the classified point clouds and geomorphic change could be combined with hydrological information. Linking the observed erosion and deposition patterns with flow velocity could help better capture the effect of hydrology on sediment movement. Integrating these results with surface velocity measurements from LSPIV from UAV surveys could provide a link between flow conditions and sediment dynamics, using the same drone flights. Obtained data could help retrieve channel geometry, and linking it with hydrological measurements could also be interesting for hydraulic studies.

7 Conclusion

This thesis developed and evaluated a classification framework for RGB point clouds derived from UAV-SfM surveys to support geomorphic monitoring of Alpine rivers. The approach adapted the 3DMASC methodology to RGB and topographic data and introduced a rolling protocol to evaluate model performance across repeated surveys.

The results show that light-invariant RGB features had a larger influence on classification performance than the choice of classifier. Once these features were included, the four tested models showed only small differences in performance, with XGBoost achieving the highest results. The final round of the rolling protocol, trained on the largest number of surveys, reached an Overall Accuracy of 85.5% and a macro F1 score of 85.2%. The per-class results also showed that water and vegetation were particularly well classified, while sediment classes and large woody debris remained more difficult to distinguish.

The SHAP analysis showed that the model relies on different combinations of spectral and geometric features depending on the class and spatial scale. Important descriptors were mainly found at small scales of 0.5–1 m and at the largest scale of 6.5 m, indicating the complementary role of local surface properties and broader spatial structure. The analysis also provided a more interpretable view of the model and helped relate its predictions to observable characteristics of the river surface.

The combination of point cloud classification with M3C2 further showed the potential of the framework for geomorphic monitoring. Between June and July 2024, the study area experienced 6,621 m³ of erosion and 10,424 m³ of deposition, resulting in a net deposition of 3,803 m³. While M3C2 quantified the magnitude and spatial distribution of geomorphic change, the classification provided information on the surface classes associated with these changes. The two approaches therefore provide complementary information for describing river evolution.

Overall, this thesis shows that RGB UAV-SfM point clouds can provide a practical and accessible data source for 3D geomorphological classification and river monitoring. The proposed framework provides a basis for repeated surveys and for further applications on other river reaches, while also highlighting the need for additional work on generalizability, class labeling and the effects of acquisition conditions.

8 AI statement

Two AI tools were used during this project. GitHub Copilot was mostly used for coding and for assistance with CloudCompare and QGIS, whereas Claude was used to help structure and revise the thesis text.

References

- [1] F. Comiti et al. “Glacier melt runoff controls bedload transport in Alpine catchments”. en. In: *Earth and Planetary Science Letters* 520 (Aug. 2019), pp. 77–86. ISSN: 0012821X. DOI: 10.1016/j.epsl.2019.05.031.
- [2] Emma K. Raven, Stuart N. Lane, and Louise J. Bracken. “Understanding sediment transfer and morphological change for managing upland gravel-bed rivers”. en. In: *Progress in Physical Geography: Earth and Environment* 34.1 (Feb. 2010), pp. 23–45. ISSN: 0309-1333, 1477-0296. DOI: 10.1177/0309133309355631.
- [3] Rocko A. Brown and Gregory B. Pasternack. “Hydrologic and topographic variability modulate channel change in mountain rivers”. en. In: *Journal of Hydrology* 510 (Mar. 2014), pp. 551–564. ISSN: 00221694. DOI: 10.1016/j.jhydrol.2013.12.048.
- [4] Agata Walicka et al. “Terrestrial Laser Scanning for the Detection of Coarse Grain Size Movement in a Mountain Riverbed”. en. In: *Water* 11.11 (Oct. 2019), p. 2199. ISSN: 2073-4441. DOI: 10.3390/w11112199.
- [5] Daniel Vázquez-Tarrió et al. “Particle transport in gravel-bed rivers: Revisiting passive tracer data”. en. In: *Earth Surface Processes and Landforms* 44.1 (Jan. 2019), pp. 112–128. ISSN: 0197-9337, 1096-9837. DOI: 10.1002/esp.4484.
- [6] S. Lallias-Tacon, F. Liébault, and H. Piégay. “Step by step error assessment in braided river sediment budget using airborne LiDAR data”. en. In: *Geomorphology* 214 (June 2014), pp. 307–323. ISSN: 0169555X. DOI: 10.1016/j.geomorph.2014.02.014.
- [7] Joseph M. Wheaton et al. “Accounting for uncertainty in DEMs from repeat topographic surveys: improved sediment budgets”. en. In: *Earth Surface Processes and Landforms* 35.2 (Feb. 2010), pp. 136–156. ISSN: 0197-9337, 1096-9837. DOI: 10.1002/esp.1886.
- [8] Joseph M. Wheaton et al. “Morphodynamic signatures of braiding mechanisms as expressed through change in sediment storage in a gravel-bed river”. en. In: *Journal of Geophysical Research: Earth Surface* 118.2 (June 2013), pp. 759–779. ISSN: 2169-9003, 2169-9011. DOI: 10.1002/jgrf.20060.
- [9] Maarten Bakker et al. “Morphological Response of an Alpine Braided Reach to Sediment-Laden Flow Events”. en. In: *Journal of Geophysical Research: Earth Surface* 124.5 (May 2019), pp. 1310–1328. ISSN: 2169-9003, 2169-9011. DOI: 10.1029/2018JF004811.
- [10] Christian E. Torgersen et al. “Riverscape approaches in practice: perspectives and applications”. en. In: *Biological Reviews* 97.2 (Apr. 2022), pp. 481–504. ISSN: 1464-7931, 1469-185X. DOI: 10.1111/brv.12810.
- [11] David J. Milan et al. “Filtering spatial error from DEMs: Implications for morphological change estimation”. en. In: *Geomorphology* 125.1 (Jan. 2011), pp. 160–171. ISSN: 0169555X. DOI: 10.1016/j.geomorph.2010.09.012.
- [12] Gottfried Mandlbürger et al. “Topo-Bathymetric LiDAR for Monitoring River Morphodynamics and In-stream Habitats—A Case Study at the Pielach River”. en. In: *Remote Sensing* 7.5 (May 2015), pp. 6160–6195. ISSN: 2072-4292. DOI: 10.3390/rs70506160.
- [13] Bharat Sharma Acharya et al. “Unmanned Aerial Vehicles in Hydrology and Water Management: Applications, Challenges, and Perspectives”. en. In: *Water Resources Research* 57.11 (Nov. 2021), e2021WR029925. ISSN: 0043-1397, 1944-7973. DOI: 10.1029/2021WR029925.
- [14] Jakub Śledziowski. “UAV Lidar vs. Structure-from-Motion (SfM) Photogrammetry for Coastal Dune Monitoring: A Two-Year Multi-Temporal Case Study from the Southern Baltic Sea”. en. In: *Photogrammetric Engineering & Remote Sensing* (2026). ISSN: 0099-1112. DOI: 10.14358/PERS.26-00015R2.
- [15] Quinn W. Lewis and Bruce L. Rhoads. “LSPIV Measurements of Two-Dimensional Flow Structure in Streams Using Small Unmanned Aerial Systems: 2. Hydrodynamic Mapping at River Confluences”. en. In: *Water Resources Research* 54.10 (Oct. 2018), pp. 7981–7999. ISSN: 0043-1397, 1944-7973. DOI: 10.1029/2018WR022551.

- [16] G. Bodart et al. “Synthetic River Flow Videos for Evaluating Image-Based Velocimetry Methods”. en. In: *Water Resources Research* 58.12 (Dec. 2022), e2022WR032251. ISSN: 0043-1397, 1944-7973. DOI: 10.1029/2022WR032251.
- [17] Zhihong Wu et al. “Floodwater Extraction from UAV Orthoimagery Based on a Transformer Model”. en. In: *Remote Sensing* 16.21 (Oct. 2024), p. 4052. ISSN: 2072-4292. DOI: 10.3390/rs16214052.
- [18] Uyen Nguyen et al. “Mapping vegetation types in semi-arid riparian regions using random forest and object-based image approach: A case study of the Colorado River Ecosystem, Grand Canyon, Arizona”. en. In: *Ecological Informatics* 50 (Mar. 2019), pp. 43–50. ISSN: 15749541. DOI: 10.1016/j.ecoinf.2018.12.006.
- [19] Michael J. Starek and Justin Giessel. “Fusion of uas-based structure-from-motion and optical inversion for seamless topo-bathymetric mapping”. en. In: *2017 IEEE International Geoscience and Remote Sensing Symposium (IGARSS)*. Fort Worth, TX: IEEE, July 2017, pp. 2999–3002. ISBN: 978-1-5090-4951-6. DOI: 10.1109/IGARSS.2017.8127629.
- [20] Ge Pu, Lindi J. Quackenbush, and Stephen V. Stehman. “Identifying Factors That Influence Accuracy of Riparian Vegetation Classification and River Channel Delineation Mapped Using 1 m Data”. en. In: *Remote Sensing* 13.22 (Nov. 2021), p. 4645. ISSN: 2072-4292. DOI: 10.3390/rs13224645.
- [21] Martin Weinmann et al. “Semantic point cloud interpretation based on optimal neighborhoods, relevant features and efficient classifiers”. en. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 105 (July 2015), pp. 286–304. ISSN: 09242716. DOI: 10.1016/j.isprsjprs.2015.01.016.
- [22] E. Grilli, F. Menna, and F. Remondino. “A REVIEW OF POINT CLOUDS SEGMENTATION AND CLASSIFICATION ALGORITHMS”. en. In: *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences XLII-2/W3* (Feb. 2017), pp. 339–344. ISSN: 2194-9034. DOI: 10.5194/isprs-archives-XLII-2-W3-339-2017.
- [23] Nicolas Brodu and Dimitri Lague. *3D Terrestrial lidar data classification of complex natural scenes using a multi-scale dimensionality criterion: applications in geomorphology*. en. arXiv:1107.0550 [cs.CV]. Jan. 2012. DOI: 10.48550/arXiv.1107.0550.
- [24] Mathilde Letard et al. “3DMASC: Accessible, explainable 3D point clouds classification. Application to bi-spectral topo-bathymetric lidar data”. en. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 207 (Jan. 2024), pp. 175–197. ISSN: 09242716. DOI: 10.1016/j.isprsjprs.2023.11.022.
- [25] Rui Zhang et al. “Deep-Learning-Based Point Cloud Semantic Segmentation: A Survey”. en. In: *Electronics* 12.17 (Aug. 2023), p. 3642. ISSN: 2079-9292. DOI: 10.3390/electronics12173642.
- [26] Abderrazzaq Kharroubi et al. “Multi-Context Point Cloud Dataset and Machine Learning for Railway Semantic Segmentation”. en. In: *Infrastructures* 9.4 (Apr. 2024), p. 71. ISSN: 2412-3811. DOI: 10.3390/infrastructures9040071.
- [27] Evania Chetty, Maqsooda Mahomed, and Shaeden Gokool. “A Comparative Analysis of Multi-Spectral and RGB-Acquired UAV Data for Cropland Mapping in Smallholder Farms”. en. In: *Drones* 10.1 (Jan. 2026), p. 72. ISSN: 2504-446X. DOI: 10.3390/drones10010072.
- [28] Lukas Winiwarter, Katharina Anders, and Bernhard Höfle. “M3C2-EP: Pushing the limits of 3D topographic point cloud change detection by error propagation”. en. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 178 (Aug. 2021), pp. 240–258. ISSN: 09242716. DOI: 10.1016/j.isprsjprs.2021.06.011.
- [29] Samira Nourbakhshbeidokhti et al. “A Workflow to Estimate Topographic and Volumetric Changes and Errors in Channel Sedimentation after Disturbance”. en. In: *Remote Sensing* 11.5 (Mar. 2019), p. 586. ISSN: 2072-4292. DOI: 10.3390/rs11050586.
- [30] Raphaël Miazza, Ivan Pascal, and Christophe Ancey. “Automated grain sizing from uncrewed aerial vehicles imagery of a gravel-bed river: Benchmarking of three object-based methods”. In: *Earth Surface Processes and Landforms* 49.5 (2024), pp. 1503–1514. DOI: <https://doi.org/10.1002/esp.5782>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1002/esp.5782>.

- [31] Antoine Soloy et al. *Mapping the size and shape of coarse clasts using Mask R-CNN: spatial and temporal variability over six different study sites including sea shores, and lake and river banks*. Dec. 2021. DOI: 10.13140/RG.2.2.21866.18888.
- [32] Christophe Ancey, Patricio Bohorquez, and Eric Bardou. “Sediment Transport in Mountain Rivers”. In: *Ercoftac Bulletin* 100 (Sept. 2014), pp. 37–52.
- [33] Eric Bardou and Eric Travaglini. “Rivières et flux sédimentaires dans les Alpes : adaptation de l’échelle spatiotemporelle d’analyse basée sur l’exemple de la Navisence, sud-ouest des Alpes suisses”. In: (Oct. 2015).
- [34] Eric Travaglini et al. “Analysis of Sediment Transport from Recorded Signals of Sediments in a Gravel-Bed River: Role of Sediment Availability”. In: (Sept. 2014). DOI: 10.1007/978-3-319-09054-2_98.
- [35] A. S. Woodget et al. “Quantifying submerged fluvial topography using hyperspatial resolution UAS imagery and structure from motion photogrammetry”. en. In: *Earth Surface Processes and Landforms* 40.1 (2015), pp. 47–64. ISSN: 1096-9837. DOI: 10.1002/esp.3613.
- [36] Patrice E. Carbonneau et al. “Adopting deep learning methods for airborne RGB fluvial scene classification”. In: *Remote Sensing of Environment* 251 (Dec. 2020), p. 112107. ISSN: 0034-4257. DOI: 10.1016/j.rse.2020.112107.
- [37] Dimitri Lague, Nicolas Brodu, and Jérôme Leroux. “Accurate 3D comparison of complex topography with terrestrial laser scanner: Application to the Rangitikei canyon (N-Z)”. en. In: *ISPRS Journal of Photogrammetry and Remote Sensing* 82 (Aug. 2013), pp. 10–26. ISSN: 09242716. DOI: 10.1016/j.isprsjprs.2013.04.009.
- [38] Mark Pauly, Richard Keiser, and Markus Gross. “Multi-scale Feature Extraction on Point-Sampled Surfaces”. In: *Computer Graphics Forum* 22.3 (2003), pp. 281–289. DOI: <https://doi.org/10.1111/1467-8659.00675>. eprint: <https://onlinelibrary.wiley.com/doi/pdf/10.1111/1467-8659.00675>.
- [39] M. Weinmann et al. “GEOMETRIC FEATURES AND THEIR RELEVANCE FOR 3D POINT CLOUD CLASSIFICATION”. en. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* IV-1/W1 (May 2017), pp. 157–164. ISSN: 2194-9050. DOI: 10.5194/isprs-annals-IV-1-W1-157-2017.
- [40] N. Vandapel et al. “Natural terrain classification using 3-d ladar data”. In: *IEEE International Conference on Robotics and Automation, 2004. Proceedings. ICRA '04. 2004*. Vol. 5. 2004, pp. 5117–5122. DOI: 10.1109/ROBOT.2004.1302529.
- [41] Timo Hackel, Jan D. Wegner, and Konrad Schindler. “Contour Detection in Unstructured 3D Point Clouds”. In: (2016), pp. 1610–1618. DOI: 10.1109/CVPR.2016.178.
- [42] M. Weinmann, B. Jutzi, and C. Mallet. “Feature relevance assessment for the semantic interpretation of 3D point cloud data”. In: *ISPRS Annals of the Photogrammetry, Remote Sensing and Spatial Information Sciences* II-5/W2 (2013), pp. 313–318. DOI: 10.5194/isprsannals-II-5-W2-313-2013.
- [43] M. Cavalli and L. Marchi. “Characterisation of the surface morphology of an alpine alluvial fan using airborne LiDAR”. In: *Natural Hazards and Earth System Sciences* 8.2 (2008), pp. 323–333. DOI: 10.5194/nhess-8-323-2008.
- [44] Johannes Stangl, Eric Rascher, and Oliver Sass. “Comparative analysis of sediment routing in two different alpine catchments”. In: *Source-to-Sink Fluxes in Undisturbed Cold Environments*. Ed. by Achim A. Beylich, John C. Dixon, and Zbigniew Editors Zwoliński. Cambridge University Press, 2016, pp. 364–377.
- [45] Tobias Ullmann and Georg Stauch. “Surface Roughness Estimation in the Orog Nuur Basin (Southern Mongolia) Using Sentinel-1 SAR Time Series and Ground-Based Photogrammetry”. In: *Remote Sensing* 12.19 (2020). ISSN: 2072-4292. DOI: 10.3390/rs12193200.
- [46] Thani Jintasuttisak et al. “Accurate Segmentation of Vegetation in UAV Desert Imagery Using HSV-GLCM Features and SVM Classification”. In: *Journal of Imaging* 12.1 (2026). ISSN: 2313-433X. DOI: 10.3390/jimaging12010009.

- [47] Tianqi Chen and Carlos Guestrin. “XGBoost: A Scalable Tree Boosting System”. In: *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*. arXiv:1603.02754 [cs.LG]. Aug. 2016, pp. 785–794. DOI: 10.1145/2939672.2939785.
- [48] Guolin Ke et al. “LightGBM: A Highly Efficient Gradient Boosting Decision Tree”. In: *Advances in Neural Information Processing Systems*. Vol. 30. Curran Associates, Inc., 2017.
- [49] Yunna Indah et al. “Comparison of Random Forest, XGBoost, and LightGBM Methods for the Human Development Index Classification”. In: *Jambura Journal of Mathematics* 7 (Jan. 2025), pp. 14–18. DOI: 10.37905/jjom.v7i1.28290.

A Appendix

A.1 Reproducibility tests -Installation check

Table 8: Reproducibility of paper’s results. Model: OpenCV with base settings: 150 decision trees, max node depth 25. FP = Full predictor, O = Optimized. The number of predictors does not match because the parameter file was not provided by the authors of 3DMASC [24]

	Model	Feature set	OA (%)	Nb predictors	Mean confidence
Paper	5 classes	FP	97.9	371	0.93
		O	97.6	18	0.93
	11 classes	FP	94.6	352	0.9
		O	94.0	37	0.9
Results	5 classes	FP	97.6	2266	0.93
		O	97.8	310	0.92
	11 classes	FP	95.8	2266	0.81
		O	95	310	0.78

A.2 Intermediate results: scales and Large Woody Debris

Table 9: Comparison between two ranges of scales (6.5m and 9.5m) and two training datasets having different proportions of the Large woody debris class (LWD). The selected XGBoost model is obtained from preliminary results (see subsection 4.1

Model	Scales	LWD class	OA (%)	LWD precision	LWD recall
XGBoost	0.5:0.5:9.5	Equal	85.6	97.4	71
		Wood 70%	85.34	97	69.2
	0.5:0.5:6.5	Equal	84.1	97.4	68.94
		Wood 70%	85.5	96.1	71.5

A.3 Hyperparameters for each model

Table 10: Hyperparameter grid search ranges for each machine learning model

Hyperparameter	XGBoost	LightGBM	Random Forest	OpenCV RTrees
<i>Tree structure parameters</i>				
max_depth	[4, 6]	[-1]	[None, 20]	[10, 20]
num_leaves	—	[31, 63]	—	—
<i>Ensemble parameters</i>				
n_estimators	[800]	[600, 800]	[300, 600]	—
n_trees	—	—	—	[150, 300]
<i>Learning / splitting parameters</i>				
learning_rate (η)	[0.05, 0.1]	[0.05, 0.1]	—	—
subsample	[0.8, 1.0]	[0.8, 1.0]	—	—
colsample_bytree	[0.8, 1.0]	[0.8, 1.0]	—	—
min_child_weight	[1.0]	—	—	—
min_samples_split	—	—	[2, 5]	—
min_samples_leaf	—	—	[1, 2]	—
min_sample_count	—	—	—	[2, 5]
<i>Regularization parameters</i>				
reg_lambda (L_2 , fixed)	1.0	1.0	—	—
reg_alpha (L_1 , fixed)	0.0	—	—	—
<i>Feature parameters</i>				
max_features	—	—	["sqrt"]	—
max_categories	—	—	—	[15]
n_active_vars	—	—	—	[0]

A.4 RGB Point Clouds ROI

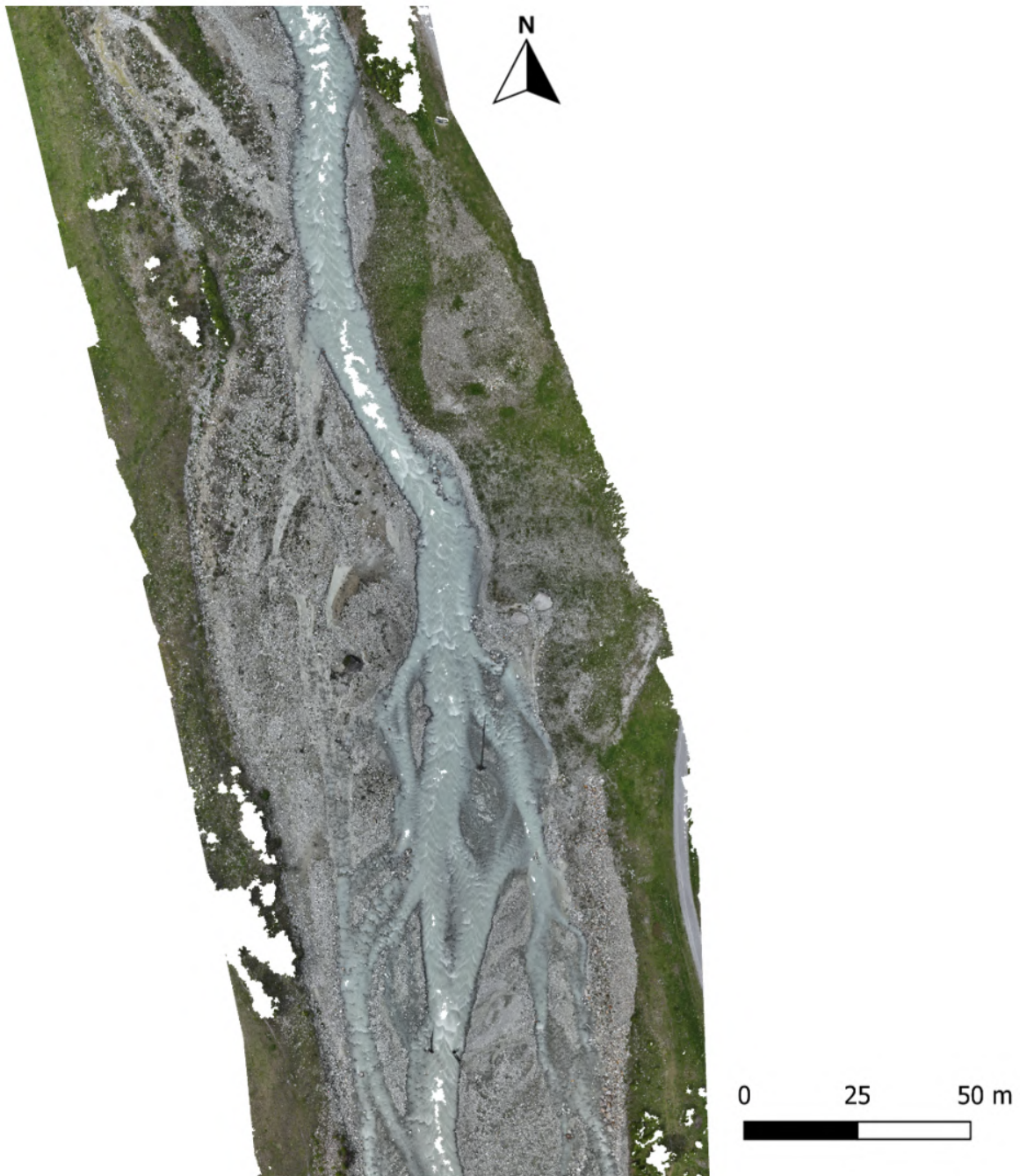


Figure 24: Original RGB point cloud for the ROI acquired on June 2024. Reach = 270m

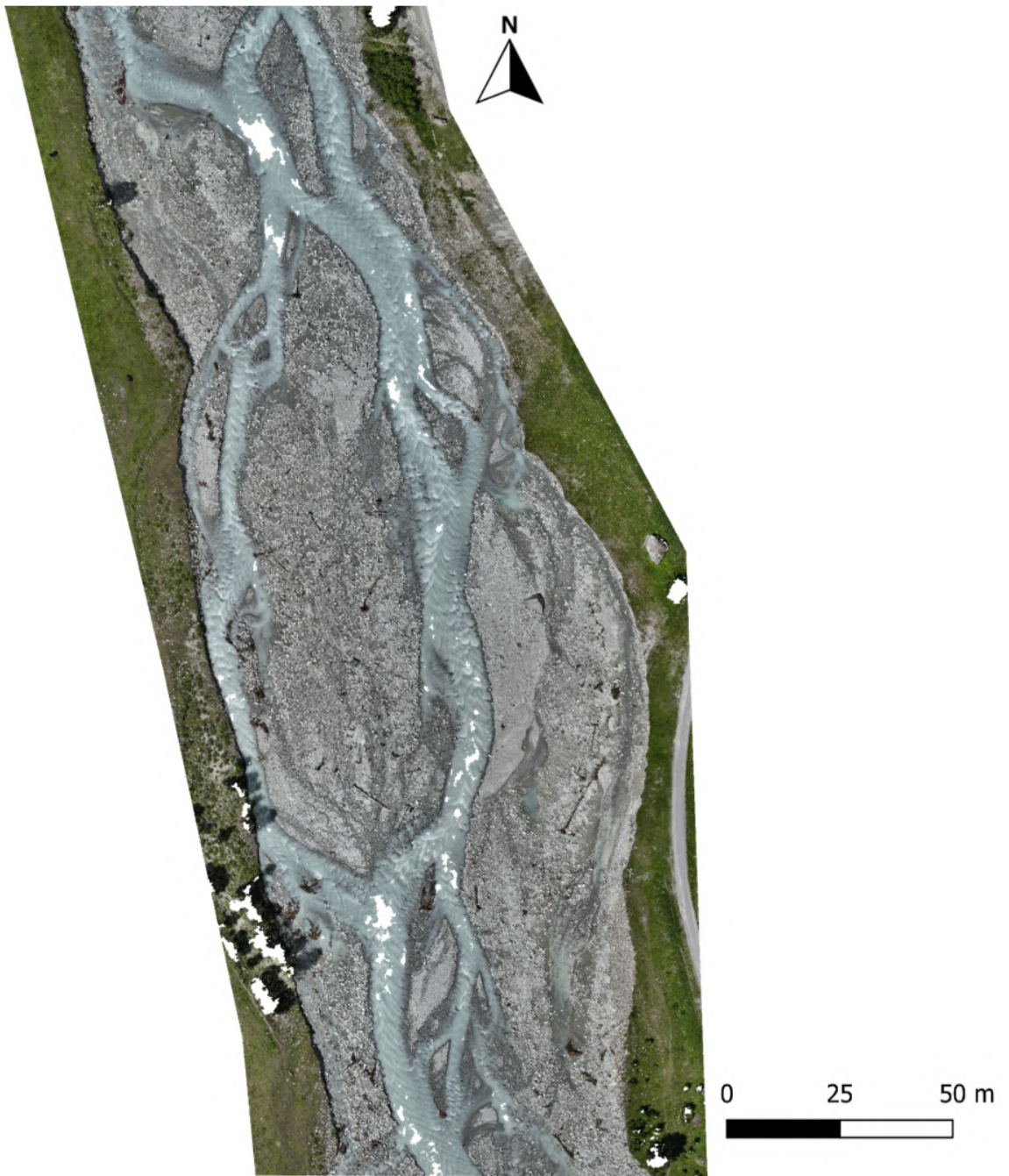


Figure 25: Original RGB point cloud for the ROI acquired on July 2024. Reach = 270m

PROTOCOL FOR CLASSIFICATION FRAMEWORK

PREREQUISITES

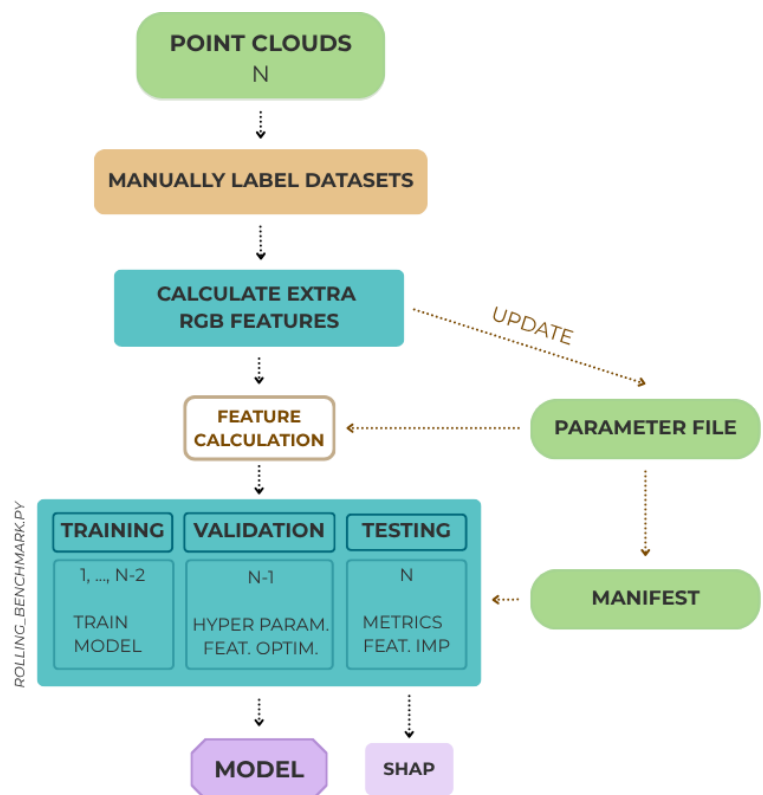
Executing this framework requires the following software and data inputs:

- Multiple point clouds (PCs) capturing the same region across different dates ($N \geq 3$ datasets).
- A working Python environment with all required dependencies installed.
- CloudCompare installed on the computer.
- The lidar_platform repository downloaded from GitHub: [p-leroy/lidar_platform](https://github.com/p-leroy/lidar_platform)

Note that this pipeline is currently not optimized for excessively large datasets and may fail on heavy point clouds.

DATA PREPARATION AND LABELING

1. **Alignment and Base Features:** Align the datasets spatially. Within CloudCompare, extract raw color channels by navigating to Edit > Colors > Convert to Scalar Field > R, G, B channels.
2. **Spectral Indexing:** Generate the required illumination-invariant spectral features by running the `add_scalar_fields_invariant.py` script.
3. **Data Export:** Save the processed PC in LAZ format. It is critical that the scalar fields (SFs) are ordered correctly, starting with index 0. Ensure this by configuring the LAS Saving Options menu and arranging the fields under Extra fields (Extra Bytes VLRs).
4. **Segmentation and Labeling:** Use CloudCompare's segmentation tool to isolate patches of each geomorphic class. Add a new SF named "Classification" to these segments and assign the corresponding integer class value.
5. **Class Balancing:** Merge all spatial segments belonging to the same class, then apply spatial subsampling to isolate 2,000 points per class.
6. **Final Merge:** Merge all class clouds together into a single labeled dataset.



Order of spectral features in `parameter_file.txt`:

	SF0	SF1	SF2	SF3	SF4	SF5	SF...
Features	R	G	B	NORM_R	NORM_G	NORM_B	Additional feats

7. **Parameter Configuration:** Create a *parameter_file.txt* following 3DMASC documentation (see provided example). Define the point clouds (PC1, PCX) and spatial scales using the format *first:step:last*. Append the custom RGB descriptors to the feature list using the syntax feature: SF#_....

PC1 = Point cloud that characterizes the scene of interest, used to compute the features on PCX.

PCX = Point cloud that is classified. It has to be within the extent of PC1. It can be a subsample of PC1.

8. **Manifest Creation:** Construct a *manifest.yaml* file to serve as the directory map for the benchmark. This file must contain the absolute paths to the CloudCompare executable, the *parameter_file.txt*, and both the original and labeled point clouds, ensuring both point clouds have the correctly ordered SFs.
9. **Command Line Execution:** Execute the framework via the terminal by passing the manifest and runtime arguments to the main script.

A standard execution call is structured as follows:

```
python rolling_benchmark.py --manifest "../manifests/manifest_Zinal.yaml" --category plats --models
xgboost,lightgbm --params-key RGB_invariant --prefer-cached --save-models
```

The *rolling_benchmark.py* script outputs:

- Input point clouds with every calculated feature as a SF ("_WITH_FEATURES_category_#feats.sbf)
- Model(s) of the last round of training.
- Global and per-class feature ranking and evaluation metrics (+ best model hyperparameters) for each round.

10. **ROI Classification:** The model can be used to classify any PC. The repository contains an example of PC classification with the XGBoost model in *run_XGBoost_classification.py*

OPTIONAL: SHAP ANALYSIS

11. **SHAP execution:** The repository contains a Jupyter notebook *round_shap_analysis.ipynb* that allows to perform SHAP analysis on a given round of the Classification Framework.

More information : [fmalvido/PDM-CREALP: 3DMASC-inspired ML algorithms to classify river geomorphology](https://github.com/fmalvido/PDM-CREALP)

```
#clouds
cloud: PC1=PC1.sbf
cloud: PCX=PCX.sbf

#core points
core_points: PCX

scales: 0.5:0.5:9.5

#Z-based features (slope-robust)
feature: Z_SCx_STD_PC1
feature: Z_SCx_SKEW_PC1
feature: Zmin_SCx_PC1
feature: Zmax_SCx_PC1
feature: ZRange_SCx_PC1

#Geometric features (slope-invariant)
feature: DIP_SCx_PC1
feature: ROUGH_SCx_PC1
feature: PCA3_SCx_PC1
feature: PCA2_SCx_PC1
feature: PCA1_SCx_PC1
feature: ANISO_SCx_PC1
feature: SPHER_SCx_PC1
feature: LINEA_SCx_PC1
feature: PLANA_SCx_PC1
feature: CURV_SCx_PC1

#Custom indices (computed at each scale)
feature: SF3_SCx_MEAN_PC1
feature: SF4_SCx_MEAN_PC1
feature: SF5_SCx_MEAN_PC1
feature: SF6_SCx_MEAN_PC1
feature: SF7_SCx_MEAN_PC1
feature: SF8_SCx_MEAN_PC1
feature: SF9_SCx_MEAN_PC1
feature: SF10_SCx_MEAN_PC1
```