Supplementary Information

From point cloud to digital elevation model: Airborne topo-bathymetric LiDAR processing over the 28 km Ardèche River Gorges

Ron Nativ[1*], Dimitri Lague[1,2], Paul Leroy[2], Laure Guerit[1], Thomas Bernard[1,3], Mathilde Letard[4], Daniel Girardeau-Montaut[5], Vincent Godard[6], Rodolphe Cattin[7], Olivier Payrastre[8] and Philippe Steer[1]

[1]Univ Rennes, CNRS, Géosciences Rennes, UMR 6118, 35000, Rennes, France
[2]Univ Rennes, CNRS, Plateforme LiDAR topo-bathymétrique, OSERen, UAR 3343, 35000, Rennes, France
[3]Università di Trento, Italy
[4]Technical University of Munich, TUM School of Engineering and Design, Department of Aerospace and Geodesy, 80333, München, Germany
[5]CloudCompare
[6]Aix Marseille Univ, CNRS, IRD, INRAE, CEREGE, Aix-en-Provence, France
[7] Université de Montpellier, CNRS, Géosciences Montpellier, UMR 5243, , Montpellier, France
[8] Univ Gustave Eiffel, GERS-LEE, F-44344 Bouguenais, France

* correspondence: ron.nativ@univ-rennes.fr

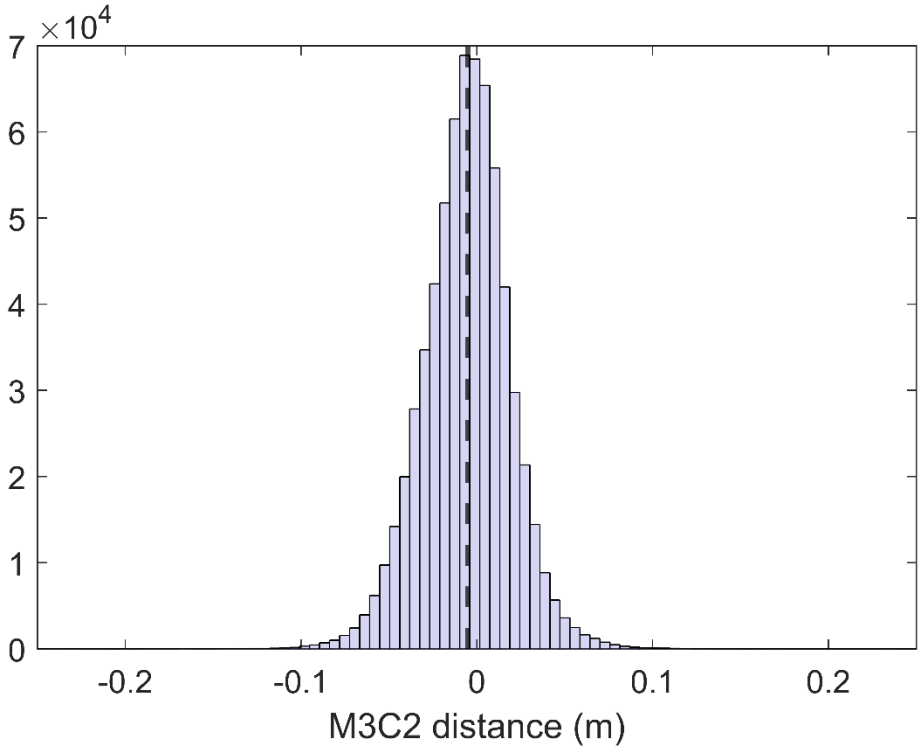

Figure S1: A distribution of M3C2 distances between the TB LiDAR and LiDAR HD datasets over stable features such as roads and buildings, located both above (in elevation) and along the Ardèche River's path. Following alignment and filtering to retain only morphologically stable surfaces, the differences between the two datasets exhibit an approximately Gaussian distribution, with a mean ± standard deviation of -0.51 ± 8.11 cm. The analysis is based on a total of 671,798 points.

Dominant feature analysis for 3DMASC classification of 532 nm bathymetry versus volume echoes

This approach differs from our initial classification, which used the full set of 3DMASC features (Table S1). Here, we take advantage of the manually cleaned volume and bed echoes in PC_{532} to identify the smallest set of features and spatial scales that best distinguish the two classes (bathymetry versus volume echoes). The goal of this analysis is to guide future users who wish to perform the same type of classification on other fluvial datasets.

Training and validation sets were constructed from spatially random samples of classified 532-nm LiDAR points (bathymetry and volume echoes). We selected 8,000 points and intermittently segmented the gorge into sections a few kilometers long. Training data were compiled from non-contiguous reaches, and the validation set was assembled in the same way, ensuring no spatial overlap between the training and unseen validation data. We extended the range of spatial scales up to 10 m and applied the embedded backward feature selection approach described by Letard et al. (2024). In practice, this consists of iteratively reducing the number of features based on feature importance calculated at each random forest training step, and stopping when the overall classification accuracy begins to decline. We show the confusion matrix for this classification in Table S2. The procedure yielded a minimal set of nine features that optimally separates volume echoes from the channel bed (Table S3).

The overall accuracy reached by this simplified classifier is 0.93. Interestingly, using only this reduced set of 9 features with the initial training and validation dataset we used for the first classification results in an overall accuracy of 0.96.

Consistent with Letard et al. (2024), the distance to the water surface is a dominant feature, along with the third principal component (PCA3), the local slope of the 532 nm point cloud, and the mean number of returns within a spherical neighborhood. However, unlike Letard et al. (2024), features combining both the 532 nm and 1064 nm point clouds were not important. This result suggests that, for this task, a classification based solely on the 532 nm point cloud and a water surface reference may be sufficient. In our case, the water surface is derived from the 1,064 nm channel, but it could also be estimated from the 532 nm channel (Höfle et al., 2009).

In contrast to Letard et al. (2024), features describing the relative elevation of the core point with respect to the minimum (Z_{min}) or maximum (Z_{max}) elevation within a spherical neighborhood were

important. The dominant spatial scales were 2.5, 6, 8, and 10 m. The discrepancy with the results of Letard et al. (2024) is not unexpected, as the class diversity in their analysis was much higher (5 or 11/14 classes) than in this study (2 classes).

Table S1: Confusion matrix for the initial 3DMASC bathymetry classification (section 3.3.2), with overall accuracy of 95.0%.

	Predicted as volume echoes	**Predicted as bathymetry**	**Precision**	**Recall**	**F1-score**
Volume Echoes	**98.9%**	**1.1%**	**0.92**	**0.99**	**0.95**
Bathymetry	**8.9%**	**91.1%**	**0.99**	**0.91**	**0.95**

Table S2: Confusion matrix for the second, minimized 3DMASC bathymetry classification (section 3.3.2) with overall accuracy of 92.7%.

	Predicted as volume echoes	**Predicted as bathymetry**	**Precision**	**Recall**	**F1-score**
Volume Echoes	**91.7%**	**8.3%**	**0.93**	**0.92**	**0.92**
Bathymetry	**6.4%**	**93.6%**	**0.92**	**0.94**	**0.93**

Table S3: Final 3DMASC classifier features ranked by decreasing feature importance. In the following, CTX is the reference water surface (context cloud), C3HD is the 532 nm bathymetric point cloud. All features but the first one are evaluated within a spherical neighborhood of diameter "Scale". More details on features in Letard et al. (2024)

	Feature name	Importance	Cloud	Scale (m)	Explanation
1	DZ4_SC0_CTX_0	0.159495	Water Surface	-	Vertical distance to the reference water surface
2	Zmax_SC10_C3HD	0.125973	C3HD	10	Vertical distance to the point of maximum elevation within the spherical neighborhood
3	Zmin_SC8_C3HD	0.121404	C3HD	8	Vertical distance to the point of minimum elevation within the spherical neighborhood
4	Zmin_SC10_C3HD	0.115741	C3HD	10	Vertical distance to the point of minimum elevation within the spherical neighborhood
5	Zmin_SC6_C3HD	0.101133	C3HD	6	Vertical distance to the point of minimum elevation within the spherical neighborhood
6	Dip_SC8_C3HD	0.100607	C3HD	8	Local slope of the best fit plane within the spherical neighborhood
7	Zmax_SC8_C3HD	0.100449	C3HD	8	Vertical distance to the point of maximum elevation within the spherical neighborhood
8	PCA3_SC2.5_C3HD	0.088046	C3HD	2.5	3rd eigenvalue of the principal component analysis
9	NbRet_SC10_MEAN_C3HD	0.087151	C3HD	10	Mean number of returns