



MOX-Report No. 66/2020

**A comparison between machine learning and functional
geostatistics approaches for data-driven analyses of solid
transport in a pre-Alpine stream**

Didkovsky, O.; Ivanov, V.; Papini, M.; Longoni, L.;
Menafoglio, A.

MOX, Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

mox-dmat@polimi.it

<http://mox.polimi.it>

A comparison between machine learning and functional geostatistics approaches for data-driven analyses of solid transport in a pre-Alpine stream

Oleksandr Didkovsky^{1,2}, Vladislav Ivanov³, Monica Papini³,
Laura Longoni³, Alessandra Menafoglio^{1*},

¹ MOX-Department of Mathematics, Politecnico di Milano, Milano, Italy

² Center for Analysis, Decisions and Society, Human Technopole, Italy

³ Department of Civil and Environmental Engineering
Politecnico di Milano, Milano, Italy

* alessandra.menafoglio@polimi.it

Abstract

We address the problem of providing data-driven models for sediment transport in a pre-Alpine stream in Italy. We base our study on a large set of measurements collected from real pebbles, traced along the stream through Radio Frequency IDentificator (RFID) tags after precipitation events. We propose and evaluate two classes of data-driven models – based on machine learning and functional geostatistics approaches respectively – to predict the probability of movement of single pebbles within the stream. The first class is built upon gradient boosting decision trees, and allows one to estimate the probability of movement of a pebble, based on the pebbles' geometrical features, river flow rate, locations, and subdomain types. The second class is built upon functional kriging, a recent geostatistical technique which allows one to predict a functional profile –i.e., the movement probability of a pebble, as a function of the pebbles geometrical features or of the stream's flow rate– at unsampled locations in the study area. Although grounded on different perspectives, both these models aim to account for two main sources of uncertainty, namely (i) the complexity of river's morphological structure, and (ii) the highly-nonlinear dependence between probability of movement, pebble's size and shape, and the stream's flow rate. We extensively compare the performances of the two methods in terms of classification accuracy, and show that, although these techniques are grounded on different perspectives, an overall consistency appears between the methods suggesting that both approaches may provide valuable modeling frameworks for the problem at hand. We finally discuss on the use of the developed models in a bottom-up strategy, which starts with the prediction/classification of a single pebble and then integrates the results into a forecast of the grain-size distribution of mobilized sediments.

1 Introduction

Bedload transport has been recognized as a phenomenon that has a significant role in a range of applications with non-negligible environmental and societal impacts, including agriculture (Haddadchi et al. 2014), reservoir siltation (de Miranda and Mauad (2015), Longoni et al. (2016b)), urban planning (Dotterweich 2008), riverine species' habitat (Wharton et al. 2017), river-structure interactions (Pizarro et al. 2020), and flood risk management (Radice et al. (2016), Mazzorana et al. (2013)). Bedload transport studies have demonstrated that the dynamics of the process are largely dependent on the hydraulic parameters of the stream (e.g. Hassan and Bradley (2017), Vázquez-Tarrío et al. (2019)), while the effects of sediment transport are particularly prominent in mountain streams due to the abundant availability of sediment material and the swift time of concentration leading to significant sediment mobility even for events of short duration, such as several tens of hours (Sear et al. (1995), Stover and Montgomery (2001), Lane et al. (2007), Longoni et al. (2016a)).

Individual pebble tracing has been outlined as an innovative method which allows for the collection of bedload transport field data, which could provide insights in the dynamics of the process at a variety of temporal and spatial scales. Radio Frequency Identification (RFID) transponders (a.k.a. Passive Integrated Transponders or PIT tags) have been used as sediment tracers and deployed in field and flume experiments to understand particle transport. Both active and passive tracers have been used by a number of authors for pebble tracking (e.g. Cassel et al. (2017)). While the former feature higher detection ranges and thus a lower loss rate, the latter are significantly less expensive and thus a larger sample of tracer-equipped pebbles could be created. Recent reviews on passive tag pebble tracking can be found in (Hassan and Bradley 2017), (Vázquez-Tarrío et al. 2019) (Ivanov et al. 2020a). The possibility to monitor the sample with a desired frequency permits the correlation of quantities such as the pebbles' mobility, displacement, and velocity to river discharge and meteorological events' parameters.

Pebble tracing data is generally processed so as to analyze trends in travelled distances, virtual velocity, and proportion of mobile pebbles. Those control parameters are to be put in relation to variables considered key drivers of sediment transport such as river discharge, as well as predisposing factors such as pebbles' dimensions or, less commonly, the local morphological conditions (e.g. Ferguson et al. (2017), Vázquez-Tarrío et al. (2019), Cain and MacVicar (2020), Ivanov et al. (2020a)). The proportion of mobile pebbles within the period of observation provides an indication on the mobilizing capacity of the stream during a given event. This parameter was analysed in the work of Papangelakis and Hassan (2016), who established a linearly increasing trend with respect to the total excess flow energy expenditure over an entire season with a quite good fit ($R^2 = 0.78$ and $R^2 = 0.72$, respectively for two investigated reaches), while its relation to the peak flow discharge demonstrated a weaker relationship. No dependency was established between the proportion of mobile pebbles and their size. Further, Ferguson et al. (2017) report a weakly increasing trend of pebble mobility with increasing peak flow rate, observed at the event-scale for 6 events. Instead, Ivanov et al. (2020a) did not find any correlation between dimensionless peak flow rate and the ratio of mobile particles over a dataset including 18 event observations. This difference in results from studies carried out at different timescales highlights the intermittency of the process as well as the multifaceted nature of sediment mobilization, where factors such as sediment size and morphology hinder a clear trend to be established at the event-scale level. The discrepancies between results obtained by different authors suggest that the dynamics of the process can vary significantly when pebble tracing data are analyzed at the event scale, while general trends of bedload behaviour become more obvious when seasonal or yearly data are considered. It is likely that the multifaceted nature of pebble mobility

renders it hard to describe with a simple regression, typically used to relate pebble tracking data to control variables. More complex nonlinear models could therefore be able to incorporate the variety of factors affecting the mobility of pebbles at the event scale.

Advanced analytical approaches which may allow to model the complex phenomena taking place in sediment transport may be grouped at least in two classes (i) purely physics-based approaches, or (ii) highly non-linear, data-driven approaches. In the former case, systems of partial-differential equations (PDE) are used to model the dynamics of the flow, and consistently assess the sediment transport (see, e.g. [Vetsch et al. \(2017\)](#)). In this case, field data can be used to calibrate the PDE, both in terms of providing sensible input parameters (e.g. [Bakke et al. \(2017\)](#), [Gatti et al. \(2020\)](#)), or to validate the model outputs (e.g. [Brambilla et al. \(2020\)](#)). Critical points of this class of methods typically lie in the numerical complexity of solving the PDEs, in the data assimilation process, and in the uncertainty quantification of the model, which often require the development of *ad-hoc* techniques. In this work, the focus shall be on the latter approach instead. Data-driven methods can be used to build *empirical* models for the sediment transport, in which data are used directly to infer on the connection between the sediment transport and the stream/bedload characteristics – without relying on the physical laws governing the system. Data-driven models have the advantage of being typically characterized by a lower number of input parameters to be calibrated, beside being often naturally suitable to effectively perform uncertainty quantification (e.g., via resampling methods, [Friedman et al. \(2001\)](#)).

Amongst the data-driven approaches available in the literature, we here consider the comparison of two viewpoints to the problem of predicting the probability of pebbles’ movement, namely (a) a machine learning approach based on boosting methods, and (b) a functional geostatistics framework. In the first case, a model for the probability of pebbles’ movement is built based on decision trees applied iteratively, in the framework of gradient boosting decision trees (see, e.g., [Friedman et al. \(2001\)](#)). Note that the iterative construction of the trees precisely allows one to build a highly non-linear model for the relation between the probability of movement of single pebbles and the characteristics of the pebbles themselves (e.g., shape, size), and of the stream (e.g., flow, geomorphology). In the second case, a functional data analysis (FDA, [Ramsay and Silverman \(2005\)](#)) approach is used to reconstruct the non-linear functional relation between the probability of pebbles’ movement and their characteristics (i.e., shape, size). These functional forms –which can be estimated only locally– are then predicted at unsampled locations along the river, by relying on the theory of object-oriented spatial statistics (O2S2, [Menafoglio and Secchi \(2017\)](#)) which provides a methodological framework to analyze functional observations distributed in space (e.g., via kriging). These two different viewpoints to the problem shall be here compared in terms of actual error in validation analyses (both in a cross-validation setting and on an independent dataset), and the results interpreted from the geomorphological standpoint, highlighting the strengths and limitations of both approaches.

The remaining of this work is organized as follows. Section [2](#) presents the available the study area, and the available data in terms of pebbles characteristics and position, stream flow and river geomorphology. Section [2.2](#) describes a preliminary analysis of the dataset, to highlight its key features and introduce the concept of *typical rainfall event*, which will be instrumental to the application of the data-driven approaches considered in this work. These will be introduced in Section [3](#), and applied to the data in Section [4](#). Section [5](#) discusses the application of the proposed approaches on an independent dataset, collecting pre- and post-event granulometric distributions at a number of sites in the study region, highlighting the critical points of this process. Finally, Section [6](#) will provide a discussion and draw conclusions of the work. All the data analyses here presented were performed using the software R ([R Core Team 2020](#)); source codes to reproduce the analyses are freely available

at the link: github.com/alexdidkovskyi/YP_Paper.

2 The Field Case

2.1 Data Description

The investigation carried out in this work is based on field data collected in the hydrographic basin of the Caldene river. This area was subject to an extensive study in recent years (Ivanov et al. (2016a), Ivanov et al. (2017), Papini et al. (2017), Ivanov et al. (2020a), Ivanov et al. (2020b), Gatti et al. (2020)) aimed to assess the hydrogeological instability and hazard within the region. The data available for the present study come from four main sources, which are illustrated in greater details below: geomorphological characteristics of the domain, sediment information (pebbles' size, dimension, etc), pebbles' locations, and river flow information. All these sources of information were independently measured. Data on domain characteristics and pebbles' dimensions are the only static information, while the other sources are dynamic, and strictly related with the sediment transport phenomenon.

The domain The hydrographic basin of the Caldene river (Fig. 1) covers an area of 28 km^2 and collects an average yearly rainfall depth of around 1400 mm . The main stream is 11 km long and outlets into the Como Lake after its passage through the town of Lecco. As in most pre-Alpine environments, active geomorphic processes include colluvial and fluvial transport responsible for the yield and further propagation of sediment downstream (Ivanov et al. 2016b). The steep slopes characterizing the stream and the limited time of concentration promote the rapid development of flood waves that are capable of transporting large amounts of sediment. The gradient of the river varies in the range $10 - 40 \%$ in the upstream portion of the basin, and $1 - 5.5 \%$ in its lower part. The channel width is typically less than 10 m . The sediment grain size distribution extends from fine sand to boulders of metric dimensions. The discharge at the downstream end of the basin ranges from $0.2 \text{ m}^3/\text{s}$ in normal conditions up to peak values of more than $100 \text{ m}^3/\text{s}$. The river reach that is the on focus in this work (henceforth referred to as the *domain*), extends for about 1 km from the confluence of the Caldene with its main tributary (Fig. 1).

From a geomorphological point of view, the domain is characterized by several *subdomains*. Morphological units identified in the reach are the following. A *cascade* zone characterized by a swift and shallow tumbling flow, disturbed by the presence of coarse sediment. Downstream, the channel transitions into a *step-and-pool* zone characterized by longitudinal steps composed of large clasts that separate the consecutive pools which contain finer grain sediment. The stream in this zone alternates from swift over the steps to slow within the pools. Finally, the monitored domain ends with a *plane-bed* area which is a flat relatively featureless bed with a lower gradient that allows for the undisturbed flow of the stream. The domain is laterally confined by a *bank* zone that is often vegetated and the stream flow here is rather slow with respect to the centre of the channel. Along the entire monitored reach, there is the formation of longitudinal and side *bars* which effectively represent sediment build-up zones. Those zones typically act both as source and deposition zone during moderate and high flow events. The reach is further characterized by the presence of large boulders of metric dimensions. The morphological units typically have a compound nature and they consist of a set of disjoint morphological sectors. These morphological sectors are depicted in Figure 1 as individual polygons.

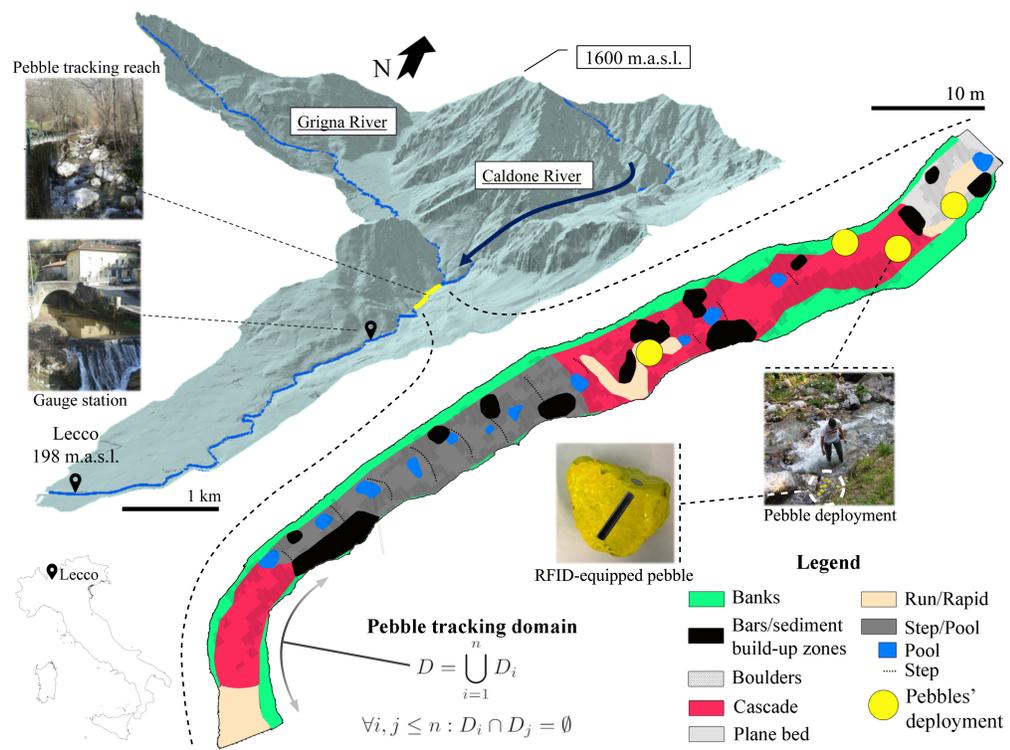


Figure (1) The hydrographic catchment of the Caldone river. Monitored domain and subdomains.

Indicator	Mean	Sd
a-axis	108.57	25.08
b-axis	79.40	17.24
c-axis	53.46	12.49
Nominal diameter	76.60	14.34
Weight	726.66	404.73
Elongation	0.75	0.13
Platyness	0.51	0.12
Sphericity	0.72	0.09

Table (1) Pebbles' primary and secondary indicators: summary statistics



Figure (2) Typical pebbles included in the dataset.

Sediment information Although, in general, a complete characterization of the shape of a pebble may require a complex representation, in our study this is summarized by its primary and secondary indicators. The primary indicators are the 3 main dimensions of the pebble (in cm), and its weight. These dimensions are computed as the length of the pebble along its three main axes, named a-, b-, and c-axis, these lengths being in a decreasing order. The secondary indicators are derived from the primary ones; they are elongation (b/a), platyness (c/a), sphericity ($((\frac{c^2}{ab})^{1/3})$) and nominal diameter ($((abc)^{1/3})$). Typically, these indicators are correlated; for instance, the weight is strongly correlated with the nominal diameter. Thus, summaries, or only part of the indicators can be used for a more efficient characterization of the shape and dimension of the pebbles (see Section 2.2). An illustrative example of typical pebbles belonging to the study is reported in Figure 2. Table I reports the mean and standard deviation of the primary and secondary indicators for the set of 664 pebbles considered for this study.

Pebbles scattering The limited cost of the RFID tags allowed for their insertion in the 664 pebbles considered in this study. Before deployment into the river, the pebbles were drilled, each one

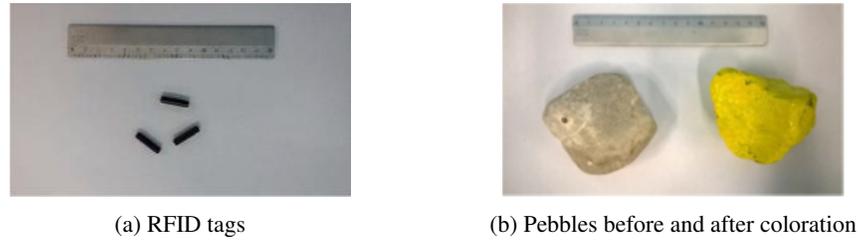


Figure (3) Experimental setup: inclusion of RFID and coloring of pebbles

was equipped with an RFID tag, and they were finally painted in a bright yellow color for visual aid as illustrated in Fig. 3. The weight and dimensions (a-, b-, and c-axis) of each pebble were recorded and associated to the respective RFID unicode. The deployment in the river (Fig. 1) was carried out in several tranches and the movement of the pebbles was monitored with a portable antenna after each significant rainfall event along the period 06/2016 - 09/2018. The successive position of the pebbles was recorded on a photorealistic model of the reach (Fig. 1). The unicode contained in each transmitter allowed each pebble to be attributed with a position before and after a flood event. Detailed explanation of the experimental procedure can be found in Papini et al. (2017).

River flow data During the period 2016-2018, the pebble samples were surveyed after 28 precipitation events. The river flow rate just downstream of the monitored reach was recorded during the event. Values of river discharge are obtained from the observation of water depth at the gauge station with a hourly resolution (Fig. 4) and the respective depth-discharge curve (detailed information can be found in Ivanov et al. (2020a), Ivanov et al. (2020b)). In this analysis, peak values of water depth and river flow rate were considered for each event. Ivanov et al. (2020a) identified a dimensionless discharge threshold for sediment mobility on the basis of a subset of the data presented in this work. The threshold value was then translated into water depth. This allowed for the definition of event duration, taken as the duration over which the water depth, and respective river discharge, remained above the threshold value. According to the type of event, its duration can range from a hour (the highest observation frequency) to as long as several days. Mobilizing events could be outlined as two general types – high peak discharge and short duration, and events with a limited intensity but a longer duration. Since the water depth threshold is empirically estimated, a sensitivity analysis was carried out on the threshold definition as it directly impacts the duration of each event – the higher the threshold, the shorter the event’s duration and vice versa. A record of rainfall depth is available for the period of observation. However this datum is collected at the downstream end of the hydrographic basin – at a considerable distance away from the monitored reach (Fig. 1).

2.2 Data exploration and pre-processing

To construct the dataset for models’ training out of initial raw data, each data source is pre-processed separately. Data pre-processing consists of: (1) data selection and treatment of missing values, and (2) dimensionality reduction of pebbles’ and flow’s indicators. Step (1) aims at cleaning the dataset, concerning particularly the management of missing data due to the fact that not all the pebbles could be found after the rainfall events. In this regard, we observe that around 20% of the pebbles were

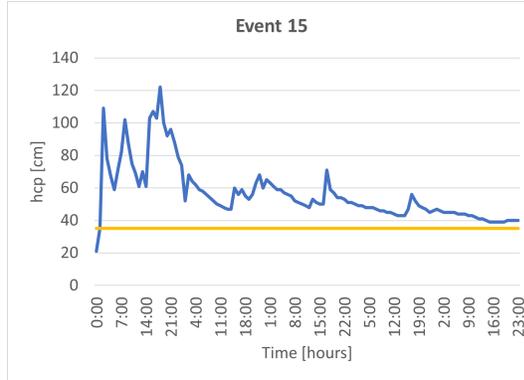


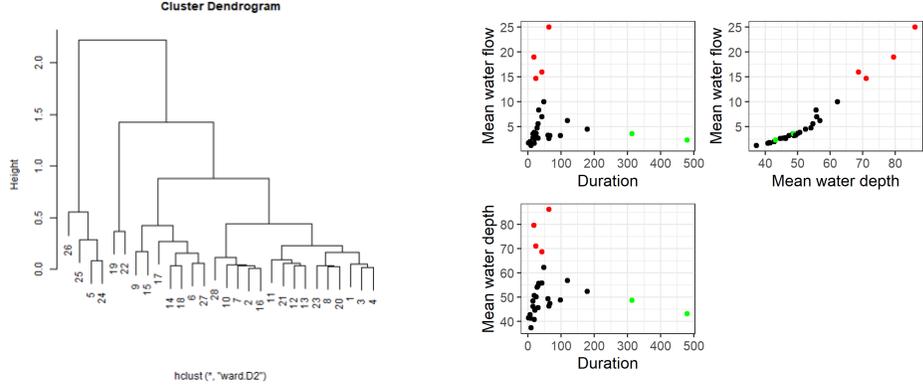
Figure (4) River depth during mobilizing event 15 (27/09/2017-30/09/2017)

registered during nonconsecutive events (i.e., 20% of the pebbles have at least one missing value). For instance, several pebbles were lost for three consecutive events and then found at their respective initial place. This could be due to a temporary increase in water depth (and consequently, in the distance between a pebble and the antenna during the survey). However, given that their position did not change, these pebbles can be assumed as still during all these events when they were lost. Treatment of missing data is thus performed through the following rules:

1. If a pebble is lost for $N \geq 1$ events and then found after the $(N + 1)$ -th event at the same location, it is considered as steady at the same place along all the $N + 1$ events (thus marked as *not-moved* at all the events).
2. If a pebble is lost for $N \geq 1$ events and then found at a different location, the partial information about this pebble is not used.
3. If a pebble is found upriver, we interpret this as a positioning error and the data point is removed.

Furthermore, in order to isolate erroneous data, we also use a simple heuristic allowing to identify observations with a potential positioning issue. Note that the domain is characterized by a slope from upstream to downstream, and, consequently, a downslope propagating river flow. Hence, assuming one directional flow, the expected pebble displacement is in the direction of the flow. We thus exclude all observations that are associated with an upstream movement and displacement larger than 1 m. This heuristic identified 65 observations with positioning issues out of 2200.

Step (2) (i.e., dimensionality reduction) was separately performed on pebbles' and flow's indicators. Focusing on pebbles dimensions, we consider the primary indicators (a-axis, b-axis, c-axis), and perform principal component analysis (PCA) to filter out the redundancy within this set of information. For the same purpose, secondary indicators are not considered further for the analysis, as strongly correlated with primary ones. The first PC (hereinafter *PC1*) is responsible for 77% of the variance of the data, while the second PC (hereinafter *PC2*) explains an additional 13% of variance. Interpretation of the loading of *PC1* ($e_1 = (0.84, 0.48, 0.24)^T$) suggests a strong association of *PC1* with overall pebbles' size (the higher the score, the larger the pebble). In turn, the *PC2* appears associated with the elongation of the pebble ($e_2 = (0.54, -0.71, -0.44)^T$) – the higher the score, the



(a) Rainfall Events Dendrogram. Numbers are IDs of mobilizing events

(b) Events' clusters

Figure (5) Rainfall Events Clustering

more elongated the pebble is. The weight of the pebbles appears to be strongly correlated with $PC1$ (correlation: $\rho = 0.87$), and is thus excluded from the predictors to avoid collinearity.

Concerning flow data, exploration of the dataset suggests the presence of three macro-groups of mobilizing events, namely typical events (T), short intense events (SI), and long mild events (LM). These clusters are clearly evidenced when applying hierarchical clustering, see, e.g., the results obtained with Euclidean distance and Ward linkage reported in Figure 5. Here, groups T, SI, LM are reported as black, green and red symbols, respectively. It is noticeable that the typical events (T) share a good degree of similarity in terms of river flow data, beside representing 20 out of 28 rainfall events (corresponding to 1594 out of the 1989 pebbles' observations).

Dimensionality reduction of the river flow data is based on PCA of the scaled values of (i) mobilizing event duration (hours), (ii) average river depth (cm), and (iii) average water flow ($m^3/hour$), when all the groups of events are considered together. These variables were scaled using min-max normalization, i.e., they were separately scaled to range in the $[0, 1]$ interval. The first PC of the flow data, named $PC1_{flow}$, explains 71% of variability and is interpreted as a contrast between duration and flow characteristic ($v_1 = (0.11, -0.69, -0.71)^T$) – high scores being associated with short intense events, low scores with long and less intense events. The second PC, $PC2_{flow}$, is responsible for an additional 28% of variance and it is strongly associated with duration ($v_2 = (0.99, 0.01, 0.01)^T$) – high scores being representative of higher durations.

In the following, only pre-processed data shall be considered for our analyses, each observation being built out of the following set of variables: PCs of pebble' dimensions, PCs of flow data (for each event), pebble' locations (after each event) and associated geomorphological domain. Pebbles' locations are used to compute their displacement after an event as the Euclidean distance between their positions before and after the event. The measured displacement d_{ij} of the i -th pebble after the j -th event is then used to classify it as *moved* ($d_{ij} > 0$) or *not-moved* ($d_{ij} = 0$).

3 Methods

In this section, we describe the two classes of methods that shall be here considered for the classification problem on sediment transport data. Results of data analysis are reported in the following Section 4.

3.1 XGBoost: forecasting pebbles’ movement from a scalar perspective

Forecasting sediment transport during multiple mobilizing events can be considered as a classification problem for a set of single pebbles, the target classes being *moved* (M) and *not-moved* (NM). It is thus natural to frame this problem in the context of two-class classification methods, that allow one to estimate the probability of movement of single pebbles based on pebbles’ characteristics, locations and flow data. Denote by $(\mathbf{x}_1, y_1), \dots, (\mathbf{x}_n, y_n)$ the set of n available observations, where y_i is a target variable and \mathbf{x}_i is a vector of features linked to the observation i . In this case $y_i \in \{0, 1\}$ (NM or M) and $\mathbf{x}_i \in R^p$, p being the number of features.

The training process of the classifier is typically based on minimization of the cost function $J(\theta)$ over a set of parameters θ , in a parameter space Θ . In the context of gradient boosting (Chen and Guestrin 2016, Ke et al. 2017), the objective functional is written as $J(\theta) = L(\theta) + \Omega(\theta)$, where $L(\theta)$ is a training loss and $\Omega(\theta)$ is a regularization term which constraints the model complexity and prevents overfitting. In the case of two-class classification, the logistic loss can be selected as training loss:

$$L(\theta) = \sum_i [-y_i \log(\hat{p}_i(\theta)) - (1 - y_i) \log(1 - \hat{p}_i(\theta))],$$

where $\hat{p}_i(\theta)$ is the predicted probability for the observation i given the parameters $\theta \in \Theta$, and $\log(\cdot)$ is the natural logarithm. Note that, to express the probability of movement for each pebble as a function of the available predictors, one may consider a very general functional, characterized by the desired degree of complexity.

Training GBdt Gradient Boosting decision trees (GBdt) are amongst the most-common approaches to train non-linear classifiers based on a set of features. They allow to break down the non-linear dependencies of the classifier into an extensive set of binomial rules, represented as binary decision trees. Various implementations of GBdt exist (e.g., XGBoost, Chen and Guestrin (2016) or LightGBM Ke et al. (2017)), the main difference relying in the way decision trees are built.

In this work, we focus on XGBoost, which nowadays is amongst the most used boosting methods, particularly to deal with relatively small datasets, with a moderate number of categorical variables. XGBoost consists of creating a set of weak classifiers $f_t(\mathbf{x}_1)$, each f_t belonging to a space of binary decision trees F . Given the $(t - 1)$ -th tree, the t -th tree is built upon the residuals of the prediction from the previous tree, i.e.,

$$\hat{y}_i^t = \sum_{k=1}^t f_k(\mathbf{x}_i) = \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i).$$

At the step t , the objective function $J(\theta)$ is thus decomposed as

$$\begin{aligned} J^{(t)} &= \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t)}\right) + \sum_{i=1}^t \Omega(f_i) \\ &= \sum_{i=1}^n l\left(y_i, \hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i)\right) + \Omega(f_t) + \text{const}, \end{aligned} \quad (1)$$

where $l\left(y_i, \hat{y}_i^{(t)}\right)$ is the value of the loss function for the i -th prediction at the t -th step and $\text{const} = \sum_{i=1}^{t-1} \Omega(f_i)$. Note that the term $\sum_{i=1}^{t-1} \Omega(f_i)$ in (1) is constant because, at step t , $(t-1)$ trees have been already elaborated, and their are kept fixed for in the construction of the t -th tree. Note that, in (1), the dependence of $\hat{y}_i^{(t-1)}$, f_t and $J^{(t)}$ on θ was dropped just for simplicity of notation.

In the case of two-class classification, the predicted probability \hat{p}_i is typically obtained using a *sigmoid* (i.e., *logistic*) function, i.e., $\hat{p}_i = S(\hat{y}_i) = \frac{1}{1+e^{-\hat{y}_i}}$. Consistently, the predicted probability at step t is obtained as $\hat{p}_i^t = S(\hat{y}_{i-1} + f_t(\mathbf{x}_i))$. Thus, the loss function at the step t can be explicitly written as

$$l\left(y_i, \hat{y}_i^{(t)}\right) = \sum_{i=1}^n \left[y_i \log\left(1 + e^{-(\hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i))}\right) + (1 - y_i) \log\left(1 + e^{(\hat{y}_i^{(t-1)} + f_t(\mathbf{x}_i))}\right) \right].$$

Minimization of the cost functional $J^{(t)}$, for $t = 1, 2, \dots$, then yields the construction of a cascade of trees, which jointly build the predicted probabilities and, ultimately, the classifier – obtained by appropriate thresholding of the predicted probability \hat{p} .

Finally, we recall the expression *gradient boosting* in GBdt refers to the usage of a second order approximation with respect to ultimate prediction $\hat{y}_i^{(t-1)}$; we refer the reader to (Chen and Guestrin 2016) for further details on this point.

Hyperparameters optimization The XGBoost model has a number of hyperparameters. These control, e.g., the proportion of features or observations that is used at the t -th step, the depth of the trees, and the learning rate. Here, finding the global optimum for the loss function is extremely hard, as the objective functional is highly non-linear and non-convex. To increase the model accuracy, one can consider fine-tuning of hyperparameters or their Bayesian optimization (Akiba et al. 2019). Although it can provide a notable improvement in terms of errors, hyperparameters optimization is a time-consuming process. Nevertheless, it is worth noting that, even with the default settings of hyperparameters¹, a XGBoost model often outperforms competitor models, such as generalized linear models, random forest, or support vector machines.

In the following, to reduce the training time of our model, we focus only on the optimization of the parameter *max_depth* which controls the maximum depth of the tree (i.e., the maximum number of steps between a root of the tree and any tree node). In this work, the selection of *max_depth* is performed using $B = 7$ repeated K-fold cross-validation (CV) procedures (Rodriguez et al. 2009). The B repetitions shall be here used to stabilize the result with respect to possible artifacts due to the splitting of the dataset in folds (see also Section 3).

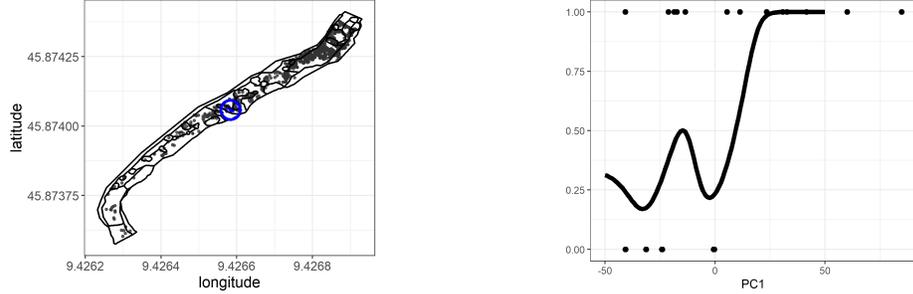
¹The extensive list of hyperparameters and their default values can be found at xgboost.readthedocs.io/en/latest/parameter.html

3.2 Functional Kriging: forecasting pebbles' movement from a functional perspective

In this section, we consider a different viewpoint to the problem of forecasting the probability of movement of pebbles along the stream, which grounds upon the theory of functional geostatistics. The approach is based on the assumption that the dependency between pebbles' dimensions, river flow data, pebbles' locations, and probabilities of movement can be modeled as a continuous function with respect to all these features. This means that one may consider the data as observations from a continuous *functional* surface, relating the value of the features \mathbf{x} to the probability of movement $p(\mathbf{x})$. Note that such surface may be evolving along the stream domain D (i.e., $p(\mathbf{x}) = p(\mathbf{x}; s)$), because of its composite nature. Being able to reconstruct the surfaces $p(\mathbf{x}; s_i)$, for the observation sites s_i in the domain D , could thus potentially lead to a dataset of *functional* observations, to be then projected over unsampled locations along the stream in a (*functional*) kriging setting. Inference on $p(\mathbf{x}) = p(\mathbf{x}; s)$ could be of a particular interest from the application viewpoint, as it would allow one to provide a direct and interpretable characterization of the *bedload* drivers and predisposing conditions for sediment transport. In the following, we discuss the operational steps which we follow to realize this idea in the following analyses.

Reconstruction of functional profiles In general, the estimation of the multi-dimensional surface $p(\mathbf{x}; s_i)$ could require an enormous amount of data. We here cope with the complexity of this estimation problem by (i) reducing the dimensionality of the vector of inputs \mathbf{x}_i , and (ii) using a local neighborhood $N(s_i)$ of location s_i to build the estimate $\hat{p}(\mathbf{x}; s_i)$. Note that both steps could be partially avoided in the presence of a larger database, in terms of events and pebbles. For step (i), we reduce the vector \mathbf{x} of features by considering only the first PC of the pebble characteristics (*PC1*), and by averaging the effect of the flow over the typical mobilizing events only (events T , see Section 2.2). This allows us to simplify the problem to the analysis of univariate functional profiles $p(PC1; s)$, indexed by the spatial index s in D . Note that summarizing the information of the pebbles' characteristic through *PC1* is justified by virtue of the high proportion of variance explored by this PC, whereas the second choice is motivated by the observation that typical mobilizing events appear similar from the flow viewpoint (see Section 2.2). Further justification on this choice is provided in Section 4.1

For step (ii), we consider a local estimate of $\hat{p}(\mathbf{x}; s_i)$, based on a spatial neighborhood of s_i . Note that these probability curves need to be estimated from sparse observations, the term *sparsity* referring both to the spatial dimension and to the variable *PC1*. In fact, focusing on a single pebble (i.e., on a single value for *PC1*), data are realizations of Bernoulli random variables, for which a limited number of realizations (i.e., events) are observed. Similarly, when focusing on a single location s_i , no more than 3 observations are typically available. To estimate $\hat{p}(PC1; s_i)$, we thus use not only the observations related to the location s_i , but also those from a neighboring zone $N(s_i)$, where $N(s_i)$ is a circle of radius $r > 0$ centered at s_i – the hyperparameter $r > 0$ being fixed by CV in a range of candidates ($r \in \{3, 5, 7\}$ m). Note that such neighborhoods are also constrained to belong to the same geomorphological subdomain as s_i , to preserve the domains' characteristic through the estimation procedure. To reduce the estimation bias induced by the consideration of neighboring data, only locations s_i with at least $n_{min} = 12$ observations in $N(s_i)$ shall be considered. Moreover, whenever the neighborhood $N(s_i)$ contains more than $n_{max} = 30$ observations, we build the estimate $\hat{p}(PC1; s_i)$ upon the n_{max} closest observations. This allows to balance the bias-variance trade-off affecting the estimate of $p(PC1; s_i)$, adjusting for the different spatial density of the observations. The



(a) Neighborhood zone for a point $s_i = (9.426, 45.874)$. Radius of the circle $r = 5$ m (b) Curve estimated using N-W approach. Kernel bandwidth $h = 20$

Figure (6) Construction of functional profiles $\hat{p}(PC1, s_i)$ from raw pebbles' data. (a) Pebbles' locations are indicated as grey points; the location s_i is indicated as a blue point, the blue circle being the boundary of $N(s_i)$. (b) black symbols indicate the binary observations (0 for NM, 1 for M); the solid lines indicates the estimated curve $\hat{p}(PC1, s_i)$.

parameters n_{min}, n_{max} were both selected by CV within a range of candidates ($n_{min} \in \{5, 7, 10, 12\}$, $n_{max} \in \{15, 20, 25, 30, 35\}$).

Figure 6 depicts an illustration of the curves generation process, highlighting a location $s_i = (9.24, 45.87)$ (marked by a blue point in Fig. 6a), the neighborhood $N(s_i)$ considered for the estimate (marked as a blue circle in Fig. 6b), and the associated estimate of $\hat{p}(PC1; s_i)$ (black curve in Fig. 6b). This latter curve was obtained by Nadaraya–Watson kernel regression (Nadaraya 1964; Watson 1964), using a Gaussian kernel K with bandwidth parameter h , i.e.,

$$\hat{p}(PC1 = x, s_i) = \frac{\sum_{j: s_j \in N(s_i)} K_h(x - x_j) y_j}{\sum_{j: s_j \in N(s_i)} K_h(x - x_j)}$$

where the x_j 's are the values of $PC1$ for the observed pebbles in $N(s_i)$ and the y_j 's are their associated binary outcomes (0 for NM, 1 for M; black symbols in Fig. 6b). The kernel bandwidth was set to $h = 20$, to balance the roughness of the curve with its capability to adapt to the data. For the estimation of $p(PC1, s_i)$, $i = 1, \dots, n$, a common support I is defined as the range of values of $PC1$ in the training data, i.e., $I = [PC1_{min}, PC1_{max}]$. Only the curves observed on the whole interval I are used during training procedure.

Functional geostatistics for probability curves From a mathematical standpoint the (estimated) relation $\hat{p}(PC1, s_i)$ between the probability of movement of a pebble in s_i and its $PC1$ can be interpreted as a functional data point and analysed in the framework of object-oriented spatial statistics (O2S2, Menafoglio and Secchi (2017)). Similarly as in scalar geostatistics (Cressie 2015), in O2S2 the set of functional data $\hat{p}(PC1, s_i)$, $i = 1, \dots, n$, is modeled as a partial observation of a functional random field $\{\hat{p}(PC1, s), s \in D\}$. Here, typical goals are modeling of the spatial dependence and spatial prediction (i.e., kriging). Given that the probability curves $\hat{p}(PC1, s_i)$ are constrained in their values in $[0, 1]$, we shall consider as data the following logit transformation of these curves

$$\chi_{s_i} = \log \frac{\hat{p}(PC1, s_i)}{1 - \hat{p}(PC1, s_i)} = \text{logit } \hat{p}(PC1, s_i),$$

where $\tilde{p}(PC1, s_i) = 1 - \varepsilon$ if $\hat{p}(PC1, s_i) = 1$ and $\tilde{p}(PC1, s_i) = \varepsilon$ if $\hat{p}(PC1, s_i) = 0$, ε being a small threshold allowing for the definition of the logit function when $\hat{p}(PC1, s_i) = 0$ or $\hat{p}(PC1, s_i) = 1$.

For a location s in D , we consider χ_s as a random element of the functional space L^2 of square integrable functions, and decompose χ_s into the sum of a linear drift term m_s and a second-order stationary residual δ_s , such that (Menafoglio et al. 2013)

$$\begin{aligned} \chi_s &= m_s + \delta_s, \\ m_s &= \mathbb{E}[\chi_s] = \sum_{l=0}^L a_l \cdot f_l(s), \\ \delta_s &\quad \text{s.t.} \quad \mathbb{E}[\delta_s] = 0; \quad \text{Cov}(\delta_{s_1}, \delta_{s_2}) = \mathbb{E}[\langle \delta_{s_1}, \delta_{s_2} \rangle_{L^2}] = C(s_1 - s_2). \end{aligned} \quad (2)$$

In (2), the parameters a_l are functional coefficients in L^2 , f_l are known spatial regressors, and $C(\cdot)$ is the (stationary) *trace-covariogram* of the residual field, which represents the functional counterpart of the classical covariance function (Cressie 2015). In this work, the spatial regressors that will be considered are the binary variables d_k , indicating whether the location s_i belongs to the k -th geomorphological subdomain ($d_k(s_i) = 1$) or not ($d_k(s_i) = 0$).

In this setting, our goal is to build an optimal prediction $\chi_{s_0}^*$ of the function χ_{s_0} at the unobserved location s_0 , based on the available data. This would ultimately allow one to (i) obtain a prediction $p^*(PC1; s_0)$ for the probability curve $p(PC1; s_0)$ as $p^*(PC1; s_0) = \text{logit}^{-1}(\chi_{s_0}^*)$, and (ii) yielding a classification for the pebbles' movement in the river domain, e.g., by thresholding $p^*(PC1; s_0)$. To this end, one may formulate a functional kriging (FK) predictor, that is the best linear unbiased combination of the observed data, $\chi_{s_0}^* = \sum_{i=1}^n \lambda_i^* \chi_{s_i}$. Here, the λ_i^* 's are scalar coefficients that minimize the variance of prediction error under unbiasedness, i.e.,

$$\begin{aligned} \min_{\lambda_1, \dots, \lambda_n} & \mathbb{E}[\|\chi_{s_0}^* - \chi_{s_0}\|^2] \\ \text{s. t.} & \quad \mathbb{E}[\chi_{s_0}^* - \chi_{s_0}] = 0. \end{aligned} \quad (3)$$

Similarly as in scalar geostatistics, problem (3) admits a unique solution, which can be obtained by solving a linear system depending on the covariance between elements of the random field – as determined by the trace-covariogram – and on the regressors f_l (see, e.g., Menafoglio et al. (2013)). Building good estimates of the trace-covariogram is thus crucial for the prediction problem. In this setting, one may alternatively estimate the *trace-variogram* of the field

$$2\gamma(s_1 - s_2) = \mathbb{E}[\|\delta_{s_1} - \delta_{s_2}\|^2],$$

which again has the same interpretation as its classical counterpart. The trace-variogram can be estimated by fitting a valid variogram family (e.g., Matérn, spherical) to the empirical variogram from the (estimated) residuals

$$2\hat{\gamma}(h) = \frac{1}{N(h)} \sum_{(s_i, s_j) \in N(h)} \|\hat{\delta}_{s_i} - \hat{\delta}_{s_j}\|^2,$$

where $N(h)$ is the set of pairs of locations (s_i, s_j) approximately separated by a vector h , $h \approx s_i - s_j$. Methods and algorithms for an effective estimation of the residuals and of the trace-variogram have been extensively studied in the literature; we refer the reader, e.g., to Menafoglio et al. (2013; 2016) Menafoglio and Secchi (2017).

3.3 Error metrics and models' validation

In this section we introduce the methodology which will be used to compare the performances of the two proposed perspectives, when used to infer on the sediment transport classification problem.

Error metrics The error metrics which will be used in the following are: accuracy, precision, recall, F1-score, and AUC (Powers 2011) –their definition being recalled below. All these metrics are widely used to evaluate and compare classification methods. Denote by P (Positive) the number of pebbles that moved, and by N (Negative) those that did not move. In the set of pebbles predicted to move, we call TP (True Positive) the number of pebbles which actually moved, and FP (False Positive) those that did not move. In the set of pebbles predicted not to move, we call TN (True Negative) those that actually stayed still, and FN (False Negative) those that moved instead. We then define the error metrics as

- Accuracy: $\frac{TP+TN}{TP+FP+TN+FN}$;
- Precision: $\frac{TP}{TP+FP}$;
- Recall: $\frac{TP}{TP+FN}$;
- F β -score: $\frac{(1+\beta^2)*precision*recall}{\beta^2*precision+recall}$ (typically $\beta = 1$ and the score is called F₁ score (Dice 1945 Sørensen 1948)).

The metric AUC is finally defined as the area under the ROC curve, which compares the true positive rate with the false positive rate when varying the threshold used to build the classification from the predicted probability (see, e.g., Friedman et al. (2001)).

Threshold setting The outcome of both the proposed approaches is the probability of movement $p^*(\mathbf{x};s)$, during a rainfall event, for a particular pebble at a given location s . Hence, part of the models' post-processing is to select, in an *optimal* way, a threshold α , such that the pebble is classified as M (moved) for $p(\mathbf{x};s) \geq \alpha$, or NM (not-moved) for $p^*(\mathbf{x};s) < \alpha$. Setting this threshold can be performed by cross-validation, using as optimality criterion the F_1 score defined above (i.e., selecting α which maximises $F_1(\alpha)$ (Dice 1945 Sørensen 1948)). Alternatively, one can consider the maximization of the Youden's J criterion, based on the index $J = \frac{TP}{TP+FN} + \frac{TN}{TN+FP} - 1$ (Youden 1950). Given that balancing precision and recall is task-specific, in the following we shall consider the results for optimal values of the threshold α set by using both Youden's J and the F_1 -score.

Validation of the XGBoost approach To validate the machine learning approach based on XGBoost presented in Section 3.1, we consider a K-fold CV approach based on the following scheme.

0. Initialize the hyperparameters: XGBoost hyperparameters (particularly the depth of the trees max_depth), the set I_α of candidate thresholds α ;
1. Split the pebbles into K folds;
2. Perform CV iteration. For $k = 1, \dots, K$

- (a) Split the data into training and test sets, the test set being the k -th fold;
- (b) Build the XGBoost classifier based on the vectors of features \mathbf{x}_i , $i = 1, \dots, n_{train}$, of the data within the training set;
- (c) Obtain $p^*(\mathbf{x}_j; s_j)$, $j = 1, \dots, n_{test}$, for the pebbles in the test set, based on their actual features \mathbf{x}_j and location s_j ;

Result of the CV iteration: $p^*(\mathbf{x}_i; s_i)$, for $i = 1, \dots, n$ (each estimated when the i -th observation is left out of the training sample).

3. Select the optimal threshold α_b within I_α :
 - (a) Based on the results at step 2., for each $\alpha \in I_\alpha$, classify the pebbles as M or NM by thresholding $p^*(PC1_i, s_i)$, $i = 1, \dots, n$, through α ;
 - (b) Select the optimal $\alpha_b \in I_\alpha$, i.e., that associated with the optimal score (F_1 or Youden's J).
4. Calculate the error metrics from the set of classifications at step 3(a), for the optimal value α_b .
5. Repeat the steps 1.-4. for $B = 7$ different splits in K folds.

The threshold α^* used for the final classifier is selected as the average of the thresholds α_b obtained for the $B = 7$ repetitions of the CV. This CV procedure is also used to set the hyperparameters of the method, as illustrated in Section [3.1](#).

Validation of Functional approach To validate the functional approach presented in Section [3.2](#), we consider a K-fold CV approach, similar to that discussed above, and based on the following scheme.

0. Initialize the hyperparameters: the bandwidth h of the N-W kernel, the radius r of the neighborhood, the tolerance ε , the set I_α of candidate thresholds α ;
1. Split the pebbles into K folds;
2. Perform CV iteration. For $k = 1, \dots, K$:
 - (a) Split the data into training and test sets, the test set being the k -th fold;
 - (b) Generate the curves $\hat{p}(PC1_i, s_i)$, $i = 1, \dots, n_{train}$ (from the training subset data only);
 - (c) Perform the geostatistical analysis; and build the prediction $p^*(PC1_j, s_j)$, $j = 1, \dots, n_{test}$, for the pebbles in the test set, based on their actual values $PC1_j$ and location s_j ;

Result of the CV iteration: $p^*(\mathbf{x}_i; s_i)$, for $i = 1, \dots, n$ (each estimated when the i -th observation is left out of the training sample).
3. Select the optimal threshold α_b within I_α :
 - (a) For each $\alpha \in I_\alpha$, classify the pebbles as M or NM by thresholding $p^*(PC1_i, s_i)$, $i = 1, \dots, n$ through α ;
 - (b) Select the optimal $\alpha_b \in I_\alpha$, i.e., that associated with the optimal score (F_1 or Youden's J).
4. Calculate the error metrics from the set of classifications at step 3(a), for the optimal value α_b .

5. Repeat the steps 1.-4. for $B = 7$ different splits in K folds.

The threshold α^* used for the final classifier is again selected as the average of the thresholds α_b obtained for the $B = 7$ repetitions of the CV. Given that during a CV iteration the curves $\hat{p}(PC1, s)$ are generated from the training set only, the value of $PC1_i$ for an observation in the test set may be out of the support of $p^*(PC1, s_i)$. In this case, the probability is calculated as the $p^*(x^*, s_i)$, x^* being the nearest value of $PC1$ within the support ($x^* = PC_{min}$ if $PC1_i < PC_{min}$, or $x^* = PC_{max}$ if $PC1_i > PC_{max}$).

4 Results

In this section, we illustrate the results of the data analyses performed according to the methodologies described in Section 3. We shall first apply the approaches separately, and then compare their results. We shall highlight the limitations of both models in terms of precision and recall, with particular reference to the morphological zones where one model outperforms the other and viceversa.

4.1 Results for XGBoost

The aim of this subsection is twofold. First, we aim to show the results and performances of XGBoost for the problem at hand. Second, we aim to verify the impact of dimensionality reduction – through the PCA presented in Section 2 – on the performances of the classifier. To do so, we distinguish the results in terms of (i) type of rainfall event (all events or typical events T) and (ii) dimensionality of the feature vector. In the latter case, we shall focus on two options, including within the model (a) all the pebbles' features, their locations and flow data (named *all-features*), or (b) only the pebbles' locations and the data PCs (named *PCs*): $PC1$, $PC2$, $PC1_{flow}$, $PC2_{flow}$ (see Section 2). This set of analyses serves also as a support to the dimensionality reduction needed to develop the functional approach discussed in Section 4.2. All the results presented in this section were obtained using the R package (Grujic and Menafoglio 2017).

All events Based on a 5-fold CV analysis repeated $B = 7$ times, we set the maximum depth of the trees to $max_depth = 7$ when all the features are used, and $max_depth = 6$ when PCs are used instead.

The error metrics for XGBoost are presented in Table 2. Tables 3, 4 report the average confusion matrices of the XGBoost model, when this is based on all the features (Table 3) or only on the PCs (Table 4). Here, the threshold α for the classification was built by optimization of the F_1 metric (see Section 3.3) – the average being $\alpha^* = 0.643$ for the case of all the features, $\alpha^* = 0.544$ for the PCs only. One may notice a that the PCs case appears to be associated with a higher accuracy and precision with respect to the all-features case. In general, the all-features case appears to present a lower number of FP, but a higher number of FN, thus yielding a general slight overestimation of the sediment transport w.r.t. the PCs case. These analyses anyway suggest that representing the features through the PCs only does not represent a significant loss of information for the purpose of classification.

Typical events We now aim to study the impact of the flow data ($PC1_{flow}$ and $PC2_{flow}$) on XGBoost models when calibration is based on typical events only. Note that typical events appear to

	AUC	ACC	Precision	Recall	F1-score
All features	0.94	0.859	0.755	0.858	0.803
Only PCs	0.939	0.87	0.803	0.81	0.806

Table (2) Results of XGBoost when all the rainfall events are considered. The optimal threshold for classification was set by optimization of the F_1 score.

		Actual				Actual	
		not-moved	moved			not-moved	moved
Predicted	not-moved	570.3	185.4	Predicted	not-moved	538.7	132.4
	moved	94.7	1138.6		moved	126.3	1191.6

Table (4) PCs: Average confusion matrix across CV repetitions

be similar in terms of flow (see Section 2). Therefore, one may argue that it is reasonable to suppose that, in this setting, exclusion of $PC1_{flow}$ and $PC2_{flow}$ should not significantly affect the prediction power of the models. We thus compare two models, obtained by training XGBoost either on PCs data ($PC1$, $PC2$, $PC1_{flow}$ and $PC2_{flow}$) or $PC1$ only, still considering in both models the locations and the subdomains binary features (i.e., $d_k(s)$).

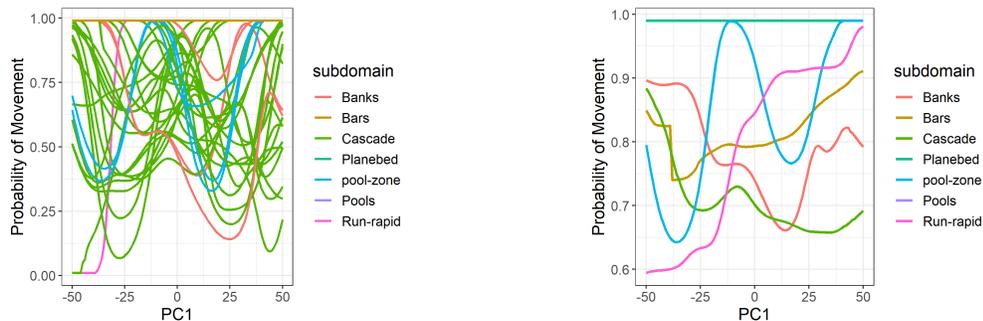
Similarly as in the previous paragraph the optimal depth of both models was selected using $B = 7$ repetitions of K-fold CV, with $K = 5$. The results of the procedure are $max_depth = 6$ for the model trained on PCs and $max_depth = 5$ in the second case. The F_1 -thresholds estimated for two models are, in both cases, $\alpha^* = 0.663$, suggesting that a balance between FP and FN is preserved. The error metrics of both models are reported in Table 5. One may note that the model based on PCs attains better results, although the difference in performance is limited compare to the significant reduction in the input dimensionality. The main source of gain in accuracy for the first model relies in FNs. According to Tables 6 and 7, the average absolute difference between FNs ($\Delta FN = 277.1 - 188 = 89.1$) in the two model is more than 3.5 times bigger than the absolute difference in terms of FPs ($\Delta FP = 96.1 - 70.7 = 25.4$). Reducing the dimensionality of the inputs thus turns in an overestimation of the incidence of the NM class (i.e., the model tends to underestimate the amount of mobilized sediment). This tendency shall be confirmed by the results of the functional approach, which are discussed in the next section.

	AUC	Accuracy	Precision	Recall	F1-score
PCs	0.925	0.838	0.739	0.891	0.808
Only PC1	0.873	0.766	0.652	0.842	0.733

Table (5) Results of XGBoost when typical rainfall events are considered. The optimal threshold for classification was set by optimization of the F_1 score.

		Actual				Actual	
		not-moved	moved			not-moved	moved
Predicted	not-moved	538.3	188	Predicted	not-moved	512.9	277.1
	moved	70.7	797		moved	96.1	707.9

Table (7) PC1: Average confusion matrix across CV repetitions



(a) A sample of 50 estimated probability curves $\hat{p}(PC1_i; s_i)$ using N-W kernel Regression

(b) Averages of the estimated curves within seven subdomains

Figure (7) Estimated probability curves: data and mean within groups. In panel (b), the scale on the y-axis was set to $[0.6, 1]$ to better appreciate the difference in $\bar{p}(PC1; D_k)$ between groups.

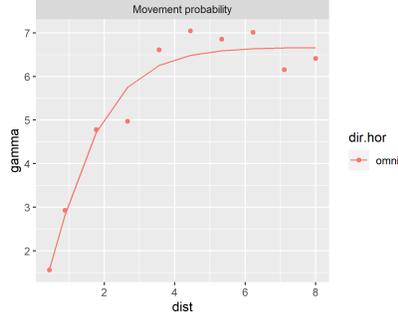
4.2 Results for the functional case

In this subsection, we illustrate the results of the analyses when these are based on the functional perspective described in Section 3.2. Recall that the functional approach is based on the consideration of just the feature $PC1$, and of the observations related with typical events (T) only. Moreover, the main hyperparameters for the method are (see also Sec. 3.2): the minimum/maximum number of points to generate a curve, set to $n_{min} = 12$; $n_{max} = 30$; the support of the curves, set to $[PC1_{min}; PC1_{max}] = [-50, 50]$; the tolerance, set to $\epsilon = 0.01$; and the kernel bandwidth, set to $h = 20$. Moreover, in the following we shall consider a Bessel model for the calibration of the variogram. All the results presented in this section were obtained by using the R package `fdagstat` (Grujic and Menafoglio 2017).

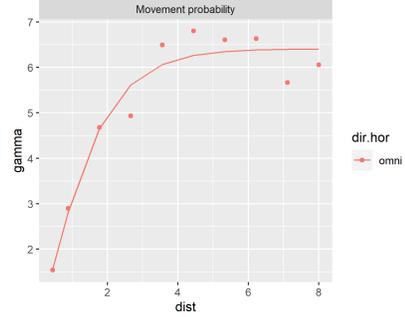
A 5-fold CV analysis run as described in Sec. 3.3 suggests to set the radius of the neighborhoods to $r = 5$. These parameter settings allowed us to estimate the curves $\hat{p}(PC1; s_i)$ for the sample

	AUC	ACC	Precision	Recall	F1-score
All features	0.844	0.756	0.654	0.773	0.708

Table (8) Results for the functional approach, when setting the optimal threshold according to the F_1 score.



(a) Variogram with a drift dependent on the geomorphological subdomains



(b) Variogram with a drift dependent on the geomorphological subdomains

Figure (8) Variograms

location s_i , $i = 1, \dots, n$. A subset of this dataset of functional profiles is reported in Figure 7. One may observe a notable variability in the shape of the curves, suggesting a highly non-linear dependence between the probability of movement and the pebbles' characteristics, which varies over space in a non-trivial fashion. Figure 7b displays the means $\bar{p}(PC1; D_k)$ of the probability curves within the geomorphological subdomains D_k , $k = 1, \dots, 7$. More precisely, these curves were computed by back-transforming the sample mean of the logit-transformations of the curves $\hat{p}(PC1; s_i)$, i.e.,

$$\bar{p}(PC1; D_k) = \text{logit}^{-1} \left(\frac{1}{n} \sum_{s_i \in D_k} \text{logit} \hat{p}(PC1; s_i) \right),$$

the transformation logit and $\hat{p}(PC1, s_i)$ being defined as in Section 3.2. Such curves are thus representative of the mean values m_s assumed by the object $\chi_s = \text{logit}(\hat{p}(\cdot; s))$ within the subdomains. One may notice a relatively high variability across groups, which suggests to consider the binary variables $d_k(s)$ ($d_k(s) = 1$ if $s \in D_k$, $d_k(s) = 0$ otherwise) in the model for the drift term. However, CV analyses suggest that slightly better performances are obtained when using a stationary approach instead, which is that discussed hereafter. For the sake of completeness, Figure 8 reports the variograms of the residuals (estimated as described in Section 3.2), when these refer to a stationary drift term (i.e., m_s is spatially constant; Figure 8a), or to a drift dependent on the geomorphological subdomains through the variables $d_k(s)$'s (Figure 8b). Both variograms are compatible with the residuals' stationarity; selection of the stationary model is thus based on the CV results.

Table 9 reports the confusion matrix of the method (averaged over the CV repetitions), which suggests that the classifier built upon the functional approach tends to be associated with a relatively high number of false positive (FP: 250 out of 1594 pebbles), which is however consistent with those associated with XGBoost based on $PC1$ for typical events, discussed in Section 4.1. The next section provides further discussion and comparison between these approaches.

4.3 Comparison between the two perspectives

We now proceed with a comparative analysis between XGBoost and FK results. For the purpose of coherency between the information that is used for training, we here compare the XGBoost model

		Actual	
		not-moved	moved
Predicted	not-moved	470.7	250.1
	moved	138.3	734.9

Table (9) Results of the functional approach: Average confusion matrix across CV repetitions

		AUC	Accuracy	Precision	Recall	F1
FK	F1	0.844	0.756	0.654	0.773	0.708
	J		0.759	0.659	0.769	0.709
XGBoost	F1	0.873	0.766	0.652	0.842	0.733
	J		0.767	0.654	0.842	0.734

Table (10) Results for the functional approach and XGBoost, when setting the optimal threshold according to the F_1 score or the Youden’s J

trained on *typical events* only (as presented in the Section 4.1) and the FK model calibrated on the same data (see Section 4.2).

We first perform a comparison of the models based on $PC1$ only, which is representative of the performances of the methods based on similar inputs. The first two lines of Table 10 report the classification performances, assessed by $B = 7$ repetitions 5-fold CV, of the functional predictor. Here, the first line corresponds to a threshold α^* set by optimization of the F_1 criterion ($\alpha^* = 0.613$), whereas the second line to the optimization of the Youden’s J ($\alpha^* = 0.604$). The last two lines of Table 10 are referred to the analogous quantities related to the XGBoost model trained on $PC1$ only, which are associated with a threshold $\alpha^* = 0.663$ (approximately the same α^* being set for both Youden’s J and F_1 criteria). Results presented in Table 10 suggest that all the four settings are practically equivalent in terms of accuracy (ca 76%) and precision (ca 65%). The main differences are related to AUC, which is slightly better in the XGBoost case (87%) than in FK (84%), indicating an overall better ordering of probabilities. Moreover, recall is higher in XGBoost than in FK (84% vs 77%), indicating a better performance for the former in terms of FPs, that is also observed in Tables 7-9. A comparison of the performances by subdomains is provided in Table 11. Here, one can note that XGBoost has overall a better performance, e.g., in *bars* zone, although FK proves better in a number of subdomains, e.g., within the *step-pool* zone. It is worth mentioning that the number of pebbles places in bars zone is more than twice those in step-pool zones, coherent with observed differences in absolute values related of FPs and FNs.

We finally compare the methods in terms of local CV errors, displayed in Figures 9, 10. Both figures display visualizations of the results of 7-repeated CV. Figure 9 represents the CV results for each pebble separately, the colors being associated with the number of times the single pebble was correctly classified along the $B = 7$ repetitions of the 5-fold CV. Figure 9 displays the average accuracy within the subdomains D_k identified according to the local morphology of the riverbed. Graphical inspection of Fig. 9 suggests that, although the two models appear similar in terms of the error metrics, slightly different patterns are observed in their errors. For instance, XGBoost is

Subdomain	Number of Pebbles	Average number of misclassified pebbles		Average proportion of misclassified pebbles	
		XGBoost	FK	XGBoost	FK
Banks	208	64.57	65.86	0.31	0.32
Bars	159	18.71	28.14	0.12	0.18
Cascade	727	248.43	249.71	0.34	0.34
Plane bed	103	0.86	0	0.01	0
Pools	12	1.14	1	0.1	0.08
Run-rapid	322	26.86	30.71	0.08	0.1
Step pool-zone	63	10.43	8.29	0.17	0.13

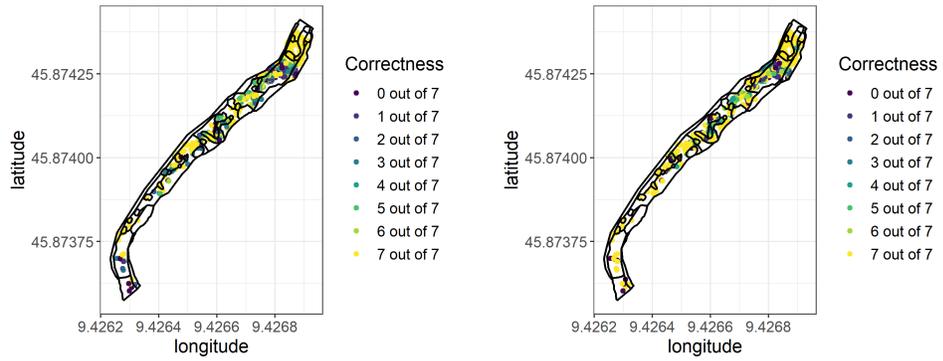
Table (11) XGBoost vs FK: misclassified pebbles for each subdomain

associated with noticeably less correct predictions for the left bottom corner, while, on average, its predictions are of high quality in the central and upper part of the domain. Similarly, observing Figure 10 one may notice that the main difference between the two models appears in the center-left of the domain, and in its left bottom part.

Even though the comparison of the models based on $PC1$ suggests an overall consistency of the results obtained with the two approaches, one should note that, when using all the PCs, improved results were obtained with XGBoost (see Section 4.1). This suggests that the input simplification needed to build the dataset of probability profiles prior to FK may have induced a loss of predictive power with respect to a scalar approach based on state-of-the-art machine learning methods. On the other hand, the functional approach clearly allows for a direct interpretation of the relation between the tendency of pebbles to moved, and their characteristics, as further highlighted in the next Section 4.4. This is a clear advantage to XGBoost, whose interpretability still appears to be limited. Finally, it is worth mentioning that additional analyses on the same dataset – not discussed here for brevity – showed that FK outperforms other ‘more standard’ statistical methods such as GLM, anyway suggesting the validity of the approach in the framework of model based-statistical classification.

4.4 A geomorphological interpretation of the results

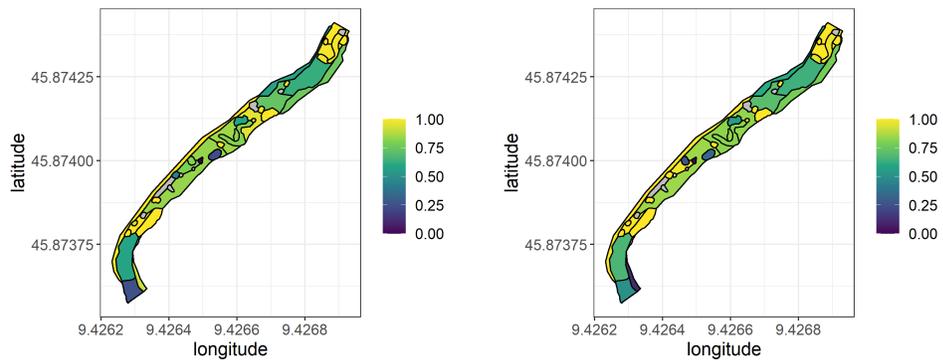
Looking at the results from a geomorphological point of view, the zones where the predictive models encounter difficulties in correctly classifying the probability of movement appear to be the *banks* and *cascade* with around 30% of misclassified observations by both models. A particular concentration of misclassified cases can be identified in the central section of the reach under investigation. This zone is characterized by a complex morphology where the presence of a large boulder forces the stream into a rapid s-curve trajectory, unlike the surrounding environment. The upstream end of



(a) XGBoost 7-repeated CV correctness map

(b) FK 7-repeated CV correctness map

Figure (9) CV error maps



(a) XGBoost. Average accuracy of 7-repeated CV by polygon (b) FK. Average accuracy of 7-repeated CV by polygon

Figure (10) CV: Average accuracy by *morphological sector*

the reach is also characterized by a high concentration of misclassified cases. This area is close to the placement location of a large number of pebbles. Hence this could be attributed to the effect of initial pebble deployment that would likely behave in an odd manner with respect to sediment that has already undergone some settlement. The concentration of misclassified cases in the *banks* zone is not surprising as well. Those zones are marginally affected by the flow rate during low to moderate flow rate events and a slight increase/decrease in water depth could determine if a pebble will be affected by the flow or not. The definition of the boundaries of those zones could be somewhat ambiguous due to e.g. the presence of vegetation in summer and its absence in winter.

As a results of the functional kriging approach, we obtained the probability of pebble movement as a function of pebble's dimensions in the different morphological units (10). This outcome illustrates a general similarity between *banks*, *bars*, and *cascade* zones, while there is a considerable difference in the predicted values for *plane bed*, *run/rapid*, and *step and pool* zone. For the former three subdomains there appears to be an agreement that very small and very large pebbles tend to have a similar probability of movement while average-sized ones are slightly less likely to be mobilized. Such a difference is present also in the case of the *step and pool* zone although it is much more pronounced in the former cases. The effect of pebble size on mobility rates has often been under discussion by several authors, arguing that the sediment mobility could be independent on the sediment dimensions in some morphological conditions. For instance, according to [Liedermann et al. \(2013\)](#) coarser particles are harder to mobilize yet, once mobilized, they may travel even farther than smaller ones. [Ferguson et al. \(2017\)](#) attributed the size-selectiveness of the sediment mobility to the different types of channel morphology – a finding that finds correspondence in the present work. For instance, the probability of pebble movement in a *plane bed* morphological unit appears to be entirely sediment size independent, while in a *run/rapid* local morphology a strong size-dependency can be observed, where smaller particles are characterized by a lower probability of movement and vice versa. [Church and Hassan \(1992\)](#) noted that smaller particles are characterized by a higher likelihood to be trapped when the stream channel is composed of large grains and this could strongly influence the dependency of the sediment mobility on the grain's size. The estimated mean probability of movement is above 50% in all the cases, which indicates that during moderate flow events the bedload mobility is quite pronounced in the presented range of pebble dimensions.

5 Application to an independent dataset

To enrich the comparison between the two considered approaches, we here apply them on an independent dataset, that did not participate in the training illustrated in the previous sections. Here an Eulerian approach was adopted to observe the mobility of pebbles in a fixed spatial reference, as opposed to the Lagrangian approach used to track each single sediment particle along the river course.

Red pebbles' data The observation zones represent 30cm x 30cm squares within which the riverbed was painted in red. The observation zones were captured before and after a flood event in order to identify the number of mobilized pebbles and their size. The size of individual pebbles was estimated from the images using the automatic object detection software *Basegrain* ([Detert et al. 2012](#)), that allowed us to estimate the size of a- and b-axis of each detected red pebble. Figure 11 illustrates the steps of the tracking process. The available dataset consists of 7 sets of observations that were gathered in correspondence of three events (events 11, 14, and 16, that took places during



Figure (11) (a) Red pebbles painted before an event, (b) post-event image of the same area; (c) estimation of the dimensions of the remaining pebbles

the year 2017²), which can be considered as typical in terms of associated flow. Besides assessing the capability of the proposed methods to deal with this type of data, this case is used to show a potential approach to the application of the models on *sets* of sediment particles (instead of single particles).

Limitations of RP data A technical limitation of this measurement campaign is the difficulty in finding a dry portion of the riverbed that can be painted, typically during low flow. Moreover, in the case of “red pebbles” (RPs), all the pebbles from the outlined zones were considered for the measurements, while, in the “yellow pebbles” (YPs) case, only the particles big enough for insertion of an RFID tag were used. This inevitably turns in a selection bias for both cases, which renders the two analyses only partially comparable. In fact, some substantial differences are present between the grain size distributions of the two datasets. For instance, the RPs are in general much smaller than the YPs used to calibrate the models, the nominal diameters of the latter being, on average, 70 mm shorter than the diameters of the former (see Table 12).

Another limitation relies in the fact that the estimation of RP measurements is elaborated based on the 2D projections of the original 3D objects (see Fig. 11c). This hinders the computation of the three axes (a-,b-,c-axis), since one of the axes (presumably the c-axis) remains covered. In fact, the estimate of the two visible axes may be itself associated with a non-negligible uncertainty. Moreover, the visible dimensions of RP gathered before mobilizing events may not correspond to the ones gathered after the events, because pebbles tend to rotate or move –even without location change– influenced by the flow. Hence, these data cannot be used to find a one-to-one correspondence between particles before and after the events and to verify whether they moved or not, but rather they can be used to assess joint summaries about the set of particles (e.g., granulometric distributions).

Applicability of the models Application of XGBoost and FK models on the RP data should take into account the specific features (and limitations) of this database. For instance, concerning FK, one should note that just 3% of red pebbles has $PC1$ larger than -50 (when estimating the c-axis as $3/4$ of the b-axis). Hence, the remaining 97% of data would be given the prediction $p^*(-50, s)$, which is associated to all the particles with a size leading to $PC1 < -50$ (see Section 3.2). In fact, application of our models to RP data requires particular care, as one should pay close attention when testing models out of the range of training data, no matter of the approach they are using. For the

²Event 11 occurred between 26/04 and 30/04; Event 14 between 06/06 and 07/06; Event 16 between 01/09 and 02/09

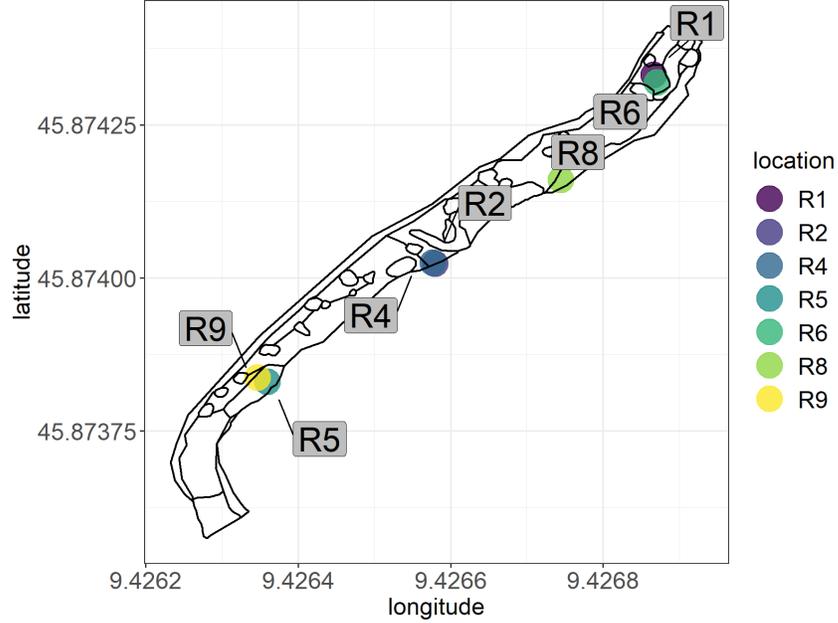


Figure (12) Locations of RP data

purpose of this study, we shall thus consider a slight modification of the models described in Section 4 which aims to render the features of the YP training set as *compatible* as possible with those of the RP test set, by using pebbles' diameters.

A classification approach based on particles' diameters To cope with the lack of correspondence between pre-mobilizing and post-mobilizing event data, we here consider a variant of the classifiers built in Section 4 which is based on the *nominal diameter*, defined as $d_i = \sqrt{a_i * b_i}$. Note that the range of d_i for YP is [43.87, 160.16]; the range of diameters in RP is found in Table 12. One can note that a partial overlap is attained between the diameters of YPs and RPs. In XGBoost, this variable is considered as input, together with the location of the pebbles and the flow data – the model being trained on typical events only. In the functional approach, the probability profile $p(d_i; s)$ is estimated at the sampled locations, and then projected via functional kriging at unknown sites along the river domain.

The application of the model to the RP data is performed as follows. Given a square region R_j and a mobilizing event e , $j \in \{1, 2, 4, 5, 6, 8, 9\}$, $e \in \{11, 14, 16\}$, we call $S_{je}^- = \{S_1, \dots, S_{n_{je}^-}\}$ the set of n_{je}^- red pebbles in R_j before the event e , and $S_{je}^+ = \{S_1, \dots, S_{n_{je}^+}\}$ the set of n_{je}^+ red pebbles which are still in R_j after the event e . Table 12 reports the values of n_{je}^- for the regions and events available in the dataset. Considering a single pebble $S_i \in S_{je}^-$, we may estimate its probability p_i^* of movement, based on the set of features associated with the considered pebble and one of the calibrated models (XGBoost or FK). To describe the joint probability of movement of the set of pebbles in S_{je}^- , we

Location	Event	Subdomain	Total	Number of pebbles with PC1 >-50	Diameters Before Event	Diameters After Event
R1	11	Plane bed	341	15	[8.64, 85.08]	[20.90, 81.39]
R2	11	Cascade	351	2	[8.53, 68.87]	[20.58, 66.92]
R4	14	Cascade	447	7	[8.60, 91.77]	[34.38, 69.48]
R5	14	Bars	275	1	[8.53, 67.29]	[9.13, 64.02]
R6	16	Plane bed	248	5	[8.20, 90.60]	[13.64, 92.99]
R8	16	Cascade	245	15	[8.67, 101.53]	[9.39, 103.82]
R9	16	Bars	259	11	[9.19, 125.68]	[9.23, 66.39]

Table (12) Summary of red pebbles (RPs) data

can then consider their joint law, which, under independence, reads: $\mathbf{p}_{je}^* = \prod_{i=1}^{n_{je}^-} p_i^*$. This allows to simulate a set of realizations from such distribution, to be then compared with the actual observations in S_{je}^+ . Comparison of the empirical distribution of the particle diameters –a.k.a. particle-size distributions (PSDs)– with the actual PSD after the events allows to evaluate the capability of the models to adapt to this type of data.

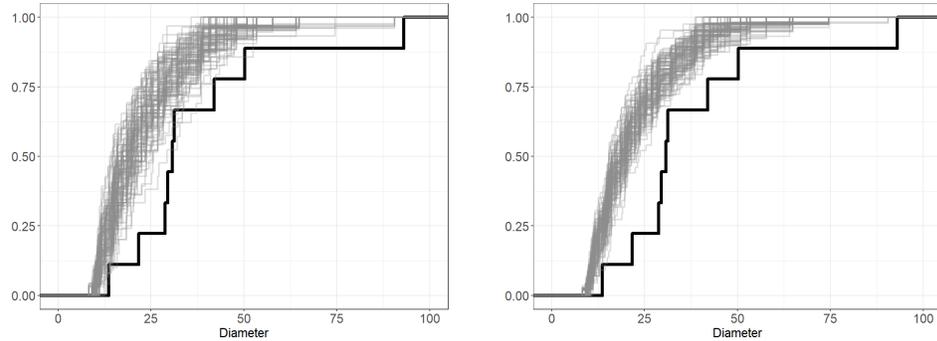
Results and comparison As a way of example, Figure 13 displays a set of $M = 100$ empirical cumulative distribution functions (ECDFs) of the diameters of the particles found in location R_6 after event 16 (grey lines). These are compared with the ECDF of the PSD estimated from the pebbles in $P_{6,16}^+$, depicted as black lines.

Graphical inspection of Figure 13 suggests that both methodologies fail to correctly represent the displacement of the pebbles with relatively small diameters (between 8 and 30 mm), possibly due to the partial incompatibility of the data in the YP and RP datasets. In particular, both models appears to be associated with an underestimation of the mobility of small particles and an overestimation of that of large particles (see also Table 13). Nevertheless, the XGBoost approach seems to be associated with a slightly higher variability of the estimated PSDs, particularly for relatively large diameters. The cloud of simulated PSDs seems thus able to *get closer* to the actual observation than for FK. Looking at the results displayed in Table 13, one may observe that, overall, both approaches result in a significant overestimation of the proportion of stationary pebbles w.r.t. the data, particularly for R4 and R5.

A quantitative overall comparison between the simulated and actual PSDs can be obtained by computing the Wasserstein distance (see, e.g., Villani (2008)) between these distributions, which is obtained as

$$d(\mathcal{P}_m, \mathcal{P}_{obs}) = \left(\int_0^1 (\mathcal{P}_m^{-1}(t) - \mathcal{P}_{obs}^{-1}(t))^2 dt \right)^{1/2},$$

$\mathcal{P}_m, \mathcal{P}_{obs}$ being the PSD of the m -th simulation (cumulative distribution functions), $m = 1, \dots, 100$, and the observed post-event PSD, respectively, and $\mathcal{P}_m^{-1}, \mathcal{P}_{obs}^{-1}$ the respective quantile functions. Table 13 reports the average Wasserstein distance, i.e., $1/m \sum_{m=1}^M d(\mathcal{P}_m, \mathcal{P}_{obs})$, for the observed



(a) ECDFs of realizations sampled from estimated probabilities of movement obtained using XGBoost (b) ECDFs of realizations sampled from estimated probabilities of movement obtained using FK

Figure (13) Sampled ECDFs of distributions pebbles that didn't move, and actual PSD, at location R_6 .

regions R_j . One may note a small discrepancy between the approaches – XGBoost performing only slightly better than FK.

Besides the efforts made to render the YP and RP datasets compatible, these analyses suggests that the YP data are only partially informative on the phenomenon described by the RP data. This reflects on poorer performances of the models calibrated on the former dataset, when applied to the latter one. More in general, the limitations highlighted within the section may point to directions of improvement for the design of experiment of future campaigns, if these are intended to support the construction of models better representative of the joint behavior of sediment particles within the regions R_j 's.

6 Discussion and Conclusions

In this work we proposed a comparison of two methodological approaches to the problem of prediction of sediment transport in a pre-Alpine region. To this end, we considered state-of-the-art machine learning and geostatistical methods, namely XGBoost (Chen and Guestrin 2016) and functional kriging (FK, (Menafoglio and Secchi 2017)). Results on real data suggest that both approaches have good performances on the problem at hand, with a slightly better predictive power of XGBoost over FK. Nonetheless, FK yields more interpretable results, since it allows for an explicit prediction of the probability of movement of a single pebble at a site s as a function of its dimension (characterized through $PC1$). These results are in agreement with the general pros/cons evidenced in the comparison between machine-learning and model-based statistical methods, where the former are often able to outperform the latter in terms of prediction power, while the latter are typically associated with higher interpretability.

However, the methods developed in this work are subject to a number of limitations, that should be taken into account when applying the calibrated models to independent data. Beside the limitations for the use of the models for independent data only partially compatible with our training set (e.g., for RP data, see Section 5), one should note that not all the regions of the spatial domain are

Location	Number of still pebbles			Proportion of still pebbles			Average Wasserstein Distance	
	Actual	XGBoost	FK	Actual	XGBoost	FK	XGBoost	FK
R1	18	3.71	7.09	0.053	0.011	0.021	37.657	36.87
R2	21	269.95	134.19	0.06	0.769	0.382	16.374	16.294
R4	12	180.08	143.49	0.027	0.403	0.321	29.956	30.248
R5	129	54.37	58.44	0.469	0.198	0.213	4.021	4.215
R6	9	26.97	41.89	0.036	0.109	0.169	20.209	19.537
R8	30	89.6	27.42	0.122	0.366	0.112	13.367	20.925
R9	23	90.26	55.11	0.089	0.348	0.213	15.31	15.404

Table (13) Comparison between XGBoost and FK on red pebbles data

densely covered by the observations, which turns in a spatially varying uncertainty in predictions. Analyses have shown that the pebble location is indeed informative on its probability to move. For instance, if one trains the XGBoost model without the feature s_i , a relevant decrease in accuracy and AUC would be observed (on all the events, with all the other features: AUC = 0.915, Accuracy = 0.815, $F_1 = 0.758$, compared with the results in Table 2). As such, higher degrees of uncertainty are associated with the areas of the domains where data are sparse. Note that FK is associated with a measure of prediction uncertainty (named *kriging variance*, see e.g. Menafoglio et al. (2013)), which also accounts for the data sparsity. This is not the case of XGBoost, where measures of uncertainty are only indirectly available (e.g., through CV analyses).

Another aspect which is worth mentioning concerns the river flow data. In XGBoost models, flow data are used as features, and these appear to be associated with improved accuracy w.r.t. the cases where flow data are not considered (see Table 2). However, one should note that values of flow data are only available after the end of the rainfall event. To use the model in a real-world application aimed to forecast sediment transport, one should thus consider *predicted* flow data, or perform a *scenario-based* analysis. In both cases, an additional degree of uncertainty would characterize the final forecast on sediment transport. In this sense, although the fact that FK does not account for flow data may appear as a limitation of the approach –yielding to decreased accuracy– this may not be really the case in the actual use of the model.

Nevertheless, an intrinsic limitation of FK is the need of estimating the functional profile $p(\mathbf{x}, s)$ from raw data, which in turns imposes a limitation on the dimensionality of the feature vector \mathbf{x} . This aspect led us to assume all the typical events to be *similar* from the flow viewpoint, and work in an *average* flow scenario. In this respect, larger databases may allow for a higher dimensionality of \mathbf{x} than that considered in this work, to partially account for the dynamic of the event being considered. Indeed, even if the average characterizations of mobilizing events are similar, events’ dynamics could vary drastically, with a possible impact on the actual probability of movement of sediment particles.

It is worth noting that the majority of mobilizing events are characterized by moderate-flow. That is, events sufficiently energetic to provoke sediment mobility, but not exceptional. While the dataset includes observations of pebble mobility during high-flood events, these data are limited to

four events, effectively participating in the XGBoost analysis through the PCA on river flow data, but excluded from the FK analysis due to its above-mentioned limitations. On the other hand, the distinct separation and analysis of high-flow data would be quite interesting as it has been pointed out that the major portion of bedload transport is contained in extreme flood events (Coppus and Imeson (2002), Turowski et al. (2009)). In fact, the differences in pebble mobility in the different morphological units could become negligible considering that particularities of each unit would be overcome by the energy of the flow. For instance, *banks* that are typically marginally affected by the stream would experience a higher discharge. The estimation of the probability of pebble movement presented here suggests that, during moderate flow events, the mobilized YPs sample is on average between 50% and 100% depending on the morphological unit, while looking at RPs –characterized by a considerably smaller grain size distribution– the proportion of mobile pebbles is over 90% with the exception of *bars*, where this value is around 55%. While those two parameters are obtained through different methodologies and based on different (only partially comparable) datasets, they effectively correspond to the same conceptual quantity – the *proportion of motion* (Ballio et al. 2018). The combination of those two parameters indicates that (i) there appears to be a general dependency of pebble’s mobility on the grain size and (ii) moderate flow or ‘typical’ events contribute significantly to the dynamics of bedload transport. A further step in the investigation would be to try and quantify this effect in terms of the distances travelled by the pebbles, as well as their velocity (or virtual velocity). Analyses in this direction are currently ongoing, and will be the scope of future work.

References

- Akiba T, Sano S, Yanase T, Ohta T, Koyama M (2019) Optuna: A next-generation hyperparameter optimization framework. In Proceedings of the 25th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, 2623–2631
- Bakke P D, Sklar L S, Dawdy D R, Wang W C (2017) The design of a site-calibrated parker–klingeman gravel transport model. *Water* 9(6):441
- Ballio F, Pokrajac D, Radice A, Hosseini Sadabadi S A (2018) Lagrangian and eulerian description of bed load transport. *Journal of Geophysical Research: Earth Surface* 123(2):384–408
- Brambilla D, Papini M, Ivanov V I, Bonaventura L, Abbate A, Longoni L (2020) Sediment yield in mountain basins, analysis, and management: The smart-sed project. In *Applied Geology*, Springer, 43–59
- Cain A, MacVicar B (2020) Field tests of an improved sediment tracer including non-intrusive measurement of burial depth. *Earth Surface Processes and Landforms*
- Cassel M, Dépret T, Piégay H (2017) Assessment of a new solution for tracking pebbles in rivers based on active rfid. *Earth Surface Processes and Landforms* 42(13):1938–1951
- Chen T, Guestrin C (2016) Xgboost: A scalable tree boosting system. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, KDD '16, New York, NY, USA: Association for Computing Machinery, ISBN 9781450342322, 785–794
- Church M, Hassan M A (1992) Size and distance of travel of unconstrained clasts on a streambed. *Water Resources Research* 28(1):299–303
- Coppus R, Imeson A (2002) Extreme events controlling erosion and sediment transport in a semi-arid sub-andean valley. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group* 27(13):1365–1375
- Cressie N (2015) *Statistics for spatial data*, Revised Edition. John Wiley & Sons
- de Miranda R B, Mauad F F (2015) Influence of sedimentation on hydroelectric power generation: Case study of a brazilian reservoir. *Journal of Energy Engineering* 141(3):04014016

- Detert M, Weitbrecht V, et al. (2012) Automatic object detection to analyze the geometry of gravel grains—a free stand-alone tool. In *River flow*, Taylor & Francis Group London, 595–600
- Dice L R (1945) Measures of the amount of ecologic association between species. *Ecology* 26(3):297–302
- Dotterweich M (2008) The history of soil erosion and fluvial deposits in small catchments of central europe: deciphering the long-term interaction between humans and the environment—a review. *Geomorphology* 101(1-2):192–208
- Ferguson R, Sharma B, Hodge R, Hardy R, Warburton J (2017) Bed load tracer mobility in a mixed bedrock/alluvial channel. *Journal of Geophysical Research: Earth Surface* 122(4):807–822
- Friedman J, Hastie T, Tibshirani R (2001) *The elements of statistical learning, volume 1*. Springer series in statistics New York
- Gatti F, Menafoglio A, Togni N, Bonaventura L, Brambilla D, Papini M, Longoni L (2020) A novel downscaling procedure for compositional data in the aitchison geometry with application to soil texture data. *arXiv preprint arXiv:200706856*
- Grujic O, Menafoglio A (2017) *fdagstat, an R package*. R package version 1.0
- Haddadchi A, Nosrati K, Ahmadi F (2014) Differences between the source contribution of bed material and suspended sediments in a mountainous agricultural catchment of western iran. *Catena* 116:105–113
- Hassan M A, Bradley D N (2017) Geomorphic controls on tracer particle dispersion in gravel-bed rivers. *Gravel-Bed Rivers: Process and Disasters* 167
- Ivanov V, Brambilla D, Longoni L, Papini M (2016a) Long term analysis of sediment transport in a pre-alpine river. volume 3, cited By 1
- Ivanov V, Radice A, Papini M, Longoni L (2020a) Event-scale pebble mobility observed by rfid tracking in a pre-alpine stream: a field laboratory. *Earth Surface Processes and Landforms* 45(3):535–547
- Ivanov V, Zanchi B, Papini M, Radice A, Barazzetti L, Zanzi L, Longoni L (2020b) Integration of low-cost technological solutions for flow rate and bedload transport monitoring in the caldone river, italy. In *River Flow 2020 : Proceedings of the 10th Conference on Fluvial Hydraulics (Delft, Netherlands, 7-10 July 2020)*, 910
- Ivanov V I, Brambilla D, Longoni L, Arosio D, Papini M (2017) Rfid-aided sediment transport monitoring—laboratory and preliminary field test results. In *Workshop on World Landslide Forum, Springer*, 623–630
- Ivanov V I, Brambilla D, Longoni L, Papini M (2016b) Long term analysis of sediment transport in a pre-alpine river. *International Multidisciplinary Scientific GeoConference: SGEM* 3:171–178
- Ke G, Meng Q, Finley T, Wang T, Chen W, Ma W, Ye Q, Liu T Y (2017) Lightgbm: A highly efficient gradient boosting decision tree. In *Advances in neural information processing systems*, 3146–3154
- Lane S, Tayefi V, Reid S, Yu D, Hardy R (2007) Interactions between sediment delivery, channel change, climate change and flood risk in a temperate upland environment. *Earth Surface Processes and Landforms: The Journal of the British Geomorphological Research Group* 32(3):429–446
- Liedermann M, Tritthart M, Habersack H (2013) Particle path characteristics at the large gravel-bed river danube: results from a tracer study and numerical modelling. *Earth Surface Processes and Landforms* 38(5):512–522
- Longoni L, Ivanov V, Brambilla D, Radice A, Papini M (2016a) Analysis of the temporal and spatial scales of soil erosion and transport in a mountain basin. *Italian Journal of Engineering Geology and Environment* 16(2):17–30, cited By 9
- Longoni L, Papini M, Brambilla D, Barazzetti L, Roncoroni F, Scaioni M, Ivanov V I (2016b) Monitoring riverbank erosion in mountain catchments using terrestrial laser scanning. *Remote Sensing* 8(3):241
- Mazzorana B, Comiti F, Fuchs S (2013) A structured approach to enhance flood hazard assessment in mountain streams. *Natural hazards* 67(3):991–1009
- Menafoglio A, Secchi P (2017) Statistical analysis of complex and spatially dependent data: a review of object oriented spatial statistics. *European journal of operational research* 258(2):401–410

- Menafoglio A, Secchi P, Dalla Rosa M (2013) A universal kriging predictor for spatially dependent functional data of a hilbert space. *Electron J Statist* 7:2209–2240
- Menafoglio A, Secchi P, Guadagnini A (2016) A class-kriging predictor for functional compositions with application to particle-size curves in heterogeneous aquifers. *Mathematical Geosciences* 48(4):463–485
- Nadaraya E A (1964) On estimating regression. *Theory of Probability & Its Applications* 9(1):141–142
- Papangelakis E, Hassan M A (2016) The role of channel morphology on the mobility and dispersion of bed sediment in a small gravel-bed stream. *Earth Surface Processes and Landforms* 41(15):2191–2206
- Papini M, Ivanov V I, Brambilla D, Arosio D, Longoni L (2017) Monitoring bedload sediment transport in a pre-alpine river: An experimental method. *Rendiconti Online Societa Geologica Italiana* 43:57–63
- Pizarro A, Manfreda S, Tubaldi E (2020) The science behind scour at bridge foundations: A review. *Water* 12(2):374
- Powers D M (2011) Evaluation: from precision, recall and f-measure to roc, informedness, markedness and correlation
- R Core Team (2020) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria
- Radice A, Longoni L, Papini M, Brambilla D, Ivanov V I (2016) Generation of a design flood-event scenario for a mountain river with intense sediment transport. *Water* 8(12):597
- Ramsay J O, Silverman B W (2005) *Functional data analysis*. Springer Science+Business Media
- Rodriguez J D, Perez A, Lozano J A (2009) Sensitivity analysis of k-fold cross validation in prediction error estimation. *IEEE transactions on pattern analysis and machine intelligence* 32(3):569–575
- Sear D, Newson M, Brookes A (1995) Sediment-related river maintenance: the role of fluvial geomorphology. *Earth Surface Processes and Landforms* 20(7):629–647
- Sørensen T J (1948) A method of establishing groups of equal amplitude in plant sociology based on similarity of species content and its application to analyses of the vegetation on Danish commons. I kommission hos E. Munksgaard
- Stover S, Montgomery D (2001) Channel change and flooding, skokomish river, washington. *Journal of Hydrology* 243(3-4):272–286
- Turowski J M, Yager E M, Badoux A, Rickenmann D, Molnar P (2009) The impact of exceptional events on erosion, bedload transport and channel stability in a step-pool channel. *Earth Surface Processes and Landforms* 34(12):1661–1673
- Vázquez-Tarrío D, Recking A, Liébault F, Tal M, Menéndez-Duarte R (2019) Particle transport in gravel-bed rivers: Revisiting passive tracer data. *Earth Surface Processes and Landforms* 44(1):112–128
- Vetsch D, Siviglia A, Ehrbar D, Facchini M, Kammerer S, Koch A, Peter S, Vonwiller L, Gerber M, Volz C, et al. (2017) System manuals of basement, version 2.7. laboratory of hydraulics, glaciology and hydrology (vaw), eth zurich
- Villani C (2008) *Optimal transport: old and new*, volume 338. Springer Science & Business Media
- Watson G S (1964) Smooth regression analysis. *Sankhyā: The Indian Journal of Statistics, Series A* :359–372
- Wharton G, Mohajeri S H, Righetti M (2017) The pernicious problem of streambed colmatation: a multi-disciplinary reflection on the mechanisms, causes, impacts, and management challenges. *Wiley Interdisciplinary Reviews: Water* 4(5):e1231
- Youden W J (1950) Index for rating diagnostic tests. *Cancer* 3(1):32–35

MOX Technical Reports, last issues

Dipartimento di Matematica
Politecnico di Milano, Via Bonardi 9 - 20133 Milano (Italy)

- 65/2020** Di Gregorio, S.; Vergara, C.; Montino Pelagi, G.; Baggiano, A.; Zunino, P.; Guglielmo, M.; Fu
Prediction of myocardial blow flow under stress conditions by means of a computational model
- 64/2020** Fiz, F.; Viganò, L.; Gennaro, N.; Costa, G.; La Bella, L.; Boichuk A.; Cavinato, L.; Sollini, M.
Radiomics of Liver Metastases: A Systematic Review
- 63/2020** Tuveri, M.; Milani, E.; Marchegiani, G.; Landoni, L.; Torresani, E.; Capelli, P.; Sperandio, N.;
HEMODYNAMICS AND REMODELING OF THE PORTAL CONFLUENCE IN PATIENTS WITH CANCER OF THE PANCREATIC HEAD: A PILOT STUDY
- 62/2020** Massi, M. C.; Ieva, F.
Representation Learning Methods for EEG Cross-Subject Channel Selection and Trial Classification
- 61/2020** Pozzi, S.; Redaelli, A.; Vergara, C.; Votta, E.; Zunino, P.
Mathematical and numerical modeling of atherosclerotic plaque progression based on fluid-structure interaction
- 60/2020** Lupo Pasini, M; Perotto, S.
Hierarchical model reduction driven by a Proper Orthogonal Decomposition for parametrized advection-diffusion-reaction problems
- 59/2020** Massi, M.C.; Franco, N.R; Ieva, F.; Manzoni, A.; Paganoni, A.M.; Zunino, P.
High-Order Interaction Learning via Targeted Pattern Search
- 58/2020** Beraha, M.; Pegoraro, M.; Peli, R.; Guglielmi, A
Spatially dependent mixture models via the Logistic Multivariate CAR prior
- 57/2020** Regazzoni, F.; Quarteroni, A.
An oscillation-free fully partitioned scheme for the numerical modeling of cardiac active mechanics
- 56/2020** Botti, L.; Botti, M.; Di Pietro, D. A.;
A Hybrid High-Order method for multiple-network poroelasticity