A regional early warning system for debris flows

Michel Ponziani1*, Denise Ponziani1, Andrea Giorgi2, Hervé Stevenin2, and Sara Maria Ratto2

1CIMA Foundation, Savona, Italia
2Centro Funzionale della Valle d’Aosta, Dipartimento Protezione Civile, Valle d’Aosta, Italia

Abstract. In this study, we have developed a predictive model for debris flows using machine learning techniques on a detailed dataset composed by a variety of geomorphological and hydro-meteorological variables. The variables of the dataset were collected from daily measured and modelled data for all of the drainage basins in which at least one debris-flow event was generated during the time period considered (2009-2019). The performances of the models obtained with different machine learning techniques were evaluated with the ROC analysis. The most suitable model was then experimentally implemented in the existing early warning system of the Aosta Valley Region. The model provides daily values of debris-flow probability (DFP) for individual basins, based on the input geo-morphological and hydro-meteorological variables. These results can be used to issue specific debris-flow alerts at the scale of the alert areas of the region.

1 Introduction

Forecasting, monitoring and surveillance activities of meteorological phenomena in the Aosta Valley region are managed by the Centro Funzionale, with the aim of evaluating the hydrogeological risk. The region is divided into four alert areas, which are homogeneous in terms of meteorological conditions and soil instabilities. The hydrogeological alerts are released daily by a colour code (red, yellow, orange or green) based on forecast scenarios and the severity and spatial extent of the phenomena.

A debris-flow indicator (DFI) based on thresholds of expected rainfall and recorded temperature was previously developed and implemented [1] to issue specific debris-flow alerts caused by summer thunderstorms, which are high-intense rainfall of short duration. In this study, we present a new debris-flow predictive model (DFM) obtained from machine learning algorithms and a large dataset of several geo-morphological and hydro-meteorological variables. The data were collected for the years from 2009 to 2019, in which 121 summer debris-flow events occurred in 91 different drainage basins of the Aosta Valley region. The model derived from this analysis was then experimentally implemented in the early warning system (EWS) of the region.

2 Debris-flow dataset

The data to develop the model were collected in the summers from 2009 to 2019, based on the availability of consistent and continuous data of the hydrological model Continuum [2-3]. During this period, 121 debris-flow events were recorded [4] for the total of 91 different drainage basins of the region (Figure 1).

Fig. 1. The four alert areas of the region with the weather stations and the debris-flow basins considered in this study.

A dataset of 15 different geo-morphological and hydro-meteorological variables was built collecting daily values of the variables for the period considered and for each basin included in this study (Table 1).

The geo-morphological variables, which are constant over time for each basin, were obtained from a 10m DEM and from the lithological map of the region. The altitude was included with the mean, the maximum and the minimum values for each basin. The slope gradient and the orientation were obtained as the averages of the respective values within the basin area; the first was calculated from the tangent of the angle of the surface to the horizontal, whereas the latter from the deviation of the orientation from the North. The lithological map was used to derive the percentage of the basin area covered by alluvial deposits, slope debris, glacial deposits, landslide deposits, glacier and rock.
The hydro-meteorological variables are time dependent and hence defined with a daily temporal resolution consistent with the information available for the debris-flow events. The mean and the maximum precipitation in each basin were calculated from daily raster maps obtained with the GRISO algorithm [5] from the rain gauges distributed over the region (Figure 1). The daily precipitation was chosen in this study as it includes the short-duration thunderstorms (typically 2 to 4 hours) that trigger the summer debris flows in the region.

The snow cover area (SCA) as well as the mean and the maximum soil moisture content were obtained from the hydrological model Continuum [2-3], which was calibrated and validated for the Aosta Valley region using hydro-meteorological, hygrometric and satellite data. The freezing level of the last 10 days describes the effect of persistent high temperatures [1] and it was calculated from the weather stations distributed within the region.

### 3 Methods

Five different machine learning techniques were used to formulate the debris-flow predictive models: the logistic regression model (LR), the decision tree (DT), the random forest (RF), the K-nearest neighbours (KNN), and the support vector machine (SVM). The variable to be modelled was the debris flow in each specific basin from 0 to 1. In the dataset, this variable can be either 0 (no debris flow) or 1 (debris-flow recorded) and it is unequally distributed due to the low number of debris-flow events recorded in the period considered. Therefore, four resampling techniques were applied to balance the dataset: undersampling, oversampling [6], SMOTE [7] and ROSE [8].

The dataset was split into two parts: the training dataset (80%) to develop the model, and the testing dataset (20%) to assess the model performance. The best-performing model was determined with the Receiver Operating Characteristic (ROC) analysis [9] and the calculation of the corresponding Area Under the Curve (AUC), which is an indicator of the classifier performance ranging between 0 and 1. ROC curves are obtained plotting the relationship between the true positive rate (TPR) and the false positive rate (FPR), defined as follows:

\[
TPR = \frac{TP}{TP+FN} \quad (1)
\]
\[
FPR = \frac{FP}{FP+TN} \quad (2)
\]

where the true positives (TP) are the debris flows predicted correctly, the false positives (FP) are the missed debris flows, the false negatives (FN) are the false alarms, and the false negatives (FN) are the non-debris flows predicted correctly.

### 4 Results

Figure 2 shows the ROC curves obtained applying five machine-learning algorithms on the testing dataset. The best-performing model was produced by the logistic regression, with a corresponding AUC of 0.85. All of the resampling techniques that were then applied on the testing dataset improved considerably the performance of the model. The final best-performing model, which was selected to be implemented in the EWS of the region, was the logistic regression with ROSE resampling, which scored an AUC of 0.95.

![ROC curves for the five different machine-learning algorithms applied to the testing dataset: logistic regression (LR), random forest (RF), K-nearest neighbours (KNN), decision tree (DT), and support vector machine (SVM).](image)

The variable importance for the selected model is shown in Figure 3. The precipitation and the mean freezing level are the variables that weight the most in the debris-flow prediction, followed by the snow cover area of the basin. This study only included specific debris-flow basins, therefore geo-morphological variables such as the slope gradient were found to be less important by the prediction algorithms.

The application of the model on the testing dataset showed that the debris-flow events of the past were predicted with a mean probability of 77%, whereas the days with no debris flows had a mean probability of 25%.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>Altitude [m a.s.l.]</td>
<td>DEM</td>
</tr>
<tr>
<td>Slope gradient [°]</td>
<td>DEM</td>
</tr>
<tr>
<td>Orientation [°]</td>
<td>DEM</td>
</tr>
<tr>
<td>Alluvial deposits [%]</td>
<td>Lithological map</td>
</tr>
<tr>
<td>Slope debris [%]</td>
<td>Lithological map</td>
</tr>
<tr>
<td>Glacial deposits [%]</td>
<td>Lithological map</td>
</tr>
<tr>
<td>Landslide deposits [%]</td>
<td>Lithological map</td>
</tr>
<tr>
<td>Glacier [%]</td>
<td>Lithological map</td>
</tr>
<tr>
<td>Rock [%]</td>
<td>Lithological map</td>
</tr>
<tr>
<td>Precipitation (Mean) [mm]</td>
<td>Rain gauges</td>
</tr>
<tr>
<td>Precipitation (Max) [mm]</td>
<td>Rain gauges</td>
</tr>
<tr>
<td>Freezing level last 10 days [°C]</td>
<td>Weather stations</td>
</tr>
<tr>
<td>Snow Cover Area [%]</td>
<td>Hydrological model</td>
</tr>
<tr>
<td>Soil moisture (Mean) [%]</td>
<td>Hydrological model</td>
</tr>
<tr>
<td>Soil moisture (Max) [%]</td>
<td>Hydrological model</td>
</tr>
</tbody>
</table>

The application of the model on the testing dataset showed that the debris-flow events of the past were predicted with a mean probability of 77%, whereas the days with no debris flows had a mean probability of 25%.
values. The overall best-performing model was selected to be experimentally implemented in the EWS of the Aosta Valley region as an improvement of the existing debris-flow indicator (Ponziani et al. 2020). Compared to conventional rainfall thresholds, the model developed with machine-learning algorithms includes a large number of variables that have an impact on the resulting debris-flow probability. The most important variables of the model were found to be the precipitation, the freezing level of the last ten days and the snow cover area. This confirms that intense rainfalls and prolonged high temperatures increase the probability of debris flows triggered by thunderstorms in the Alpine region of Aosta Valley, while the presence of snowpack considerably reduces the debris-flow probability.

The debris-flow model was implemented in the EWS of the Aosta Valley region to provide a daily debris-flow probability for the 91 selected basins. The input daily variables are obtained from the weather forecast bulletin (precipitation), the hydrological model (snow cover area and soil moisture content) and the weather stations (mean freezing level of the previous ten days). However, even though the debris-flow probability is computed for 91 basins, the results can be operatively used at the moment only at the scale of the alert area.

References