DELUIGI Nicola (2018) : Data-driven analysis and mapping of the potential distribution of mountain permafrost

Abstract

In alpine environments, mountain permafrost is defined as a thermal state of the ground and it corresponds to any lithosphere material that is at or below 0°C for at least two years. Its degradation is potentially leading to an increasing rock fall activity and sediment transfer rates. During the last 20 years, knowledge on this phenomenon has significantly improved thanks to many studies and monitoring projects, revealing an extremely discontinuous and complex spatial distribution, especially at the micro scale (scale of a specific landform; tens to several hundreds of metres).

The objective of this thesis was the systematic and detailed investigation of the potential of data-driven techniques for mountain permafrost distribution modelling. Machine learning (ML) algorithms are able to consider a greater number of parameters compared to classic approaches. Not only can permafrost distribution be modelled by using topo-climatic parameters as a proxy, but also by taking into account known field permafrost evidences. These latter were collected in a sector of the Western Swiss Alps and they were mapped from field data (thermal and geoelectrical data) and ortho-image interpretations (rock glacier inventorying). A permafrost dataset was built from these evidences and completed with environmental and morphological predictors. Data were firstly analysed with feature relevance techniques in order to identify the statistical contribution of each controlling factor and to exclude non-relevant or redundant predictors. Five classification algorithms, belonging to statistics and machine learning, were then applied to the dataset and tested: Logistic regression (LR), linear and non-linear Support Vector Machines (SVM), Multilayer perceptrons (MLP) and Random forests (RF). These techniques inferred a classification function from labelled training data (pixels of permafrost absence and presence) to predict the permafrost occurrence where this was unknown.

Classification performances, assessed with AUROC curves, ranged between 0.75 (linear SVM) and 0.88 (RF). These values are generally indicative of good model performances. Besides these statistical measures, a qualitative evaluation was performed by using field expert knowledge. Both quantitative and qualitative evaluation approaches suggested employing the RF algorithm to obtain the best model. As machine learning is a non-deterministic approach, an overview of the model uncertainties is also offered. It informs about the location of most uncertain sectors where further field investigations are required to be carried out to improve the reliability of permafrost maps.

RF demonstrated to be efficient for permafrost distribution modelling thanks to consistent results that are comparable to the field observations. The employment of environmental variables illustrating the micro-topography and the ground characteristics (such as curvature indices, NDVI or grain size) favoured the prediction of the permafrost distribution at the micro scale. These maps presented variations of probability of permafrost occurrence within distances of few tens of metres. In some talus slopes, for example, a lower probability of occurrence in the mid-upper part of the slope was predicted. In addition, permafrost lower limits were automatically recognized from permafrost evidences. Lastly, the high resolution of the input dataset (10 metres) allowed elaborating maps at the micro scale with a modelled permafrost spatial distribution, which was less optimistic than traditional spatial models. The permafrost prediction was indeed computed without recurring to altitude thresholds (above which permafrost may be found) and the representation of the strong discontinuity of mountain permafrost at the micro scale was better respected.

Key words: mountain permafrost mapping, environmental modelling, machine learning, spatial data mining.