

EGU23-180, updated on 14 Apr 2024

<https://doi.org/10.5194/egusphere-egu23-180>

EGU General Assembly 2023

© Author(s) 2024. This work is distributed under the Creative Commons Attribution 4.0 License.



## A Random Forest approach to quality-checking automatic snow-depth sensor measurements

**Giulia Blandini**<sup>1,2</sup>, Francesco Avanzi<sup>2</sup>, Simone Gabellani<sup>2</sup>, Denise Ponziani<sup>2</sup>, Hervé Stevenin<sup>3</sup>, Sara Ratto<sup>3</sup>, and Luca Ferraris<sup>2,1</sup>

<sup>1</sup>DIBRIS, University of Genoa, Genova, Italy

<sup>2</sup>CIMA Research Foundation, Savona, Italy

<sup>3</sup>Centro Funzionale Valle D'Aosta

Advanced environmental technologies have made available an increasing amount of data from remote sensing satellites, and more sophisticated ground data. Their assimilation into dynamic models is progressively becoming the most frequent, and conceivably the most successful, solution to estimate snow water resources. Models reliability is therefore bounded to data quality, which is often low in mountain, high-elevation, and unattended settings. To add new value to snow-depth sensor measurements, we developed a machine-learning algorithm to automatize the QA/QC procedure of near-surface snow depth observations collected through ground stations data. Starting from a consolidated manual classification, based on the expert knowledge of hydrologists in Valle D'Aosta, a Random Forest classifier was developed to discriminate snow cover from grass or bare ground data and detect random errors (e.g., spikes). The model was trained and tested on Valle d'Aosta data and then validated on 3 years of data from 30 stations on the Italian territory. The F1 score was used as scoring metric, being it most suited to describe the performances of a model in case of a multiclass imbalanced classification problem. The model proved to be robust and reliable in the classification of snow cover and grass/bare ground discrimination (F1 values above 90%), yet less reliable in random error detection, mostly due to the dataset imbalance. No clear correlation with single year meteorology was found in the training domain, and the promising results from the generalization to a larger domain corroborates the model robustness and reliability. This machine learning application of data quality assessment provides more reliable snow ground data, enhancing the quality of snow models.