

EGU2020-4243

<https://doi.org/10.5194/egusphere-egu2020-4243>

EGU General Assembly 2020

© Author(s) 2020. This work is distributed under the Creative Commons Attribution 4.0 License.



Comparison of Machine Learning Techniques Powering Flood Early Warning Systems. Application to a catchment located in the Tropical Andes of Ecuador.

Paul Munoz¹, Johanna Orellana-Alvear^{1,2}, Jörg Bendix², and Rolando Célleri^{1,3}

¹Departamento de Recursos Hídricos y Ciencias Ambientales, Universidad de Cuenca, Cuenca 010150, Ecuador

²Laboratory for Climatology and Remote Sensing (LCRS), Faculty of Geography, University of Marburg, D-35032 Marburg, Germany

³Facultad de Ingeniería, Universidad de Cuenca, Cuenca 010150, Ecuador

Flood Early Warning Systems have globally become an effective tool to mitigate the adverse effects of this natural hazard on society, economy and environment. A novel approach for such systems is to actually forecast flood events rather than merely monitoring the catchment hydrograph evolution on its way to an inundation site. A wide variety of modelling approaches, from fully-physical to data-driven, have been developed depending on the availability of information describing intrinsic catchment characteristics. However, during last decades, the use of Machine Learning techniques has remarkably gained popularity due to its power to forecast floods at a minimum of demanded data and computational cost. Here, we selected the algorithms most commonly employed for flood prediction (K-nearest Neighbors, Logistic Regression, Random Forest, Naïve Bayes and Neural Networks), and used them in a precipitation-runoff classification problem aimed to forecast the inundation state of a river at a decisive control station. These are “No-alert”, “Pre-alert”, and “Alert” of inundation with varying lead times of 1, 4, 8 and 12 hours. The study site is a 300-km² catchment in the tropical Andes draining to Cuenca, the third most populated city of Ecuador. Cuenca is susceptible to annual floods, and thus, the generated alerts will be used by local authorities to inform the population on upcoming flood risks. For an integral comparison between forecasting models, we propose a scheme relying on the F1-score, the Geometric mean and the Log-loss score to account for the resulting data imbalance and the multiclass classification problem. Furthermore, we used the Chi-Squared test to ensure that differences in model results were due to the algorithm applied and not due to statistical chance. We reveal that the most effective model according to the F1-score is using the Neural Networks technique (0.78, 0.62, 0.51 and 0.46 for the test subsets of the 1, 4, 8 and 12-hour forecasting scenarios, respectively), followed by the Logistic Regression algorithm. For the remaining algorithms, we found F1-score differences between the best and the worse model inversely proportional to the lead time (i.e., differences between models were more pronounced for shorter lead times). Moreover, the Geometric mean and the Log-log score showed similar patterns of degradation of the forecast ability with lead time for all algorithms. The overall higher scores found for the Neural Networks technique suggest this algorithm as the engine for the best

forecasting Early Warning Systems of the city. For future research, we recommend further analyses on the effect of input data composition and on the architecture of the algorithm for full exploitation of its capacity, which would lead to an improvement of model performance and an extension of the lead time. The usability and effectiveness of the developed systems will depend, however, on the speed of communication to the public after an inundation signal is indicated. We suggest to complement our systems with a website and/or mobile application as a tool to boost the preparedness against floods for both decision makers and the public.

Keywords: Flood; forecasting; Early Warning; Machine Learning; Tropical Andes; Ecuador.