Geophysical Research Abstracts
Vol. 21, EGU2019-9692-1, 2019
EGU General Assembly 2019

Short-term Extreme Flow Forecasting in an Tropical Andean Mountain Catchment—Development of a Step-Wise Methodology Based on the Random Forest Algorithm

Paul Muñoz (1,2), Johanna Orellana-Alvear (1,3), Patrick Willems (2), Rolando Célleri (1,4)

(1) Universidad de Cuenca, Departamento de Recursos Hídricos y Ciencias Ambientales, Cuenca, Ecuador, (2) Department of Civil Engineering, Hydraulics Section, KU Leuven, 3001 Leuven, Belgium, (3) Laboratory for Climatology and Remote Sensing, Faculty of Geography, University of Marburg, 35032 Marburg, Germany, (4) Facultad de Ingeniería, Universidad de Cuenca, Cuenca 010150, Ecuador

Flash-flood and drought forecasting has emerged worldwide due to the catastrophic socio-economical impacts these hazards might cause and the expected increase of their frequencies in the future. In mountainous areas, drought and flash-flood forecasting is challenging considering that information other than precipitation and discharge is not commonly available due to budget constraints, remoteness of the study areas and more importantly due to extreme spatio-temporal variability of precipitation. This is especially true for the tropical Andes in South America, which is the longest and widest cool region in the tropics yet its snowline is present from altitudes higher than 5800 meters above the sea level. Consequently, a simple, although useful, approach is the development of precipitation-runoff forecasting models. While process-based models are hard to implement, there is a potential to use machione learnin techniques, such as the random forest algorithm (based on multiple decision trees) due to its simplicity, robustness and capacity to deal with complex data structures. Here a step-wise methodology is proposed to derive parsimonious models accounting for both hydrological functioning of the catchment (e.g., input data, representation of antecedent moisture conditions) and random forest procedures (e.g., sensitivity analyses, dimension reduction, optimal input composition). The methodology was applied to develop short-term forecasting models of varying time duration (4, 8, 12, 18 and 24 hours) for a catchment representative of the Ecuadorian Andes. Results show that derived parsimonious models can reach validation efficiencies (Nash-Sutcliffe coefficient) from 0.761 (4-hour) to 0.384 (24-hour) for optimal inputs composed only by features accounting for 80% of the model's outcome variance. Regardless of the forecasting horizon, model efficiencies were higher for extreme low flows, in contrast, extreme peak flows prediction presented higher model residuals and bias. Improvement in the prediction of extreme peak flows was demonstrated (extreme value analysis) by including precipitation information in contrast to the use of pure autoregressive models.