



Bayesian Networks and Data Driven Models for Estimating Extreme River Discharges Case Study: Magdalena-Cauca Basin, Colombia

Ahmed Nasr (1), Leonardo Alfonso (2), and Oswaldo Morales Nápoles (3)

(1) Faculty of Engineering, Cairo University, Cairo, Egypt (ahmed.ali.nasr@cu.edu.eg), (2) IHE Delft Institute for Water Education, Delft, The Netherlands (l.alfonso@un-ihe.org), (3) Faculty of Civil Engineering and Geosciences, Delft University of Technology, Delft, The Netherlands (O.MoralesNapoles@tudelft.nl)

Floods are among the most threatening natural hazards that exist and the leading cause of deaths due to natural disasters worldwide. The quantification of hazards related to the increment of water levels and consequent flood extent requires an adequate estimation of extreme discharges. Traditionally, extreme river discharges are obtained by univariate frequency analysis of peak discharges that are obtained either from discharge observations or from hydrological models. However, either of these two options requires an important amount of historical data.

The Magdalena-Cauca river basin in Colombia, as in many basins in the developing world, has insufficient amount of available records of discharge observations in several locations. Therefore, the quantification of hazards becomes a challenging problem, particularly in completely ungauged areas. To overcome this problem, we used and proposed different methods that are computationally inexpensive to indirectly estimate extreme discharges from different and easy to obtain data sources. One of them, a stochastic model based on Bayesian Networks (BNs), developed for data-rich basins and previously applied in Europe and the contiguous USA is used in this study. A Bayesian Network is a direct acyclic graph that consists of nodes that represent random variables and arcs that represent the dependence structure between these variables. To validate the performance of the Bayesian Network model in the Magdalena- Cauca basin, two different deterministic data driven models that weren't explored earlier for this purpose, based on lazy learning (k-nearest neighbours) and eager learning (M5 Model Trees) are proposed and used. In addition, the BN model performance is evaluated with regional frequency analysis.

The inputs of all the models utilize the same data, related to topography, land-use characteristics and local climate. In particular, these inputs, seven in total, are: 1) the steepness of the catchment; 2) the catchment area; 3) the annual maximum of daily precipitation and snowmelt; 4) the maximum runoff coefficient of the catchment; 5) the percentage of the catchment area covered by marshes, 6) lakes and 7) built-up areas. All the models establish relations among them and the annual maximum of daily river discharges, in such a way that it can be predicted for different inputs.

Results show that all the models perform generally well in estimating annual peak discharges. All the models perform well when used in big sub-catchments and tend to underperform in small ones. Univariate frequency analysis applied on the peak discharges obtained by all the models show that the output of the BN model outperforms other models in estimating extreme discharges with certain exceedance probabilities. In this context, the BN model is used to evaluate the effect of climate change projection scenarios (RCP4.5 and RCP8.5) on extreme river discharges. Interestingly, the results contradict a recent study based on a distributed hydrological model built with free global products.