EGU23-844, updated on 16 May 2023
https://doi.org/10.5194/egusphere-egu23-844
EGU General Assembly 2023
© Author(s) 2023. This work is distributed under
the Creative Commons Attribution 4.0 License.

Streamflow Estimation in Ungauged Catchments in Brazil using Machine Learning Approaches

Rafael Barbedo, Mino Sorribas, and Walter Collischonn
Universidade Federal do Rio Grande do Sul, Instituto de Pesquisas Hidraulicas, Porto Alegre, Brazil
(rbarbedofontana@gmail.com)

Knowing river flows in space and time is fundamental for several hydrological and environmental applications. One of the greatest challenges in hydrology, however, is having this information at every river stretch, as we can only focus our resources in obtaining measurement at particular sites. Several research initiatives have been developed over the next years to address this problem, a notorious one being the prediction in ungauged basins (PUB) by the International Association of Hydrological Sciences (IAHS).

One of the most used approaches for PUB is using catchment descriptors – such as elevation, slope, land cover, and soil types – in statistical (data-driven) models to estimate hydrological signatures – such as mean annual streamflow, flow-duration curves, and high/low flows. There is a wide range of statistical methods that can be used in this regard, either by grouping catchments of similar characteristics and applying regression equations, using geospatial interpolation techniques, among others. In recent years, regression techniques based on Machine Learning (ML) approaches have been extensively developed, presenting great results in all areas of knowledge. In hydrological sciences, particularly for PUB, the potential of using these techniques is enormous, and yet, they have not been much explored.

In this context, we've built a ML regression modelling pipeline to estimate mean annual flows and low flows, and tested it in several different catchments covering the whole of Brazil, using different models to compare the results. The pipeline consists in (1) collecting environmental data for the catchments, (2) selecting the best descriptors, (3) tunning the hyperparameters of the ML model, (4) evaluating the performance of the model, (5) computing the importance of the predictors, and (6) assessing the uncertainty of the estimations. Also, the pipeline is model-independent, i.e., it can be applied to any ML regression model.

We evaluated results against consistent streamflow data from 1069 gauges spread across the country that cover distinct characteristics, using 100-fold cross validation, obtaining R^2 scores of ~0.8 for mean annual flows and ~0.7 for low flows, for all ML models except multiple linear regression, which didn't present good results. Average and low precipitations were the main drivers for predicting both flow variables, although using these alone didn't yield in good metrics. Other important predictors were linked to soil types, land cover, wetlands, and drainage density. We extrapolated the results to all catchments in Brazil, along with uncertainty estimations.

Acknowledgement:

The authors would like to acknowledge the financial support provided by the Brazilian Coordination for the Improvement of Higher Education Personnel (CAPES) and the Brazilian National Water and Sanitation Agency (ANA), the latter in the context of the project "Technological Cooperation for Hydrological Assessments in Brazil" (grant number: TED-05/2019-ANA). Additional acknowledgements to the Google LLC for making available the Google Earth Engine (GEE) platform, and all data providers for the global products used in this study.