

EGU2020-8596

<https://doi.org/10.5194/egusphere-egu2020-8596>

EGU General Assembly 2020

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



Regional flood frequency estimation for the contiguous USA using Artificial Neural Networks

Valeriya Filipova¹, David Leedal², and Anthony Hammond¹

¹JBA Risk Management, Hydrology & Statistics, United Kingdom of Great Britain and Northern Ireland

(valeriya.filipova@jbarisk.com)

²Kalibrate Technologies Ltd, United Kingdom

We have recently demonstrated the utility of a machine learning-based regional peak flow quantile regression model that is currently providing flood frequency estimation for the re/insurance industry across the contiguous US river network. The scheme uses an artificial neural network (ANN) regression model to estimate flood frequency quantiles from physical catchment descriptors. This circumvents the difficult-to-justify assumption of homogeneity required by alternative 'region of hydrological similarity' approaches. The structure of the model is as follows: the output (dependent) variable is a set of peak flow quantiles where the distributions used to derive the quantiles were parameterised from observations at 4,079 gauge sites using the USGS Bulletin 17C extreme value estimation method (notable for its inclusion of pre-instrumental flood events). The features (regressors) for the model were formed from 25 catchment descriptors covering; geometry, elevation, land cover, soil type and climate type for both the gauged sites and the catchments related to a further 906,000 ungauged sites where peak flow quantile estimation was undertaken. The feature collection requires massive computational resource to achieve catchment delineation and GIS processing of land-use, soil-type and precipitation data.

This project integrates many modelling and computational science elements. Here we focus attention on the ANN modelling component as this is of interest to the wider hydrology research community. We pass on our experience of working with this modelling approach and the unique challenges of working on a problem of this scale.

A baseline multiple linear regression model was generated, as were several non-linear alternative formulations. The ANN model was chosen as the best approach according to a root mean square error (RMSE) criterion. Alternative ANN formulations were evaluated. The RMSE indicated that a single hidden layer performed better than more complex multiple hidden layer models. Variable importance algorithms were used to assess the mechanistic credibility of the ANN model and showed that catchment area and mean annual rainfall were consistently identified as dominant features in agreement with the expectations of domain experts together with more subtle region-specific factors.

The results of this study show that ANN models, used as part of a carefully configured large-scale computational hydrology project, produce very useful regional flood frequency estimates that can

be used to inform flood risk management decision-making or drive further hydrodynamic 2D-modelling and are appropriate to the ever-increasing scale of contemporary hydrological modelling problems.