REGIONAL FLOOD FREQUENCY ESTIMATION FOR THE CONTIGUOUS USA USING ARTIFICIAL NEURAL NETWORKS (ANN).

Valeriya Fillipova[1], David Leedal[2], and Anthony Hammond[1]
1 JBA Risk Management, Hydrology & Statistics, United Kingdom of Great Britain and Northern Ireland (valeriya.filipova@jbarisk.com)
2 Kalibrate Technologies Ltd, United Kingdom

Background

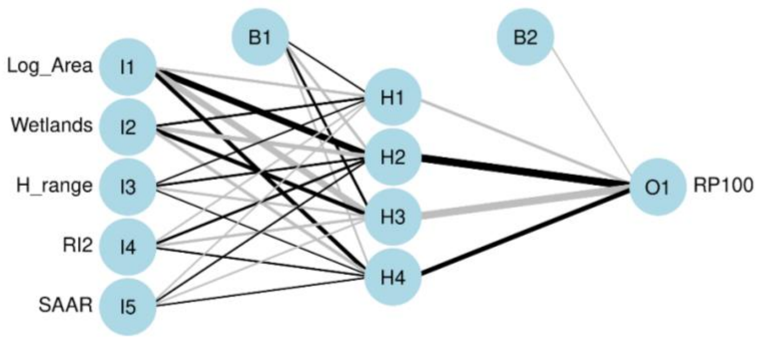
This study utilises machine learning to estimate flood quantiles at ungauged catchments for the contiguous USA, which are later used for developing high resolution 5 metre flood maps. Additionally, we assess the optimal ANN architecture by comparing one hidden layer to multiple hidden layers and evaluate the regional performance of the models. Although several different return levels were calculated, we present results only for RP100 (100-year flood).

Method

For computational reasons, the study area was divided into 15 regions and included a total of 4,079 gauges. These regions represent major watersheds but not necessarily hydrologically similar regions. Each region was processed following a general method that can be described in the following steps:

1. Estimate extreme flow quantiles at randomly chosen gauged sites (75% of the total number of sites) to calibrate the ANN models for every region
2. Extract and select from 25 different catchment descriptors (related to area, elevation, land cover, soil types and climate) to use as the ANN inputs
3. Validate the models on the remaining 25% of catchments
4. Apply the models in ungauged catchments to estimate extreme flow quantiles

In order to determine the required complexity of the ANN models, we fitted single hidden layer (fig. 1), and two and three hidden layer MLP (Multilayer Perceptron) ANNs using the R packages caret (Kuhn 2008) and neuralnet (Fritsch 2019). The models were optimised using a backpropagation algorithm.



I - Input vectors; catchment descriptors

H - Hidden layer or layers; in this case there are 4 nodes in the hidden layer

B1, B2 - Bias vectors

O - Output; calculated flood quantiles

Figure 1: An example of ANN model.

Assessment of model configurations

The results for R2 and RMSE (fig. 2) suggest that models with only one hidden layer perform similarly to deeper networks that have two or three hidden layers. More complex ANN models do not provide any advantage at this point and further testing is needed to determine whether these results are dependent on the optimisation algorithm.

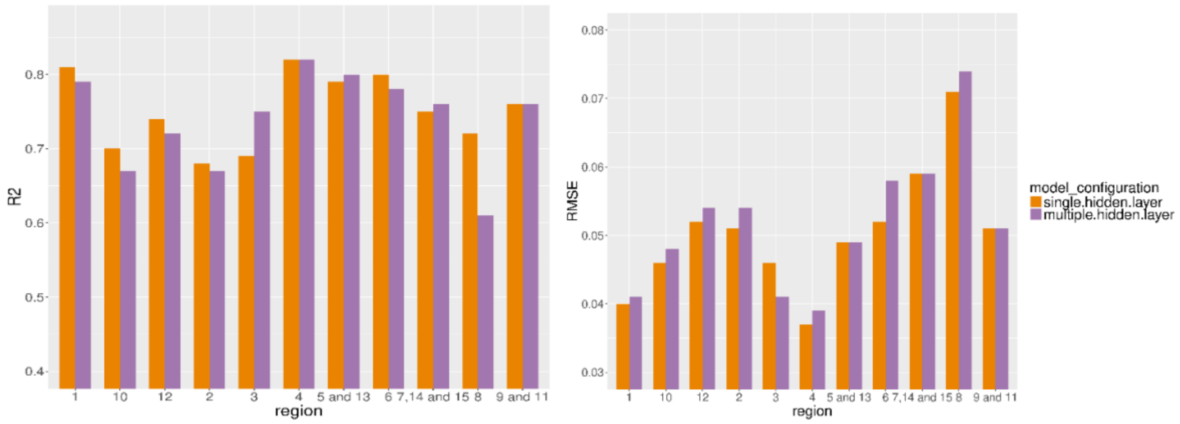


Figure 2: Results extracted from training different ANN models.

Regional Differences

The median percentage error for RP100 is 37%, however there are regional variations (fig. 3). Larger errors are found in catchments that have arid (Köppen index BWh and Bsk) and humid continental (Köppen index Dfa) climates and in catchments with large lake percent. We can hypothesise that there is more uncertainty in fitting the flood frequency curve for these gauges.

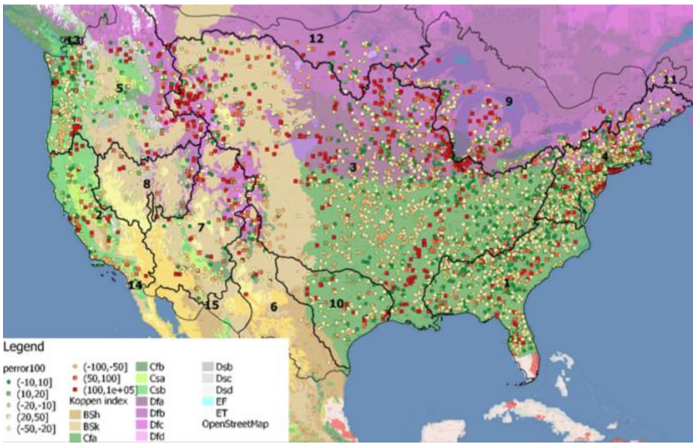


Figure 3 (right): Percentage error for the validation and training sets.

Conclusion

These results demonstrate that ANNs can be used to provide reliable regional flood frequency estimates in a large study area. In particular, the ability of ANNs to model non-linear relationship presents a significant advantage because there is no need for delineating hydrologically similar regions. In addition, simpler one-hidden-layer ANNs can provide similar performance compared to more complex models.

References:

Kuhn, M. 2008. “caret Package.” Journal Of Statistical Software, 28(5), 1–26.

Günther, F. and Fritsch, S. 2010. Neuralnet: Training of neural networks, R J., doi:10.32614/rj-2010-006

Valeryia and team will be available to answer your questions and comments via live chat during the EGU2020 event.