Hydrol. Earth Syst. Sci. Discuss., 12, C2765–C2768, 2015
www.hydrol-earth-syst-sci-discuss.net/12/C2765/2015/

Hydrology and Earth System Sciences
Discussions
Open Access

HESSD
12, C2765–C2768, 2015

Interactive Comment

Interactive comment on "Estimation of flood warning runoff thresholds in ungauged basins with asymmetric error functions" *by* E. Toth

Anonymous Referee #1

Received and published: 22 July 2015

Review of the manuscript *Estimation of flood warning runoff thresholds in ungauged basins with asymmetric error functions* by *E. Toth*, submitted to the Journal *Hydrology and Earth System Sciences*

In this manuscript, the author estimates the median annual flood discharge $Q_{T=2yr}$ in a set of Italian catchments in a cross-validation framework. The method chosen for the regionalization is an Artificial Neural Network (ANN), where the loss function is chosen to be asymmetric, in this case penalizing the overestimation errors.

While there is value in the study performed by itself, as the method applied (ANN with non-symmetrical loss function) seems to be novel in this particular hydrological application, the paper needs to undergo **major changes** in order to be useful for the hydrological community in general and to be accepted for publication in HESS.

General Comments

- What are the advantages of using a biased estimate of $Q_{T=2yr}$?

In principle, there is no possible criticism to the way the methodology has been applied, as all the required steps and precautions have been taken when calibrating –or *training*–, validating and testing the ANN with asymmetric loss function. What concerns me is the actual usefulness of applying a method that inherently produces biased estimates of a quantity of interest (here, the median annual flood discharge $Q_{T=2yr}$).

For example, just by thinking about (probably) the simplest approach possible to the problem proposed. One could apply multiple linear regression as a prediction model for $Q_{T=2yr}$ with the 3 first Principal Components (PC's) of the catchment attributes as explanatory variables. Expressions for the uncertainty bounds of the predicted $Q_{T=2yr}$ are available in many statistical handbooks. In fact, one can estimate the full conditional distribution $f_{Q_2}(q|PC_1, PC_2, PC_3)$. With this information, the modeler can choose whatever level of acceptable risk (or type I error), and give e.g. the 95% quantile for the estimated $Q_{T=2yr}$. This kind of statement is also particularly useful in a floods warning framework, as the one discussed in this manuscript, as the modeler is transmitting not only an estimate of $Q_{T=2yr}$, but also a probabilistic assessment of it, even if it is also an estimate.

In the methodology proposed, the estimate will be systematically biased (it's in the nature of the non-symmetrical loss function), and the modeler will have no knowledge about the type I error associated to it, while in the simpler methodology described above there is a full probabilistic description of the estimate of $Q_{T=2yr}$.

I am not familiar myself with ANN's, but I am sure that there are numerical methods –

maybe Monte Carlo simulations – to derive uncertainty bounds of the estimates for the case of a symmetrical quadratic loss function (unbiased, and variance minimizer) .

In this case, what will be the difference(s) between the proposed approach and taking an upper (e.g 95%) uncertainty bound? What the advantages, if any? These points should be explicitly addressed in the manuscript.

- Regional Flood Frequency analysis is not regression

In a couple of locations in the text, page 6014 line 14–29 and page 6030 lines 10–18, there seems that there is the direct association between Regional Flood Frequency analysis to Regression with catchment attributes (regression or related techniques like ANN's).

The reader should not get this idea. Even in one of the references that the author uses, Bloeschl et al. 2013, the central point is to remove this –on the other hand very usual– association between Regional methods in hydrology and "just" regression.

The author could rephrase the aforementioned sections, by saying that Regression-like techniques (where ANN's can be included) are one of the many possible approaches for predicting floods in ungauged basins.

- Relative error could be also very valuable

For assessing the performance of several variants of the proposed method, the measures MAE and RMSE are proposed, both functions of the error ϵ. Given the large range of discharges considered in the study, it could be also very valuable to report additionally boxplots of the relative errors $\epsilon_{rel} = (Q_{obs} - Q_{pred})/Q_{obs}$, and the analogous measures Root Mean Squared *Normalized* Error (RMSNE), and Mean Absolute *Normalized* Error (MANE), both functions of ϵ_{rel}.

Full Screen / Esc

Printer-friendly Version

Interactive Discussion

Discussion Paper

Detailed Comments

Page 6013, Line 27: is “real-world” here “real-time”?

Page 6017, Lines 17–19: The author defines here the error as the observed minus the predicted value. To my knowledge, in runoff prediction in ungauged basins, it is almost a consensus to define error as predicted minus observed. If the author wants to define it here inversely, a stronger warning to the reader should be given, in order to avoid confusions.

Pages 6021–6023: Maybe adding an schematic figure with the structure of the selected ANN could help the reader.

Page 6027, line 4: is here “scour” the Q_2?

Page 6027, line 9: I think “prudence” is not the right word here. Maybe “tendency to over/underestimate”?

Page 6029, line 7: ... the errors are not negligible...

Page 6031, line 13: Again, “prudentially” is not the right word here.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 12, 6011, 2015.

HESSD

12, C2765–C2768, 2015

Interactive Comment