Hydrol. Earth Syst. Sci. Discuss.,
https://doi.org/10.5194/hess-2020-430-AC1, 2021

Hydrology and Earth System Sciences
Open Access
Discussions

Interactive comment on "Resampling and ensemble techniques for improving ANN-based high streamflow forecast accuracy" *by* Everett Snieder et al.

Everett Snieder et al.

usman.khan@lassonde.yorku.ca

Received and published: 1 January 2021

AC3-General. This paper explored the potential of data-driven models such as ANN for improving the accuracy of high flow estimation through integrating resampling and ensemble techniques. For this exercise, three resampling techniques: random undersampling (RUS), random oversampling (ROS), and SMOTER; and four ensemble techniques: randomized weights and biases, bagging, adaptive boosting (AdaBoost), least-squares boosting (LSBoost) were systematically combined to show the improvement in the forecast accuracy in terms of reducing the timing and amplitude error. This paper used the hourly river stage data along with other meteorological data collected

from Bow and Don River basins, Canada to demonstrate the proposed modelling approaches. While many previous papers have already reported the potential application of several ensembles and resampling methods to improve the forecast accuracy of data-driven models, this paper claims that the implementation of ROS, and new approaches for SMOTER, LSBoost, and SMOTER-AdaBoost are the new addition. The paper is well written and interesting to the researchers of hydrology. However, the paper needs some more clarity, which I have marked below.

RC3-General. Thank you for the positive comments on our manuscript. We have addressed the comments for better clarity below.

RC3-1. Since the variation in the streamflow is evident, I do not know the usage of word imbalance is correct or not in this context.

AC3-1. The term 'imbalance' is widely used in machine learning literature to describe imbalance of labels for classification problems [1]. Many studies have extended the use of this term to regression problems [2]. In simple terms, "imbalanced" is defined as the "existence of an over-representation of a given class(es) or numeric value interval(s), over another" [3].

AC3-2. You have selected the 80th percentile to segregate the peak flow data from the entire dataset. I agree that ANN models are completely dependent on the choice of data. Still, it would be interesting to see the effect of selecting any other values (70th and 90th percentile), at least for a few cases.

RC 3-2. Thank you for this recommendation. A formal grid-search and subsequent discussion on HF thresholds (ranging from 50 to 90th percentile flow) has been added to the revised manuscript (please see Fig 11 and associated text the attached supplement). In this new analysis, we compare the Bagging ensemble with SMOTER-Bagging in order to quantify the relative effects of resampling for various high flow thresholds.

RC3-3. How to choose the model HF, TF for the unknown data for the future forecast?

AC3-3. It is not necessary to know the HF or TF of the future, unknown data - we assume these values to be constant for the simulations. With sufficient historic data, the temporal variance of the high flow threshold is assumed to be negligible. This can be confirmed using statistical bootstrapping: bootstrapping the 80th percentile flow (n = 10,000) returns a standard deviation of 0.0046 and 0.0031 for the Bow and Don River, respectively. Alternatively, as recommended in the text, a high flow threshold could be chosen based on physical characteristics (such as the stage at which a river exceeds its banks).

RC3-4. How do you define highly imbalanced flow datasets?

AC3-4. “Highly imbalanced” is commonly used in literature [4-7]; however, the authors are not aware of a quantitative definition. For that reason, the term “highly” has been removed from the text and instead we simply state “imbalanced”.

RC3-5. Line 25: “One cause of low model accuracy on high flows is the scarcity of representative sample observations available with which to train such models.” Add one or two references

AC3-5. We appreciate this recommendation and have added the following reference [8].

RC3-6. Line 30: “As a result, studies that assess models using traditional performance metrics risk overlooking deficiencies in high flow performance.” I agree with this point. However, separating high flow hydrograph from the dataset and evaluate the model performance using the traditional indices would still reveal the actual model performance. This should be mentioned.

AC3-6. We appreciate this insightful suggestion. In fact, this is exactly what we do in our manuscript; a fixed threshold is used to separate high and typical stage values. This simple approach was favoured over more complex methods of high flow performance assessment such as for hydrograph extraction, which isolate separate hydrological events and baseflow. Large hydrological events during low baseflow periods, which occur in highly seasonal watersheds such as the Bow, are not necessarily relevant to flood forecasting, as the event peak stage may not attain flood-level stage. Comparatively, a small event during high baseflow may reach a higher peak stage. Thus, we found that simply separating high and typical stages based on a fixed threshold is more appropriate than more complex hydrograph extraction methods for applications such as high stage forecasting. We will include a discussion on this in the updated manuscript. In our previous research [9, 10] on improving high flow performance, we demonstrate hydrograph extraction and the use of more complex, event-specific performance measures; however, this approach was abandoned for the aforementioned reasons.

RC3-7. Line 40: Improving the accuracy of high flow forecasts has been the focus of many studies. Several studies have examined the use of preprocessing techniques to improve model performance. Reference is required. I would suggest adding Kasiviswanathan et al. (2015).

AC3-7. Thank you for this suggestion, we have added this citation in the manuscript. Preprocessing methods used in several works cited in the original draft: [11] evaluates statistical transformations and [12] evaluates a multi-model approach based on cluster-based preprocessing that evaluate preprocessing techniques for improving high flow performance.

RC3-8. Line 85: The Bow and Don River watersheds are the focus of this research. You may consider deleting this line.

AC3-8. Thank you for this recommendation. The focus of the research is indeed the ensemble and resampling methods. The sentence has been modified as follows: “The Bow and Don Rivers are featured as case studies in this research to evaluate methods for improving the accuracy of high stage data-driven forecasts”

RC3-9. Authors refer to the stage as flow. This should be corrected.

AC3-9. Thank you for this recommendation. In our original submission, we refer to flow forecasting but opt to use stage data in our models, as stage in forecasts are more useful than flow for indicating flooding and stage-discharge curves are readily available for both rivers. We understand how this may cause confusion and have changed all uses of ‘flow’ to ‘stage’, where appropriate.

RC3-10. It would be interesting to see how the peak flow of Bow River in the years 2005 and 2013 forecasted by these models. Similarly, for Don River.

AC3-10. The results for the Bow River for 2005 were included in the original manuscript (Fig. 3). Individual performance for this calendar year is reported in Table RC3-1 (please refer to attached supplement).

We currently do not have access to the 2013 Bow River data but will consider adding it to the analysis for the final revised paper. For the Don River, the high resolution data needed for our research is unavailable for 2005 and 2013 for some of the hydrometeorological stations used in this research.

However, note that the current dataset is highly imbalanced as shown in Fig. 2 in the original manuscript. Thus, we think the data we used is sufficient for the analysis to demonstrate the effects of resampling/ensemble with respect to the data imbalance problem, even without the additional data for 2005 and 2013 for each river.

RC3-11. Why stage data, why not directly for the discharge data?

AC3-11. Stage data is used for several reasons. Foremost, it is more relevant, compared to discharge, for an early flood warning system, which is the anticipated application of this research. Next, stage is measured directly, whereas flow is calculated based on an uncertain stage-discharge relationship. References to flow or discharge have been corrected to flow, following RC3-9.

References [1] M. Galar, A. Fernandez, E. Barrenechea, H. Bustince, and F. Herrera, “A review on ensembles for the class imbalance problem: Bagging-, boosting-, and

hybrid-based approaches,” IEEE Transactions on Systems, Man and Cybernetics Part C: Applications and Reviews, vol. 42, no. 4. pp. 463–484, 2012.

[2] P. Branco, L. Torgo, and R. P. Ribeiro, “A survey of predictive modeling on imbalanced domains,” ACM Computing Surveys, vol. 49, no. 2. Association for Computing Machinery, p. 31, 01-Aug-2016.

[3] N. Moniz, P. Branco, L. Torgo, and B. Krawczyk, “Evaluation of Ensemble Methods in Imbalanced Regression Tasks,” Proc. Mach. Learn. Res., vol. 74, pp. 129–140, 2017.

[4] M. Galar, A. Fernández, E. Barrenechea, and F. Herrera, “EUSBoost: Enhancing ensembles for highly imbalanced data-sets by evolutionary undersampling,” Pattern Recognit., vol. 46, no. 12, pp. 3460–3471, Dec. 2013.

[5] G. Haixiang, L. Yijing, J. Shang, G. Mingyun, H. Yuanyue, and G. Bing, “Learning from class-imbalanced data: Review of methods and applications,” Expert Systems with Applications, vol. 73. Elsevier Ltd, pp. 220–239, 01-May-2017.

[6] P. Branco, L. Torgo, and R. P. Ribeiro, “A survey of predictive modeling on imbalanced domains,” ACM Computing Surveys, vol. 49, no. 2. Association for Computing Machinery, p. 31, 01-Aug-2016.

[7] O. Loyola-González, J. F. Martínez-Trinidad, J. A. Carrasco-Ochoa, and M. García-Borroto, “Study of the impact of resampling methods for contrast pattern based classifiers in imbalanced databases,” Neurocomputing, vol. 175, pp. 935–947, 2016.

[8] N. Moniz, P. Branco, and L. Torgo, “Resampling strategies for imbalanced time series forecasting,” Int. J. Data Sci. Anal., vol. 3, no. 3, pp. 161–181, May 2017.

[9] E. J. Snieder, “Artificial Neural Network-Based Flood Forecasting: Input Variable Selection and Peak Flow Prediction Accuracy,” Master’s Thesis, York University, 2019. Available at: http://hdl.handle.net/10315/36792 [10] U. T. Khan and E. Snieder, “Assessment of peak flow error metrics and error weights for data-driven riverine flow fore-

casting models.," in Geophysical Research Abstracts, Proceedings of the European Geosciences Union General Assembly, 2019, vol. 21.

[11] K. P. Sudheer, P. C. Nayak, and K. S. Ramasastri, "Improving peak flow estimates in artificial neural network river flow models," Hydrol. Process., vol. 17, no. 3, pp. 677–686, Feb. 2003.

[12] W. Wang, P. H. A. J. M. V. Gelder, J. K. Vrijling, and J. Ma, "Forecasting daily streamflow using hybrid ANN models," J. Hydrol., vol. 324, no. 1–4, pp. 383–399, 2006.

Please also note the supplement to this comment:
https://hess.copernicus.org/preprints/hess-2020-430/hess-2020-430-AC1-supplement.pdf

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., https://doi.org/10.5194/hess-2020-430, 2020.