Regional median flood estimation with generalized additive models: model selection across durations

Danielle M. Barna[a,c], Kolbjørn Engeland[a], Thomas Kneib[e], Thordis L. Thorarinsdottir[b,d], Chong-Yu Xu[c]

[a] *Norwegian Water Resources and Energy Directorate, P.O. Box 5091 Majorstua, NO-0301 Oslo, Norway.*
[b] *Norwegian Computing Centre, P.O. Box 114 Blindern, NO-0314 Oslo, Norway*
[c] *Department of Geosciences, University of Oslo, P.O. Box 1047 Blindern, NO-0316 Oslo, Norway.*
[d] *Department of Mathematics, University of Oslo, Oslo, Norway.*
[e] *Chair of Statistics, University of Göttingen, Göttingen, Germany.*

Abstract

1. Introduction

A common challenge in engineering design for retention-based applications is the need for duration-specific design values at ungauged locations. Since these applications focus on total storage capacity, they often require frequency estimates of average streamflow over pre-determined durations. Moving-window analyses (alternatively termed sustained flood flow, N-day flood or flood-duration-frequency analyses) enable frequency analysis of the period with the highest average streamflow over n consecutive hours; see, for example, Lamontagne et al. (2012); Kennedy et al. (2014); Lind et al. (2020); Barna et al. (2023); Ouarda et al. (2006); Javelle et al. (2002); Balocki and Burges (1994). Frequency analysis is typically performed separately for each pre-determined duration. Some approaches (e.g., Barna et al. (2023); Ouarda et al. (2006); Javelle et al. (2002)) attempt to scale across durations by fitting an 'average' distribution and estimating a scaling parameter. However, this scaling relies on strict assumptions that are sometimes unsupported by empirical analysis, especially when applied over broad regions (Barna et al., 2023; Kennedy et al., 2014). Separate frequency analyses for each duration are the more common approach in regional applications and are the focus of this paper.

To extend these estimates to ungauged locations, regression models are typically used. Moving-window analyses often model a wide range of durations (e.g., Lamontagne et al. (2012) models durations from 1 to 30 days) by developing a single best parametric regression model and re-estimating the coefficients for each duration. Relying on a single parametric form is practical and is theorized to help maintain consistency across durations (Lind et al., 2020).

However, this approach assumes that the same parametric form applies to all durations. This may not fully account for the fact that different processes drive high average streamflow over short vs long durations. For example, high average streamflow over short durations is often driven by intense rainfall events, while longer durations are influenced by processes like sustained snowmelt. These processes, and their interactions

Email address: daba@nve.no (Danielle M. Barna)

with catchment descriptors used in regression modeling, may be so fundamentally different that simply re-estimating coefficients may not capture the variations in relationships across durations.

Despite this, the assumption of a single parametric form is rarely challenged due to the complexities of developing and testing regional regression models on many durations. Instead, practitioners sometimes limit event types in moving-window analyses (Lamontagne et al., 2012) and/or allow for differing amounts of cross-correlation between stations when re-estimating regression coefficients for each duration (Lamontagne et al., 2012; Kennedy et al., 2014; Lind et al., 2020). The latter typically requires a two-step estimation procedure as high cross-correlations can occur between gauging stations in moving-window analyses, particularly at longer durations (Kennedy et al., 2014).

This paper investigates whether use of a single parametric form limits the predictive performance of a regional regression model applied to multiple durations. We examine whether there are statistically significant differences in consistency, predictive accuracy, reliability, and fitted relationships across different durations and different types of regression models.

Of course, establishing a need for different models at different durations is of limited usefulness if it is not practical to build the models. Developing a regional regression model is a substantial effort. Hydrologic datasets of catchment and climate characteristics often contain a large number of potential—and potentially collinear—predictors, making predictor selection a challenge. Improving model fit often involves selecting appropriate polynomial terms and transformations of the chosen predictors, which can be a labor-intensive process. And models used to produce design values must be tested on a wide variety of performance metrics that assess predictive accuracy, reliability, and uncertainty at ungauged locations.

Therefore, implementation potential played a large role when considering what types of regression models to assess. We considered several options, all of which have established use in regional hydrologic studies.

Fully parametric regression models, like linear, log-linear, or nonlinear models, are the current state of practice in moving window analyses and are well-established in regional flood frequency analysis in general (Robson and Reed, 1999; Pandey and Nguyen, 1999). These models are straightforward to interpret and assess but require significant work to develop an appropriate parametric form. Additionally, they rely on careful selection of uncorrelated predictors, a major part of the analysis. In contrast, fully data-driven (non-parametric) models, such as the machine learning models of, for example, Aziz et al. (2014); Laimighofer et al. (2022b); Jarajapu et al. (2022); Haddad and Rahman (2020) or Esmaeili-Gisavandani et al. (2023), are highly adaptable, can handle large, collinear predictor sets, and do not need a defined parametric form. They are powerful methods when predictive accuracy is the only concern. But these models generally lack the distributional assumptions needed for uncertainty and reliability assessments. A middle ground can be found in semi-parametric regression models, such as generalized additive models (GAMs) (see, e.g., Chebana et al. (2014); Msilini et al. (2022); Rahman et al. (2018)), which model predictor-response relationships non-parametrically while still specifying the response distribution, enabling full statistical assessments of accuracy, reliability, and uncertainty.

We develop a GAM for regional median flood prediction at multiple durations. Some moving window analyses use the regression model to predict a distribution parameter for frequency analysis (Lamontagne et al., 2012; Lind et al., 2020), while others predict flood quantiles directly (Kennedy et al., 2014). In this investigative study, we focus on predicting the median annual maximum flood (index flood, as in Dalrymple (1960)) which can serve as both a quantile and, through the reparametrization of Castro-Camilo et al. (2022), the location parameter of the generalized extreme value distribution.

Just as with a standard parametric model in moving window analyses, we specify the GAM once and re-estimate at each duration. However, unlike the parametric model, the GAM's explanatory component–the predictor-response relationship–adapts as we re-estimate, offering a data-driven comparison and visualization of fitted relationships and estimation uncertainty across durations. For the predictive component of our model assessment, we compare the models' predictions to observed data. The models' predictive power at ungauged locations is assessed through a cross validation study, where reliability and accuracy are evaluated through comparison between predictions and holdout data. Finally, predictions at ungauged locations are assessed for statistically significant deviations from duration consistency.

In our context (design values for retention-based applications) reliability and uncertainty analyses are highly relevant. As noted earlier, this limits our reliance on machine learning. However, machine learning models are powerful and popular methods. For the sake of completeness, we include machine learning in our analysis to the extent possible. In addition to the GAM, we develop also a gradient-boosted tree ensemble (XGBoost) for median flood prediction and use it to provide a benchmark value for predictive accuracy. XGBoost has established use in hydrology (Zounemat-Kermani et al., 2021) and is applied in, for example, Laimighofer et al. (2022a) and Ni et al. (2020). Additionally, studies have shown tree-based models like XGBoost to be effective in identifying small, non-redundant predictor sets from larger, sometimes highly correlated, predictor sets (see, e.g., Alsahaf et al. (2022), Galelli and Castelletti (2013), and Prasad et al. (2017)). We implement the algorithm of Galelli and Castelletti (2013) and assess its effectiveness for our specific study.

Our study area is Norway, where we compare the GAM to an existing index flood model for Norwegian catchments. The models are compared on a set of durations ranging from 1 hour to 30 days. Shorter durations are relevant to Norway's many small hydropower plants, while longer durations apply to larger reservoirs. The focus on annual maxima aligns with Norwegian flood guidelines. The following research questions will be addressed:

1. Can the GAM achieve comparable or improved performance compared to the benchmark model across different durations?

2. Can we identify and describe duration-specific differences in how catchment covariates influence the median flood? How impactful are these differences? (i.e. if we ignore them, what is the impact on predictive performance?).

The remainder of the paper is organized as follows: section ?? introduces the flood data and catchment descriptors. Section ?? presents an outline of the study design. Section ?? presents the GAM used in this study and summarizes the chosen predictor selection approach. This section also summarizes the two reference models and the evaluation methods used to assess all models in the study. The results section ?? presents the predictive performance and model reliability results as well as the functional relationships identified by the GAM. The paper finishes with a discussion (section ??) and conclusions (section ??).

References

Alsahaf, A., Petkov, N., Shenoy, V., Azzopardi, G., 2022. A framework for feature selection through boosting. Expert Systems with Applications 187, 115895.

Aziz, K., Rahman, A., Fang, G., Shrestha, S., 2014. Application of artificial neural networks in regional flood frequency analysis: a case study for australia. Stochastic environmental research and risk assessment 28, 541–554.

Balocki, J.B., Burges, S.J., 1994. Relationships between n-day flood volumes for infrequent large floods. Journal of Water Resources Planning and Management 120, 794–818.

Barna, D.M., Engeland, K., Thorarinsdottir, T.L., Xu, C.Y., 2023. Flexible and consistent flood–duration–frequency modeling: A bayesian approach. Journal of Hydrology 620, 129448.

Castro-Camilo, D., Huser, R., Rue, H., 2022. Practical strategies for generalized extreme value-based regression models for extremes. Environmetrics , e2742.

Chebana, F., Charron, C., Ouarda, T.B., Martel, B., 2014. Regional frequency analysis at ungauged sites with the generalized additive model. Journal of Hydrometeorology 15, 2418–2428.

Dalrymple, T., 1960. Flood-Frequency Analyses. Manual of Hydrology Part 3. Flood-flow techniques. Usgpo 1543-A, 80. URL: `http://pubs.usgs.gov/wsp/1543a/report.pdf`.

Esmaeili-Gisavandani, H., Zarei, H., Fadaei Tehrani, M.R., 2023. Regional flood frequency analysis using data-driven models (m5, random forest, and anfis) and a multivariate regression method in ungauged catchments. Applied Water Science 13, 139.

Galelli, S., Castelletti, A., 2013. Tree-based iterative input variable selection for hydrological modeling. Water Resources Research 49, 4295–4310.

Haddad, K., Rahman, A., 2020. Regional flood frequency analysis: evaluation of regions in cluster space using support vector regression. Natural Hazards 102, 489–517.

Jarajapu, D.C., Rathinasamy, M., Agarwal, A., Bronstert, A., 2022. Design flood estimation using extreme gradient boosting-based on bayesian optimization. Journal of Hydrology 613, 128341.

Javelle, P., Ouarda, T.B., Lang, M., Bobée, B., Galéa, G., Grésillon, J.M., 2002. Development of regional flood-duration–frequency curves based on the index-flood method. Journal of Hydrology 258, 249–259.

Kennedy, J.R., Paretti, N.V., Veilleux, A.G., 2014. Methods for Estimating Magnitude and Frequency of 1-, 3-, 7-, 15-, and 30-day flood-duration flows in Arizona. Technical Report. US Geological Survey.

Laimighofer, J., Melcher, M., Laaha, G., 2022a. Low-flow estimation beyond the mean–expectile loss and extreme gradient boosting for spatiotemporal low-flow prediction in austria. Hydrology and Earth System Sciences 26, 4553–4574.

Laimighofer, J., Melcher, M., Laaha, G., 2022b. Parsimonious statistical learning models for low-flow estimation. Hydrology and Earth System Sciences 26, 129–148.

Lamontagne, J.R., Stedinger, J.R., Berenbrock, C., Veilleux, A.G., Ferris, J.C., Knifong, D.L., 2012. Development of regional skews for selected flood durations for the Central Valley Region, California, based on data through water year 2008. Technical Report. US Geological Survey.

Lind, G.D., Lamontagne, J.R., Stonewall, A.J., 2020. Development of regional skew coefficients for selected flood durations in the Columbia River Basin, northwestern United States and British Columbia, Canada. Technical Report. US Geological Survey.

Msilini, A., Charron, C., Ouarda, T.B., Masselot, P., 2022. Flood frequency analysis at ungauged catchments with the gam and mars approaches in the montreal region, canada. Canadian Water Resources Journal/Revue canadienne des ressources hydriques 47, 111–121.

Ni, L., Wang, D., Wu, J., Wang, Y., Tao, Y., Zhang, J., Liu, J., 2020. Streamflow forecasting using extreme gradient boosting model coupled with gaussian mixture model. Journal of Hydrology 586, 124901.

Ouarda, T.B., Cunderlik, J.M., St-Hilaire, A., Barbet, M., Bruneau, P., Bobée, B., 2006. Data-based comparison of seasonality-based regional flood frequency methods. Journal of Hydrology 330, 329–339.

Pandey, G.R., Nguyen, V.T.V., 1999. A comparative study of regression based methods in regional flood frequency analysis. Journal of Hydrology 225, 92–101.

Prasad, R., Deo, R.C., Li, Y., Maraseni, T., 2017. Input selection and performance optimization of ann-based streamflow forecasts in the drought-prone murray darling basin region using iis and modwt algorithm. Atmospheric Research 197, 42–63.

Rahman, A., Charron, C., Ouarda, T.B., Chebana, F., 2018. Development of regional flood frequency analysis techniques using generalized additive models for australia. Stochastic Environmental Research and Risk Assessment 32, 123–139.

Robson, A., Reed, D., 1999. Flood Estimation Handbook. Vol. 3: Statistical Procedures for Flood Frequency Estimation. Institute of Hydrology.

Zounemat-Kermani, M., Batelaan, O., Fadaee, M., Hinkelmann, R., 2021. Ensemble machine learning paradigms in hydrology: A review. Journal of Hydrology 598, 126266.