ISBN: 978-1-7636843-3-1

Flood Frequency Analysis: Exploring the Role of Statistical Models in Engineering Education and Practice

Laura Rima[1], Ridwan S M H Rafi[2], Ataur Rahman[3] and Khaled Haddad[4]

[1]PhD student, School of Engineering, Design and Built Environment, Western Sydney University, Penrith, NSW 2751, Australia
[2]Research Assistant, North South University, Bangladesh
[3]Professor of Water and Environmental Engineering, School of Engineering, Design and Built Environment, Western Sydney University, Penrith, NSW 2751, Australia
[4]Independant Researcher, Sydney, Australia
Corresponding author's E-mail: 19274818@student.westernsydney.edu.au

Abstract

Flood presents a significant risk to communities globally as flood damage is increasing due to climate change. To effectively reduce flood damage, scientists use a term 'design flood', which refers to a flood discharge linked to annual exceedance probability. This study investigates the role of several statistical models adopted in design flood estimation using flood and catchment data from 88 catchments in eastern Australia. The first objective of this study is to compare the effectiveness of two statistical modeling approaches, generalised additive models (GAM) and log-log regression in estimating flood quantiles. Generalised Extreme Value (GEV) distribution with L-moments was employed where the mean, the coefficient of variation and the coefficient of skewness were adopted as the response variables and the catchment characteristics as predictor variables. This study reveals that GAM performs better than log-log regression technique in capturing the variability of flood quantile estimates. The second objective of this study is to illustrate the learning aspects of the adopted statistical models. It is found that most of the students do not understand the fundamentals of these statistical modelling techniques and often reach inappropriate conclusions. The findings of this study will assist students and junior researchers to understand the assumptions related to statistical flood modelling approaches and how this affects decision making in sustainable flood plain management.

Keywords: Flood, sustainability, education, probability, learning.

1. INTRODUCTION

Floods are highly destructive natural disasters that lead to loss of life and cause severe damage to infrastructure, agriculture, and property (Stojkovic et al., 2017). The increasing frequency and intensity of flooding is linked to climate change resulting in tragic economic and social consequences (Hartmann et al., 2013). Hence, effective flood management has become a key priority for many regions in the world. Design flood is a significant concept in flood management, which refers to a specific flood magnitude associated with a given annual exceedance probability (AEP). It plays a vital role in infrastructure development and flood risk management. Accurate flood estimation is fundamental for reducing these risks and protecting communities. Several techniques have been used to estimate design floods by analysing annual maximum flood (AMF) data, catchment characteristics and predicting flood quantiles. As a result, flood frequency analysis (FFA) is a widely applied approach for estimating design floods, which rely on sufficient streamflow data (in terms of both the quantity and quality) at sites of interest. In ungauged catchments when data is limited or unavailable, regional flood frequency analysis (RFFA) in used (Souaissi et al., 2023; Micevski et al., 2015). Generalised Additive Models (GAM) and log-log regression models have showed efficacy in many RFFA studies (Singh et al., 2021; Rahman et al., 2018; Chebana et al., 2014; Yan et al., 2019).

Beside the technical analysis, many civil engineering students struggle with understanding water engineering concepts, especially in advanced hydrology courses. These courses are becoming more important due to the increasing frequency of floods (Shrestha et al., 2018). Ruddell et al. (2015) provided a historical overview of engineering hydrology education, highlighting the challenges encountered in the hydrology education in the 21st century, such as the need for global teamwork and shared learning tools. Shrestha et al. (2018) discussed different techniques and IT tools aimed at improving student engagement and learning in hydrology education. They highlight successful examples based on effective teaching methods. Chowdhury (2019) discussed challenges in engineering education, including the growing demand for work-based learning, online education, and aligning educational programs with industry requirements.

On the other hand, women, particularly doing PhD studies, face several challenges that can affect their ability to finish in time. These include personal issues related to marriage, family, and health, as well as problems with funding, and research opportunities (Maher et al., 2004; Islam and Pavel, 2011). These challenges make it harder to manage multiple responsibilities and can lead to feelings of isolation and stress. As a result, some women may choose to leave their doctoral studies before completion. Gill et al. (2018) identified specific obstacles faced by women engineers and proposed educational strategies to help them succeed in their careers. They underlined the need to improve workplace culture to promote fairness, diversity, and retain more women in engineering. Arthur et al. (2022) studied the experiences of women engineering students in cooperative programs and emphasised the importance of supportive relationships in helping women learn and grow in engineering. Mbonyiryivuze et al. (2023) reviewed the challenges faced by women PhD candidates and suggest that they should be provided with extra counseling support. Universities and supervisors also should create a better environment to help women succeed and complete their research degrees.

This study briefly explores and compare the application of GAM and log-log regression approaches in RFFA using Generalised Extreme Value (GEV) distribution with L-moments. Furthermore, this study examines the educational journey of learning these models, focusing on the challenges faced by students, particularly women in engineering, and the skills learnt through the PhD research journey. It explores the educational implications of statistical flood modeling in engineering practices. It also highlights the difficulties faced by women in balancing academic and family responsibilities. By sharing author’s experience, this study aims to enhance learning outcomes.

The findings of this research are intended to provide valuable insights for both academic and professional audiences. This study aims to address the comparative effectiveness of GAM and log-log regression in flood quantile estimation while identifying educational challenges faced by the first author. By combining technical analysis with educational insights, this research seeks to contribute to more accurate and sustainable flood management practices and enhance the learning experience for engineering students, particularly women in the field.

2. STUDY AREA AND DATA

This study examines 88 catchments in New South Wales (NSW), Australia, with catchment areas ranging from 8 to 1010 km² and an average size of 325 km². The record lengths for AMF data span from 25 to 89 years, with an average of 48 years. Eight key catchment characteristics were used as predictor variables: catchment area (Area), design rainfall intensity ($I_{6,2}$), mean annual rainfall (MAR), shape factor (SF), mean annual evapotranspiration (MAE), stream density (SDEN), slope (S1085), and forest coverage (Forest). The GEV distribution with L-moments was used to derive flood quantiles for AEPs of 1 in 2, 5, 10, 20, 50, and 100, denoted as Q_2, Q_5, Q_{10}, Q_{20}, Q_{50}, and Q_{100}, respectively. Two statistical modeling approaches, GAM and log-log regression, were employed to estimate the parameters of the GEV distribution and predict flood quantiles. GAM, known for their flexibility, were used to model non-linear relationships between the predictor variables and response variables, such as the mean (equivalent to GEV location parameter), coefficient of variation (LCV from L-moments equivalent to GEV scale), and coefficient of skewness (LSK from L-moments, equivalent to GEV shape) of flood data. The models' performances were then compared based on several evaluation statistics to assess their ability to capture the variability of the flood quantiles.

3. METHODOLOGY

3.1. GEV Distribution with L-moments:

This study used GEV distribution with L-moments, as recommended by Hosking and Wallis (1997), to model extreme events like floods, known for their robustness against outliers. The GEV distribution, introduced by Jenkinson (1955), is given by:

$$F(x) = \exp\left\{ -\left[1 - \frac{\kappa(x-\mu)}{\alpha} \right]^{1/\kappa} \right\} \tag{1}$$

where μ is the location parameter, α the scale parameter, and κ the shape parameter which are based respectively on the mean, variance, and skewness of the data, respectively.

3.2. Generalised Additive Model (GAM):

GAM is a statistical technique that can capture non-linear relationships between predictor variables and the response variable while maintaining an additive structure. It allows for greater flexibility in capturing complex patterns in the data. Mathematically, a GAM can be expressed as follows:

$$g(Y) = b_0 + \sum_{j=1}^{p} f_j\left(x_{i,j}\right) + \varepsilon_i \tag{2}$$

Where Y_i is the response variable, b_0 is a constant term, $f_j\left(x_{i,j}\right)$ represents non-linear functions for each predictor variable and ε_i is the error term.

3.3. Log-Log Linear Regression Model:

The log-log multiple linear regression model represents a linear relationship between the predictor (X) and response variables (Y) in the following form:

$$log(Y) = b_0 + b_1(\log(X_1)) + b_2(\log(X_2)) + \cdots + \varepsilon_i \tag{3}$$

Where Y is the response variable (such as mean, LCV and LSK); X_1, X_2 are the predictor variable (such as catchments characteristics); b_0, b_1, b_2 are the regression coefficients to be estimated; and ε_i represents the error term.

3.4. Evaluation Statistics:

The models were validated using a leave-one-out (LOO) validation and assessed using various statistics, including the median ratio of predicted to observed quantiles, median relative error (Median REr%), mean squared error (MSE), relative root mean square error (rRMSE%), mean bias (BIAS), and relative mean bias (rBIAS%). These metrices provide a comprehensive evaluation of model performance.

4. RESULTS AND DISCUSSION

In this study, both GAM and log-log regression were employed to estimate the parameters of the GEV distribution, focusing on the mean, LCV, and LSK. A total of 256 models were explored using eight catchment characteristics as predictor variables for each GEV parameter. The best models were selected based on R^2, Generalised Cross-Validation (GCV) scores (for GAM models), Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), and model complexity. LOO validation was conducted to examine model performance. The most suitable GAM and log-log regression models were chosen for each GEV parameter based on significant predictor variables and model statistics. Both model types were then assessed for quantile prediction, with the best combinations evaluated for their ability to balance predicted and observed flood quantiles.

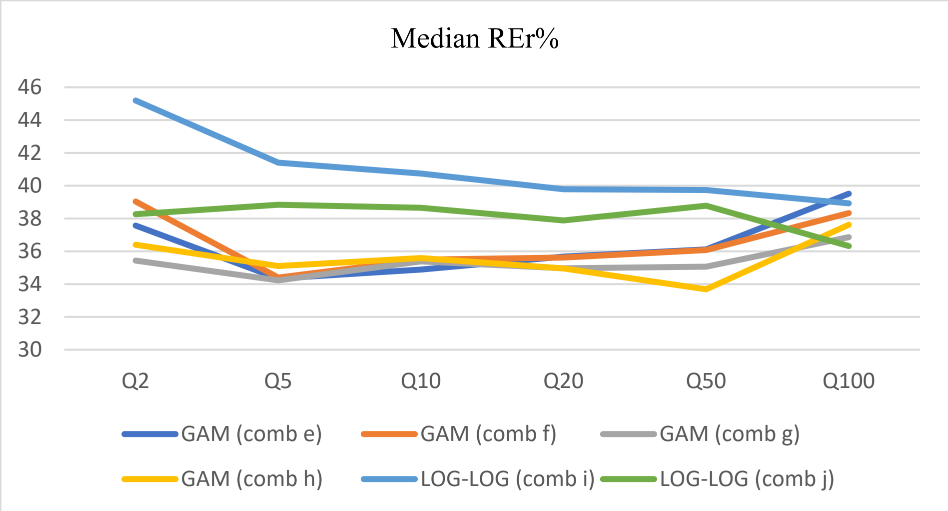

Figure 1. Median REr% values for different combinations for Q_2, Q_{10}, Q_{20}, Q_{50}, Q_{100}

The analysis results revealed that, in general, GAMs outperformed log-log regression models for return periods between 2 and 50 years. For the Q_{100} quantile, the performance of both models was comparable. Figure 1 illustrates the comparison of median REr% between GAM and log-log regression models across various quantiles (Q_2 to Q_{100}). It can be shown that GAM models (e, f, g, h) exhibit lower median REr% values (33%-39%) while log-log regression models (i, j) demonstrate slightly higher relative errors (35%-45%) indicating larger prediction errors. Additionally, GAM models demonstrate lower error variability, bias, and RMSE, making them a more accurate and reliable option for flood quantile estimation, particularly for shorter return periods.

Consequently, GAM models offer more precise flood quantile predictions compared to log-log regression models, particularly for lower return periods (Q_2 to Q_{50}) showing a superior ability to capture flood magnitude variability and reduce prediction errors.

5. LEARNING STATISTICAL MODELS

As a researcher, the first author has identified several key challenges in learning statistical flood modeling during educational workshops and meetings. Effective research requires a deep understanding of the basics of statistical models, such as the assumptions of the GEV distribution and the importance of L-moments as compared to ordinary product moments. Gap in this knowledge may prevent these models from being applied correctly in practical situations. Another challenge is applying these models to real-world data, where students need to be able to prepare data, fit models, and interpret results accurately. It is also important for students to recognise the sensitivity of statistical models to outliers and the quality and size of data. Many misconceptions need to be addressed, such as the belief that more complex models always give better results. Misunderstanding in these areas can lead to wrong conclusions and poor decisions in flood management. This underscores the need to enhance both understanding and application skills in statistical flood modeling.

6. MY PHD JOURNEY: CHALLENGES AND GROWTH

Coming from a French educational background, I initially started my Master of Philosophy. After nine months, I completed my confirmation of candidature (CoC) and upgraded to a Doctor of Philosophy (PhD) program. At the beginning of my PhD journey, my technical writing was not good enough, I had to manage family commitments with little children, and I faced difficulties with health issues, which took a considerable amount of time. However, with the constant support and guidance from my supervisors, I managed to address these challenges and have reached the end of my candidature.

One of the first critical tasks was conducting a comprehensive literature review to gain a deep understanding of my research topic, recent findings, various approaches, and identifying gaps in existing studies. After completing the review, data collation and preparation became essential, as accurate and up to date data are crucial for producing reliable results. Throughout my PhD journey, I have developed valuable skills that significantly enhanced my academic and professional skills. Working with large datasets, I became more familiar in data analysis techniques, including regression and LOO validation, and gained a good knowledge in software tools such as FLIKE, MAPINFO, SPSS, and EASYFIT. My programming skills improved as I have developed proficiency in languages like R, MATLAB, and FORTRAN. Additionally, I gained a deep understanding of statistical methods and kept my literature review up to date by regularly following new journal publications in my field.

My academic writing skills has also advanced, enabling me to write conference papers, reports, and journal articles effectively. Time management was essential as I have balanced research with personal responsibilities as a mother, wife, and student. Regular meetings with my supervisors enhanced my communication skills, and their guidance kept me focused and motivated. Attending seminars and conferences have allowed me to present my work, receive feedback, and network with colleagues and experts. Even small advice from experienced researchers helped improve my research methods. These experiences have greatly enriched my PhD journey and equipped me with skills for future success.

7. LEARNING JOURNEY IN EDUCATION AS A WOMAN IN ENGINEERING

Learning and career pathway in civil engineering presents unique challenges for women due to various social and cultural factors. From my early days as a civil engineer, working on construction sites where men were the majority, I faced the challenge of proving my skills and showing confidence in a male-dominated field. As careers progress, these challenges often grew. Now, as a PhD researcher, balancing academic work with personal responsibilities, such as pregnancy, motherhood, and raising children, has become even more complex with the ability to complete multi tasks and under pressure. Sometimes, important tasks such as meeting research deadlines, attending seminars, workshops, and conferences, overlap with childcare duties, managing household activities, and ensuring my children's health and well-being. There are times when I need to work late nights to make up for time spent at doctor appointments or school events. The struggling to maintain this balance is a common issue for women in STEM (science, technology, engineering, and mathematics) disciplines, where the dual responsibilities can result in stress and exhaustion (Ceci and Williams, 2011). Moreover, the gender gap in technical fields can lead to feelings of isolation and pressure to constantly prove their abilities (Etzkowitz et al., 2000). These difficulties reveal that women in engineering need better support and assistance. It also taught me how important it is to develop resilience, determination and to think differently when solving problems (Powell and Sang, 2013).

8. CONCLUSION

Using AMF data from 88 catchments in NSW, Australia, this study compares two statistical modeling approaches, including GAM and log-log regression, within the framework of the GEV distribution and L-moments. GAM models outperform log-log regression models, particularly for lower return periods while both techniques offer same accuracy for Q_{100}. Additionally, the study highlights the importance of understanding model fundamentals and explores the educational challenges faced by students, particularly women in engineering, in learning and applying these statistical models. Future researchers are advised to focus on developing critical skills required for flood design to succeed in their research and practice.

ACKNOWLEDGMENTS

Authors would like to acknowledge Western Sydney University for the excellent research environment and support for this research.

REFERENCES

Arthur, B., Guy, B., Armitage, E., LaBarre, M., & O'Connor, S. (2022). “Difficult but worth it”: Exploring the Experiences of Women in Engineering during Co-op. *Experiential Learning & Teaching in Higher Education*, *5*(1), 55-62.

Ceci, S. J., & Williams, W. M. (2011). Understanding current causes of women's underrepresentation in science. *Proceedings of the National Academy of Sciences*, *108*(8), 3157-3162.

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*(6), 2418-2428.

Chowdhury, R. (2019). Industry-Practice-Based Engineering Hydrology Education at USQ, Australia. *Education Sciences*, *9*(3), 213.

Etzkowitz, H., Kemelgor, C., Neuschatz, M., & Uzzi, B. (2000). *Athena unbound: The advancement of women in science and technology*. Cambridge University Press.

Gill, J., Sharp, R., Mills, J., & Franzway, S. (2008). I still wanna be an engineer! Women, education and the engineering profession. *European Journal of Engineering Education*, *33*(4), 391-402.

Hartmann, D. L., Tank, A. M. K., Rusticucci, M., Alexander, L. V., Brönnimann, S., Charabi, Y. A. R., ... & Zhai, P. (2013). Observations: atmosphere and surface. In *Climate change 2013 the physical science basis: Working group I contribution to the fifth assessment report of the intergovernmental panel on climate change* (pp. 159-254). Cambridge University Press.

Islam, M. N., & Pavel, T. (2011). Factors contributing towards dropouts at undergraduate level: An analysis. *ASA University Review*, *5*(1), 97-113.

Jenkinson, A. F. (1955). The frequency distribution of the annual maximum (or minimum) values of meteorological elements. *Quarterly Journal of the Royal meteorological society*, *81*(348), 158-171.

Maher, M. A., Ford, M. E., & Thompson, C. M. (2004). Degree progress of women doctoral students: Factors that constrain, facilitate, and differentiate. *The Review of Higher Education*, *27*(3), 385-408.

Mbonyiryivuze, A., Dorimana, A., Nsabayezu, E., & Nyirahabimana, P. (2023). Challenges Affecting Women PhD Candidates for Completion of Doctoral Educations: A Synthesis of the Literature. *African Journal of Educational Studies in Mathematics and Sciences*, *19*(1), 123-134.

Micevski, T., Hackelbusch, A., Haddad, K., Kuczera, G., & Rahman, A. (2015). Regionalisation of the parameters of the log-Pearson 3 distribution: A case study for New South Wales, Australia. *Hydrological Processes*, 29(2), 250-260.

Powell, A., & Sang, K. J. C. (2013). Gender and work-life balance in academia: Experiences of women academics and their perspectives on policy. *Gender, Work & Organization*, *20*(3), 253-272.

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.

Ruddell, B. L., & Wagener, T. (2015). Grand challenges for hydrology education in the 21st century. *Journal of Hydrologic Engineering*, *20*(1), A4014001.

Shrestha, S. (2018). Hydrology Education in Australia: Challenges & Opportunities. In *1 st International Conference on Advancements in Engineering Education (iCAEED-2018)* (p. 96).

Singh, J., Ghosh, S., Simonovic, S. P., & Karmakar, S. (2021). Identification of flood seasonality and drivers across Canada. *Hydrological Processes*, *35*(10), e14398.

Souaissi, Z., Ouarda, T. B., St-Hilaire, A., & Ouali, D. (2023). Regional frequency analysis of stream temperature at ungauged sites using non-linear canonical correlation analysis and generalized additive models. *Environmental Modelling & Software*, *163*, 105682.

Stojković, M., Prohaska, S., & Zlatanović, N. (2017). Estimation of flood frequencies from data sets with outliers using mixed distribution functions. *Journal of Applied Statistics*, *44*(11), 2017-2035.

Yan, L., Li, L., Yan, P., He, H., Li, J., & Lu, D. (2019). Nonstationary flood hazard analysis in response to climate change and population growth. Water, 11(9), 1811.