Proc. 3rd International Conference on Advancements in Engineering Education (iCAEED2024)
20-23 Nov 2024, Sydney, Australia
ISBN: 978-1-7636843-3-1

Challenges in application of statistical techniques in hydrology: A case study on regional flood modelling in southeast Australia

Ali Ahmed [1], Sadia Tasneem [2], Ataur Rahman [3]

[1]PhD student, Western Sydney University, Kingswood, Penrith, NSW, Australia
[2]Research Assistant, Ahsanullah University of Science and Technology, Dhaka, Bangladesh
[3]Professor, Western Sydney University, Kingswood, Penrith, NSW, Australia
Corresponding author's E-mail: 19938136@student.westernsydney.edu.au

Abstract

Regional flood frequency analysis (RFFA) is the most widely used statistical technique to estimate design floods in an ungauged location using flood data recorded at nearby stream gauging stations. The application of statistical techniques in hydrology, particularly in RFFA modelling is challenging in Australia due to a higher degree of non-homogeneity. Statistical tools and techniques involve planning, designing, sampling, data collection, analysing (descriptive and inferential), interpretation, and reporting to conduct statistical hydrologic research. Besides these there are lots of statistical and mathematical terms, concepts, and formulas, which are arithmetically complex and their physical interpretation in relation to rainfall-runoff process is relatively difficult. Hence, students who are not good at maths and hydrology face challenges to overcome these difficulties. Generally exploratory data analysis and visualisation, generalising the study results, scaling up the achieved results, and merging the findings in software engineering are tricky in statistical hydrology. In southeast Australia, the use of statistical techniques in RFFA is challenging, which might be due to highly variability and non-stationarity in hydrological data. This paper presents a case study using data from 88 catchments in New South Wales (NSW) where regression-based methods are used to develop prediction equations that can be applied to ungauged catchments in NSW to estimate design floods. This paper also highlights the learning aspects of statistical hydrology by the first author, which was based on the student-centred learning approach.

Keywords: Statistical techniques, hydrology, regional flood, student-centred approach

1. INTRODUCTION

In hydrological design, like frequency analysis of flood data, proper model selection, and parameter estimation are crucial where the statistical methods have high importance. However, hydrological data analysis such as designing, planning, forecasting, and evaluating with appropriate prediction model under changing climate is a challenging task (Chen et al., 2022). Application of statistics in undergraduate engineering courses is limited, which makes it difficult to understand the contents of statistical hydrology course (Zhan et al., 2010).

More than four decades ago American Statistical Association noticed the importance of statistical knowledge in engineering industry (Boardman et al., 1980). While the students feel bored about probability and statistics courses, Godfrey (1686) made this interesting and enjoyable by using unrelated methods (coin-tossing, card-playing dice-rolling) to illustrate the concept of this subject and students were able to comprehend the subject contents (Godfrey, 1986). Thus, it is important to illustrate the

complex issues of statistics by using examples with real data and simulation to make students understandable more realistically (Franklin & Kader, 2006).
Zhan et al. (2010) integrate statistics effectively for the Electronic Engineering Technology at Texas A & M University and showed that “learning-by-using” is identified as the best approach to taught students effectively. To teach statistical hydrology, it was showed that “blended learning approach (face-to-face workshop and pre-recorded materials)” is pretty much effective at Western Sydney University, Sydney, Australia (Rahman et al., 2018).

Flood is a natural disaster, which causes significance loss, damage and disruption (Zalnezhad et al., 2022). Under changing climate conditions, floods are becoming more severe and frequent (FitzGerald et al., 2010). To reduce flood damage, design flood is widely used in hydrologic design, which is associated with an annual exceedance probability (AEP) or return period (T). For ungauged catchments, regional flood frequency analysis (RFFA) is used to estimate design floods in which, regional homogeneity is an important aspect (Šimková, 2017; Zhang & Stadnyk, 2020). Many studies, for example, Chebana & Ouarda (2007), Dalrymple (1960), Hosking & Wallis (1993), and Ouarda (2017) evaluated the homogeneity in RFFA adopting various statistical techniques as noted in Ahmed et al. (2024). Moreover, to investigate the accuracy of design flood estimates in RFFA many researchers adopted quantile regression technique (QRT), e.g., Aziz et al. (2016), Haddad et al. (2012), Rahman et al. (2011), Rahman et al. (2020), Rahman & Rahman (2020), and Taylor et al. (2011). This study applies QRT to a flood dataset of New South Wales (NSW) as a part of first ahthor’s doctoral research. This also highlights the learning aspects of the first author in his doctoral research in statistical hydrology.

2. METHODS AND MATERIALS

2.1 STUDY AREA

The state NSW, a part of southeast Australia, is selected as the study area. A total of 88 gauged catchments across NSW was selected in this study, which are mostly natural and are not affected by any major land use change. The Australian Rainfall and Runoff (ARR) recommended an upper limit of 1000 km2 as catchment area for small to medium sized catchments (Rahman et al., 2019), and hence our catchments were selected based on this guide. The recommended record length of flood data is 20 years (Rahman et al., 2015). A study conducted by Zalnezhad et al., (2022) mentioned that the streamflow data quality of NSW is good. In addition, the selected catchments from NSW mostly satisfy the above criteria. Eight catchment characteristics (Table 1) are selected for this study. Figure 1 shows the study area with the spatial distribution of the selected catchments.

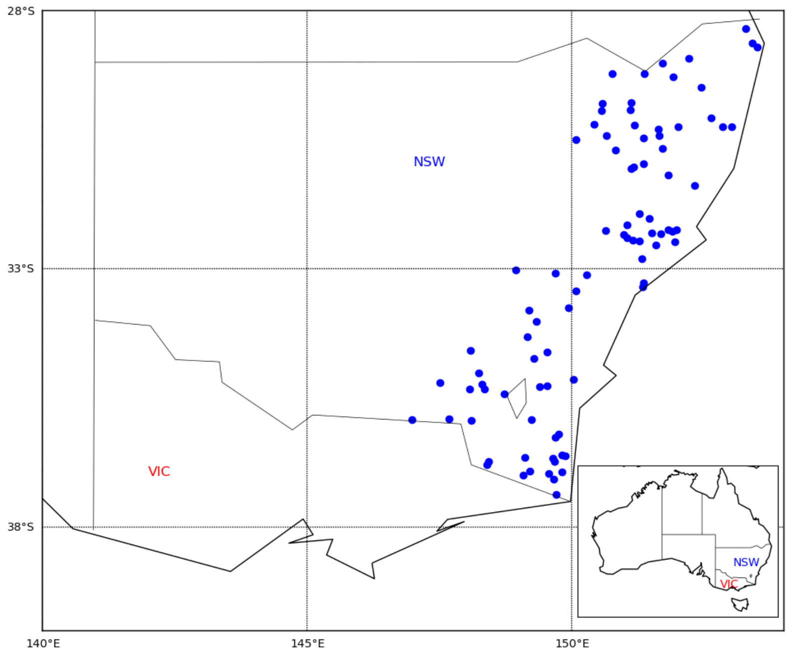

Figure 1. Spatial distribution of the selected stations in New South Wales (NSW)

2.2 METHODOLOGY

Adopting split sample validation technique, the selected 88 sites are split into two groups. One group comprising 75% of sites (n = 66) treated as the training data set and the remaining 25% (n = 22) are taken as the test data set. Then Hosking and Wallis (1993) test statistics; discordancy measure (D_i), heterogeneity measure (H_i), and goodness-of-fit measure (Z-statistics) are calculated. Based on L-moments coefficient the H_i-statistics are estimated (Hosking & Wallis, 1993). Integrating ordinary least square (OLS) regression, the QRT is adopted to develop prediction equations to estimate design floods. Design flood is associated with an annual exceedance probability (AEP). Using annual maximum flood (AMF) data of the selected stations, log Pearson's Type Three (LP3) distribution was used to estimate flood quantiles (Q_T) with AEPs of 50%, 20%, 10%, 5%, 2%, and 1% (Q_2, Q_5, Q_{10}, Q_{20}, Q_{50}, and Q_{100}, respectively). The selected eight catchment and climatic characteristics (see Table 1) are used in the regression analysis to develop prediction equations. Mathematically the adopted regression model can be expressed as (equation 1):

$$log_{10}(Q_T) = b_0 + b_1 * log_{10}(\text{AREA}) + b_2 * log_{10}(I_{62}) + b_3 * log_{10}(\text{MAR}) + b_4 * log_{10}(\text{SF}) + b_5 * log_{10}(\text{MAE}) + b_6 * log_{10}(\text{SDEN}) + b_7 * log_{10}(S1085) + b_8 * log_{10}(\text{FOREST}) \qquad (1)$$

where Q_T, the flood quantile for AEP of 1 in T and b_0 is the model intercept and $b_1, b_2, b_3, \ldots, b_8$ are the regression coefficients.

Using the developed prediction equation (1) based on the model data set, the flood quantiles are estimated for the test stations. To quantify the model accuracy, relative error (RE) is estimated:

$$\text{RE} = \frac{Q_{\text{pred}} - Q_{\text{obs}}}{Q_{\text{obs}}} \times 100 \qquad (2)$$

where Q_{pred} is the predicted flood quantile using the developed prediction equation and Q_{obs} is the at-site flood quantiles estimated by fitting a LP3 distribution. The split-sample process was repeated 10 times to get a more realistic view of the model estimation error. The co-efficient of determination (R^2) for each model (quantile vs repetition) is also calculated to investigate the strength of the associated model. R Studio is used to perform the statistical analysis.

3. RESULT

3.1. Basic statistics of selected stations

Table 1 summarises the basic statistics of catchment and climatic characteristics of the selected 88 sites of the study area. The minimum and maximum area of the selected sites are 8 km2 and 1010 km2, respectively. The median is 260 km2 and the mean ± SD is 351.98 ± 281.43 km2. The highest and the lowest record length of AMF data are 89 and 25 years, respectively with a median of 44 years (mean ± SD = 47.55 ± 13.17 years). These figures for mean annual rainfall are 1953.23 mm, 626.17 mm, and 909.92 mm, respectively (mean ± SD = 1000.28 ± 304.48 mm). The design rainfall intensity for the selected sites ranges from 31.30 to 87.30 mm/h. The median and the mean ± SD of this predictor is 43.10 mm/h and 45.40 ± 11.27 mm/h, separately. The lowest and the highest mean annual evapotranspiration for the selected sites are 980.40 mm and 1543.30 mm, respectively with a mean of 1223.69 mm and standard deviation of 126.30 mm.

Table 1. Summary of basic statistics of the selected catchment characteristics (n = 88)

Catchment Characteristics and Statistics	Minimum	Maximum	Median	Mean	Standard Deviation
Catchment area (AREA, km^2)	8.00	1010.00	260.00	351.98	281.43
Rainfall intensity (I_{62}) in mm/h	31.30	87.30	43.10	45.40	11.27
Mean annual rainfall (MAR) in mm	626.17	1953.23	909.92	1000.28	304.48
Shape factor (SF)	0.26	1.63	0.77	0.76	0.21
Mean annual evapo-transpiration (MAE) in mm	980.40	1543.30	1185.55	1223.69	126.30
Stream density (SDEN) in /km	0.52	5.47	2.72	2.85	1.10
Mainstream slope ($S1085$) in m/km	1.54	49.86	9.08	12.92	10.80
Forest (FOREST) (fraction)	0.00	0.99	0.52	0.51	0.32
Record length of AMF data (years)	25.00	89.00	44.00	47.55	13.17

3.2. QRT with split sample validation

3.2.1.Summary of D_i, H_i, Z-statistics for training and test data sets

Table 2 demonstrates the descriptive summary of Hosking and Wallis (1993) test statistics for both the training and test data sets for each repetition. The test statistics are: (a) discordancy measure to detect a discordant site, b) heterogeneity measure to test the homogeneity of the proposed region, and c) goodness-of-fit measure to find the best fit distribution for the proposed region. In case of test data set (n = 22), out 10 repetitions only four have at least one discordant site. For instance, in repetition 2 (222016, D_i = 3.62 and 419029 = 3.29) and in repetition 7 (210017, D_i = 3.79 and 419029, D_i = 3.64) each have two discordant sites, whereas the repetition 1 (203002, D_i = 3.85), 6 (208001, D_i = 3.26), and 9 (203002, D_i = 3.63) has only one discordant site. In contrast, for the training data set (n = 66) every repetition has at least two discordant sites. Maximum number of discordant sites are detected for repetition 5 (222016, $D_i = 6.61$; 419029, $D_i = 5.20$; 203002, $D_i = 3.82$; 410057, $D_i = 3.12$) and 8 (222016, $D_i = 5.20$; 419029, $D_i = 5.54$; 203002, $D_i = 3.40$; 222015, $D_i = 3.01$). Irrespective of repetitions the H_1-values range from 4.19 to 9.60 for the test data set and 11.03 to 14.22 for the training data set, which indicates that the regions are highly heterogeneous. It might be due to highly variable hydrology of Australia. The goodness-of-fit measures of the proposed regions show that the Pearson Type Three (PE3) and Generalised Pareto distribution (GPA) are the best fit distribution for the proposed regions with some exceptions like Generalised Normal distribution (GNO) is also suitable for several regions.

Table 2. Summary of results of the homogeneity test

Statistics and Repetitions	D-values		H-values						Name of the distribution with Z-values (|Z| ≤ 1.64 is acceptable)									
	Test data set (n = 22)	Train data set (n = 66)	Test data set (n = 22)			Train data set (n = 66)			Test data set (n = 22)					Train data set (n = 66)				
	Discordant sites with D-values		$H1$	$H2$	$H3$	$H1$	$H2$	$H3$	GLO	GEV	GNO	PE3	GPA	GLO	GEV	GNO	PE3	GPA
1	203002 = 3.85	222016 = 5.41, 419029 = 5.02	6.27	4.32	1.96	12.90	10.06	6.31	3.23	2.37	0.70	-2.19	-0.62	9.29	7.42	4.26	-1.21	1.16
2	222016 = 3.62, 419029 = 3.29	222015 = 3.87, 208001 = 3.45, 204026 = 3.05	7.23	4.58	1.82	12.45	9.90	6.50	4.34	3.49	1.63	-1.57	0.40	8.69	6.79	3.72	-1.60	0.54
3	-	222016 = 4.76, 419029 = 4.35, 203002 = 4.17	8.04	6.01	3.40	11.61	8.99	5.28	4.71	3.54	1.71	-1.45	-0.29	8.62	6.89	3.72	-1.76	0.95
4	-	222016 = 5.37, 419029 = 4.28, 222015 = 3.08	5.62	5.95	4.78	12.64	9.06	4.92	5.06	4.12	2.23	-1.04	0.79	8.40	6.54	3.42	-1.97	0.32
5	-	222016 = 6.61, 419029 = 5.20, 203002 = 3.82, 410057 = 3.12	8.34	6.30	3.35	11.58	9.15	6.07	4.25	3.29	1.53	-1.52	0.00	9.10	7.21	3.94	-1.71	0.83
6	208001 = 3.26	222016 = 5.16, 419029 = 4.49, 203002 = 3.53	4.19	4.38	4.75	14.22	10.13	4.66	3.93	3.01	1.15	-2.05	-0.24	9.15	7.25	4.09	-1.36	0.93
7	210017 = 3.79, 419029 = 3.64	222016 = 4.53, 203002 = 3.70	8.12	5.70	3.26	11.52	8.86	5.29	5.18	4.17	2.28	-0.99	0.69	8.46	6.60	3.44	-2.04	0.39
8	-	222016 = 5.20, 419029 = 5.54, 203002 = 3.40, 222015 = 3.01	8.49	8.16	5.76	11.63	8.08	4.44	5.52	4.55	2.60	-0.76	1.13	8.32	6.43	3.28	-2.18	0.14
9	203002 = 3.63	222016 = 4.62, 419029 = 4.55	6.62	4.88	2.93	12.57	9.68	5.93	4.75	3.72	1.77	-1.59	0.16	8.72	6.88	3.75	-1.68	0.70
10	-	222016 = 4.46, 419029 = 4.10	9.60	7.50	3.91	11.03	8.35	5.41	5.60	4.18	2.44	-0.58	-0.15	8.41	6.74	3.47	-2.19	0.88

Generalized Logistic (GLO), Generalized Extreme Value (GEV), Generalized Normal (GNO), Pearson Type 3 (PE3), Generalized Pareto (GPA)

Table 3 illustrates the absolute median relative error (AMRE) values generated from the developed prediction equations for flood quantile estimates with the respective coefficient of determination (R^2) values and H_1-statistic. It also examines the association between the AMRE and H_1-values.

The lowest AMRE values (25.11%) for Q_{10} are found in repetition 5 when the H_1- and R^2-values are 8.34 and 0.63, respectively. However, having almost similar values of H_1 (8.12) and R^2 (0.69) the highest AMRE values (75.46%) are found for Q_{100} in repetition 7. It indicates that there is no association between the AMRE values and the heterogeneity measures. This finding is similar to previous studies such as Ahmed et al. (2024). With the lowest R^2-values (0.52) the estimates AMRE value is 35.00% for Q_{100} and having the highest R^2-values (0.79) the calculated AMRE value is 44.36% for Q_5.

Table 2. Summary statistics of absolute median relative error and R^2 values for different repetitions

	Repetition

		1	*2*	*3*	*4*	*5*	*6*	*7*	*8*	*9*	*10*
H_1-statistics	H_1	6.27	7.23	8.04	5.62	8.34	**4.19**	8.12	8.49	6.62	**9.60**
Absolute median relative error (%)	Q_2	41.86	47.08	40.59	61.32	48.34	37.61	43.43	47.68	45.21	44.65
	Q_5	31.89	36.15	34.30	49.83	29.37	49.77	44.36	38.94	37.11	41.60
	Q_{10}	27.66	33.25	38.83	53.03	**25.11**	49.35	52.39	49.44	42.54	48.25
	Q_{20}	31.71	36.57	46.05	58.62	38.03	49.25	57.09	55.86	45.35	51.21
	Q_{50}	39.01	44.90	55.49	58.76	28.94	52.71	67.08	65.61	58.58	58.05
	Q_{100}	44.12	54.91	62.37	64.06	35.00	60.15	**75.46**	71.96	62.37	63.43
R^2 values	Q_2	0.74	0.75	0.70	0.78	0.68	0.78	0.77	0.72	0.65	0.76
	Q_5	0.73	0.75	0.70	0.79	0.66	0.77	**0.79**	0.75	0.67	0.78
	Q_{10}	0.70	0.74	0.68	0.77	0.63	0.75	0.78	0.74	0.64	0.77
	Q_{20}	0.66	0.71	0.64	0.74	0.59	0.72	0.76	0.73	0.61	0.75
	Q_{50}	0.60	0.68	0.60	0.70	0.55	0.67	0.72	0.70	0.57	0.71
	Q_{100}	0.56	0.65	0.57	0.66	**0.52**	0.63	0.69	0.67	0.54	0.68

4. ENGINEERING EDUCATION ASPECTS

The learning of statistical hydrology is challenging as noted by Rahman et al. (2018) due to its empirical nature. The first author of this article has Bachelor and Masters degree in statistics from a reputed university in Bangladesh, and has over 20 years of experiences in an international organisation where he was involved with health-related statistical data analysis. He faced few challenges in his PhD study in statistical hydrology area. Firstly, he had to learn basic hydrological terms and concepts, which were new to him. Secondly, he had to learn advanced statistical hydrologic techniques such as L-moments based homogeneity test. Thirdly, he had to learn R-program to carry out the necessary statistical analysis for the selected data set. Fourthly, he had to learn scholarly writing to meet international standards e.g. writing of Q1 journal articles.

The following learning strategies were found to be effective for the first author: (i) Regular meetings with the Principal Supervisor; (ii) Attending lectures of Advanced Statistical Hydrology subject in WSU; (iii) Learning R-program from one fellow PhD student and one former Research Fellow of the Principal Supervisor; and (iv) Addressing comments of the draft papers and thesis chapters provided by the supervisors and co-authors.

One of the major difficulties faced by the first author included delay in travel from Bangladesh to Australia due to COVID time travel restriction. It should be noted that the first author has published two Q1 journal articles with the assistance of the supervisory panel, and most importantly, he submitted his PhD thesis in three years' time.

The Principal Supervisor of the first author found it challenging to teach hydrology principles to the first author who had no background in this subject. To bridge the knowledge gap, both the student and the supervisor had to work closely.

The learning of the first author was achieved by student-centred learning as highlighted by Thompson et al. (2012) and Ngambeki et al. (2012) where the student (first author) took the lead in learning in particular mastering of R-program and scholarly writing.

5. CONCLUSION

This study presents development and testing of a RFFA method for NSW state. It applies QRT and Hosking and Wallis method and finds that homogeneous regions cannot be established in NSW. The Pearson Type 3 distribution is found to be the best fit distribution for the proposed regions. The median relative error values for the developed prediction equations vary from 25% to 75%. The first author of the paper faced few challenges in learning statistical hydrology as he did not study hydrology in his Bachelor and Masters studies. A student-centred approach was found to be effective for the student in learning statistical hydrology and scholarly writing during his doctoral study.

REFERENCES

Ahmed, A., Khan, Z., & Rahman, A. (2024). Searching for homogeneous regions in regional flood frequency analysis for Southeast Australia. *Journal of Hydrology: Regional Studies*, *53*, 101782.

Ahmed, A., Yildirim, G., Haddad, K., & Rahman, A. (2023). Regional Flood Frequency Analysis: A Bibliometric Overview. *Water*, *15*(9), 1658.

Aziz, K., Haque, M. M., Rahman, A., Shamseldin, A. Y., & Shoaib, M. (2016). Flood estimation in ungauged catchments: application of artificial intelligence based methods for Eastern Australia. *Stochastic Environmental Research and Risk Assessment*, *31*(6), 1499-1514. https://doi.org/10.1007/s00477-016-1272-0

Chebana, F., & Ouarda, T. B. (2007). Multivariate L-moment homogeneity test. *Water Resources Research*, *43*(8). https://doi.org/DOI: 10.1029/2006wr005639

Chen, Y., Wang, D., Liu, D., Li, B., & Sharma, A. (2022). Statistics in Hydrology. *Water*, *14*(10), 1571.

Dalrymple, T. (1960). *Flood-frequency analyses, manual of hydrology: Part 3*.

FitzGerald, G., Du, W., Jamal, A., Clark, M., & Hou, X. Y. (2010). Flood fatalities in contemporary Australia (1997–2008). *Emergency Medicine Australasia*, *22*(2), 180-186.

Haddad, K., Rahman, A., & Stedinger, J. R. (2012). Regional flood frequency analysis using Bayesian generalized least squares: a comparison between quantile and parameter regression techniques. *Hydrological Processes*, *26*(7), 1008-1021. https://doi.org/10.1002/hyp.8189

Hosking, J. R. M., & Wallis, J. R. (1993). Some statistics useful in regional frequency analysis. *Water Resources Research*, *29*(2), 271-281. https://doi.org/10.1029/92wr01980

Ngambeki, I., Thompson, S., Troch, P., Sivapalan, M., & Evangelou, D. (2012). Engaging the students of today and preparing the catchment hydrologists of tomorrow: student-centered approaches in hydrology education. *Hydrology & Earth System Sciences Discussions*, *9*(1).

Ouarda, T. B. M. J. (2017). *Handbook of Applied Hydrology, Second Edition, Chapter 77, Regional Flood Frequency Modeling*. McGraw-Hill Education. https://www-accessengineeringlibrary-com.ezproxy.uws.edu.au/content/book/9780071835091

Rahman, A., Haddad, K., Kuczera, G., & Weinmann, E. (2019). *Regional Flood Methods. In: Australian Rainfall and Runoff, Chapter 3, Book 3, edited by Ball et al,* . Geoscience Australia, Commonwealth of Australia, Engineers Australia. http://www.arr-software.org/arrdocs.html

Rahman, A., Haddad, K., Rahman, A. S., & Haque, M. (2015). *Australian Rainfall and Runoff Revision Project 5: Regional Flood Methods: Database Used to Develop ARR RFFE Technique: Stage 3 Report* (0858259400).

Rahman, A., Haddad, K., Zaman, M., Kuczera, G., & Weinmann, P. E. (2011). Design Flood Estimation in Ungauged Catchments: A Comparison Between the Probabilistic Rational Method and Quantile Regression Technique for NSW. *Australasian Journal of Water Resources*, *14*(2), 127-139. https://doi.org/10.1080/13241583.2011.11465381

Rahman, A., Kordrostami, S., & Purdy, D. (2018). Statistical hydrology teaching using a blended learning approach. In *Blended Learning in Engineering Education* (pp. 1-13). CRC Press.

Rahman, A. S., Khan, Z., & Rahman, A. (2020). Application of independent component analysis in regional flood frequency analysis: Comparison between quantile regression and parameter regression techniques. *Journal of Hydrology*, *581*, Article 124372. https://doi.org/10.1016/j.jhydrol.2019.124372

Rahman, A. S., & Rahman, A. (2020). Application of Principal Component Analysis and Cluster Analysis in Regional Flood Frequency Analysis: A Case Study in New South Wales, Australia. *Water*, *12*(3), 1-26. https://doi.org/10.3390/w12030781

Šimková, T. (2017). Homogeneity testing for spatially correlated data in multivariate regional frequency analysis. *Water Resources Research*, *53*(8), 7012-7028. https://doi.org/10.1002/2016wr020295

Taylor, M., Haddad, K., Zaman, M., & Rahman, A. (2011). Regional flood modelling in Western Australia: Application of regression based methods using ordinary least squares. 19th International Congress on Modelling and Simulation - Sustaining Our Future: Understanding and Living with Uncertainty, MODSIM2011 Perth, WA 2011,

Thompson, S., Ngambeki, I., Troch, P., Sivapalan, M., & Evangelou, D. (2012). Incorporating student-centered approaches into catchment hydrology teaching: a review and synthesis. *Hydrology and Earth System Sciences*, *16*(9), 3263-3278.

Zalnezhad, A., Rahman, A., Vafakhah, M., Samali, B., & Ahamed, F. (2022). Regional Flood Frequency Analysis Using the FCM-ANFIS Algorithm: A Case Study in South-Eastern Australia. *Water*, *14*(10), 1608. https://doi.org/10.3390/w14101608

Zhan, W., Fink, R., & Fang, A. (2010). Application of statistics in engineering technology programs. *American Journal of Engineering Education*, *1*(1), 65-78.

Zhang, Z., & Stadnyk, T. A. (2020). Investigation of Attributes for Identifying Homogeneous Flood Regions for Regional Flood Frequency Analysis in Canada. *Water*, *12*(9), Article 2503. https://doi.org/10.3390/w12092570