Proc. 3rd International Conference on Water and Environmental Engineering (iCWEE2022)
27-30 Nov 2022, Sydney, Australia
ISBN: 978-0-6456692-0-6

Degree of Heterogeneity versus Prediction Error in Regional Flood Frequency Analysis: A Case Study for Victoria, Australia

Ali Ahmed[1], Xiao Pan[1], Khaled Haddad[2], Zaved Khan[3] and Ataur Rahman[4]

[1]PhD candidate, Western Sydney University, Kingswood, NSW, Sydney, Australia
[2]Drainage, Marine & Floodplain Engineer, City of Canada Bay, NSW, Sydney, Australia
[3]Adjunct Research Fellow, Western Sydney University, Kingswood, NSW, Sydney, Australia
[4]Professor, Western Sydney University, Kingswood, NSW, Sydney, Australia

Corresponding author's email: aliahmed2007@gmail.com

Abstract

In flood management and hydraulic infrastructure design, flood risk assessment is needed. To estimate flood quantiles accurately at an ungauged catchment Regional Flood Frequency Analysis (RFFA) is widely adopted. In RFFA, the homogeneity of a region refers to the state of similar flood responses, which is mostly the reflection of similar flood and catchment characteristics. This study examines the homogeneity of 113 gauged catchments in Victoria, Australia. The selected catchments are divided into two groups by drainage division and then subdivided each of them into two sub-regions. Hosking and Wallis (HW) test statistics (H) are applied, and few sites are detected as discordant. H_1-statistics are relatively low (ranging from 3.6 to 20.2) in the sub-groups but highest (26.6) in Victoria as a single region, which indicates that these regions were highly heterogeneous. A log-log model is used to develop prediction equations using ordinary least squares regression (OLS). To check the relative accuracy of the developed RFFA models a leave-one-out (LOO) is adopted. It is found that the degree of heterogeneity does not have any direct effect on the accuracy of design flood estimates. More investigation is needed to better understand the association between the degree of regional heterogeneity and model accuracy in RFFA.

Keywords*:* Regional flood frequency analysis, Heterogeneity, Design flood estimation, Relative errors, Australian Rainfall and Runoff (ARR), Australia.

1 INTRODUCTION

In RFFA, the delineation of homogeneous regions is one of the main steps (Šimková, 2017; Zhang & Stadnyk, 2020) and it is important (Lettenmaier et al., 1987). A design flood is a flood discharge associated with an annual exceedance probability (AEP). A reliable design flood estimation is essential because in hydraulic infrastructure design underestimation of design flood increases flood damage and overestimation amplifies the capital cost (Ouarda, 2017). Burn (1988) quoted that Hosking et al. (1985a) and Lettenmaier et al. (1987) addressed the effects of heterogeneity of sites within a formed region by applying different types of regional flood estimators. Regional homogeneity is the vital assumption of RFFA (Castellarin et al., 2008; Fill & Stedinger, 1995a; Masselot et al., 2017) and in general the more accurate flood quantile estimate can be achieved through developed RFFA model forming relatively more homogeneous regions. In southeastern Australia, using data from 104 catchments, Bates et al. (1998) showed that the combination of physical and climatic characteristics has an impact on homogeneous region identification. However, they did not assess the impacts of regional homogeneity on the accuracy of design flood estimates. This study aims to fill this research gap.

2 MATERIALS AND METHODS

2.1 Study area and data

The state of Victoria (VIC) in Australia is the focus of this study. A total of 113 gauged catchments across Victoria is used. The selected catchments are mostly natural and are not affected by any major land use change. Figure 1 illustrates the study area and selected catchments

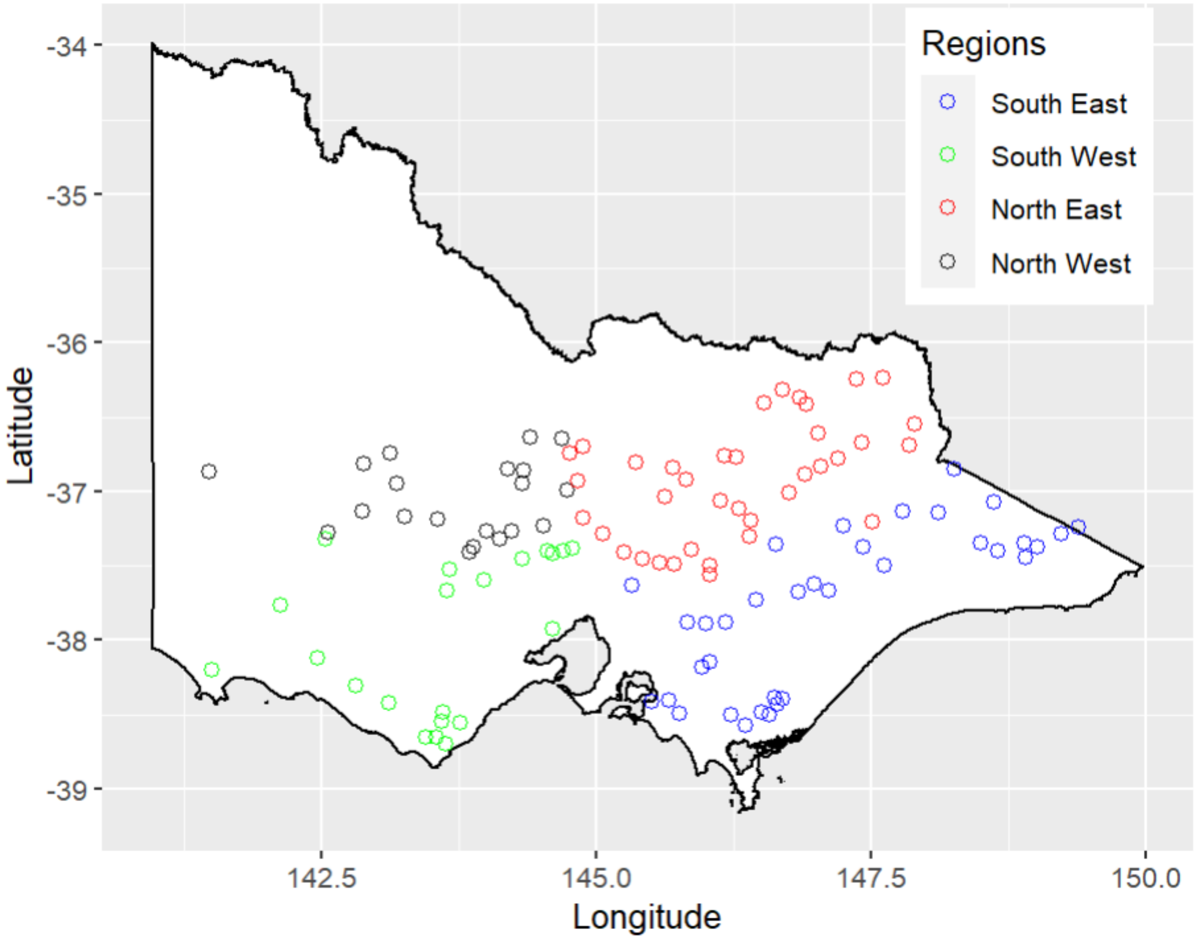

Figure 1. Spatial distribution of selected 113 catchments in Victoria, Australia

Table 1 summarized the basic summary statistics of catchment and climatic characteristics of the selected 113 catchments. The maximum and minimum areas of the selected catchments were 997.0 km^2 and 3.0 km^2, respectively with a median value of 282.0 km^2 (standard deviation 246.9 km^2). The highest mean annual rainfall for the selected catchments is 1760 mm and the lowest one is 484 mm with a mean and standard deviation of 932.6 mm and 320.2 mm, respectively. The record length of annual maximum flood data ranged from 26 to 67 years with a mean of 46.2 years (SD was 5.5 years).

Table 1. Descriptive statistics of the selected catchment characteristics data

	Catchment Characteristics, Victoria, n = 113								
Statistics	Catchment area (AREA, km2)	Rainfall intensity (*I*62) in mm/h	Shape factor (SF)	Stream density (SDEN) in /km	Mean annual rainfall (MAR) in mm	Mean annual evapo-transpiration (MAE) in mm	Forest (FOREST) (fraction)	Mainstream slope (*S*1085) in m/km	Record length of AMF data (years)
Minimum	3.00	24.60	0.28	0.52	484.40	925.90	0.01	0.80	26.00
Maximum	997.00	46.70	1.43	4.25	1760.80	1155.30	1.00	69.91	67.00

Median	282.00	33.70	0.79	1.44	881.40	1030.70	0.65	9.59	46.00
Mean	320.00	34.30	0.79	1.52	932.60	1035.60	0.59	13.40	46.15
Standard Deviation	246.93	5.29	0.22	0.53	320.24	42.96	0.35	12.35	5.47

2.2 Methodology

Initially, a single region was formed considering all the selected 113 catchments. Then Hosking and Wallis discordancy (*Di*) and *H*-statistics were calculated for this region. Here *Hi*-statistics were measured based on the coefficient of *L*-moments (Hosking & Wallis, 1993). To develop a prediction equation ordinary least square (OLS) regression (Quantile regression technique - QRT) was adopted for this region. Flood quantiles (Q_T) with annual exceedance probabilities (AEPs) of 50%, 20%, 10%, 5%, 2%, and 1% (Q_2, Q_5, Q_{10}, Q_{20}, Q_{50}, and Q_{100}, respectively) were considered as dependent variables in the OLS regression analysis. Q_T values were estimated by fitting a log Pearson's Type Three (LP3) distribution to the annual maximum flood series of each selected catchment. Eight catchment and climatic predictors (log_{10} scale) were used in the regression analysis (Table 1). The adopted regression model can be expressed by equation (1),

$$log_{10}(Q_T) = b_0 + b_1*log_{10}(AREA) + b_2*log_{10}(I62) + b_3*log_{10}(SF) + b_4*log_{10}(SDEN) + b_5*log_{10}(MAR) + b_6*log_{10}(MAE) + b_7*log_{10}(FOREST) + b_8*log_{10}(S1085) \quad (1)$$

Here, Q_T is the flood quantile for ARI of *T* years, b_0 is the model intercept and b_1, b_2, b_3, …, b_8 are the regression coefficients.

To validate the model, a LOO validation technique was adopted. Relative error (RE) was calculated in terms of predicted and observed quantile values by equation (2).

$$RE = ((Qpred – Qobs)/Qobs)*100 \quad (2)$$

Searching for a more homogeneous region, the selected catchments were divided into two regions by their drainage division (DD), i.e drainage division II (DD II) and drainage division IV (DD IV). Basically, DD II and DD IV are the south and north part of Victoria state, respectively. Thereafter, these two parts were subdivided into two regions as shown in Figure 1. Finally, we got four regions two from each drainage division. The assumed regions were Southeast and Southwest from DD II and Northeast and Northwest from DD IV. Following the previous steps, discordancy and *H*-statistics were calculated for these four regions. Applying OLS regression prediction equations were developed and RE values were estimated based on LOO validation. Finally, the relationship between heterogeneity (*H*-statistics) and RE was examined to get better flood quantile estimates. R Studio was used to perform the analysis.

3 RESULTS AND DISCUSSION

3.1 Region formation and homogeneity testing

Victoria as a single region has six discordant (*Di* > 3) sites. The sites are 226222, 226402, 227236, 235205, 405205, and 405209 with *Di* values 5.04, 3.74, 3.09, 5.02, 4.40 and 3.30, respectively. In the south Victoria region, there are two discordant sites (226222, *D* = 3.41 and 235205, *D* = 4.13) and 226222 sites remain discordant in the sub-group (southeast Victoria) with *D* values 3.42, but no discordant site in the southwest Victoria region. Likewise, in north Victoria 405205 (*D* = 5.80) and 405209 (*D* = 3.16) are found discordant and these two sites remain discordant in the sub-group of north Victoria (northeast Victoria region) but with a different *Di* value (4.23 and 3.05, respectively). The only discordant site in northwest Victoria is 415217 (*D* = 3.70).

Figure 2 illustrates the comparison of *H*-statistics among different candidate regions by drainage division in Victoria, Australia. Hosking and Wallis heterogeneity measure reveals that the assumed regions are highly heterogeneous. This heterogeneity measure also indicates that the *H*-statistics values get lower in the sub-regions compared to the single region. In the two assumed regions based on drainage division (south and north portion of Victoria), the estimated H_1-statistics are 16.7 to 20.2, respectively. It is 26.6 in Victoria as a single region. In the sub-regions of south Victoria, the H_1-statistics range from 9.6 to 14.7, whereas in north Victoria it is 3.6 to 13.4. Interestingly, there is a decreasing trend of *Hi*-statistics (H_1, H_2, *and* H_3) when the region is subdivided into smaller groups, except in the northwest part of Victoria. No pattern is observed among the *Hi*-statistics over the regions.

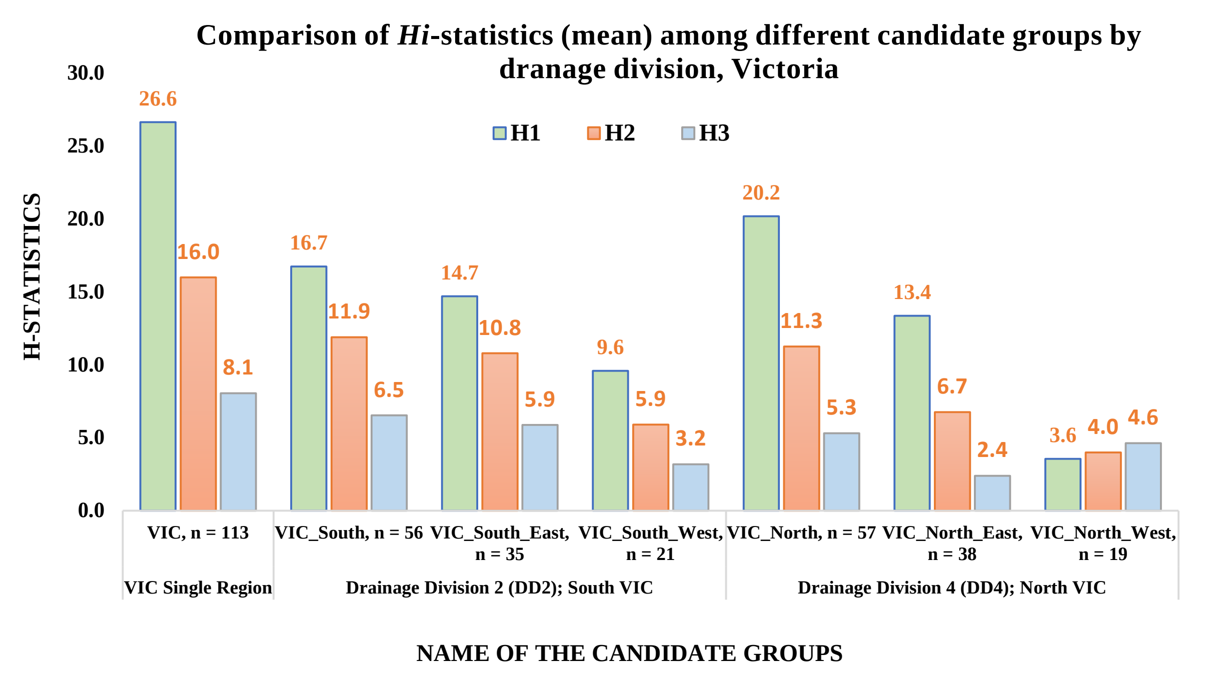

Figure 2. H-statistics for different candidate groups in Victoria, Australia

Table 2 demonstrates Z-distribution values for the candidate regions. A distribution is said to be acceptable if the Hosking and Wallis |Z|-statistic value is ≤ 1.64. It is clear from table 2 that Pearson Type Three (PE3) and Generalized Pareto (GPA) distributions are suitable distributions for all the assumed regions except the northwest region.

Table 2. Z-statistic values for candidate regions in Victoria, Australia

Name of the candidate groups		Name of the distribution with Z-values (|Z| <= 1.64, fit well/acceptable)				
		Generalized Logistic (GLO)	Generalized Extreme Value (GEV)	Generalized Normal (GNO)	Pearson Type 3 (PE3)	Generalized Pareto (GPA)
Victoria as a single region	VIC Single Region, n=113	14.07	10.33	6.62	0.16	-0.49
Drainage Division 2 (DD2); South Victoria	South Victoria, n = 56	9.96	7.37	4.75	0.19	-0.13
	Southeast Victoria, n = 35	7.52	5.48	3.42	-0.15	-0.44
	Southwest Victoria, n = 21	6.88	5.23	3.53	0.58	0.44
Drainage Division 4 (DD4); North Victoria	North Victoria, n = 57	9.99	7.27	4.64	0.07	-0.53
	Northeast Victoria, n = 38	6.64	4.08	2.26	-0.94	-2.81
	Northwest Victoria, n = 19	8.31	7.18	5.21	1.81	3.37

3.2 Degree of heterogeneity versus relative error

Figure 3 shows the association between absolute median relative error (ABSMedRE) of quantile estimates and the degree of heterogeneity (mean value of H_1) for all the assumed regions.

Forming Victoria as a single region comprising 113 catchments, the estimated mean H_1 value is 26.6. For this region, the ABSMedRE ranges from 32.7% (Q_2) to 40.5% (Q_{100}) among different quantiles. Splitting the 113 catchments by drainage division into two regions; south Victoria (n = 56) and north Victoria (n = 57) the H_1 value reduces gradually, but the ABSMedRE does not decline remarkably. The H_1 values for these two regions are 16.8 and 20.2, respectively. Compared to Victoria as a single region, the ABSMedRE in north Victoria goes down approximately by 8-15% (32.7% vs 30.2% for Q_2 and 40.5% vs 34.5% for Q_{100}), but in the south Victoria region it gets up sharply for higher return period from 32.5% for Q_2 to almost 54% for Q_{100} (raised by 1-33%; 32.7% vs 32.5% for Q_2 and 40.5% vs 54% for Q_{100}). Southeast, a sub-region of south Victoria, the ABSMedRE value ranges from 29.6% (Q_5) to 44.5% (Q_{100}) with the highest H_1 value (14.69) within the sub-regions. In contrast, with relatively lower H_1 statistics (9.59) the ABSMedRE rises to its highest position (61.1% for Q_2 and 75% for Q_{100}) in the southwest of Victoria. These are the highest values among all the regions. The ABSMedRE in northeast Victoria ranges from 31.5% (Q_2) to 42.14% (Q_{100}) with a higher H_1 value (13.35). However, with the lowest H_1 statistic (3.56) in the northwest of Victoria, the ABSMedRE value ranges from 32.5% (Q_2) to 46.5% (Q_{100}). In this sub-region, the ABSMedRE attains the lowest value (21.4%) for Q_5 with the least H_1 value. In figure A1 these variations are visualised more clearly through some boxplots considering the maximum, average, and minimum H_1-statistics. It indicates that in RFFA the degree of heterogeneity does not impact the relative error of prediction by QRT in the study area.

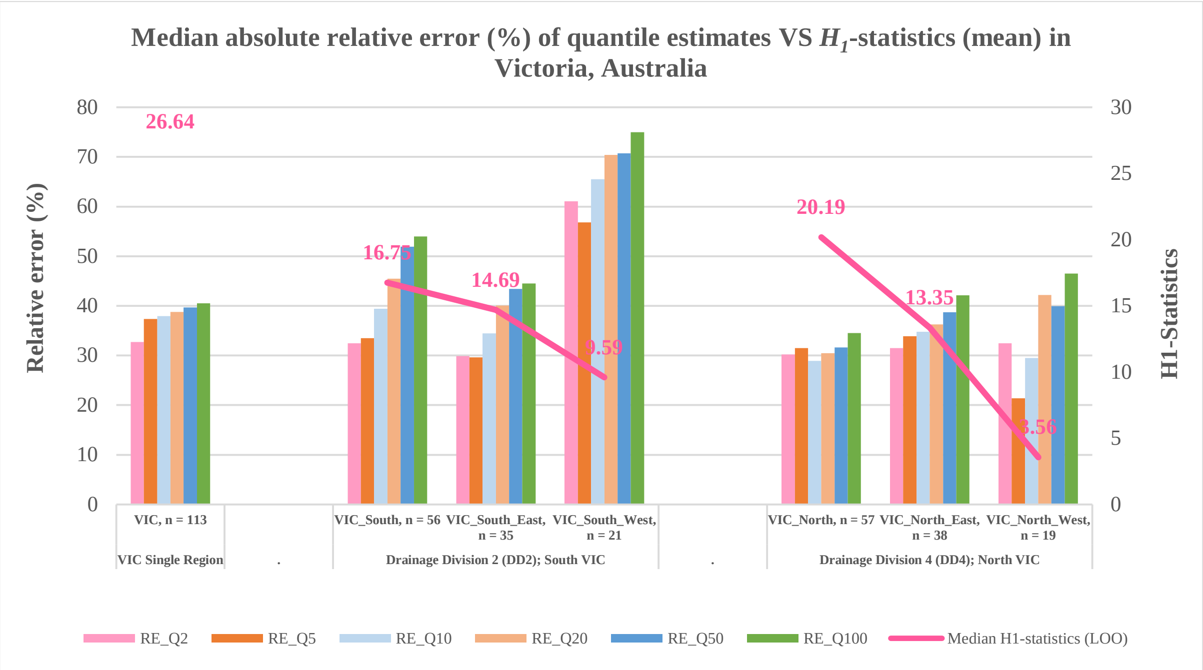

Figure 3. Median absolute relative error vs H_1-statistics (mean) for different regions in Victoria

Figure 4 illustrates the comparison of R^2 values of the developed regression model for flood quantiles among assumed regions. Usually, higher relative error is associated with the lower R^2 values (see table A1) as expected.

In the assumed regions, for Victoria as a single, north Victoria and northeast of Victoria the R^2 values remain relatively similar. For these three regions R^2 values are almost stable and range from 0.53 (Q_{100}) to 0.70 (Q_2), 0.56 (Q_{100}) to 0.67 (Q_2), and 0.57 (Q_{100}) to 0.69 (Q_2), respectively. Likewise, in the south, southeast, southwest, and northwest part of Victoria, the R^2 values are higher than the previous three regions. The maximum and minimum R^2 values for these regions are 0.81 (Q_2) and 0.63 (Q_{100}), 0.88 (Q_2) and 0.70 (Q_{100}), 0.80 (Q_5) and 0.71 (Q_{100}), 0.89 (Q_5) and 0.83 (Q_{100}), respectively. With the increase of return period the R^2 values show a declining trend for all the assumed regions.

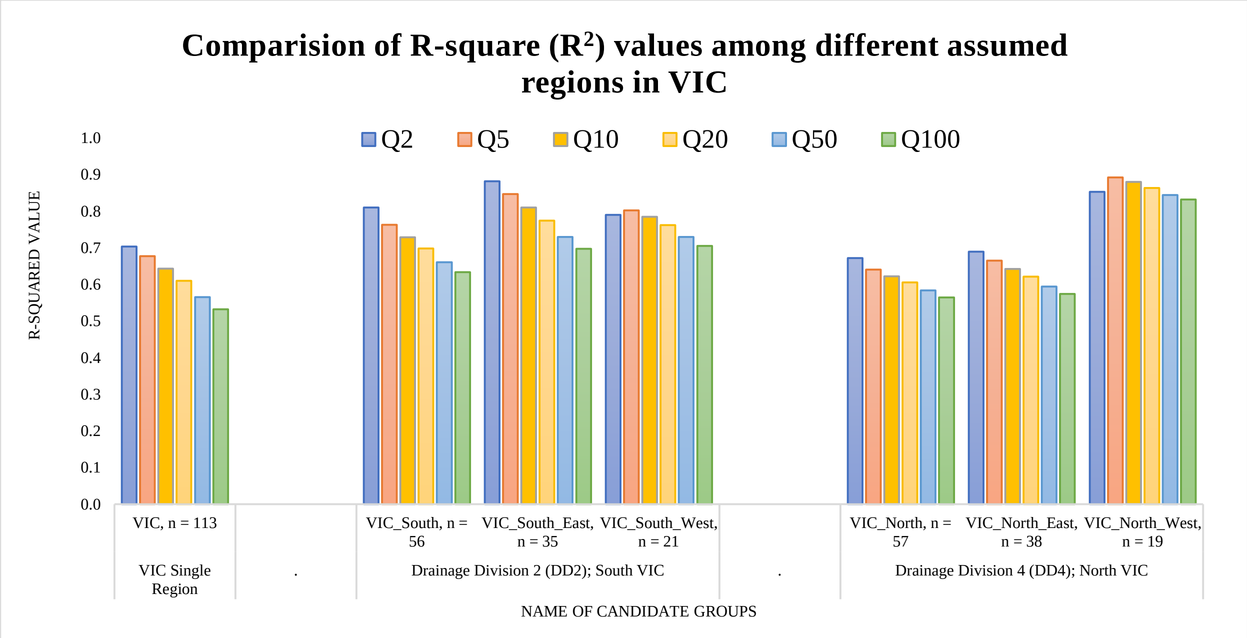

Figure 4. Comparison of R-square values of developed flood quantile models for assumed regions in Victoria, Australia

4. CONCLUSION

This study investigates the association between the degree of heterogeneity and model error (relative error – RE) in RFFA. Hosking and Wallis statistical measures are used to assessing the degree of heterogeneity. The key finding of this study is that the degree of heterogeneity does not have any impact on the relative error of flood quantile estimate. It is quite surprising and does not meet the notion that the smaller the degree of heterogeneity, the more accurate is the flood quantile estimation. The finding of this study should be verified by applying other RFFA methods like the index flood method (IFM).

ACKNOWLEDGMENTS

The authors acknowledge Victorian Government to provide streamflow data and Western Sydney University to support this study.

REFERENCES

Castellarin, A., Burn, D., & Brath, A. (2008). Homogeneity testing: How homogeneous do heterogeneous cross-correlated regions seem? *Journal of Hydrology*, *360*(1-4), 67-76.

Fill, H. D., & Stedinger, J. R. (1995a). Homogeneity tests based upon Gumbel distribution and a critical appraisal of Dalrymple's test. *Journal of Hydrology*, *166*(1-2), 81-105.

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

Lettenmaier, D. P., Wallis, J. R., & Wood, E. F. (1987). Effect of regional heterogeneity on flood frequency estimation. *Water Resources Research*, *23*(2), 313-323. https://doi.org/10.1029/WR023i002p00313

Masselot, P., Chebana, F., & Ouarda, T. B. (2017). Fast and direct nonparametric procedures in the L-moment homogeneity test. *Stochastic Environmental Research and Risk Assessment*, *31*(2), 509-522.

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

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

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

APPENDIX

Table A1. Summary of R^2 values and absolute relative errors (RE) for different candidate groups for different quantiles (Q_T), Victoria, Australia

Name of candidate groups		Criteria	Quantiles					
			Q_2	Q_5	Q_{10}	Q_{20}	Q_{50}	Q_{100}
VIC Single Region	VIC as a single region, n = 113	***R-square values***	*0.70*	*0.68*	*0.64*	*0.61*	*0.57*	*0.53*
		Absolute RE (%)	32.74	37.38	37.94	38.80	39.66	40.51
Drainage Division 2 (DD2); South VIC	South VIC, n = 56	***R-square values***	*0.81*	*0.76*	*0.73*	*0.70*	*0.66*	*0.63*
		Absolute RE (%)	32.45	33.47	39.44	45.48	51.91	53.95
	Southeast VIC, n = 35	***R-square values***	*0.88*	*0.85*	*0.81*	*0.77*	*0.73*	*0.70*
		Absolute RE (%)	29.87	29.60	34.44	40.04	43.38	44.48
	Southwest VIC, n = 21	***R-square values***	*0.79*	*0.80*	*0.78*	*0.76*	*0.73*	*0.71*
		Absolute RE (%)	61.07	56.83	65.53	70.42	70.71	75.00
Drainage Division 4 (DD4); North VIC	North VIC, n = 57	***R-square values***	*0.67*	*0.64*	*0.62*	*0.61*	*0.58*	*0.56*
		Absolute RE (%)	30.18	31.49	28.94	30.44	31.65	34.51
	Northeast VIC, n = 38	***R-square values***	*0.69*	*0.66*	*0.64*	*0.62*	*0.59*	*0.57*
		Absolute RE (%)	31.47	33.88	34.76	36.28	38.71	42.14
	Northwest VIC, n = 19	***R-square values***	*0.85*	*0.89*	*0.88*	*0.86*	*0.84*	*0.83*
		Absolute RE (%)	32.48	21.35	29.52	42.17	39.96	46.50

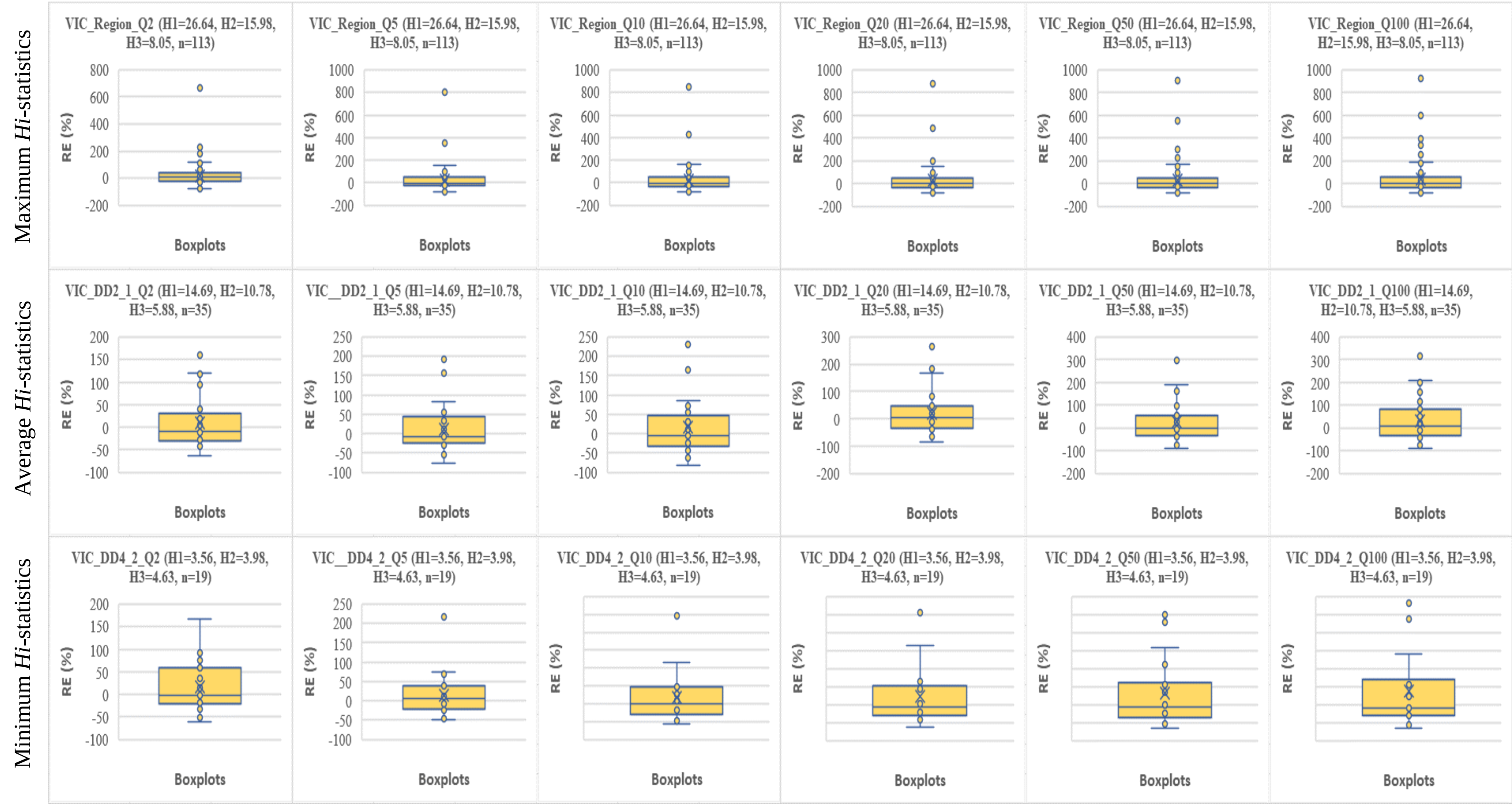

Figure A1. Boxplots of RE for different quantiles of three different candidate regions having maximum, average and minimum Hi-statistics, VIC, Australia