**Hydrology and
Earth System
Sciences**

Examination of homogeneity of selected Irish pooling groups

S. Das and C. Cunnane

Department of Engineering Hydrology, National University of Ireland, Galway, Ireland

Abstract. Flood frequency analysis is a necessary and important part of flood risk assessment and management studies. Regional flood frequency methods, in which flood data from groups of catchments are pooled together in order to enhance the precision of flood estimates at project locations, is an accepted part of such studies. This enhancement of precision is based on the assumption that catchments so pooled together are homogeneous in their flood producing properties. If homogeneity is assured then a homogeneous pooling group of sites lead to a reduction in the error of quantile estimates, relative to estimators based on single at-site data series alone. Homogeneous pooling groups are selected by using a previously nominated rule and this paper examines how effective one such rule is in selecting homogeneous groups. In this paper a study, based on annual maximum series obtained from 85 Irish gauging stations, examines how successful a common method of identifying pooling group membership is in selecting groups that actually are homogeneous. Each station has its own unique pooling group selected by use of a Euclidean distance measure in catchment descriptor space, commonly denoted d_{ij} and with a minimum of 500 station years of data in the pooling group. It was found that d_{ij} could be effectively defined in terms of catchment area, mean rainfall and baseflow index. The study then investigated how effective this selected method is in selecting groups of catchments that are actually homogenous as indicated by their L-Cv values. The sampling distribution of L-CV (t_2) in each pooling group and the 95% confidence limits about the pooled estimate of t_2 are obtained by simulation. The t_2 values of the selected group members are compared with these confidence limits both graphically and numerically. Of the 85 stations, only 1 station's pooling group members have all their t_2 values within the confidence limits, while 7, 33 and 44 of them have 1, 2 or 3 or more, t_2 values outside the confidence limits. The outcomes are also compared with the heterogeneity measures H1 and H2. The H1 values show an upward trend with the ranges of t_2 values in the pooling group whereas the H2 values do not show any such dependency. A selection of 27 pooling groups, found to be heterogeneous, were further examined with the help of box-plots of catchment descriptor values and one particular case is considered in detail. Overall the results show that even with a carefully considered selection procedure, it is not certain that perfectly homogeneous pooling groups are identified.

1 Introduction

It is widely accepted that a short annual flood (AM) series is inadequate for the estimation of design floods of large return periods. Regionalization (FSR, 1975), i.e. pooling analysis (FEH, 1999), is one of the possible methods used to provide a framework for design floods. In pooling analysis flood data are pooled from other gauging stations that possess similar hydrological behaviours to the at-site station. A very common way to implement regional/pooling is the index flood method proposed by Dalrymple (1960). The estimation of Q_T, T-year flood, based on this approach involves derivation of a growth curve which shows the relation between X_T and the return period T where $X_T = Q_T/Q_I$ and Q_I is the index flood at the site of interest. Generally the mean (FSR, 1975) or median (FEH, 1999) of the at-site AM flood series is taken as the index flood. It is assumed that the $X_T - T$ relation is the same at all sites in a homogeneous pooling group. The identification of a homogeneous pooling group is therefore important in pooling analysis. Lettenmaier et al. (1987); Stedinger and Lu (1995) and Hosking and Wallis (1997) among other researchers have demonstrated that a successful pooling analysis requires a homogeneity criterion to be satisfied.

However very recently Kjeldsen and Jones (2009) have approached this in a different way.

An examination of homogeneity is normally used to assess whether a proposed group of sites is homogeneous or not. Examination of the homogeneity of regions/pooling groups is usually based on a statistic that relates to the formulation of a frequency distribution model, e.g. the coefficient of variation, CV (Wiltshire, 1986; Fill and Stedinger, 1995) and/or skew coefficient, g, their L-moment equivalents (Chowdhury et al., 1991; Hosking and Wallis, 1997) or of dimensionless quantiles such as the 10-year event (Dalrymple, 1960; Lu and Stedinger, 1992). Hosking and Wallis (1993, 1997) proposed homogeneity tests based on L-moment ratios such as L-CV alone (H1) and L-CV and L-skewness jointly (H2) which are widely used in flood frequency analysis although the former one is recommended by these authors for having better power to discriminate between homogeneous and heterogeneous regions. Very recently, a similar conclusion has been drawn by Viglione et al. (2007) when they compared several homogeneity tests. They stated that the H1 test is ahead of all others when the L-skewness is lower than 0.23. They further concluded that the H2 as a homogeneity test lacks power. These findings certainly indicate that the heterogeneity among the sites in a group is mainly due to variations in the sample L-CVs. However, one of the main assumptions of these tests is that the true regional distribution is kappa. For that reason and others Hosking and Wallis (1997) recommended that though the heterogeneity statistic is constructed like a significance test it should not be used in that way. They, Hosking and Wallis (1997, p. 70), further stated that "...*a significance test is of doubtful utility anyway, because even a moderately heterogeneous region can provide quantile estimates of sufficient accuracy for practical purposes. Thus a test of exact homogeneity is of little interest.*" In this paper a graphical way of examining the homogeneity of a pooling group is presented which is based on L-CV , i.e. t_2. The main idea behind the approach is the comparison of the variability of t_2 from each site in the pooling group with that expected (un-weighted average pooled t_2) supposing the differences between sites to be due to sampling error. The pooling groups are identified by the Region of Influence (ROI) approach. The population distribution is GEV (with $k = -0.05$, $k = 0.0$, $k = +0.03$), rather than Kappa as suggested by Hosking and Wallis (1997), and was based on the GEV's descriptive ability of the annual maximum data series of Ireland.

The outline of the paper is structured as follows: the next section describes the procedure used to obtain growth factors and flood quantiles in the context of flood frequency pooling analysis. This is followed by a description of procedures to select pooling variables for similarity distance measure (d_{ij}) in the context of formation of pooling groups using the ROI approach. A graphical way of examining homogeneity of pooling groups obtained by the ROI approach is then presented. Then the analysis of the examination procedure is summarised and finally a selected number of heterogeneous pooling groups are reviewed with the help of Box-plots of catchment descriptors.

2 Estimation of pooled growth factors and flood quantiles

The growth factor X_T is the factor which when multiplied by the index flood Q_I , gives the flood magnitude of return period T, Q_T, as in Eq. (1)

$$Q_T = Q_I \times X_T \tag{1}$$

The relationship between X_T and T is often referred to as the growth curve. When a growth curve is obtained by pooling the information from sites of a pooling group, it is called the pooled growth curve. Qmed is used as the index flood in this study where Qmed is the median of the annual maximum series.

In this study the pooled growth curve is obtained using the approach based on the method of L-moments. The L-moments developed by Hosking (1990) are based on probability weighted moments (PWMs) introduced by Greenwood et al. (1979). With this approach the derivation of a growth curve in a pooling group involves the following key steps:

1. computation of at-site and pooled L-moment ratios

2. selection of a suitable form of distribution and estimation of its parameters by the method of L-moments.

L-moments are calculated and then the dimensionless L-moment ratios t_2 and t_3 are calculated for each site. Pooled L-moment ratios for the target site, i, are then computed using the following equation:

$$t^{(i)R} = \frac{\sum_{j=1}^{M} w_{ij} \, t^{(j)}}{\sum_{j=1}^{M} w_{ij}} \tag{2}$$

where $t^{(j)}$ is the L-moment ratio (either t_2 or t_3) for the j-th most similar site and w_{ij} is a weighting term.

Weights can be related to a site's record length and/or a site's d_{ij} values. Recently a more complex way of assigning weights is proposed by Kjeldsen and Jones (2009) although they state that only a little has been gained in the flood estimation procedure using the new approach. In this study w_{ij} is taken as 1. Choice of unweighted averages was guided by the observations made by Hosking and Wallis (1997, p.90), namely "The calculation of regional averages by weighting the sites proportionally to their record lengths is not essential. If the region is exactly homogeneous, then a good approximation of the variance of $t(i)$ is proportional to $n(i) - 1$, and in this case weighting the sites proportionally to their record lengths minimizes the variance of the regional average t^R. If the region is heterogeneous, it is possible that weighting proportionally to record length may give undue influence to sites that have frequency distributions markedly different from the region as a whole and that also have long records".

The Generalised Extreme Value (GEV) has been selected as the pooled distribution function. The selection of the GEV distribution is explained in Sect. 4. The values t_2^R, t_3^R are equated to expressions for these quantities written in terms of the distribution's unknown parameters (expressed in dimensionless form) and the resulting equations are solved for the unknown parameter values. The dimensionless GEV growth curve (X_T) is defined by two parameters k and β:

$$X_T = 1 + \frac{\beta}{k}\left((\ln 2)^k - \left(\ln \frac{T}{T-1}\right)^k\right) \tag{3}$$

where T is the return period.

The two parameters k and β are estimated from the sample L-CV, t_2, and sample L-skewness, t_3, as follows (Hosking and Wallis, 1997)

$$k = 7.8590\,c + 2.9554\,c^2 \tag{4}$$

in which

$$c = \frac{2}{3+t_3} - \frac{\ln 2}{\ln 3} \tag{5}$$

$$\beta = \frac{k\,t_2}{t_2\big(\Gamma(1+k) - (\ln 2)^k\big) + \Gamma(1+k)\big(1 - 2^{-k}\big)} \tag{6}$$

where Γ denotes the complete gamma function.

3 Formation of pooling groups using Region Of Influence (ROI) approach

The Region of Influence (ROI) approach of formation of a pooling group is considered to be the most appropriate and meaningful way of delineating a pooling group. The technique developed by Burn (1990), involves the identification of a region of influence i.e. a separate pooling group for each gauging station in a region. The identification of a pooling group consists of selecting stations that are hydrologically similar to the site of interest. Similarity is measured generally by a Euclidean distance measure in catchment descriptor space.

The effective identification of a pooling group in a ROI approach is governed by two important criteria: the choice of appropriate site descriptors as pooling variables and the size of a group in terms of number of sites and station years included. Burn (1990) investigated a number of options to determine a threshold value based on the d_{ij} values to define a cut-off for the inclusion of stations in the ROI method for a target site. However, a more practical way of choosing an appropriate size of a pooling group was presented by FEH (1999). They investigated a range of pooling group sizes and decided on adoption of the $5T$ rule, namely that the total number of station years of data to be included when estimating the T year flood should be at least $5T$. The adoption of such a rule was a compromise. If too few stations are included the precision of the Q_T estimate is sacrificed whereas if far too many stations are included then the assumption of homogeneity may be compromised. Hosking and Wallis (1997) however show that a small departure from homogeneity can be tolerated so that having too few stations included may be less desirable than having slightly too many. They also suggested not to use more than 20 sites in a group as little gain in the accuracy of quantile estimates is obtained by using more than about 20 sites in a group. Recently, Kjeldsen and Jones (2009) found that a fixed pooling group consisting of 500 station years performed well for a range of return periods. In relation to identifying site descriptors as pooling variables, careful consideration is necessary as to which form of catchment descriptors are to be used in a ROI method of pooling analysis. In the next subsection an investigation of selecting pooling variable for the Irish case is described in detail.

3.1 Choice of catchment descriptors on effectiveness of ROI distance measures

The general form of the similarity measure used for selecting members of a pooling group is defined by

$$d_{ij} = \sqrt{\sum_{k=1}^{n} W_k \left(X_{k,i} - X_{k,j}\right)^2} \tag{7}$$

where d_{ij} is the weighted Euclidean distance from site j to site i; n is the number of attribute variables; $X_{k,i}$ is the value of the k-th variable at the i-th site and W_k is the weight applied to attribute k, reflecting its relative importance. The subscript i denotes the subject site and the subscript j denotes the j-th pooled site.

In choosing a distance measure d_{ij} a decision has to be made about which catchment descriptors are to be included in the distance measure and what weightings are to be applied to them and whether logarithms or other transformations are to be used. The FEH (1999) provided a number of useful maxims for choosing a distance measure. It recommended not to use at-site flood statistics (e.g. CV, g) as pooling variables because this might well result in groups consisting of sites that have experienced similar floods in recent history. Neither could such site flood statistics be used for ungauged catchments. Seasonality of the flood response (e.g. timing and regularity of flood events) has also been considered (Burn, 1997; Cunderlik and Burn, 2006) as a similarity measure. Seasonality statistics are obtained from observed flood series. Therefore, a similarity measure based on these could not be used for ungauged sites, without additional assumptions.

For Irish conditions two sets of catchment descriptors have been selected as potential pooling variables:

- similar variables as used in FEH i.e., AREA (catchment area), SAAR (standard average annual rainfall), BFI (baseflow index) and FARL (index of flow attenuation by reservoir and lake)

– on the assumption that homogeneity is strongly dependent on CV or L-CV, those catchment descriptors that could predict L-CV best were identified and a selection of these were used to form d_{ij}. This approach is along the lines outlined by Kjeldsen and Jones (2009).

For selecting the final set of pooling variables, FEH used pooled uncertainty measure (PUM) which is a weighted average of the squared differences between each at-site growth factor and the pooled growth factor measured on a logarithmic scale. In this part of the study a simulation procedure is used for this purpose because far fewer stations (85) than the 602 stations used for the UK study were available. The first objective is to find which combinations of FEH descriptors, which are listed in Table 2, lead to pooling groups which are most effective at exploiting the information about the flood distribution contained in the pooling groups.

The simulation procedure uses the GEV distribution for data generation which is considered to be representative of what is appropriate in Irish conditions. Hosking and Wallis (1997, p.93) suggested not to use the observed sample L-moment ratios as the population L-moment ratios of the simulated region because this would yield a simulated region that has much more heterogeneity than the actual data. Castellarin et al. (2001) addressed the issue by using a region of influence approach to estimate the at-site population values of t_2 and t_3. A similarity measure based on at-site flood statistics is used to form a group of sites for a subject site and its population values of t_2 and t_3 are considered as the corresponding pooled estimate of t_2 and t_3 for the group. Later, Gaál et al. (2008) adopted this approach in their study. A similar kind of approach is used here with a similarity measure defined as

$$\delta_{ij} = \sqrt{\left(\frac{t_{2,i} - t_{2,j}}{\sigma_{t_2}}\right)^2 + \left(\frac{t_{3,i} - t_{3,j}}{\sigma_{t_3}}\right)^2} \qquad (8)$$

which is independent of the descriptor variables being considered in Table 2. A pooling group is formed for each site using Eq. (8) and the pooled t_2 and t_3 are estimated using Eq. (2). The estimated pooled values of t_2 and t_3 are then used as population values for each site in step 2 of the simulation procedure. The simulation procedure does not consider the implications of intersite correlation among sites in a pooling group because it was found by Hosking and Wallis (1997, p.127) to be of very little consequence. The steps of the simulation procedure for selecting variables are described as follows.

1. The gauging stations in the subject site's pooling group are identified using the d_{ij} values of Eq. (7) for a set of catchment descriptors having a minimum of 5T station years of data in the pooling group.

2. Random samples are drawn from GEV populations for the subject site and for each site in the pooling group.

For each site the sample size is taken as being equal to the length of the observed historical record at the site and the parameters are estimated from the site t_2 and t_3 values obtained using the procedure described above, as in Castellarin et al. (2001) and Gaál et al. (2008).

3. The t_2 and t_3 values are obtained for each sample in the pooling group and the average of these is calculated to represent the pooled t_2 and t_3 values.

4. The pooled t_2 and t_3 values are then used to determine the pooling group's GEV growth curve parameters k and β using Eqs. (4) and (6).

5. The subject site's $\hat{X}_T$ value is calculated for $T = 50$ and 100 years respectively using Eq. (3).

6. Steps 2 to 5 are repeated 10 000 times to provide 10 000 values of $\hat{X}_T$ and the RMSE_T and BIAS_T are calculated for the subject site by the following equations:

$$\mathrm{RMSE}_T\,[\%] = \frac{1}{M}\sum_{i=1}^{M}\sqrt{\frac{1}{S}\sum_{s=1}^{S}\left(\frac{\hat{X}_{i,s}^T - X_i^T}{X_i^T}\right)^2} \times 100 \quad (9)$$

$$\mathrm{BIAS}_T\,[\%] = \frac{1}{M}\sum_{i=1}^{M}\frac{1}{S}\sum_{s=1}^{S}\left(\frac{\hat{X}_{i,s}^T - X_i^T}{X_i^T}\right) \times 100 \qquad (10)$$

where $\hat{X}_{i,s}^T$ is the estimated T-year growth factor at a site i at the s-th repetition; X_i^T is the assumed true T-year growth factor at site i; M is the number of sites in the pooling group and S is the number of repetitions.

RMSE_T and BIAS_T defined in the simulation procedure has been evaluated at 50 and 100-year return periods for each site. The eight combinations listed in Table 2 of the four variables have been tested based on RMSE_T (primarily) . In all, 85 stations have been considered for the study. The data sets that have been used in the study are summarized in Table 1. For each of these sites, a pooling group was selected from the 85 stations. Initially in the simulation procedure all weights W_k in Eq. (7) were set to unity. Figures 1 and 2 shows, in box-plot form, respectively the variation in the 100-year RMSE and BIAS values for different sets of catchment descriptors used in Eq. (7). In Table 2, the corresponding mean variation of RMSE_{100} and RMSE_{50} values, for different sets of pooling variables, is summarised. It shows that the numerical measures of effectiveness vary by very little between rows. The set of two variables, lnAREA and lnSAAR, and the set of the single variable lnAREA performed best in terms of providing the lowest RMSE_{100} values. In terms of RMSE_{50}, the set consisting of lnAREA and lnSAAR comes second best to the set consisting of lnAREA on its own. Overall, the set of variables comprised of lnAREA and lnSAAR may be considered as being
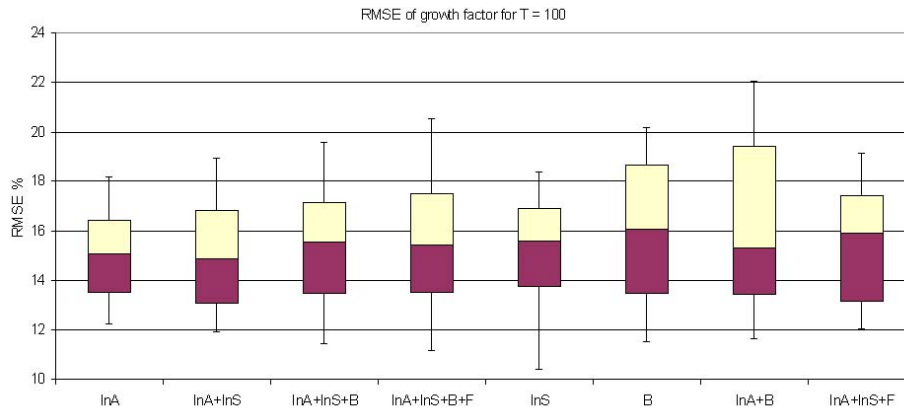
Fig. 1. Box-plot of RMSE of growth factors corresponding to 100 yr return periods for different sets of catchment descriptors used in defining the distance measure d_{ij}. Each Box-plot gives the percentiles for the frequencies 0.05, 0.25, 0.5, 0.75, 0.95.
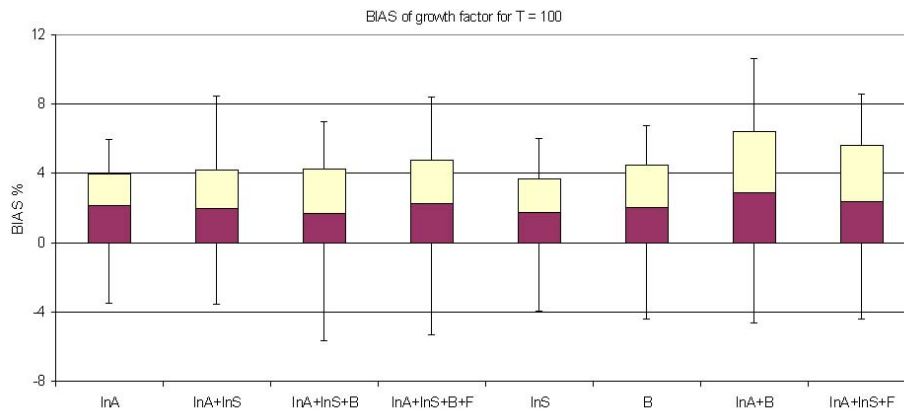

Fig. 2. Values of BIAS for the analysis summarised in Fig. 1.

the most suitable set of pooling variables for Irish conditions. However, if there is also a desire to incorporate another physical catchment effect then the BFI could be included with these two. While inclusion of just one or two catchment descriptors may indeed be best, there is an intuitive attraction in also representing some descriptor of catchment response even at the cost of a small apparent loss in effectiveness. This could be of relevance in engineering investigations where differences in catchment behaviour are considered of importance by the investigator. An extension to this investigation with varying values of weights W_k in Eq. (7) was also done, particularly for the set of variables of lnAREA, lnSAAR and BFI but the results of all variations examined are not reported in detail here. An automatic search procedure was not used but it was found, by trial and error, that the weights 1.5, 1.0 and 0.1 for lnAREA, lnSAAR and BFI respectively gave RMSE$_{100}$ = 15.22 and RMSE$_{50}$ = 12.81 which offer small improvements on the W_k = 1.0 values used in the calculations for the set of variables of lnAREA, lnSAAR and BFI. The trial and error approach involved assigning a selection of weights, varying from 0 to 3, to each of the quantities, i.e. lnAREA, lnSAAR and BFI.

Table 1. Summary of AMF data sets used in the study.

Number of stations	85
Shortest record length	17
Longest record length	55
Mean record length	36.5
Number of AMF events	3213

In the second approach a set of catchment descriptors were identified through the use of regression models of L-CV on the catchment descriptors. These descriptors were then also used as potential pooling variables. In the search for a best regression model both log-transformed and non-transformed variants of the catchment descriptors and L-CV were used. The best regression model for L-CV containing three catchment descriptors was found to be based on MSL, FORMWET and ARTDRAIN, where MSL is the mean stream length, FORMWET is a form of catchment wetness index analogous to PROPWET in FEH (1999) and ARTDRAIN is an arterial drainage index which is defined as

Table 2. Variation in the mean RMSE corresponding to $T = 100$ and 50 for different sets of pooling variables: variables as used in FEH (1999).

Variables used in model	RMSE$_{100}$%	RMSE$_{50}$%
lnAREA (lnA)	15.13	12.47
lnAREA, nSAAR (lnA + lnS)	15.11	12.77
lnAREA, lnSAAR, BFI (lnA + lnS + B)	15.52	13.22
lnAREA, lnSAAR, BFI, FARL (lnA + lnS + B + F)	15.57	13.20
lnSAAR (lnS)	15.27	13.23
BFI (B)	15.97	13.83
lnAREA, BFI (lnA + B)	16.21	13.44
lnAREA, lnSAAR, FARL (lnA + lnS + F)	15.54	12.78

Table 3. Variation in the mean RMSE corresponding to $T = 100$ and 50 for different sets of pooling variables: variables that predict L-CV best.

Variables used in model	RMSE$_{100}$%	RMSE$_{50}$%
lnMSL	15.12	12.59
lnMSL, FORMWET	15.19	12.82
lnMSL, FORMWET, ARTDRAIN	15.23	12.81
FORMWET	16.89	15.41
ARTDRAIN	15.80	13.97
lnMSL, ARTDRAIN	15.09	12.80

% of catchment area affected by arterial drainage improvements. These descriptors were identified from a pool of twenty five catchment descriptors made available by the Irish Office of Public Works. The R^2 value of the best available model is a modest 29%.

These identified catchment descriptors were also assessed by the above simulation procedure. The RMSE$_T$ values for $T = 50$, 100 are listed in Table 3 for six combinations of the three variables. The set of two variables, lnMSL and ARTDRAIN, and the set of the single variable lnMSL performed best in terms of providing the lowest RMSE$_{100}$% values.

Both approaches described above provide similar outcomes in terms of RMSE$_{100}$%. This may be partly due to the relatively weak relations identified for predicting L-CV ($R^2 = 0.29$). A regression of L-CV on the other set's catchment descriptors (AREA, SAAR, BFI, FARL) also yields a weak relation for predicting L-CV ($R^2 = 0.21$). Since both sets of catchment descriptors can predict L-CV only in a weak manner, and both approaches are similar in RMSE it is concluded that neither approach is clearly superior to the other.

4   Procedure for examination of homogeneity

A homogeneity test is used to assess whether a proposed group of sites is homogeneous or not. A homogeneous group of sites leads to a reduction in the error of quantile estimators relative to estimators based on single at-site data series alone which is the main goal of a regional flood frequency analysis. A homogeneity test was introduced by Dalrymple (1960). Other tests were introduced by Wiltshire (1986), Lu and Stedinger (1992), Fill and Stedinger (1995) and Hosking and Wallis (1993, 1997).

A simulation procedure, using graphical presentation of key results is applied in this study to examine homogeneity of pooling groups that were formed using the ROI technique. GEV distributions with 3 different shape parameter values ($k = -0.05$, $k = 0.0$ (EV1), $k = +0.03$) are used in

the simulation procedure. The GEV, and its special case the EV1, have a history of usage in Ireland since publication of the Flood Studies Report (FSR, 1975, p.173–174, Table 2.38, Fig. 2.14, Vol. I). More recently, a national study sponsored by the Office of Public Works, Dublin, based on annual maximum flood data of 110 stations, with average length of record 37 years and with a quarter of them between 50 and 55 years, has indicated that GEV and EV1 distributions are suitable parents for the majority of Irish flood series (Das, 2010, Ch. 3). This conclusion is based on visual examination of probability plots and numerical scores assigned to them, on classical goodness of fit tests and on L-Moment Ratio diagrams, such as Fig. 10 which shows that GEV/EV1 looks more suitable as a parent than other 3-parameter distributions tested such as Generalised logistic and Lognormal 3. While the 4 parameter Kappa distribution has been recommended by Hosking and Wallis (1997) as a parent for simulation studies, this choice was sometimes found to be problematical because of numerical difficulties and estimation failures during the parameter estimation process and as a result GEV was selected as parent distribution in this study.

The steps of the simulation procedure are as follows:

1. The gauging stations in the subject site's pooling group are identified using d_{ij} values obtained by the following equation having a minimum of 500 station years of data in the pooling group and satisfying the $5T$ rule for the 100 year quantile.

$$d_{ij} = \qquad\qquad\qquad\qquad\qquad\qquad (11)$$

$$\sqrt{1.5\left(\frac{\ln A_i - \ln A_j}{\sigma_{\ln A}}\right)^2 + \left(\frac{\ln \text{SAAR}_i - \ln \text{SAAR}_j}{\sigma_{\ln \text{SAAR}}}\right)^2 + 0.1\left(\frac{\text{BFI}_i - \text{BFI}_j}{\sigma_{\text{BFI}}}\right)^2}$$

The weights 1.5, 1.0 and 0.1 are those reported in Sect. 3 above.

2. The t_2 is obtained for each site in the pooling group and the average, without weights, of these is calculated to represent the pooled average t_2 (t_2^R).
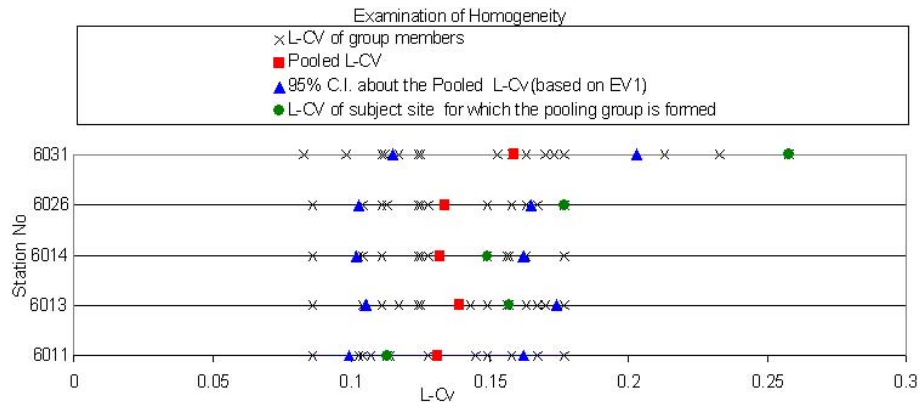
Fig. 3. Examination of homogeneity (EV1 case).

Table 4. Summary of events outside the confidence limits for 85 pooling groups.

events outside the CL (m)	GEV ($k = -0.05$)	EV1 ($k = 0$)	GEV ($k = +0.03$)
	No. of groups (%)	No. of groups (%)	No. of groups (%)
0	2 (2)	1 (1)	0 (0)
1	13 (15)	7 (8)	7 (8)
2	12 (14)	8 (9)	8 (9)
3	24 (28)	25 (29)	23 (27)
>3	34 (40)	44 (52)	47 (55)

Table 5. Summary of heterogeneity measure, H1 and H2 for 85 pooling groups.

Heterogeneity measure	% of groups with heterogeneity < 2	% of groups with heterogeneity < 4
H1	5	22
H2	38	86

is also noted whether the t_2 of the subject site is outside the confidence limits (CL).

4.1 Analysis

The procedure described above is applied for each of the 85 stations. Each station had its own unique pooling group. The sample values of t_2 for the stations in the group, t_2^R and the CL about t_2^R are displayed in Fig. 3 for five stations. The summary statistics of the procedure are given in tabular form in Table 4. In addition to that the heterogeneity measures, H1 and H2, described in Appendix A, for each group is calculated and a summary of these measures is reported in Table 5.

The following observations and findings are obtained from the analysis.

3. Random samples are drawn from GEV distributions with 3 different shape parameter values ($k = -0.05$, $k = 0.0$ (EV1), $k = +0.03$) using the t_2^R as the population value to construct a 95% confidence interval for t_2^R. These population shape parameters, $k = -0.05$, $k = 0.0$ (EV1) and $k = 0.03$, are selected in this context which correspond to L-skewness ≈ 0.21, 0.17 and 0.15 respectively, this being the range relevant for Ireland. The sample size is taken as being equal to the average record length of the observed historical record at the gauging sites and the parameter values are estimated from the value of the t_2^R. The 95% confidence interval is constructed assuming that the samples t_2^R values are normally distributed. While the L-CV values may not be perfectly normally distributed Viglione's (Viglione, 2010) results show that the departure from normality is not severe for the range of L-CV and L-skewness values that are observed in Irish conditions. Hence the normality assumption was made in the calculation of confidence intervals.

4. The number of stations in the selected pooling group whose t_2 values fall outside the confidence interval (the attribute termed here as m) is counted and reported. It

1. Table 4 lists how many stations, m fall into the categories of zero value outside the CL, one value outside the CL, 2 values outside the CL, 3 values outside the CL or more than 3 outside the CL. In all, for the case of EV1, only one station (1%) was in the first category while 52% of stations were in the latter category. This information in the form of proportion of values, m/N where N is the number of stations in the pooling group, falling outside the CL, ranges from 0.08 to 0.08 when $m = 1$, from 0.13 to 0.18 when $m = 2$, from 0.2 to 0.3 when $m = 3$ and from 0.25 to 0.55 when $m > 3$. m/N
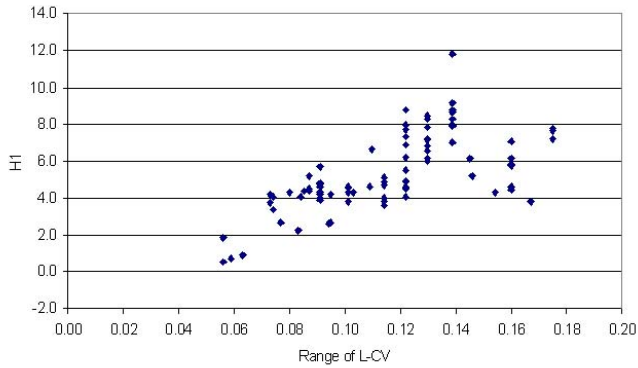
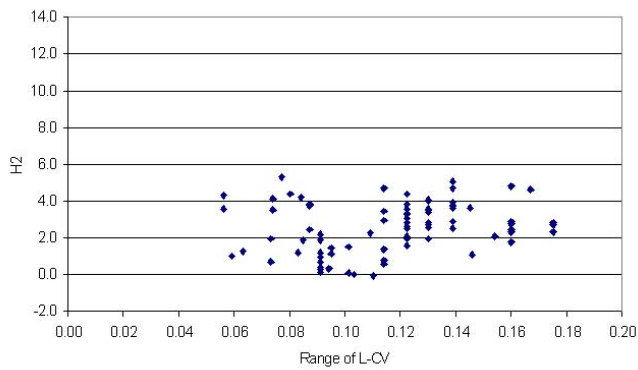**Fig. 4.** H1 plotted versus range of L-CV. Each point represents a pooling group.

Fig. 5. H2 plotted versus range of L-CV. Each point represents a pooling group.

values for GEV ($k = -0.05$) and GEV ($k = +0.03$) are broadly similar.

2. From Table 4, it is seen that as the shape parameter increases from $k = -0.05$ to $+0.03$ the number of cases where $m > 3$ increases from 33 to 47.

3. In 27 groups (32% of groups) the t_2 of the subject site was outside the CL for the case of EV1. The corresponding numbers for the case of negative shaped GEV and for the case of positive shaped GEV are 27 and 28 respectively. All the 27 stations of the EV1 case were also in the latter cases.

4. Table 5 summarises the results of H1 and H2 for the 85 pooling groups. 22% of groups have a H1 value lower than 4.0. The percentage increases to 86% when the same criterion is set for H2 and that is very similar to what was found for the UK pooling groups (FEH, 1999, p. 176).

5. The range of t_2 values, max t_2–min t_2, was calculated for the 85 pooling groups. The average range of t_2 for the 85 pooling groups was 0.11 with a minimum value
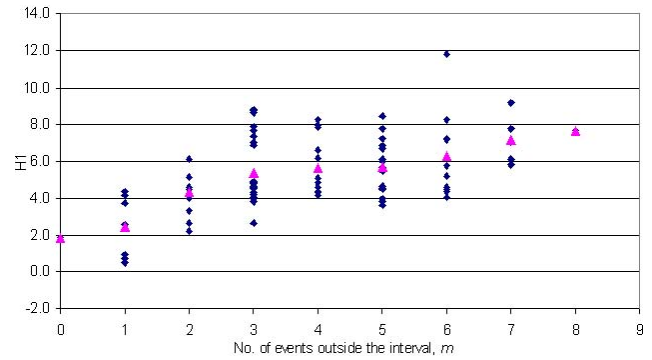

Fig. 6. H1 plotted versus m. Each point represents a pooling group.

of 0.06 and a maximum value of 0.18. Figure 4 shows a plot between H1 values and ranges of L-CV values for the 85 groups. The plot shows an upward trend, implying that a high H1 value can be expected when the t_2 values in a pooling group have a large range, which can be expected in the absence of homogeneity. A similar plot is drawn for H2 in Fig. 5, showing no obvious trend, implying that a low H2 value may be obtained for a pooling group which is in fact a heterogeneous group.

6. Figure 6 shows a plot between H1 and m. Different values of H1 occur for a particular m value and that is reasonable as the memberships of the groups in those cases are different even though they may have some overlap. However, the average values, marked by triangles in the plot, show an increase of H1 with m, i.e. the higher the number of t_2 values of group members outside the CL, the higher the value of H1 that can be expected. If a H1 value less than 4.0 is considered as a good criterion for testing homogeneity, then in this approach it is required that fewer than $m = 2$ values fall outside the confidence limit, i.e. $m/N \leq 0.15$.

7. Figure 7 shows a plot between H1 and $d_{ij,\max}$ of the pooling groups. The $d_{ij,\max}$ is defined here as the distance associated with the group member which just qualified as a member of the pooling group. The plot shows an upward trend to some extent, implying that a low H1 value can be expected for a low $d_{ij,\max}$ value, which is an implicit assumption of a ROI pooling scheme. However in many cases, low $d_{ij,\max}$ values, even those below 1.0, can lead to a high value of H1 suggesting that the assumption may not always be true particularly for Irish conditions. A similar plot is drawn in Fig. 8 between $d_{ij,\max}$ and m. The plot leads to a similar conclusion to that for Fig. 7. While a low value of $d_{ij,\max}$ is desirable, it is noted that even low values of $d_{ij,\max}$ can occur where a significant number of group members' t_2 values falls outside the CL.
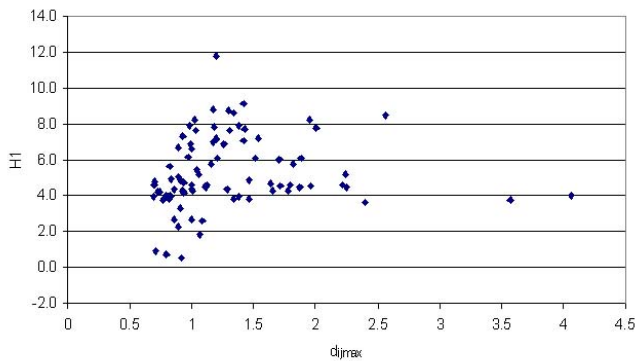
Fig. 7. H1 plotted versus $d_{ij,\max}$. Each point represents a pooling group.

Fig. 8. m plotted versus $d_{ij,\max}$. Each point represents a pooling group.

5 Investigation of selected heterogeneous pooling groups

The investigation has been carried out on those 27 cases where the pooling groups are heterogeneous and in which the t_2 of the subject site lies outside the confidence limits. The investigation mainly focuses on identifying any inappropriateness among group members that would cause the pooling groups to be heterogeneous. In this context, FEH (1999, 3, Fig. 16.9) documented a detailed review system, providing an example. That system mainly considers two attributes: (1) whether the subject site has any special qualities that need to be taken into account and (2) whether any of the pooled sites has catchment descriptors that are particularly different from those of the subject site. Sites in the pooling group can be investigated using several characteristics including at-site flood statistics and catchment descriptors. Statistics in a pooling group such as discordancy measure (Hosking and Wallis, 1997) and the distance measure (d_{ij}) can also be used to investigate sites in the pooling group. In this part of the study, four catchment descriptors, namely, AREA, SAAR, BFI, FARL; and the distance measure (d_{ij}) are taken into account in the investigation process. The first three of the catchment descriptors, AREA, SAAR and BFI,were already used for initial selection of sites for a pooling group. In the investigation procedure, sites are reviewed with the help of Box-plots and a summary table and in some cases, with the help of the 'examination of homogeneity' chart described in Sect. 4. Four Box-plots of catchment descriptors, such as AREA, SAAR, BFI and FARL, are constructed to show the subject site in the context of the pooling group. For each of these catchment descriptors, the placement of numerical values for sites in the pooling group is displayed against a backdrop of the relative frequency of the 85 sites considered in this study. This facilitates the identification of any particularly inappropriate sites. In the summary table, statistical properties such as t_2, t_3 and d_{ij} values of sites in a pooling group are listed as shown in Fig. 9. The investigation procedure for pooling groups of station no 6031 is described in detail as it serves as an example.
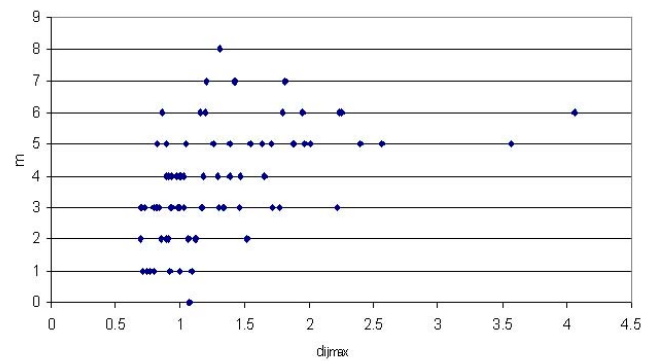
An example: station no 6031 on the River Flurry

There are 17 sites in the pooling group of which eight, including the subject site, have values which fall outside the CL, thus indicating a strongly heterogeneous group. The heterogeneity measures H1 and H2 for the group are 7.66 and 2.82 respectively. The examination of Box-plots in Fig. 9 reveals the catchment area of the subject site is small (46.2 km^2) and it is very near to the 5 percentile mark on the Box-plot of AREA. The site is not positioned at the centre of the group of gauged catchments in the pooling group. There are 5 sites on the left of the subject site and there are as many as 11 sites on the right. The attribute certainly includes some sites that have large catchment area compared to the subject site. This may lead to d_{ij} values exceeding the value 1.0 in several cases. The d_{ij} values for the last three sites are around 1.3 and these sites are among the seven other sites that fall outside the CL. The examination of the summary table on the right hand side of Fig. 9 shows that the subject site has large values of both t_2 and t_3 and that these are the largest among the group members. Hence, the conclusion can be drawn here that the pooling group in its present structure may not be ideal for that subject site 6031. Leaving out some sites at the bottom of the table might be considered in this context. The large number of sites, 17, in the pooling group is also a possible contributor to heterogeneity.

The remaining 26 pooling groups of 27 heterogeneous pooling groups were also investigated and in 8 cases the heterogeneity was due to exceptionally large or small catchment area relative to the other group members. Likewise 3 cases were similarly caused by exceptionally large SAAR or small SAAR values and 3 cases by exceptionally large BFI or small BFI values. A further 5 cases were caused by extremely low FARL values relative to other pooling group members. In 7 cases there was no obvious single cause of heterogeneity.
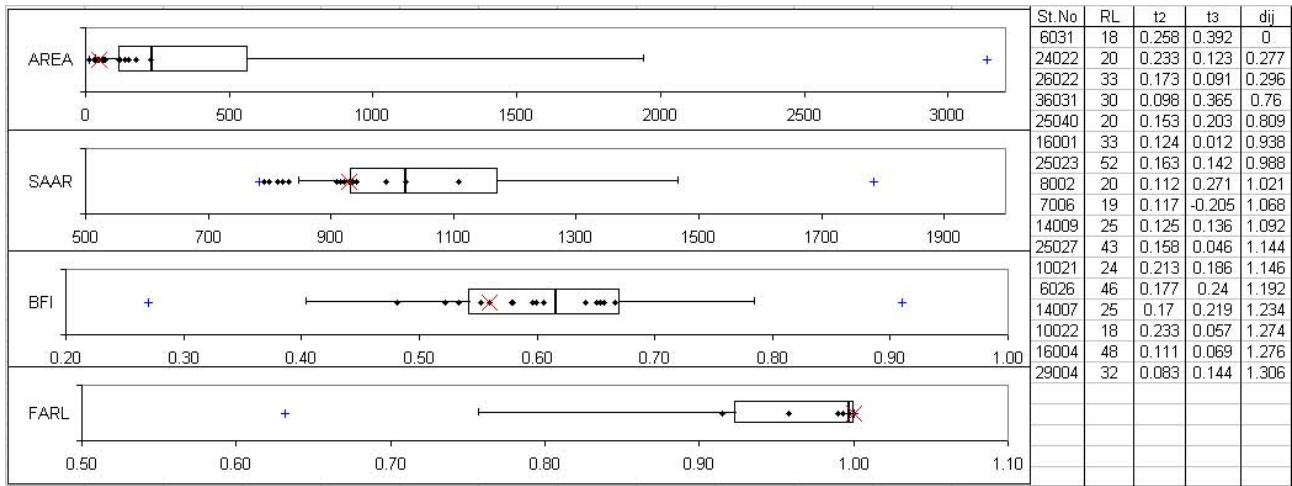
St.No	RL	t2	t3	dij
6031	18	0.258	0.392	0
24022	20	0.233	0.123	0.277
26022	33	0.173	0.091	0.296
36031	30	0.098	0.365	0.76
25040	20	0.153	0.203	0.809
16001	33	0.124	0.012	0.938
25023	52	0.163	0.142	0.988
8002	20	0.112	0.271	1.021
7006	19	0.117	-0.205	1.068
14009	25	0.125	0.136	1.092
25027	43	0.158	0.046	1.144
10021	24	0.213	0.186	1.146
6026	46	0.177	0.24	1.192
14007	25	0.17	0.219	1.234
10022	18	0.233	0.057	1.274
16004	48	0.111	0.069	1.276
29004	32	0.083	0.144	1.306

Fig. 9. Four Box-plots and a summary table for investigating a pooling group. The subject site is marked with a ×. Small dots denote sites included in the pooling group. The underlying distribution of each catchment descriptor is shown in the Box-plots. Each Box-plot gives the minimum and the maximum value (+) and percentiles for the frequencies 0.05, 0.25, 0.5, 0.75, 0.95. The summary table lists record length, t_2, t_3 and d_{ij} values for a 100-yr pooling group for subject station 6031.
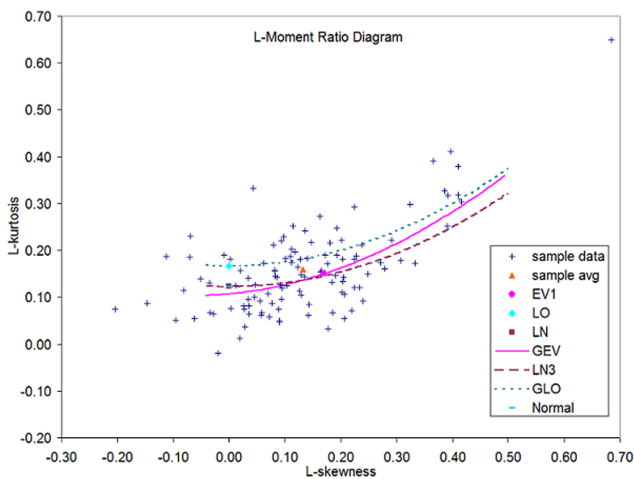

Fig. 10. L-moment ratio diagram for annual maximum flow for 110 Irish stations.

6   Conclusions

In the context of ROI pooling group based flood frequency estimation procedure, the most suitable form of distance measure d_{ij} for Irish conditions was sought. The ROI method with the suitably identified distance measure, Eq. (1), was used to form pooling groups for the subject sites. A simple graphical approach of examining homogeneity of the pooling groups was presented. The graphical approach compared the sampling variability of pooled estimates of L-CV with the L-CV of pooling group members. The approach also allowed the location of L-CV of the subject site to be viewed in the context of pooling group members, which is important in the

case of site specific pooling group. Most of the Irish pooling groups exhibited a degree of heterogeneity among the group members. A graphical approach of reviewing a heterogeneous pooling group was also presented in this context. The following conclusions were obtained from the above studies:

1. It was found that the distance measure d_{ij} could be satisfactorily defined in terms of lnAREA and lnSAAR but if there is a desire to incorporate another physical catchment effect then the BFI could be included with these two. The d_{ij} can also be defined in terms of lnMSL and ARTDRAIN.

2. A visual approach for the identification of the homogeneity of ROI pooling groups has been presented. The results are compared with the heterogeneity measures H1 and H2, obtained for those groups. Overall the results show that even with a carefully considered ROI procedure, such as using distance measure of Eq. (1), it is not certain that perfectly homogeneous pooling groups are identified. As a compromise it is recommended that a group containing more than 2 values of L-CV outside the 95% confidence limits of that variable, i.e. $m/N > 0.15$ should not be considered homogeneous.

3. A thorough investigation on 27 heterogeneous pooling groups has been carried out. In many cases, special attributes of the subject site such as extremely large or small values of AREA or of SAAR or of BFI or exceptionally low values of FARL contributed to the degree of observed heterogeneity of the pooling groups. It is deemed necessary that the subject site be positioned near the centre of the group of gauging sites, on the respective catchment descriptor axes, to which it is

hydrologically similar; but in some cases the fulfillment of that condition does not guarantee that the pooling group is homogeneous.

Appendix A

Heterogeneity test measures

The heterogeneity test measures proposed by Hosking and Wallis (1997) are based on (1) L-CV alone (the H1 statistic) and (2) L-CV and L-skewness jointly (the H2 statistic). These tests measures the sample variability of the L-moment ratios among the samples in the pooling group and compare it to the variation that would be expected in a homogeneous pooling group. The sample variability of the L-moment ratios is measured as the standard deviation of the at-site sample L-moment ratios weighted proportionally to the sites' respective record lengths. The measure of the sample variability based on L-CV alone, i.e. V_1, and L-CV and L-skewness jointly, i.e. V_2, are defined as

$$V_1 = \left[\sum_{i=1}^{M} n_i \left(t_2^i - t_2^R \right)^2 \middle/ \sum_{i=1}^{M} n_i \right]^{1/2} \quad \text{(A1)}$$

$$V_2 = \sum_{i=1}^{M} n_i \left[\left(t_2^i - t_2^R \right)^2 + \left(t_3^i - t_3^R \right)^2 \right]^{1/2} \middle/ \sum_{i=1}^{M} n_i \quad \text{(A2)}$$

where t_2^R and t_3^R are the group average of L-CV and L-skewness, respectively; t_2^i, t_3^i and n_i are the values of L-CV, L-skewness and the sample size for site i and M is the number of sites in the pooling group.

Simulation is used to establish what "would be expected" of a homogeneous group. Some 500 homogeneous groups are generated using a four-parameter kappa distribution with L-moment ratio values equal to t_2^R, t_3^R, t_4^R and the at-site mean, L1 = 1, in order to obtain the expected mean value, μ_{V_j}, and the standard deviation, σ_{V_j}, of the variability measures for a homogeneous group.

The heterogeneity measures H_j are then estimated using the expression below.

$$H_j = \frac{\left(V_j - \mu_{V_j} \right)}{\sigma_{V_j}}, \text{ for } j = 1, 2 \quad \text{(A3)}$$

Hosking and Wallis (1997) recommended using the H1 statistic over the H2 statistic as they found that the heterogeneity measure based on V_1 has better power to discriminate between homogeneous and heterogeneous regions. They suggested that a region is considered to be "acceptably homogeneous" if H1 < 1, "possibly heterogeneous" if 1 < H1 < 2, and "definitely heterogeneous" if H1 > 2. However, the H2 statistic as the heterogeneity measure was adopted by FEH (1999) for testing the homogeneity of

pooling groups as both the L-CV and L-skewness are required for fitting pooled growth curves with a Generalised Logistic (GLO) or a Generalised Extreme Value distribution (GEV). FEH (1999) revised the heterogeneity criteria based on the H2 statistics, suggesting that if 2 < H2 < 4, a region could be considered as heterogeneous whereas if H2 > 4 it could be considered as strongly heterogeneous.

References

Burn, D. H.: Evaluation of regional flood frequency-analysis with a region of influence approach, Water Resour. Res., 26, 2257–2265, 1990.

Burn, D. H.: Catchment similarity for regional flood frequency analysis using seasonality measures, J. Hydrol., 202, 212–230, doi:10.1016/S0022-1694(97)00068-1, 1997.

Castellarin, A., Burn, D. H., and Brath, A.: Assessing the effectiveness of hydrological similarity measures for flood frequency analysis, J. Hydrol., 241, 270–285, 2001.

Chowdhury, J. U., Stedinger, J. R., and Lu, L. H.: Goodness-of-fit tests for regional generalized extreme value flood distributions, Water Resour. Res., 27, 1765–1776, 1991.

Cunderlik, J. M. and Burn, D. H.: Switching the pooling similarity distances: Mahalanobis for Euclidean, Water Resour. Res., 42, W03409, doi:10.1029/2005WR004245, 2006.

Dalrymple, T.: Flood frequency analysis, Water Supply Paper 1543A, US Geol. Surv., Washington, p.51, 1960.

Das, S.: Examination of flood estimation techniques in the Irish context, Ph.D. thesis, Department of Engineering Hydrology, NUI Galway, 2010, available at: http://hdl.handle.net/10379/1688 (last access: 7 March 2011), 2011.

FEH: Flood Estimation Handbook, vol. 1–5, Institute of Hydrology, Wallingford, 1999.

Fill, H. D. and Stedinger, J. R.: Homogeneity tests based upon Gumbel distribution and a critical-appraisal of Dalrymple test, J. Hydrol., 166, 81–105, 1995.

FSR: Flood Studies Report, vol. 1, Nat. Environ. Res. Council, London, 1975.

FSU: Irish Flood Studies Update – Work Package 2.2, Office of Public Works, Dublin, 2009.

Gaál, L., Kyselý, J., and Szolgay, J.: Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia, Hydrol. Earth Syst. Sci., 12, 825–839, doi:10.5194/hess-12-825-2008, 2008.

Greenwood, J. A., Landwehr, J. M., Matalas, N. C., and Wallis, J. R.: Probability weighted moments – definition and relation to parameters of several distributions expressable in inverse form, Water Resour. Res., 15, 1049–1054, 1979.

Hosking, J.: L-moments: Analysis and estimation of distributions using linear combinations of order statistics, J. Roy. Stat. Soc., 52, 105–124, 1990.

Hosking, J. and Wallis, J. R.: Regional Frequency Analysis, Cambridge University Press, Cambridge, 1997.

Hosking, J. R. M. and Wallis, J. R.: Some statistics useful in regional frequency-analysis, Water Resour. Res., 29, 271–281, 1993.

Kjeldsen, T. R. and Jones, D. A.: A formal statistical model for pooled analysis of extreme floods, Hydrol. Res., 40, 465–480, doi:10.2166/nh.2009.055, 2009.

Lettenmaier, D. P., Wallis, J. R., and Wood, E. F.: Effect of regional heterogeneity on flood frequency estimation, Water Resour. Res., 23, 313–323, 1987.

Lu, L. H. and Stedinger, J. R.: Sampling variance of normalized GEV PWM quantile estimators and a regional homogeneity test, J. Hydrol., 138, 223–245, 1992.

Stedinger, J. R. and Lu, L. H.: Appraisal of regional and index flood quantile estimators, Stoch. Hydrol. Hydraul., 9, 49–75, 1995.

Viglione, A.: Confidence intervals for the coefficient of L-variation in hydrological applications, Hydrol. Earth Syst. Sci., 14, 2229–2242, doi:10.5194/hess-14-2229-2010, 2010.

Viglione, A., Laio, F., and Claps, P.: A comparison of homogeneity tests for regional frequency analysis, Water Resour. Res., 43, W03428, doi:10.1029/2006WR005095, 2007.

Wiltshire, S. E.: Identification of homogeneous regions for flood frequency-analysis, J. Hydrol., 84, 287–302, 1986.