Hydrol. Earth Syst. Sci. Discuss., 7, C2355–C2369, 2010
www.hydrol-earth-syst-sci-discuss.net/7/C2355/2010/

Hydrology and Earth System Sciences Discussions

Interactive comment on "Examination of homogeneity of selected Irish pooling groups" *by* S. Das and C. Cunnane

S. Das and C. Cunnane

s.das1@nuigalway.ie

Received and published: 17 September 2010

General:

The manuscript which has been prepared initially for the STAHY workshop in Toarmina, May 2010, subsequently submitted to the HESS for possible publication in the special issue "Advances in Statistical Hydrology" associated with that STAHY workshop.

In this paper pooling groups being examined are obtained by a previously adopted National flood estimation procedure for the Republic of Ireland. The emphasis is an examination of the ability of this procedure to identify groups which are homogeneous. In fact it was found that most groups have some degree of heterogeneity.

The comments by Referee 1 and Referee 2 are much appreciated and we have responded to each of them separately below and we have amended the manuscript in line with our responses. Not all of the amended or newly inserted sentences or paragraphs are included in the responses that follow. The revised manuscript is ready to be available to the referees should they so request.

Response to Referee 1:

"The paper presents a procedure to pool sites together for regional flood frequency analysis and to assess the homogeneity of the pooled groups. The only novelty in this paper is represented by the homogeneity criteria."

Response 1: The method, even though not new, had to be outlined in order to explain the context of the study, while the main focus was intended to be the study of homogeneity.

"It is a sort of graphical version of the Hosking and Wallis' H1 heterogeneity measure in which confidence bounds for the regional L-CV are defined and the number of local L-CVs lying outside these bounds is used as heterogeneity measure. I have concerns on how the confidence bounds are formed. It is assumed that the parent distribution of the variable of interest (annual flood peak) is a 2-parameter GEV distribution with the third parameter selected a-priori. This is a much more strong assumption than the kappa distribution used in the Hosking and Wallis heterogeneity measures and appears to me quite subjective."

Response 2: The Hosking-Wallis assumption of kappa has sometimes failed to produce (Hosking and Wallis, 1997, p. 67) results and they then assumed 3-parameter GLO. To some extent this could also be deemed subjective e.g. why not GEV or LN3?

"Also, the authors suggest to consider as heterogeneous the groups with more than two values of sample L-CV falling outside the confidence limits, without taking into account the number of sites forming the groups."

Response 3: Note that all groups are roughly of the same size e.g. 11-17 sites per group, and the number m is almost as informative as m/N, but use of m/N is included in the revised draft.

"I would have reasoned in terms of percentages of values of local L-CV falling outside the 95% confidence limits and I would have expected it to be around 5% for homogeneous regions, in the ideal case of correct evaluation of the confidence bounds."

Response 4: We agree. Given the group sizes of 11-17, the number equal to 5% of stations, has between 0.55 and 0.85. These are not measurable by counting in the present context where only integer values can be observed.

"Other points as the formation of the pooling groups (Section 2) and the investigation of the causes of heterogeneity (Section 4) do not differ from what suggested in other works as the FEH (1999) or Castellarin et. al (2001), among those cited by the authors."

Response 5: Yes, we agree. These has been addressed, see below.

"In general, I did not like the presentation of the methods, which is sometimes confusing and incomplete (see comments below). Even if the work has its own value as an illustration of the methods used for flood frequency regionalisation in Ireland, I deem it is not appropriate for the special issue "Advances in statistical hydrology" because of the lack of novelty. "

Response 6: However the graphical presentation of results, as in Fig. 3-9, has not been presented previously and in our view presents some concrete insight into aspects of homogeneity that are encapsulated within more complex formulae. For instance Fig.7 shows that small dij values do not guarantee homogeneity which is a prime assumption in many ROI schemes.

"More detailed comments follow. Suggestion: the abstract, specially the first part of it, looks like an introduction and could be shortened. I would avoid the citations in the abstract, if possible."

Response 7: The first 11 lines of the abstract have been removed

"Page 5102, lines 9-10: I would not say that "heterogeneity among the sites in a group is mainly due to variations in the sample L-CVs" "

Response 8: This is in accordance with Viglione et al. (2007) and is consistent with Hosking-Wallis preference for V1.

"but that "heterogeneity among the sites in a group can be detected looking at the variability of the sample L-CVs" or something like that. Page 5102, lines 23-24: I don't understand the sentence on the population distribution. Does it mean that the expected variability of t2 is not modelled assuming a kappa distribution? Please be clearer here."

Response 9: GEV, not Kappa, was used and the selection of GEV was based on the summary statistics of the region.

"Page 5104, line 10: what's Qmed? I guess it is the index flood (the median)."

Response 10: Correct but see response 12 where we state that this sentence is replaced by another

"If so, why should the same variables be used to identify the ROI groups that relate to the growth curve? Page 5104, eq. 2: is this the equation"

Response 11: Equation (2) is an equation for dij, not for Qmed (some misunderstanding here?).

"proposed for Qmed by the FSU (2009)? If yes, deem it is not necessary to state it here. If not, please clarify how is it obtained."

Response 12: The sentence "These variables have been found to be effective in explaining the observed variation in Qmed values in the FSU (2009) regression studies" has been replaced by "Such variables have been found to be effective in the ROI context, in , for example, FEH(1999)".

"Page 5104, lines 14-15: What does "the information about the flood distribution contained in the pooling groups" mean?"

Response 13: This refers to the information content of the raw data about the underlying distribution

"Page 5105, lines 14-15: Eq. 3 does not return t2 and t3 as output. The procedure of estimating the local "population" values of L-CV and L-CA (pages 5104-5105) is not properly explained."

Response 14: The local “population” values of L-CV and L-skewness is estimated using a ROI scheme. A pooling group is formed for each site using eq (3) and the pooled L-CV and L-skewness are estimated using eq (B1). The pooled estimates of L-CV and L-skewness are then used as local population values for each site. In the simulation procedure at step 2 random samples are drawn using these local population values. This has been done to avoid a simulated region that has much more heterogeneity than the actual data. Sentences to that effect have been included.

"Maybe it would help to add the equations used to estimate those values (eq. 3 just refers to the ROI pooling procedure). "

Response 15: Reference to eq 3 has been removed and replaced by “as in Castellarin et al. (2001) and Gaal et al. (2008)”.

"Which weights are used to obtain the average t2 and t3? Are they related to sample length? "

Response 16: In fact weights were not used and the word “weighted” has been replaced by “unweighted”

"Page 5106-5107, lines 17-22/1-4: looking at table 2, the natural choice would be lnAREA alone, which gives the same RMSE (if not better) than lnAREA+lnSAAR or lnAREA+lnSAAR+BFI. It is said that "there is an intuitive attraction in also representing some descriptor of catchment response even at the cost of a small apparent loss in

effectiveness". What is this intuitive attraction?"

Response 17: The following has been added “This could be of relevance in engineering investigations where differences in catchment behaviour are considered of importance by the investigator” in line 4 before the sentence beginning “An extension. . ..”

"Please extend this point, which is relevant. It represents an expert-judgement decision on top of the statistical method, which is indeed always required in practical applications. Page 5107, lines 5-10: the trial and error procedure to find the weights for equation (1) is of interest. Maybe some more space should be devoted to describe it, even qualitatively."

Response 18: The following has been added “An automatic search procedure was not used” in line 7 before the sentence beginning “It was found. . ..and BFI.” A new sentence is added, as follows in line 10 at the end of the paragraph “The trial and error approach involved assuming weights varying from a selection of 0 to 3 to each of quantities in eq (2)”.

"Pages 5107-5108, point 3: the confidence intervals for t2R are drawn from a Gumbel distribution (or a GEV, but with what selected shape parameter?)."

Response 19: The sentence “Random. . .for t2R” has been replaced by “Random samples are drawn from GEV distributions with 3 different shape parameter values (k = -0.05, k = 0 (EV1), k = +0.03) using the t2R as the population value to construct a 95% confidence interval for t2R”

"Why is it necessary to assume that the sample t2 are normally distributed? Couldn’t the empirical distribution obtained from the simulations be used instead?"

Response 20: It could have been but experience of L-CV sampling distributions obtained by simulation, is that they are reasonably symmetrical.

"More importantly, at page 5102, line 11, it was stressed that one assumption of the Hosking and Wallis heterogeneity measures was the parent kappa distribution. Here

2 parameter distributions are used instead, which is a much more strict assumption. Please comment on that."

Response 21: Ireland is a low-CV region and previous studies FSR (1975) and FSU (2009) have found that either EV1 or a small deviation from it in the form of GEV gives a good representation of Irish annual maximum floods. This is why we feel that it is not necessary to use the more general Kappa distribution.

"Page 5108, point 4: is the number of t2 outside the confidence bounds (m) a proper measure? It depends on the number of sites forming the group (say N). In general, one would expect 5% of the values to be outside the 95% confidence bounds, i.e., m/N=0.05. I would rather use m/N as a measure for the homogeneity criteria."

Response 22: m/N is included in to the revised manuscript as well as m.

"Page 5108, line 14: these k are those used to produce Table 3, while for figure 3 just k=0 is displayed, isn't it?"

Response 23: That is correct. This is already included in the legend, but will also now be included in the Figure caption.

"Page 5108, point 2: what does this point 2 imply? If a left skewed parent distribution is assumed, then the groups appear more homogeneous?"

Response 24: No, it is the other way around. The left skewed distribution is the k = +0.03 case for which the number of cases of m> 3 is 47.

"Page 5109, point 3: are the 27 sites named here also listed in table 3?"

Response 25: The 27 sites are from among the 85 considered in Table 3.

"Page 5109, point 5: here the "range of t2" is analysed in relation to H2. Does this range correspond to the 95% confidence bounds? Please define it more clearly."

Response 26: Range means “maximum t2 - minimum t2” and does not refer to any

confidence bounds. The sentence “The range of t2 values, max t2 – min t2, was calculated for the 85 pooling groups” has been included.

Response to Referee 2:

"General comments: This manuscript presents work undertaken to develop a general procedure for regional flood frequency estimation in Ireland with particular emphasis on the formation of homogeneous pooling groups. The work is heavily inspired by the methodology presented in the Flood Estimation Handbook (FEH) published by the Institute of Hydrology (1999). However, the researchers seems to have overlooked more recent developments of the FEH methodology as presented by Kjeldsen and Jones (2009) in which most aspects of the FEH pooling group methodology were revised. A key aspect of the revised FEH methodology is that the need for defining homogeneous pooling groups have been made redundant through incorporation of the between-site variation in the L-moment ratio directly into the underlying statistical model of the pooling group. In general I think the manuscript does not read well – a bit like extracts from a larger report resulting in terms not appropriately defined to the extent necessary for a standalone manuscript. In many places the authors could be more helpful to the reader and provide more guidance and background to the analysis."

Response 27: The manuscript has been prepared, under a severe limitation on length, for the STAHY workshop in Toarmina, May 2010. Hence background explanation was kept to a minimum. We are not made aware that this limitation had been relaxed for the Special Issues.

"Also, the introduction is a little light on references and acknowledgement of previous work into regional frequency analysis. A simple literature search should reveal many relevant previously published papers on this topic."

Response 28: It is never our intention to deny proper recognition for any previous published work but merely to include those which we explicitly used or quoted from. We are not unaware of the vast array of relevant literature in existence. After all this is

not a review paper. Note: The 2009 Kjeldsen and Jones paper includes 17 references, including 4 self references as compared to our 16.

"My own personal misgivings about the pooling methodology (in general) are that there are too many free parameters that must be fixed which make it difficult to ensure that an optimal procedure has been devised. For example, in this study the following choices has been made + Size of pooling group: 5T rule (adopted directly from FEH – no questions asked) + Weight of each site within pooling group: record-length weighted + Formation of pooling group: Distance measure based on lnAREA, lnSAAR and BFI-HOST (similar to FEH), with a set of weight assigned to each catchment descriptor based on trial-and-error."

Response 29: An additional paragraph has been included in Manuscript in which some of the issues raised are discussed. The question of weightings is newly discussed in another section.

"It is worth noticing here that Kjeldsen and Jones (2009) revised all the above assumptions in the FEH and as a result developed a more efficient method. However, despite all that effort it turned out that a simple weighted least square linear regression model linking the at-site L-moment ratios and catchment descriptors produced a method for predicting L-moment ratios at ungauged sites which performed about as well as the more elaborate pooling procedure. I therefore suggest that the authors test how well a set of linear regression models compare to the method presented in this paper."

Response 30: This paper, as the title implies, is primarily concerned about the homogeneity of pooling groups involved in a previously selected procedure rather than justifying the previously selected National procedure or developing a new one.

"Perhaps the authors could also compare their new method to the existing methods, for example the growth curves for Ireland published in the Flood Studies Report (FSR) published by NERC (1975)."

Response 31: Note: There was only one curve for Ireland in NERC (1975).

"In Section 2.1 a Monte Carlo simulation experiment is design to assist in the evaluation of the predictive power of different pooling group formation methods. But I am not sure why a simulation experiment is required as, to my mind, it seems an unnecessary complication. Why not simply report the RMSE and the BIAS based on the (squared) difference between at-site and pooled estimates, where the pooled estimates have been obtained as-if the site is ungauged?"

Response 32: What the Referee is saying is why not use PUM? We decided to do otherwise because we had far fewer stations (85) than the 602 stations available for the UK study.

"In my opinion the results of such an experiment would be much more transparent than the MC experiment described here. Also, there are some issues related to choice of at-site population values and the presence of intersite correlation between AM series that will have an influence on the simulated, and these effects are not discussed or even recognised."

Response 33: We did not mention inter-site correlation because it was found by Hosking and Wallis (1993, 1997, p- 127) to be of very little consequences. A sentence to that effect has been included.

"Consequently, I would suggest dropping the MC experiment in favour of a conceptually simpler cross-validation (leave-one-out) exercise. It is not obvious why the authors want BFI to be included in the distance measure when they could not detect any improvements in predictive ability when including BFI. The only reason seems to be a belief that BFI should have some explanatory power. Consequently, as compromise solution has been found where BFI is included but not given much weight."

Response 34: As noted in response to a related query by Referee 1 (Response 17), we expressed the view that this is an intuitive attraction to inclusion of BFI where we

stated “This could be of relevance in engineering investigations where differences in catchment behaviour are considered of importance by the investigator.”

"Note here that the revised FEH procedure (Kjeldsen and Jones, 2009)"

Response 35: Does reference to the “revised FEH procedure” mean that the FEH methodology has been revoked? Inclusion of FARL and FPEXT is based on a regression where R2 is only 19% - not a very convincing basis for their inclusion.

"came to the same conclusion that BFI does not add to the description of the between catchment variation in the high order L-moment ratios and, thus, was omitted from the revised similarity distance measure and replaced by FARL (lake and reservoir attenuation) and an index of upstream flood plain extent. Section 3 describes another Monte Carlo simulation experiment – this time to examine the degree of homogeneity of the pooling groups. However, here I think the authors need to be much more helpful to the readers and start by telling them what the purpose of this investigation is rather than go straight to a point-by-point description of a simulation procedure and then leave it to an interested reader to try and understand the thinking behind this exercise."

Response 36: As suggested additional introductory material has been included

"I am not confident that the test procedure described in Section 3 is entirely valid. Would it not be better to estimate the pooled L-CV as the record-length weighted average as this is a more efficient estimate of the mean than the unweighted average used in the paper? Then the standard deviation of t2R can be obtained by simulating from an assumed homogeneous pooling group (all sites have values of t2= t2R but different record length)."

Response 37: In deciding to use unweighted averages we were 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 tR. 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. For this reason, Jin and Stedinger (1989) preferred to use an unweighted average’.

"Alternatively, simply derive the analytical standard deviation of a weighted mean and assume it is normally distributed."

Response 38: Surely, the individual components of this expression would have to be obtained by simulation?

"Section 4 discusses reasons why individual pooling group might turn out to be heterogeneous and an exploratory plot is shown to illustrate how catchment descriptors vary within a particular pooling group. However, there is no real substance to this section. I think the authors should be more ambitious and include all their pooling groups into an investigation of reasons for heterogeneity. Perhaps they could define a measure of spread of catchment descriptor values within a pooling group and plot this against catchment descriptor values of subject sites (e.g. catchment area). This would enable a more systematic review of the impact of the subject site characteristics, and could potentially be used to inform a better definition of the similarity distance measure."

Response 39: In section 4 we report an examination of the 27 cases in which the subject site L-CV fell outside the confidence limits in the EV1 case. We did not envisage the type of measure suggested and we are not certain that such an exercise would prove fruitful.

"Minor comments: Abstract: the first 10-11 lines do not belong in an abstract"

Response 40: The first 11 lines of the abstract have been removed

"Page 5102, 1st paragraph: Regionalisation is not necessarily equal to pooling anal-

ysis. There are several methods for regional frequency analysis of which the pooling method is but one."

Response 41: The sentence has been amended to “Regionalization (FSR, 1975), i.e. pooling analysis (FEH, 1999), is one of the possible methods used”

"Page 5102, line 14: the terms XT, QT and T needs better defining. These are key concepts for the study and the reader should be in no doubt what they represent."

Response 42: A new section is added entitled “Estimation of the pooled growth curve” after section “Formation of pooling groups using ROI approach” where these terminology have been defined in detail. Note: The contents of Appendix B entitled “Estimation of growth factor” have been included in this section and Appendix B has been removed in the revised manuscript.

"Page 5102, line 17-18: Kjeldsen and Jones (2009) presented a version of the pooling group method which does not require homogeneous pooling groups to be formed. "

Response 43: A sentence acknowledging this has been included.

"Page 5104, line 3: (PUM) which is the squared difference between. . ."

Response 44: The word “squared” has been included

"Page 5104, line 5-7: why select only 4 catchment descriptors?"

Response 45: 4 were selected on intuitively and on previous experience, that AREA and SAAR are widely known to be effective and it was intuitively felt that the other two BFI and FARL would complement them sufficiently and parsimoniously.

"Consider defining useful descriptors by defining relevant catchment descriptors through a linear regression model between high order L-moment ratios and catchment descriptors."

Response 46: The suggested methodology is interesting but as shown by Kjeldsen

and Jones (2009) the corresponding R2 values are very low.

"Page 5104, line 9-11: The reader has to know the FEH-terminology (QMED) for this sentence to make any sense."

Response 47: We agree, but the sentence containing Qmed has replaced by an alternative, in response to a comment- by Referee 1 (see Response 12).

"Page 5104, Eq. (2): no explicit weights, Wk, are included in Eq. (2) Page 5104, Eq. (3): perhaps replace d_ij with something else (t_ij?) to distinguish from similarity measure in Eq. (2)."

Response 48: We agree. δij is used instead of dij in eq (3). Note tij could be confused with L-moment ratios.

"Page 5110, line 14: What does ‘special qualities’ refer to?"

Response 49: This is from FEH as stated and refers to any anomalies or peculiarities that the analyst might have identified.

References:

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

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

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

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

Gaal, L., Kysely, J., and Szolgay, J.: Region-of-influence approach to a frequency analysis of heavy precipitation in Slovakia, Hydrology and Earth System Sciences, 12, 825–839, 2008.

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 Resources Research, 29, 271–281, 1993.

Kjeldsen, T. R. and Jones, D. A. (2009) A formal statistical model for pooled analysis of extreme floods. Hydrology Research, 40(5), 465-480, doi: 10.2166/nh.2009.055.

Viglione, A., Laio, F., and Claps, P.: A comparison of homogeneity tests for regional frequency analysis, Water Resources Research, 43, 2007.

Interactive comment on Hydrol. Earth Syst. Sci. Discuss., 7, 5099, 2010.