A formal statistical model for pooled analysis of extreme floods

Thomas R. Kjeldsen and David A. Jones

ABSTRACT

This paper describes a formal statistical model underlying the region-of-influence method routinely used in regional frequency analysis of hydrological extremes, and is an improvement to the existing Flood Estimation Handbook (FEH) method for pooled frequency analysis of annual maximum flows in the UK. Specification of a pooling-group method requires three issues to be resolved: how to define hydrological similarity, the size of pooling groups and calculation of the pooled L-moment ratios. Because these issues are interrelated, an exploratory and iterative study has been undertaken before arriving at the final version of the method. Improvements provided by the model are: (1) that it allows an increased weight to be given to a gauged catchment when it is itself the target location and (2) it does not require identification of a homogeneous region, since the expected differences between the L-moment ratios within a pooling group are explicitly accounted for. Using annual maximum series from 602 gauged rural catchments, a comparison of candidate methods shows that the new method performs better than these others, including the FEH method. While the numerical comparison suggests that the improvement is 4%, and thus only minor, arguments are given for why this is a misleading conclusion.

Key words | annual maximum series, L-moment ratios, pooled frequency analysis, pooling groups

Thomas R. Kjeldsen (corresponding author)
David A. Jones
Centre for Ecology and Hydrology,
Maclean Building,
Crowmarsh Gifford,
Wallingford OX10 8BB,
UK
E-mail: *trkj@ceh.ac.uk*

INTRODUCTION

Flood frequency analysis in the UK is often based on statistical analysis of annual maximum peak flow data using the pooling group (or regional) procedure outlined in the Flood Estimation Handbook (FEH) published by the Institute of Hydrology (1999). The FEH pooling procedure was designed to allow users a high degree of flexibility in the analysis as well as enabling the use of data from recent flood events. The procedure is routinely used for a wide range of practical applications including flood mapping studies, flood risk assessments and the design of flood alleviation schemes.

The FEH pooling-group method is a hybrid of the index flood method (Stedinger *et al.* 1993) combined with the region of influence (ROI) approach for formation of pooling groups suggested by Burn (1990) on the basis of work by Acreman & Wiltshire (1987, 1989). Considering a particular catchment of interest, be it gauged or ungauged, an estimate of a pooled growth curve is obtained through data transfer from a group of ‘hydrologically similar’ catchments, where similarity is quantified based on a set of physical catchment descriptors. The foundation of the index flood method is that the underlying true distribution of the annual maximum peak flows at the different catchments in a pooling group are identical except for a scaling parameter denoted the index flood. The quantile function standardized by the index flood is called the growth curve.

Because pooling groups are formed based on hydrological similarity, it is implicitly assumed that the catchment descriptors used in the similarity measure can adequately explain the variability of the growth curves between the catchments, where the growth curve is determined by the high-order L-moment ratios, L-CV and L-SKEW.

doi: 10.2166/nh.2009.055

(The usefulness of these L-moment ratios is advocated by Hosking & Wallis 1997.) By subsequently assigning weights to the catchments within the pooling group based on their rank (or catchment similarity), the existing FEH methodology acknowledges that the catchments are in fact not similar. Hence the method departs from the original assumptions made in the index flood method. While these considerations might not affect the practical use of the method, they are important when developing a formal statistical framework necessary for optimizing the performance of the pooling-group method. The development of the method needs to consider the following aspects: (i) formation of pooling groups; (ii) weights within pooling groups; (iii) size of pooling group; and (iv) performance of method. These four aspects of the method are highly interdependent and ideally should be considered simultaneously, but in practice it has been necessary to adopt a more sequential approach where each of the first three aspects were analyzed in turn and once a decision made, it was carried over to analyze the next aspect.

While the ROI concept has generally been embraced in the scientific literature as a convenient method for conducting regional frequency analysis of both extreme high and low events, for example, Castellarin *et al.* (2001) and Holmes *et al.* (2002) only a few examples of general application to a larger region has been reported. More recently, Merz & Blöschl (2005) suggested that spatial proximity is a better indicator of hydrological similarity than a non-geographic distance measure based purely on catchment descriptors. A further obvious indicator of similarity is the location of gauging stations on the same part of the river network as the target site such as suggested by Gottschalk (1993) and Skøien *et al.* (2006).

This study presents the development of a procedure for using pooling groups in flood frequency analysis in the UK. Underlying the procedure is a formal statistical model describing the relationship between the statistical properties of a pooling group and the individual pooling-group members, expressed in terms of the L-moment ratios, L-CV and L-SKEW. However, to make the procedure operational it was necessary to undertake an extensive analysis of the datasets. While the reporting in the paper might give the impression of a straightforward process, the analysis was in practice very exploratory and each aspect was investigated many times based on feedback from subsequent analysis. Note that the exploratory nature of this paper necessitates the introduction of many different symbols, and a list of these is provided in Appendix A.

POOLED FREQUENCY ANALYSIS

The FEH recommends the three-parameter Generalized Logistic (GLO) distribution for flood frequency analysis in the UK. The inverse of the cumulative distribution function (cdf) for estimating the T-year event Q_T is given as

$$Q_T = \xi + \frac{\alpha}{\kappa}(1-(T-1)^{-\kappa}) = \xi\left[1+\frac{\beta}{\kappa}(1-(T-1)^{-\kappa})\right] = \xi z_T \quad (1)$$

where ξ, α, κ and $\beta = \alpha/\xi$ are GLO model parameters, T is the return period and z_T is the growth curve at T defined by the term within the square brackets in Equation (1). The parameter estimation method used in this study is that adopted by the FEH (Institute of Hydrology 1999) and is a variant of the method of L-moments described by Hosking & Wallis (1997). Given a flow series from a particular gauging station with n annual maximum series peak flow values $[q_1, \ldots, q_n]$, the location parameter ξ is estimated by equating the distribution median to the sample median:

$$\hat{\xi} = \text{median}(x_1, \ldots, x_n). \quad (2)$$

The shape parameter κ and the scale parameter β are estimated as

$$\hat{\kappa} = -t_3$$
$$\hat{\beta} = \frac{t_2\hat{\kappa}\sin(\pi\hat{\kappa})}{\hat{\kappa}\pi\sin(\hat{\kappa}+t_2) - t_2\sin(\pi\hat{\kappa})} \quad (3)$$

where t_2 and t_3 are the sample L-moment ratios L-CV and L-SKEW, respectively.

When extending the at-site analysis to a pooled frequency analysis, the FEH uses the median based index flood method, i.e. the T-year event is estimated as

$$Q_T^{(p)} = \hat{\xi} z_T^{(p)} \quad (4)$$

which has a similar structure to the at-site case in Equation (1), but the superscript p indicates that the factors are obtained from a pooled analysis. The growth curve $z_T^{(p)}$ is estimated using information from M sites deemed sufficiently hydrologically similar to the catchment of interest. The pooled growth curve is estimated by substituting the pooled L-moment ratios $t_r^{(p)}$ into Equation (3) and the pooled L-moment ratios themselves are calculated as the weighted average of the individual at-site L-moment ratios within the pooling group. Thus, for a pooling group consisting of M catchments, the pooled L-moment ratios are given as

$$t_r^{(p)} = \sum_{i=1}^{M} w_{r,i} t_{r,i} \quad (5)$$

where $w_{r,i}$ are the weights assigned to the ith catchment for L-CV ($r = 2$) or L-SKEW ($r = 3$). The FEH recommended a set of weights based on rank within the pooling group (when the catchments are ordered either in terms of catchment similarity or by user-preference) and record length. In the following, a new set of improved weights will be derived based on a statistical model of the underlying statistical structure of the pooling groups.

OTHER POTENTIAL SOURCES OF INFORMATION

For practical flood frequency estimation, the FEH (Institute of Hydrology 1999) argues that estimates of design floods in a particular catchment should be supplemented with additional information where available, e.g. in the form of evidence of pre-gauging events (historical events) or transfer of data from nearby gauged catchments. The model presented here does not allow for the inclusion of historical flood data. While this would undoubtedly be a desirable extension, no systematic and quality controlled database of historical events currently exists in the UK. The reader is referred to Hosking & Wallis (1997) for a more comprehensive discussion of the use of historical information in regional frequency analysis.

METHODS

General approach

This study takes a largely empirical approach to the question of formulating a pooling-group methodology. Many of the details of the pooling-group scheme are decided on the basis of comparisons of alternatives by testing these out on a large dataset, which is described in the following section. The comparisons are made using a particular measure of the performance of the given scheme when applied to multiple cases within the dataset: the measure of performance is described in the section after next.

Data

The dataset used in this study comprises annual maximum peak flow series from 602 gauged catchments located throughout the UK, as well as values of four catchment descriptors used for defining catchment similarity. An initial screening of the catchments was undertaken to ensure both that the stage-discharge ratings were adequate and that the catchments are effectively non-urban.

A summary of the dataset is shown in Table 1, where AREA is the catchment area in km^2, SAAR is the standard average annual rainfall (in mm) based on measurements from 1961–1990 and FARL is an index of flood attenuation due to reservoirs and lakes and varies between 0 and 1 (= no attenuation). The variable FPEXT is related to the extent of floodplains in the upstream catchment defined by the 100-year flood level adopted from an existing national floodplain map (Morris & Flavin 1990), and can take on values between 0 (= no flood plain) and 1 (= flood plain extends to entire catchment). More details of the FPEXT descriptor are given by Kjeldsen *et al.* (2008). The other three catchment descriptors are defined in the Flood Estimation Handbook (Institute of Hydrology 1999) through which they are available in electronic form for all

Table 1 | Summary of data from 602 rural catchments located throughout the UK

	Min	Mean	Max
Median annual maximum flow (m^3 s^{-1})	0.2	92.7	981.4
Record length, number of years	4	33	117
AREA (km^2)	1.6	335.1	4587.0
SAAR (mm)	558	1162	2848
FARL (−)	0.645	0.970	1.000
FPEXT (−)	0.0023	0.0611	0.2498

catchments in the UK larger than $0.5\,km^2$. Note that in the exploratory analysis, 1 – FPEXT rather than FPEXT was used to ensure compatibility with FARL, for which a value of 1 indicates no attenuation from upstream features.

The particular choice of catchment descriptors in Table 1 and their transformations used here was based on other analyses, described later.

Performance measure

The performance of the alternative pooling procedures in this study is based on a pooled uncertainty measure (PUM) defined as

$$\mathrm{PUM}_T = \left(\frac{\sum_{i=1}^{M} h_i\left(\ln z_{T,i} - \ln z_{T,i}^{(p)}\right)^2}{\sum_{i=1}^{M} h_i}\right)^{1/2} \tag{6}$$

where $z_{T,i}$ is the at-site T-year growth factor for the ith catchment (obtained from a GLO distribution fitted by L-moments), the superscript p indicates a growth factor derived from the pooling-group method and h_i is the weight assigned to the ith catchment. The rationale behind the PUM measure is that a good pooling method will, on average, produce growth curves that are close to the true growth curves for the site of interest.

The weight assigned to each catchment is defined as

$$h_i = \frac{n_i}{1 + n_i/16}, \tag{7}$$

where n_i is the record length in years, thereby reducing the importance assigned to individual catchments with long records compared to a set of weights $h_i = n_i$ more directly linked to record length. The weights in Equation (7) were selected on the basis of simulation experiments which compared the between-site variation of the at-site estimates of the growth curves with the sampling errors of these estimates for different record lengths. Statistical theory not discussed here indicates how these results can be used to define optimal weights. However, these weights would vary with return period, and the divisor 16 was found to be a compromise suitable for all return periods.

MEASURE OF CATCHMENT SIMILARITY

Similarity distance measure

This study has adopted the ROI approach for creating pooling groups tailored to each specific site of interest. By considering catchments which are similar to the site of interest (gauged or ungauged) with regards to a chosen set of catchment descriptors, it is assumed that these catchments are also 'hydrologically similar'. The expression 'hydrologically similar' means that a particular site does not violate the fundamental assumption of the index flood method, i.e. that the annual maximum flood series is generated from an underlying flood distribution with similar high-order moments (L-CV and L-SKEW). The P catchment descriptors defining the *similarity distance measure* (SDM) are called the pooling variables and the SDM itself is defined as

$$s_{ij} = \sqrt{\sum_{k=1}^{P} \omega_k \left(\frac{x_{i,k} - x_{j,k}}{\sigma_k}\right)^2} \tag{8}$$

where $x_{i,k}$ is the kth pooling variable at the ith catchment, ω_k is the weight assigned to the kth pooling variable and σ_k is the standard deviation between the kth catchment descriptor from all gauged sites.

By using an approach similar to the FEH (Institute of Hydrology 1999) and Merz & Blöschl (2005), a suitable set of catchment descriptors were identified through the use of regression models linking L-CV and L-SKEW to a set of catchment descriptors. An initial comprehensive search was conducted, which identified the three best combinations of catchment descriptors for each case of using from 1 to 10 descriptors in a regression model. The purpose here was to obtain a general idea of which descriptors were most important. Both log-transformed and non-transformed versions of the catchment descriptors were included in the search. The sample L-moment ratios themselves were not log-transformed in this experiment. A similar investigation was conducted as part of the FEH development (Institute of Hydrology 1999) where it was found that 37% and 8% of the between catchment variation of L-CV and L-SKEW, respectively, could be explained using a regression model. The FEH based the formation of pooling groups on three catchment descriptors–ln(AREA), ln(SAAR) and BFIHOST–where BFIHOST is the baseflow index (BFI) as derived from HOST soil data. The FEH noted that the models using a larger number of descriptors performed only marginally better.

In the present analysis, the best regression model for L-CV containing four catchment descriptors was found to include ln(AREA), ln(SAAR), FARL and 1-FPEXT. This particular model has an R^2 value of 44% which is comparable to the value found in the FEH. For L-SKEW, the optimal four-variable regression model shares three catchment descriptors with the L-CV model–ln(AREA), ln(SAAR) and 1-FPEXT–but used AREA (untransformed) rather than FARL. The R^2 value for the L-SKEW model is rather low ($R^2 = 19\%$), which again is comparable to the result reported in the FEH. As a result, the four parameters used in the L-CV model were adopted for use as pooling variables in the distance measure. The order of the four pooling variables represents the ranking of their significance in the regression models describing the L-moment ratios.

Assigning weights in the distance measure

The weight assigned to each of the four catchment descriptors–ln(AREA), ln(SAAR), FARL and FPEXT–for defining SDM in Equation (8) was investigated through an empirical procedure based on PUM values. At this stage in the procedure, the set of weights used within each pooling group to calculate the pooled L-moment ratios has not yet been defined. Instead, the L-moment ratio (both L-CV and L-SKEW) for each catchment is weighted according to its record length using weights given by Equation (7).

The first step in the procedure is to set all four weights ω_k in Equation (8) to unity. Next, the weight assigned to the first pooling variable was varied between 0 and 10 in steps of 0.25. For each combination of weights, pooling groups containing 17 catchments were formed for each of the 602 catchments and the resulting PUM calculated. Rescaling of each of the trial set of weights was undertaken to ensure that the weights in the SDM sum to 4 (the same as unity weights). Pooling groups of 17 catchments were used as a reasonable first guess of pooling-group size at this stage in the procedure. In the final procedure, the pooling-group size is defined as 500 annual maximum events: this, with an average record-length of 35 years, is similar to the size used in these initial investigations. Note that the methods were found to be relatively insensitive to the size of the poling group if a sufficiently large number of sites were included.

Having identified an optimal, or near-optimal, value of the weight of ln(AREA), the weight of the second catchment descriptor ln(SAAR) was allowed to vary between 0 and 10, with unity weight on FARL and FPEXT but with the weight on ln(AREA) changed from unity to the optimal weight identified in the previous step. As before, the best weight for ln(SAAR) is noted and the procedure moves on to FARL and finally FPEXT.

The results obtained from the procedure are shown in Figure 1. For each pooling variable, the final weight is set to the value that results in the minimum PUM value. It was found that only one run through the procedure was necessary to obtain a stable set of weights.

The final unscaled weights are as follows: ln(AREA) (7.0), ln(SAAR) (1.25), FARL (0.25) and FPEXT (0.50). These weights were subsequently scaled to ensure that they sum to 4 and the final SDM is:

In Figure 1, it can be observed that leaving out any of the pooling variables (i.e. when the value of the weight is set to 0) results in a substantially higher value of PUM, except for FARL. This suggests that FARL is the least important catchment descriptor in the distance measure. Thus, the

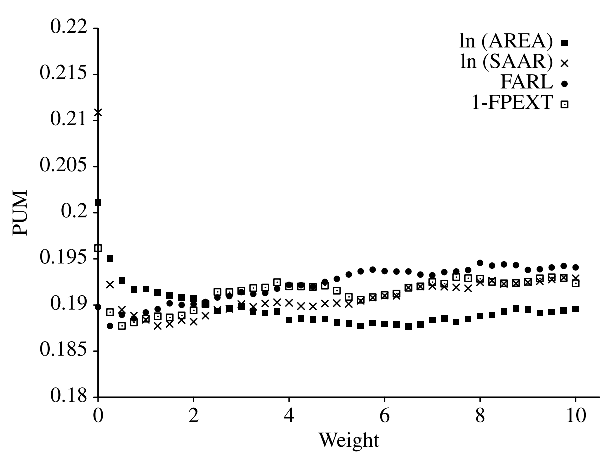

Figure 1 | PUM values for different combinations of weights in the SDM measure as determined by the empirical procedure. Note that the weights on the x-axis represent the unscaled weights.

main difference between the SDM defined in Equation (9) and the corresponding version published in the FEH is the finding that soil type has less influence on the shape of the growth curve than previously assumed. The introduction of FPEXT is believed to be a reasonable extension to the method as anecdotal evidence (supported by the regression model for L-SKEW discussed above) often suggests that the attenuating effects of upstream flood plains have a subduing effect on growth curves.

$$s_{ij} = \sqrt{3.2\left(\frac{\ln \mathrm{AREA}_i - \ln \mathrm{AREA}_j}{1.28}\right)^2 + 0.5\left(\frac{\ln \mathrm{SAAR}_i - \ln \mathrm{SAAR}_j}{0.37}\right)^2 + 0.1\left(\frac{\mathrm{FARL}_i - \mathrm{FARL}_j}{0.05}\right)^2 + 0.2\left(\frac{\mathrm{FPEXT}_i - \mathrm{FPEXT}_j}{0.04}\right)^2} \qquad (9)$$

FORMATION OF POOLING GROUPS

Weights in pooling groups

As outlined earlier, a method which assigned weights to individual members of a pooling group based on record length and rank had been developed previously as part of the FEH. This subsequently came under some criticism on the following points: (i) it depends on the rank in order of similarity to the target site within the pooling group; (ii) weights do not diminish gradually to 0 by the point at which they leave the pooling group; and (iii) the particular use of record length in the weighting factor results in undesirable high weights being assigned to low ranking very dissimilar sites with relatively long records (Morris 2003). These issues have been considered when developing a revised weighting scheme. In addition, the revised weighting scheme presented here makes a distinction between pooling groups for gauged and ungauged sites. Where the existing FEH weighting scheme only accommodates any at-site data that might be available by assigning the first rank position to the site of interest, the weights in the new methodology will more directly assign higher importance to at-site data than to data from other catchments in the pooling group. Finally, new sets of weights will be developed specifically for L-CV and L-SKEW separately, though only one pooling group will be created for each site of interest.

The new weighting scheme is developed on the basis that the weight assigned to each individual catchment in a pooling group should reflect the following ideas: (i) that there are differences in the true (but unknown) values of L-moment ratios between catchments; (ii) some of the variation between values is due to sampling error arising from individual L-moment ratios being estimated from a limited record length; and (iii) weights should reflect the distance in catchment descriptor space between the target site and sites in the pooling group. In the following section, a statistical model will be developed which has a structure for the weights in the pooling scheme representing the ideas listed above.

For a particular site, and a particular selection of the sites to be used in the pooling group, the estimate of the L-moment ratio for the target site is given by

$$t^{(P)} = \sum_{j=1}^{M} w_j t_j \qquad (10)$$

where the quantity t_j is the sample estimate of the L-moment ratio at the jth of M sites in the pooling group and w_j is the weight for that site. In the case of a gauged site, the target site would be included in the pooling group as site 0, and the summation above would start with index $j = 0$. Based on the discussion above, the initial suggested form of the weighting for the ungauged case was defined as

$$w_j \propto (\alpha + c_j + f(s_{j0};\boldsymbol{\beta}))^{-1}, \quad i = 0, \ldots, M. \qquad (11)$$

Here c_j is the sampling variance of either L-CV or L-SKEW derived using a simple formula (Equations (31) and (32)), s_{j0} is the SDM between the target site and the jth catchment in the pooling group and $f(s;\boldsymbol{\beta})$ is a function with values depending on the value of SDM s and a parameter vector $\boldsymbol{\beta}$. The function f is assumed to satisfy $f(0;\boldsymbol{\beta}) = 0$. The parameter α represents a contribution related to the variation of the true values between catchments. This form of the weights gives less weight to sites in the pooling group with short records (large sampling uncertainty) and to sites that are relatively dissimilar to the target site (those likely to have large modelling errors).

Local models

To derive the appropriate form for the function $f(s;\boldsymbol{\beta})$ in Equation (11) and for the weighting to be used for the ungauged and the gauged cases, respectively, a formal statistical model is considered which leads to weights of the required form. Essentially the same statistical model is used for both L-CV (t_2) and L-SKEW (t_3); it is therefore convenient to present the model without using a subscript of 2 or 3 to denote which L-moment ratio is being considered.

For a target site, the M sites in the overall dataset which are selected as hydrologically similar according to the pooling-group scheme being tested can be identified. This gives a set of values $\{t_j;\, j = 0, \ldots, M\}$ of the L-moment ratio, together with the distance of each catchment from the initial catchment in catchment descriptor space. Let these distances be s_{j0}: for the gauged case, catchment 0 in the pooling group will be the target site with $s_{00} = 0$. For simplicity, a notional value for the sample L-moment ratio for ungauged target sites is included.

A local statistical model is defined to represent the observed value at the jth site in the pooling group in terms of the local mean value for the pooling group plus two error terms as

$$t_j = \mu + \varepsilon_j + \eta_j \quad j = 0, \ldots, M \tag{12}$$

where μ is the local mean value for the pooling group, ε_j is the sampling error and η_j is the modelling error. Here the modelling errors represent the assumption that an individual catchment in the pooling group has a true value of the L-moment ratio which is different from the true local mean, and that this difference increases with increasing dissimilarity. The variance of this difference is described by α and $f(s;\boldsymbol{\beta})$ and is also expected to increase with increasing dissimilarity. The term 'local' is used here to emphasize that different and formally contradictory models are implied for each target site. In particular, the value of the modelling error for a given catchment differs when this catchment appears as a member of pooling groups for different target sites. Interestingly, the ability of the formal model described above in Equation (12) to allow the L-moment ratios to vary between sites within the pooling group eliminates the need to identify homogeneous regions (a selection of sites with identical higher order statistical moments). This assumption is an important feature of the theory which is usually outlined as the basis of the index flood method.

The method presented here for determining the weights of the L-moment ratios in the pooling group requires a set of assumptions about the two error terms defined in Equation (12). Firstly, the sampling error of the L-moment ratio at each site is characterized by having zero mean value and a variance defined as

$$\mathrm{var}(\varepsilon_j) = c_j, \tag{13}$$

where these sampling variances are assumed known and depend on sample size as specified later. Next, the model errors are assumed to have a zero mean and to have a variance defined as

$$\mathrm{var}(\eta_j) = \alpha + f(s_{j0};\boldsymbol{\beta}), \tag{14}$$

where α and $\boldsymbol{\beta}$ are parameters to be selected.

There is an expectation that the size of the modelling error would increase as catchments become more dissimilar and, given that catchments become increasingly dissimilar to the target site as the SDM increases, this corresponds to an expectation that the variance of the model error η_j (in the local model) should be an increasing function of SDM. When deriving the weights within a pooling group it is also assumed that none of the catchments in a pooling group will be close enough in geographical space for there to be correlation in either the modelling errors or the sampling errors. However, the later variogram analysis does make some allowance for these correlations. A previous study by the authors (Kjeldsen & Jones 2009) found that both sampling and model errors of regional hydrological models depend strongly on the geographical distance between catchments. As there is no explicit reason for the sites in the pooling group to be geographically close, the assumption of independence was introduced to simplify the analytical framework. This allows an optimal set of weights to be defined in a simple and practically useful way.

Relation to a global model

To calculate an optimal set of weights for the L-moment ratios in the pooling group, an appropriate functional form for the function $f(s;\boldsymbol{\beta})$ needs to be found. The following analysis makes use of a correspondence between the error structure of the local model in Equation (12) and analyses based on variograms, where the variogram analysis can be considered a type of global model. To illustrate this connection, consider first the difference between the observed L-moment ratio at the target site t_0 and the jth site in the pooling group t_j ($j \neq 0$) written using Equation (12)

$$\begin{aligned} t_0 - t_j &= \mu + \varepsilon_0 + \eta_0 - \mu - \varepsilon_j - \eta_j \\ &= \varepsilon_0 + \eta_0 - \varepsilon_j - \eta_j \end{aligned} \tag{15}$$

Assuming independence between the error terms, the variance for the difference in Equation (15) is then expressed as

$$\begin{aligned} E\{(t_0 - t_j)^2\} &= \mathrm{var}(\varepsilon_0) + \mathrm{var}(\varepsilon_j) + \mathrm{var}(\eta_0) + \mathrm{var}(\eta_j) \\ &= c_0 + c_j + (\alpha + f(s_{00};\boldsymbol{\beta})) + (\alpha + f(s_{j0};\boldsymbol{\beta})) \\ &= c_0 + c_j + 2\alpha + f(s_{j0};\boldsymbol{\beta}) \end{aligned} \tag{16}$$

where the variance terms defined in Equations (13) and (14) have been used. None of the terms on the right-hand side here depend upon the sites involved actually having been associated via a pooling group. It is therefore proposed to use the global variogram model

$$\begin{aligned} E\{(t_i - t_j)^2\} &= c_i + c_j + 2\alpha + f(s_{ij};\boldsymbol{\beta}) \\ &= c_i + c_j + \gamma(s_{ij}) \end{aligned} \tag{17}$$

as a means to identify α and $\boldsymbol{\beta}$. Here i and j denote any two catchments in the overall dataset. The function $\gamma(s)$ later plays the role of the variogram of the modelling errors: it has the form implied by Equation (17) because of the assumptions in the local models.

Deriving analytical expressions for the weights

It is convenient first to introduce the following notation for quantities presently regarded as known:

$$\mathrm{var}(\eta_j) = \alpha + f(s_{j0};\boldsymbol{\beta}) = b_j. \tag{18}$$

When a pooling group is considered for a given target site, the local model allows optimal weights to be found for two different cases: (i) no information for the target catchment (ungauged); and (ii) gauged data available for the target catchment (gauged). In both cases the quantity that is to be estimated is the true value at the target site τ_0 which, according to the local model in Equation (12), is defined as

$$\tau_0 = \mu + \eta_0. \tag{19}$$

Here η_0 represents the deviation of the true value at the target site from the true mean for the pooling group.

Case (i): no observed data available at the target site (ungauged)

When the target site is ungauged, the estimate of the L-moment ratio is defined as the weighted average of the L-moment ratios for each catchment in the pooling group, i.e.

$$t^{(p)} = \sum_{j=1}^{M} w_j t_j \tag{20}$$

for a set of weights $\{w_j; j = 1, \ldots, M\}$ which sum to 1. Note that no data are available at the site of interest, thus $j = 0$ is not included in the sum in Equation (20). The error in the estimate is, according to the local model,

$$\begin{aligned} e &= \tau_0 - t^{(p)} \\ &= \mu + \eta_0 - \textstyle\sum_{j=1}^{M} w_j\{\mu + \varepsilon_j + \eta_j\} \\ &= \eta_0 - \textstyle\sum_{j=1}^{M} w_j\{\varepsilon_j + \eta_j\} \end{aligned} \tag{21}$$

and the expected squared error is

$$E\{e^2\} = b_0 + \sum_{j=1}^{M} w_j^2\{c_j + b_j\}. \tag{22}$$

This is minimized, over choices of sets of weights which sum to 1, by setting

$$w_j = \frac{\{c_j + b_j\}^{-1}}{\sum_{k=1}^{M}\{c_k + b_k\}^{-1}} \tag{23}$$

Case (ii): gauged data available for the target catchment (gauged)

When gauged data are available at the target site, the estimated sample L-moment ratio t_0 has a sampling error variance of c_0. The estimate is defined as a weighted average of all elements in the pooling group

$$t^{(p)} = \sum_{j=0}^{M} w_j t_j \tag{24}$$

where, as before, the weights sum to 1. The error in the estimate is, according to the local model,

$$\begin{aligned} e &= \tau_0 - t^{(p)} \\ &= \mu + \eta_0 - \textstyle\sum_{j=0}^{M} w_j\{\mu + \varepsilon_j + \eta_j\} \\ &= \{1 - w_0\}\eta_0 - w_0\varepsilon_0 - \textstyle\sum_{j=1}^{M} w_j\{\varepsilon_j + \eta_j\} \end{aligned} \tag{25}$$

and the expected squared error is

$$E\{e^2\} = \{1 - w_0\}^2 b_0 + w_0^2 c_0 + \sum_{j=1}^{M} w_j^2\{c_j + b_j\}. \tag{26}$$

This is minimized over choices of weights summing to 1 by

$$\begin{aligned} w_0 &= \frac{b_0}{\{c_0 + b_0\}} + \frac{c_0\{c_0 + b_0\}^{-2}}{\sum_{k=0}^{M}\{c_k + b_k\}^{-1}}, \\ w_j &= \frac{c_0\{c_0 + b_0\}^{-1}\{c_j + b_j\}^{-1}}{\sum_{k=0}^{M}\{c_k + b_k\}^{-1}} \end{aligned} \tag{27}$$

where $j = 1, \ldots, M$. By inserting the weights defined in Equation (27) into Equation (24), the pooled L-moment ratio estimator can be expressed as

$$\begin{aligned} t^{(p)} &= \textstyle\sum_{j=0}^{M} w_j t_j = \frac{b_0}{\{c_0+b_0\}} t_0 + \frac{c_0}{\{c_0+b_0\}} t_*^{(p)} \\ &= \frac{b_0}{\{c_0+b_0\}}\{t_0 - t_*^{(p)}\} + t_*^{(p)} \end{aligned} \tag{28}$$

where $t_*^{(p)}$ is defined as the pooled L-moment ratio that would be obtained if the gauged observations at the target site were not treated as special. That is,

$$t_*^{(p)} = \sum_{j=0}^{M} u_j t_j \tag{29}$$

where the corresponding weights u_j are defined as

$$u_j = \frac{\{c_j + b_j\}^{-1}}{\sum_{k=0}^{M}\{c_k + b_k\}^{-1}} \tag{30}$$

for $j = 0, \ldots, M$. For the special case where the observation used for t_0 is actually not for the target site, but still one that would be in the pooling group (for example, a gauging station located just up- or downstream of the target site), it may be reasonable to revise the effect of the special observation in the above Equation (in the final form of Equation (28) and not in the formulation of the weights u_j) by increasing the effect of the observation error variance c_0, so that less weight is given to t_0. However, no formal procedure for this adjustment is provided in this paper.

FITTING THE MODELS

Estimating the variance and covariances of the sampling errors

The weights to be used in a pooling group depend on the quantities c_j which are the sample variances of the rth L-moment ratio $t_{r,j}$ for the jth site. In theory, these variances depend upon the properties of the underlying distribution of annual maxima and this distribution may vary across sites. In order to construct a simple weighting scheme, it was decided to apply the simplification of using values of c_j appropriate to a single distribution defined to be the GLO distribution whose L-moment ratios are the average of the sample estimates across the whole dataset. A simulation study based on this distribution, covering a range of sample sizes, suggested the following approximations:

$$\text{LCV}: \quad c_i = \frac{0.02609}{n_i - 1} \tag{31}$$

$$\text{L-SKEW}: \quad c_i = \frac{0.2743}{n_i - 2}, \tag{32}$$

where n_i is the record length available for the ith catchment.

The correlation of the sample L-moment ratios between sites was obtained through a bootstrapping experiment and defined as a function of geographical distance between catchment centroids:

$$\text{L-CV} \quad \text{cor}\{t_{2,i}, t_{2,j}\} = \exp(-0.030 d_{ij}) \tag{33}$$

$$\text{L-SKEW} \quad \text{cor}\{t_{3,i}, t_{3,j}\} = \exp(-0.050 d_{ij}) \tag{34}$$

where d_{ij} is geographical distance (km^2) between catchment centroids in kilometres. These correlations are not used in the formulae for the weights in the analysis above, but they are used in the analysis presented in the next subsection which determines how the weights vary with the similarity distance measure, via Equation (18).

Fitting the global variogram model

In fitting the global variogram model discussed earlier, it seemed appropriate to allow for correlation in the sampling errors. Accordingly, the contribution from a pair of catchments to a variogram analysis is defined to be an adjusted version of the squared difference of the sample L-moment ratios given by

$$g_{ij} = (t_i - t_j)^2 - \left(c_i + c_j - 2\rho_{ij}\sqrt{c_i c_j}\right) \quad (35)$$

where ρ_{ij} is the correlation of the sample estimates. We then have

$$E(g_{ij}) = 2\alpha + f(s_{ij}; \boldsymbol{\beta}) \quad (36)$$

and thus the values g_{ij}, when plotted against the SDM s_{ij}, should cluster around a line defined by Equation (36) which can be used to define a functional form for $f(s)$ and to estimate α and $\boldsymbol{\beta}$. In practice, this approach is modified to plot averages of the g_{ij} within SDM-based cells against SDM as

$$g(s) = \frac{1}{|n(s, \Delta s)|} \sum_{n(s,\Delta s)} g_{ij} \quad (37)$$

where the set $n(s, \Delta s)$ consists of the catchment pairs (i,j) located at SDM distances apart of between s and $s + \Delta s$. Here $g(s)$ is regarded as a raw estimate of the true variogram function $\gamma(s)$ of the modelling errors, as in Equation (17). In implementing the above, the correlation of the sample estimates of the rth L-moment ratio is adjusted to take into account partially overlapping record lengths to give

$$\rho_{ij} = \frac{n_{ij}}{\sqrt{n_i n_j}} \mathrm{cor}\{t_{r,i}, t_{r,j}\}, \quad (38)$$

where n_{ij} is the number of years for which the records for sites i and j overlap and where the underlying correlation is given by Equations (33) and (34).

Table 2 shows the average and maximum SDM observed when a pooling group consisting of 30 catchments was defined for each of the 602 catchments used in this study.

Based on the values of SDM in Table 2, it was decided that a maximum SDM within any pooling group would reasonably be below 4. Estimates of the variograms for L-CV and L-SKEW obtained using Equation (37) are shown in Figure 2. The shapes of the plots indicate that functional relationships can be fitted, linking the variograms to the SDM using an empirical type function

Table 2 | Values of SDM for 602 pooling groups consisting of 30 catchments each

	Distance
Average distance from target to pooling-group members (for all 602 pooling groups)	0.64
Average distance from target to most dissimilar pooling-group member (for all 602 pooling groups)	0.85
Average distance from target to pooling-group member (for pooling group with largest average SDM of all 602 pooling groups)	3.46
Maximum observed SDM value between two catchments in the data set (not in same pooling group)	11.04

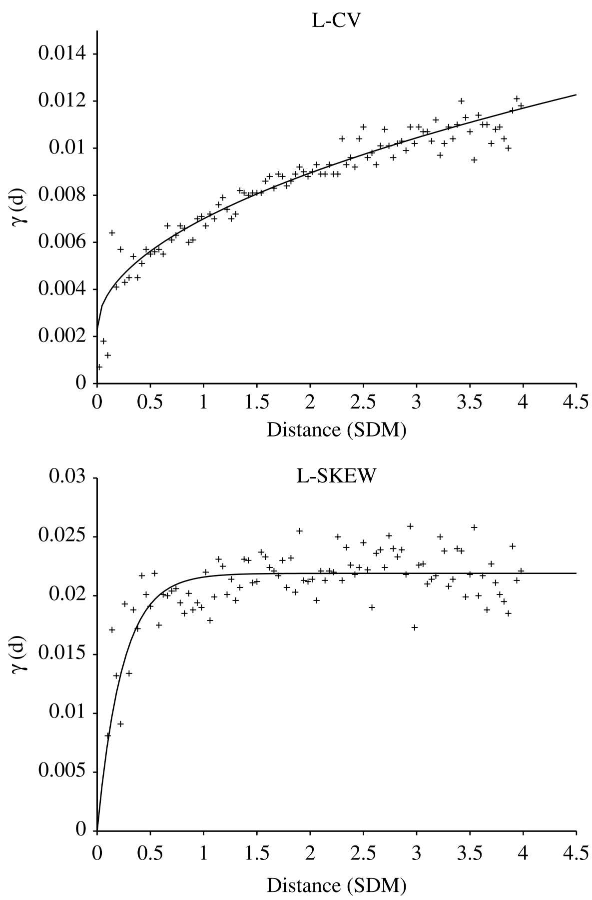

Figure 2 | Variograms for L-CV and L-SKEW plotted as a function of distance in catchment descriptor space. Maximum distance is set to 4 and subdivided into 100 intervals.

that was found to fit the data reasonably well:

$$\gamma(s) = (\beta_1\sqrt{s} + \beta_2)\left(1 - \exp\left[-\frac{s}{\beta_3}\right]\right)^{\delta} \tag{39}$$

where s is the SDM value at the middle of the interval of width Δs, $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ are model parameters and δ is a binary parameter that can be either 1 or 0, depending on whether the variogram has a nugget at distance zero or not. When using this equation for defining the parameters in Equation (36) it should be noted that, if a nugget exists, $2\alpha = \beta_2$; otherwise $\alpha = 0$. The three parameters $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)^T$ were estimated based on a least squares analysis.

In terms of the weighting of the L-moment ratios within each pooling group, the results derived here imply that the weighting parameter b_j for L-CV and L-SKEW should be defined as

$$\begin{aligned} \text{L}-\text{CV}: \quad & b_j = \left(0.0047\sqrt{s_{j0}} + \tfrac{0.0023}{2}\right) \\ \text{L}-\text{SKEW}: \quad & b_j = 0.0219\left(1 - \exp\left[-\tfrac{s_{j0}}{0.2360}\right]\right) \end{aligned} \tag{40}$$

where s_{j0} is the SDM in catchment descriptor space from the target site to the jth site in the pooling group. Combining Equations (40) with the analytical expressions for the weights in the two cases outlined above for the gauged and ungauged catchments (Equations (23) and (27)), the pooled L-moment ratios can be derived through Equation (5).

The results in Figure 2 show contrasting behaviours in the estimated variograms for L-CV and L-SKEW as functions of SDM. While the variogram for L-SKEW is essentially constant once SDM is above a low threshold, the variogram for L-CV continues to increase. This shows that values of L-CV in pairs of catchments become increasingly different with increasing SDM. A consequence of this is that the weights assigned within a pooling group will behave differently for the two L-moment ratios. Apart from differences associated with differences in record length, the weights for different catchments will decrease noticeably with increasing SDM in the case of L-CV, while remaining the same in the case of L-SKEW. Once SDM exceeds the low threshold, any catchment is equally beneficial for estimating L-SKEW at the target site. The results here also indicate that the task of identifying a set of gauged catchments with identical L-CV values (a homogeneous region), which is notionally required for the existing pooling-group scheme, is likely to be a forlorn hope. For example, Figure 2 shows that unless a pair of catchments are extremely similar, the true values of L-SKEW are likely to differ by 0.14 and the true values of L-CV will differ by at least 0.07 (these values are obtained by taking the square root of the variogram values). The formal statistical model presented here provides a way of accounting for the sizes of these differences in relation to the sampling errors in estimates of the L-moment ratios, and a way of weighting which takes into account the increasing differences between the L-moment ratios for catchments as SDM increases.

RESULTS

Size of pooling groups

The final step in the development of the pooling procedure is to determine the number of catchments to be included in a pooling group. The optimal size of a pooling group is a trade-off between the bias and variance of the estimated T-year flow. Firstly, too many sites in a pooling group will increase the probability of including sites that are different from the target site, thereby increasing the bias of the T-year event. On the other hand, including too few sites will lead to estimates with a larger variance (higher uncertainty) of the T-year estimate than necessary. Based on a series of Monte Carlo experiments, Hosking & Wallis (1997) found that little could be gained, in terms of RMSE, by using regions larger than about 20 sites.

The size of a pooling group can be measured either in terms of number of sites or total number of site-years having annual maxima. The FEH opted for the latter measure because of the large variation in record length observed in practice. The FEH introduced a $5T$ rule allowing the size of the pooling group to be determined by the target return period, e.g. for a target return period of $T = 100$ years sites should be added to the pooling group until 500 events has been reached. However, the FEH stated that this was indeed a rule-of-thumb and based on intuition rather than the outcome of a particular analysis.

In this study, the size of the pooling group was investigated based on both the number of catchments and the number of events. In both cases the PUM measure was used to assess the appropriate size. Figure 3a and b show how the PUM measure varies according to size for both cases. Here the pooling-group estimates are assessed for a number of different return periods and the conclusions reached are the same for each of these.

For both measures of pooling-group size, the PUM decreases rapidly from high values for very small pooling groups until a size of 10 sites, or about 300 site-years, is reached. Between 10 and 17 sites (or 300–500 site-years) little change in PUM can be observed after which the PUM increases for increasing pooling-group size. These results give no particular reason why one method of measuring pooling-group size should be preferred over the other. It has therefore been recommended to maintain the current measure used in the FEH, i.e. using the number of events.

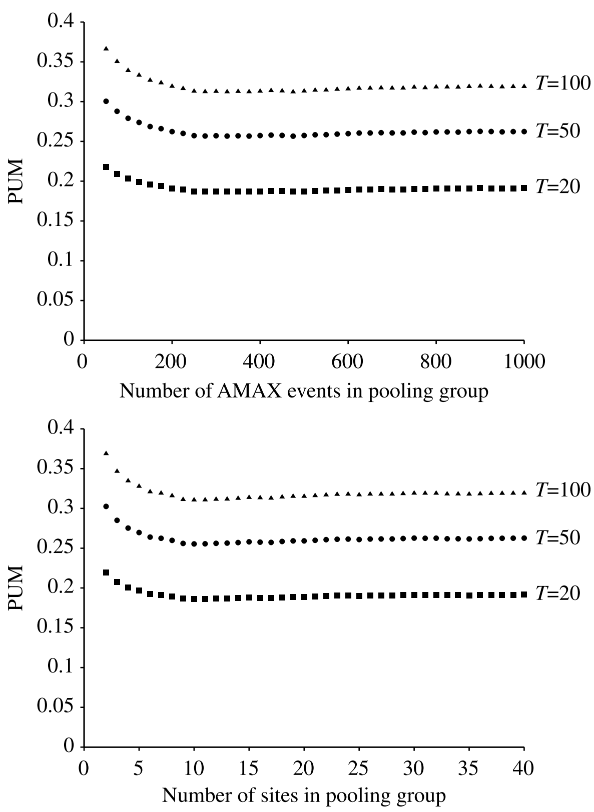

Figure 3 | PUM as a function of pooling-group size as measured by (a) number of annual maximum events and (b) number of sites. PUM was calculated for $T = 20$, $T = 50$ and $T = 100$ years.

However, while the current FEH practice recommends setting the target number of years as a function of the desired return period, the results reported here have led to the recommendation that a single pooling-group size of 500 site-years should be applied irrespective of the target return period.

Performance of pooling method

The performance of the pooling-group method needs to be assessed for two different cases: firstly performance at the ungauged site where no at-site data are available and, secondly, for the gauged site. In both cases the performance has to be compared with that of an at-site analysis and this means that the case of a gauged site is best considered separately because caution is needed in interpreting the PUM criterion when gauged data contribute both to the target value (the at-site estimate) and to the pooling-group estimate.

The ungauged site

The pooling-group method derived in the previous sections is only one candidate out of many possible procedures that could have been specified. At each step in the development different options could have been selected, which would have lead to a modified end-product. However, the development should be reasonably close to an optimal procedure (with regards to PUM) as each step in the development was justified through careful analysis of the data. To provide further evidence of the improvements made during this project, the new method developed here is compared to a series of alternatives: (i) a simplified procedure (simple weights), (ii) a single UK growth factor; (iii) pooling groups formed based on geographical distance only; (iv) the FEH methodology; and (v) regression models for L-CV and L-SKEW based on the same four catchment descriptors (pooling variables) employed in the similarity distance measure which is used to form pooling groups in the new methodology. Each of the alternatives listed above will be assessed using the PUM measure by considering each of the 602 catchments to be ungauged and calculating PUM for return periods 20, 50 and 100 years. Specifically, when a given catchment is the 'target', it is omitted from the pooling group. The results are summarized in Table 3.

From the results in Table 3 it is clear that the new method performs better than the existing FEH methodology. It is worth noting that the increase in performance gained by introducing the new weighting scheme appears rather small compared to a method using relatively simpler weights. As an empirical measure of performance, PUM is affected by sampling errors associated with constructing the at-site estimates of the growth curves for the target sites: this contributes a component to the raw prediction error for any given method which is common across all methods. This means that PUM^2 overestimates the size of the true squared-errors by an unknown amount, but one which should be the same across all methods. Thus, in the case where the target site is omitted from the pooling group, there is effectively a lower bound related to these sampling errors below which PUM cannot be reduced by any estimation method. Given this, the actual importance of the differences in the PUM values is not known but is certainly greater than that indicated by the reduction in PUM between the new and the simple weighting schemes.

The gauged site

At the gauged site, the benefit of using pooled analysis should be compared to a direct at-site analysis of the available data and should ideally consider aspects of variability and bias of the estimated design events. Based on a series of Monte Carlo experiments, Hosking & Wallis (1997) concluded that while pooled (or regional) analysis might be considered overall beneficial for a region, at-site analysis might still be preferable at individual sites. Note that they were not considering pooling schemes giving special weight to information from a gauged target site. The design of a suitable Monte Carlo experiment considering the entire UK was considered to be outside the scope of the current project. Instead, results will be reported using the PUM measure.

It can be argued that the PUM measure is not suitable for a comparison of at-site and pooling procedures, as it will favour any pooling procedure giving results close to the at-site results without consideration of variability of the estimates and, in particular, it would incorrectly judge that the best pooling-group weighting scheme would give zero weight to all catchments other than the target catchment. On the other hand, use of the PUM measure will allow a comparison between using pooled analysis at a gauged and an ungauged site. Specifically, the comparison measures how much closer the pooled estimates have moved towards the at-site estimates by the inclusion of the local information. Thus the values of PUM directly measure how much estimates would change if the purely at-site estimates of the growth curves are replaced by the pooling-group estimates which make optimal use of local information. The PUM values for the gauged case (weights calculated using Equation (29)) are shown at the bottom row of Table 3 and are lower than the corresponding estimates at the ungauged site. This indicates a substantial movement towards the purely at-site estimates.

Table 3 | Comparison of pooling methods using PUM

		$T = 20$	$T = 50$	$T = 100$
1	New method	0.1875	0.2576	0.3134
2	New method, simple weights	0.1886	0.2591	0.3152
3	Single UK growth curve	0.2164	0.2914	0.3501
4	Geographical proximity*	0.1926	0.2651	0.3226
5	FEH method†	0.1986	0.2718	0.3296
6	Regression models	0.1881	0.2598	0.3170
7	New method, gauged catchment	0.1095	0.1622	0.2062

*Pooling-group size of 700 AMAX events.
†Pooling-group size of 500 AMAX events.

DISCUSSION AND CONCLUSIONS

The particular procedure for using pooling groups in flood frequency analysis presented in this paper is based on hydrological and catchment descriptor data from the UK. However, the procedure and the results are believed to be of interest to the wider hydrological community.

The pooling procedure is developed from the need to produce a practical and auditable methodology which allows the estimation of site-specific growth curves for catchments covering a large range of catchment descriptors while, at the same time, allowing the inclusion of peak flow data not used in the development of the procedure itself. For example, peak flow recordings from flood events occurring only after the development of the procedure or a gauged record not used in this study could be included. To avoid users having to re-estimate an 'optimal procedure' every time new data becomes available, the procedure and the associated parameter values and methodological choices are assumed fixed. Because of the need to develop a method which considers each individual catchment and uses data from a limited number of hydrologically similar catchments, this study introduced the concept of a 'local' as opposed to a 'global' model. The weights used for a given target catchment are notionally optimal under the 'local' model. While the local models for different target sites are mutually contradictory they have an approximate interpretation under a 'global' model. This global model is used to determine part of the structure of the local models and therefore contributes to the specification of the weights. While it would be possible to use weighting schemes derived more directly from the 'global' model, this would lead to the use of matrix inversions and thus the weights would not be easy to interpret. An alternative example of a global model would be the use of regression models linking L-CV and L-SKEW to a set of catchment descriptors. An additional disadvantage of a global model is the potential need to re-estimate the entire model every time new data are introduced into the analysis, thereby leading to numerous alternative versions of the model depending on what data are used. Also, every new model would potentially provide a new set of estimates at any location in the country. This would make auditing and decision-making very difficult.

The sequential development of the pooling procedure considered three different aspects: (i) defining hydrological similarity; (ii) determining pooling-group size; and (iii) defining weights within a pooling group. At each step of the development the choice of methodology and estimated model parameters were based on analysis of the available dataset. Also, the choice of methodology was, to some degree, influenced by the wish to retain certain features of the existing FEH pooling procedure such as the concept of hydrological similarity as measured from catchment descriptors and measuring pooling-group size using the number of annual maximum events. The final procedure, therefore, is not strictly an optimal method, but a practical procedure found to perform reasonably well over the range of catchments included in this analysis.

On comparing the performance of the new procedure with other alternative methods that might be adopted (Table 3), the new procedure performs better (on average) at ungauged catchments than any of these methods, including the existing FEH methodology and geographical distance. Although the significance of differences in PUM values has not been assessed, the new procedure performs only marginally better than other more simple methods such as a version of the pooling procedure with simple weights or a set of regression models linking L-CV and L-SKEW to catchment descriptors. (Note that it is possible that a set of better regression models could be developed, potentially performing better than the pooling-group methodology.) However, the new procedure represents a weighting scheme for pooling groups whose formulation has been guided in an attempt to overcome criticisms of the existing FEH methodology (Morris 2003), as far as possible.

An important outcome of using the formal statistical model described here is that it allows a special form of weighting to be developed for cases where the catchment, which is the target of the analysis, does have some records of annual maximum flood available. Such a scheme gives higher weight to information that is local than to information transferred from distant but hydrologically similar catchments. The benefits of a pooled analysis at a gauged site would be a possible reduction in the variance of the *T*-year design quantiles when compared with an at-site analysis, as illustrated by Kjeldsen & Jones (2006). With regards to the transfer of data from neighbouring gauged catchments, results recently published (Kjeldsen & Jones 2007) have shown that the potential benefits of data transfer in flood frequency estimation are closely linked to the correlation of the model error between catchments. Kjeldsen & Jones (2007) showed that if proper account is not taken of the statistical structure of the model error, the use of data transfer techniques can worsen the predictive ability of a regional statistical model.

To summarize, the statistical model presented in this paper allows an optimal version of the region-of-influence (ROI) method to be developed, based on a large representative dataset, but the procedures should continue to be applicable as more data accumulates. The procedures developed in this study form the basis of forthcoming recommendations for revised versions of the FEH procedures for use in the UK.

ACKNOWLEDGEMENTS

This study was supported by the Joint Defra/EA Flood and Coastal Erosion Risk Management R&D Programme, Project SC050050, and is published with the Environment Agency's agreement. This project has benefited from the improved datasets made available through the Environment Agency HiFlows-UK project. The authors thank Professor Günther Blöschl and an anonymous reviewer for constructive comments on an earlier version of the manuscript.

REFERENCES

Acreman, M. & Wiltshire, S. E. 1987 Identification of regions for regional frequency analysis (abstract). *Eos Trans. AGU* **68**(44), 1262.

Acreman, M. & Wiltshire, S. E. 1989 The regions are dead; long live the regions. Methods of identifying and dispensing with regions for flood frequency analysis. In: L. Roalds *et al.* (eds) *FRIENDs in Hydrology*, IAHS Publ. 187, pp. 175–188.

Burn, D. H. 1990 Evaluation of regional flood frequency analysis with a region of influence approach. *Water Resour. Res.* **26**(10), 2257–2265.

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

Gottschalk, L. 1993 Correlation and covariance of runoff. *Stoch. Environ. Res. Risk Assess.* **7**(2), 85–101.

Holmes, M. G. R., Young, A. R., Gustard, A. & Grew, R. 2002 A region of influence approach to predicting flow duration curves within ungauged catchments. *Hydrol. Earth Syst. Sci.* **6**(4), 721–731.

Hosking, J. R. M. & Wallis, J. R. 1997 *Regional Flood Frequency Analysis: An Approach Based on L-moments*. Cambridge University Press, New York.

Institute of Hydrology 1999 *Flood Estimation Handbook*, Wallingford, UK.

Kjeldsen, T. R. & Jones, D. A. 2006 Prediction uncertainty in a median-based index flood method using L moments. *Water Resour. Res.* **42**, W07414.

Kjeldsen, T. R. & Jones, D. A. 2007 Estimation of the index flood using data transfer in the UK. *Hydrol. Sci. J.* **52**(1), 86–98.

Kjeldsen, T. R. & Jones, D. A. 2009 An exploratory analysis of error components in hydrological regression modelling. *Water Resour. Res.* **45**, W02407.

Kjeldsen, T. R., Jones, D. A. & Bayliss, A. C. 2008 *Improving the FEH statistical procedures for flood frequency estimation*. Environment Agency, Report SC050050/SR.

Merz, R. & Blöschl, G. 2005 Flood frequency regionalisation–spatial proximity vs. catchment attributes. *J. Hydrol.* **302**(1–4), 283–306.

Morris, D. G. 2003 *Automation and appraisal of the FEH statistical procedure for flood frequency estimation*. R&D Report to Defra, Centre for Ecology & Hydrology, Wallingford, UK.

Morris, D. G. & Flavin, R. W. 1990 *Flood risk map for England and Wales*. *IH Report* 130, Institute of Hydrology, Wallingford, UK.

Skøien, J. O., Merz, R. & Blöschl, G. 2006 Top-kriging-geostatistics on stream networks. *Hydrol. Earth Syst. Sci.* **10**, 277–287.

Stedinger, J. R., Vogel, R. M. & Foufoula-Georgiou, E. 1993 Frequency analysis of extreme events. In: Maidment, D. (ed.) *Handbook of Hydrology*. McGraw-Hill, New York (Chapter 18).

First received 3 July 2008; accepted in revised form 24 February 2009

APPENDIX A

List of symbols

α	Variation of true values of L-moment ratios between sites
b	Model error variance of L-moment ratio in pooling group
$\boldsymbol{\beta}$	Vector of model parameters linking model error and the SDM
c	Sampling variance of L-moment ratios
d	Geographical distance between catchment centroids (km)
g	Adjusted sample variogram of L-moment ratios
γ	True variogram of modelling errors in L-moment ratios
h	Weights in PUM measure
M	Number of sites in a pooling group
n	Number of annual maximum peak flow events (years)
ω	Weights in similarity distance measure (SDM)
P	Number of catchment descriptors defining hydrological similarity
q	Annual maximum peak flow in an individual year
Q_T	T-year peak flow ($m^3 s^{-1}$)
s_{ij}	Similarity distance measure between catchment i and j
σ_x	Standard deviation of one of the catchment descriptors, between All 602 catchments
T	Return period (years)
t_r	Sample L-moment ratio
$t^{(p)}$	Pooled L-moment ratio
τ_0	True L-moment ratio at target site
μ	True mean value of the L-moment ratio for a pooling group
w	Weight of individual catchments within a pooling group
x	One of a set of descriptors available for any catchments
z_T	T-year growth factor

Note there is a slight inconsistency in that α and β are also used to define the scale and normalized scale parameters of the GLO distribution as well as describing the parameters in the model linking model error variance with SDM.