Conference or Workshop Item

Kjeldsen, Thomas; Jones, David; Bayliss, Adrian; Peter, Spencer; Suresh, Surendran; Stefan, Laeger; Paul, Webster; David, MacDonald. 2008 Improving the FEH statistical method. In: *Flood & Coastal Management Conference 2008, University of Manchester, 1-3 July 2008*. Environment Agency/Defra.

IMPROVING THE FEH STATISTICAL METHOD

THOMAS KJELDSEN, MSc PhD, DAVID JONES, PhD and ADRIAN BAYLISS, MSc
Centre for Ecology & Hydrology, Wallingford

PETER SPENCER, SURESH SURENDRAN and STEFAN LAEGER
Environment Agency

PAUL WEBSTER
Hydro-Logic Ltd

DAVID MCDONALD
Black & Veatch

Key Words: Flood Estimation Handbook, flood frequency analysis, statistical method

Abstract: The FEH procedures have been adopted as standard practice, where applicable, by the Environment Agency and other bodies engaged in flood frequency estimation in the UK, and are used primarily for flood mapping studies, flood risk assessments, and the design of flood alleviation schemes. The results from the recently completed R & D project SC050050 (EA, 2008) recommend changes to the procedures contained in the flood Estimation Handbook (FEH) published by the Institute of Hydrology (1999) for use of statistical methods for flood frequency analysis in the UK.

The changes recommended arise, in part, because the HiFlows-UK project has led to the creation of a much improved database of systematically recorded flood data. Another influence on the changed procedures has been feedback from users of the FEH, both informal and formal. The changes do not deviate from the overall framework of the FEH methodology. However, most technical details of the method have been updated to improve the performance of the procedure. These updates include:

i) A new equation for estimating the median annual maximum flood (*QMED*) at ungauged catchments .
ii) An improved procedure for the use of donor catchments for estimation of *QMED* at ungauged catchments
iii) An improved procedure for the formation of pooled growth curves.

There are some cases where the analysis carried out in the science report have recommended no change to the FEH methodology these include the retention of the Generalised Logistic (GLO) distribution as default.

INTRODUCTION

The Flood Estimation Handbook (FEH) published in 1999 (Institute of Hydrology, 1999) continues to form the basis of most work on flood frequency estimation in the UK. The FEH procedures have been used for a wide range of applications such as flood mapping studies, Flood Risk Assessments, and the design of flood alleviation schemes by the Environment Agency and other bodies. It consists of five volumes describing two principal methods for conducting flood frequency analysis as well as information on electronic datasets of catchment descriptors such as catchment area and annual average rainfall. The two principal methods described in the FEH are: i) a rainfall-runoff based approach adopted largely unchanged from the original method developed in the early 1970s and published in the Flood Studies Report (FSR) by NERC (1975), and ii) a statistical method based on analysis of annual maximum series of instantaneous peak flow.

Since publication in 1999, the continuous scientific developments as well as the ongoing dialog between users of the FEH and the research team at the Centre for Ecology & Hydrology (CEH) in Wallingford have highlighted various aspects of the methods and their implementation which would benefit from further improvements. As a result, the Joint Defra/Environment Agency Flood and Coastal Management R&D program has funded a number projects aimed at improving various aspects of the two methods in the FEH. For example the “Revitalisation of the FSR/FEH rainfall-runoff method” - FD1913 (Defra, 2005) and the “Dissemination of the FSR/FEH rainfall-runoff method” - SC040029 (EA 2007) focused on improving the rainfall-runoff approach were as the recently completed project SC050050 Improving the FEH Statistical Index Flood Method and Software (EA 2008).

The aim of the SC050050 project was to improve the statistical method for flood frequency analysis reported in the FEH. The improved procedure has retained the fundamental aspects of the FEH method, i.e. using the index flood method, where the index flood is defined as the median annual maximum peak flow (*QMED*) and the dimensionless growth curve, z_T, is estimated using pooling groups. The final design peak flow value for a return period T, i.e. Q_T, is derived as the product of the index flood and the growth curve.

APPRAISAL AND SELECTION OF DATA

The development of a general statistical model for flood frequency analysis requires two types of data: i) observed peak flow data, and ii) data on physical catchment descriptors.

Peak Flow Data

The peak flow data used in this study were obtained from the HiFlows-UK project and consisted of annual maximum series of peak flow from a total of 602 rural catchments. A summary of the data set is shown in Table 1 together with a summary of the 728 rural catchments used in the development of the original FEH method.

Table 1 Summary of AMAX data sets [years of data]

	FEH	SC050050
Number of gauges	728	602
Shortest record length	2	4
Longest record length	84	117
Mean record length	22.7	32.7
Number of AMAX events	16528	19679

From the comparison of the two data sets in Table 1 it is clear that, even though the FEH used more gauging stations, the total number of annual maximum (AMAX) peak flow events is higher in the data set used in SC050050. Note that for records shorter than 14 years, the FEH used POT data to derive *QMED*, and only records with more than seven years were included in the pooled analysis (note that no POT data were used in this study). In fact, a total of only 698 sites were used in the FEH for the pooled analysis. A further comparison of the two data sets is shown in Figure 1 in the form of histograms of record length.

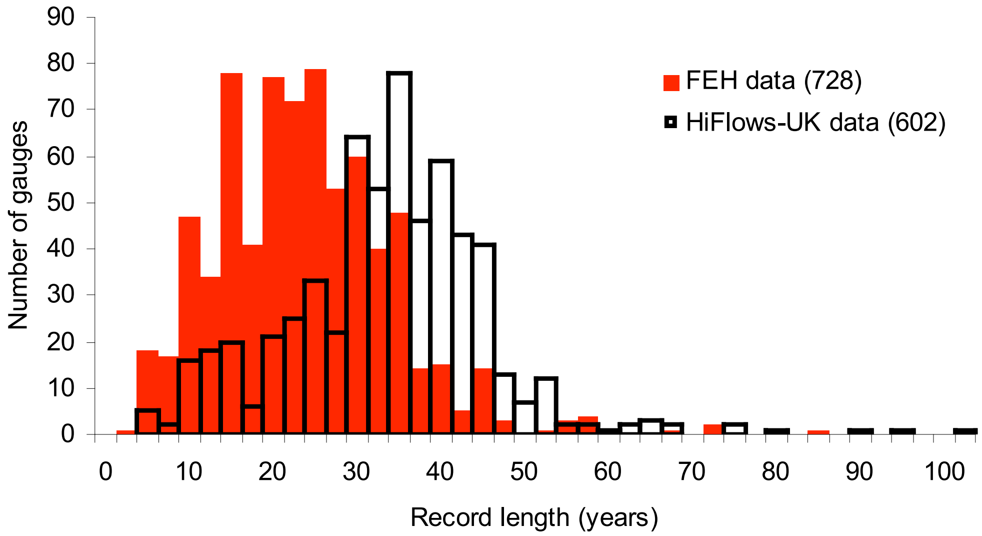

Figure 1 Histograms comparing record length of FEH and HiFlows-UK data sets.

Again, the histograms in Figure 1 illustrate the effect in HiFlows-UK of including the additional AMAX data from the end of the FEH data (at best mid-1990s) to end of water-year 2002 (which represents

October 2002 to September 2003). This increase in record length will generally reduce the sampling uncertainties of the estimates of *QMED* and of the L-moment ratios.

Catchment Descriptors
The digital catchment descriptors used in this study were mainly extracted from the FEH CD-ROM Version 2 (CEH, 2007) for each of the 602 gauged catchments. The number of catchment descriptors potentially available is large, but only a subset of variables previously found to be useful in flood studies were considered in this study. In addition to the existing descriptors available from the FEH CD-ROM, an additional descriptor describing the extent of flood plains in a catchment (*FPEXT*) was developed as part of this study and found useful when defining pooling groups. For details of the derivation of *FPEXT* values, please refer to the Science Report for project SC050050 (EA, 2008).

ESTIMATING QMED AT AN UNGAUGED SITE

Where no flood data are available at the site of interest, *QMED* has to be estimated from catchment descriptors. The estimate can subsequently be adjusted using data transfer from a nearby gauged catchment (a donor site).

A new QMED model
Through a combination of methodological developments and extensive exploratory analysis of model residuals, a revised QMED model was developed for use on rural catchment

$$QMED = 8.3062 AREA^{0.8510}\, 0.1536^{\left(\frac{1000}{SAAR}\right)} FARL^{3.4451}\, 0.0460^{BFIHOST^2} \tag{1}$$

where *AREA* is the catchment area [km^2], *SAAR* is the standard average annual rainfall [mm] based on measurements from 1961-1990, *FARL* is an index of flood attenuation due to reservoirs and lakes, and *BFIHOST* is the baseflow index derived from HOST soil data. The factorial standard error (fse) of *QMED* values estimated from this new QMED model are 1.431, which is a 7.5% reduction compared to the fse value of 1.541 reported for the original FEH QMED model. While significantly increasing the predictive accuracy of the QMED model when compared to the FEH, the new model is also analytically more simple using only four catchment descriptors compared to the six used in the QMED model reported in the FEH.

A Revised Procedure for Use of Donor Catchments
The FEH emphasises that the uncertainty of *QMED* estimated using the QMED model is generally much larger than the uncertainty of estimates obtained directly from flood data. Consequently, the FEH recommends that data transfer from nearby donor or analogue catchments should be used wherever possible. However, based on research by Kjeldsen and Jones (2007) and results obtained in project SC050050, it has become clear that the benefits from the FEH donor transfer method are generally less than previously thought. It is therefore recommended that the data transfer procedure is revised to account for the geographical distance, d_{sg}, in kilometres between the centroids of the subject site and a donor catchment as

$$QMED_{s,adj} = QMED_{s,cds}\left(\frac{QMED_{g,obs}}{QMED_{g,cds}}\right)^{a_{sg}} \tag{2}$$

where

$$a_{sg} = 0.4598\exp(-0.0200 d_{sg}) + (1-0.4598)\exp(-0.4785 d_{sg}) \tag{3}$$

And where the subscript *s* refer to the ungauged target (or subject) site and *g* the gauged donor site. The subscript *cds* refer to an estimate derived from catchment descriptors at the gauged and target sites, *obs* the observed value at the gauged donor site and *adj* the adjusted value at the ungauged target site. The donor adjustment in the form given above will automatically reduce the influence of the donor site as the geographical distance between the two catchment centroids increases. Note that currently the donor adjustment procedure is only capable of considering one donor site at the time. Further development is needed to develop a method capable of considering multiple donor sites.

AN IMPROVED METHOD FOR ESTIMATION OF POOLED GROWTH CURVES

The estimation of the growth curve is based on the pooling-group method and requires i) the formation of a pooling-group followed by ii) estimation of the pooled distribution parameters through the method of L-moments using the weighted average of the L-moment ratios within the pooling-group.

Selecting a pooling-group

As in the FEH, a pooling-group for a particular site of interest is formed by identifying a number of gauged catchments classified as hydrologically similar. The selection of catchments is based on a distance measure, measuring the distance in a catchment descriptor space defined by ln[*AREA*], ln[*SAAR*], *FARL* and *FPEXT* and calculated as

$$SDM_{ij} = \sqrt{3.2\left(\frac{\ln AREA_i - \ln AREA_j}{1.28}\right)^2 + 0.5\left(\frac{\ln SAAR_i - \ln SAAR_j}{0.37}\right)^2 + 0.1\left(\frac{FARL_i - FARL_j}{0.05}\right)^2 + 0.2\left(\frac{FPEXT_i - FPEXT_j}{0.04}\right)^2} . \tag{4}$$

The FEH recommended that the size of the pooling-group should vary according to the target return period such that the total number of AMAX events should be at least 5 times the return period (the 5T rule). However, this study found that a fixed pooling-group size consisting of 500 AMAX events performed well for a range of return periods.

Estimating the pooled growth curve

Detailed instructions of how to review and adapt the initial pooling-group were presented in the FEH. The main differences between the FEH and the revised method are due to the new weighting scheme for which the details can be found in EA (2008), however the main points of difference are summarised below. Firstly, the new weighting scheme assigns weight to each individual catchment in the pooling-group based on the distance in catchment space from the target site, as defined in equation (4), rather than on rank within the pooling-group. Hence, moving catchments up or down in the ranking order within the pooling-group will not change the weights. Secondly, the weighting scheme will differentiate between a gauged and an ungauged catchment and derive weights differently for the two cases. The new method will assign more weight to available at-site data than the FEH procedure, thus generally ensuring a closer resemblance between the single-site and pooled analysis than the FEH procedure. Finally, two separate sets of weights are used for the calculation of pooled L-CV and L-SKEW values.

COMPARISON WITH FEH RESULTS

Estimates of design peak flow values obtained from the new method can in some cases be substantially different from the estimates obtained using the original FEH procedures. The following comparisons of results are derived by using the two different procedures on the same data set, i.e. the AMAX series obtained from HiFlows-UK for the 602 rural catchments.

Comparison of QMED estimates

A comparison between estimates of QMED obtained using the new QMED equation, equation (1), and the original FEH QMED equation, is presented in Figure 2, which shows the ratio of the two estimates for each of the 602 catchments used in this study. Note that the catchments included in Figure 2 include only rural catchments, hence the relatively large geographical regions in England without any coverage. From this figure it is clear that the new QMED equation gives lower estimates of QMED in the East and South Eastern parts of England, but generally higher estimates in the Western and Northern parts of the UK.

The results in Figure 2 can be summarised by the statement that the changes in *QMED* range from factors of 0.55 to 2.01, with half being greater than 1.15 (25 per cent of the ratios are less than 1.00 and 25 per cent are greater than 1.24). Here a factor greater than one means that estimates of *QMED* from the new equation are greater than the estimates given by the equivalent FEH equation. Overall, the new estimates are larger than the FEH estimates in 75 per cent of catchments.

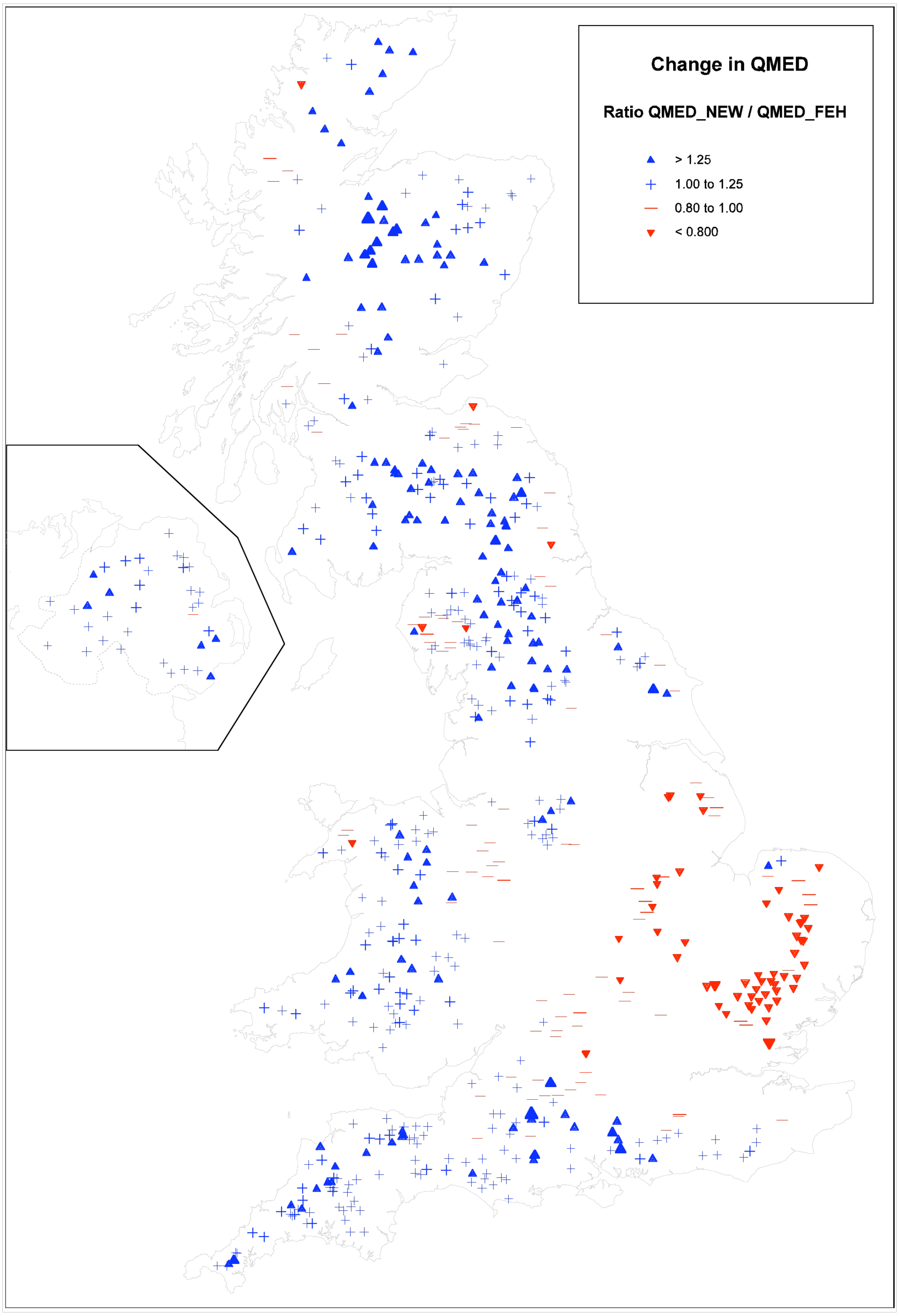

Figure 2 Comparison of *QMED* estimates obtained from catchment descriptors only using i) the new QMED model and ii) the FEH QMED model.

Comparison of results for the 100-year return period

This section presents a comparison of the differences between both the 100-year growth factors and the 100-year floods (i.e. z_T and Q_T for T=100) as estimated for an ungauged site using the pooling procedure developed in this study and the FEH methodology. Both sets of estimates are based on the HiFlows-UK data-set used in this study. The comparison assumes the subject site to be ungauged, which means that the AMAX record for each subject site is not included in its own pooling-group. The FEH methodology forms pooling-groups based on hydrological similarity as defined in the FEH (i.e. using ln[*AREA*], ln[*SAAR*] and *BFIHOST*) and each pooling-group has a target size of 500 AMAX events.

The ratio between the 100-year growth factors (z_{100}) estimated at each of the 602 catchments is shown in Figure 3. It can be seen that the growth factors obtained from the two methods are generally within ±25 per cent of each other. Also, no geographical pattern in the direction of change can readily be observed in Figure 3. The changes in the growth factors can be summarised as follows. Changes in the estimated 100-year growth factors range from ratios of 0.66 to 1.65, with half being greater than 1.00 (25 per cent of the ratios are less than 0.93, and 25 per cent are greater than 1.09). Here a ratio greater than one indicates that the new procedure produces estimates larger than the FEH procedure. These quantitative results indicate that the estimated growth curve shows little change for around half of the catchments.

The estimates of the 100-year flood quantiles obtained using the procedure developed in this study and the FEH procedure are compared in Figure 4. The new procedure (consisting of the revised regression equation for *QMED* and the revised pooling-group procedure) gives estimates of the 100-year flood that are lower than the FEH method in the east of England, but higher estimates in West England, Wales, Scotland and Northern Ireland. More quantitatively, changes in the estimated 100-year floods range from ratios of 0.48 to 2.24, with half being greater than 1.14 (25 per cent of the ratios are less than 0.97 and 25 per cent are greater than 1.32). Here a ratio greater than one indicates that the new procedure produces estimates larger than the FEH procedure.

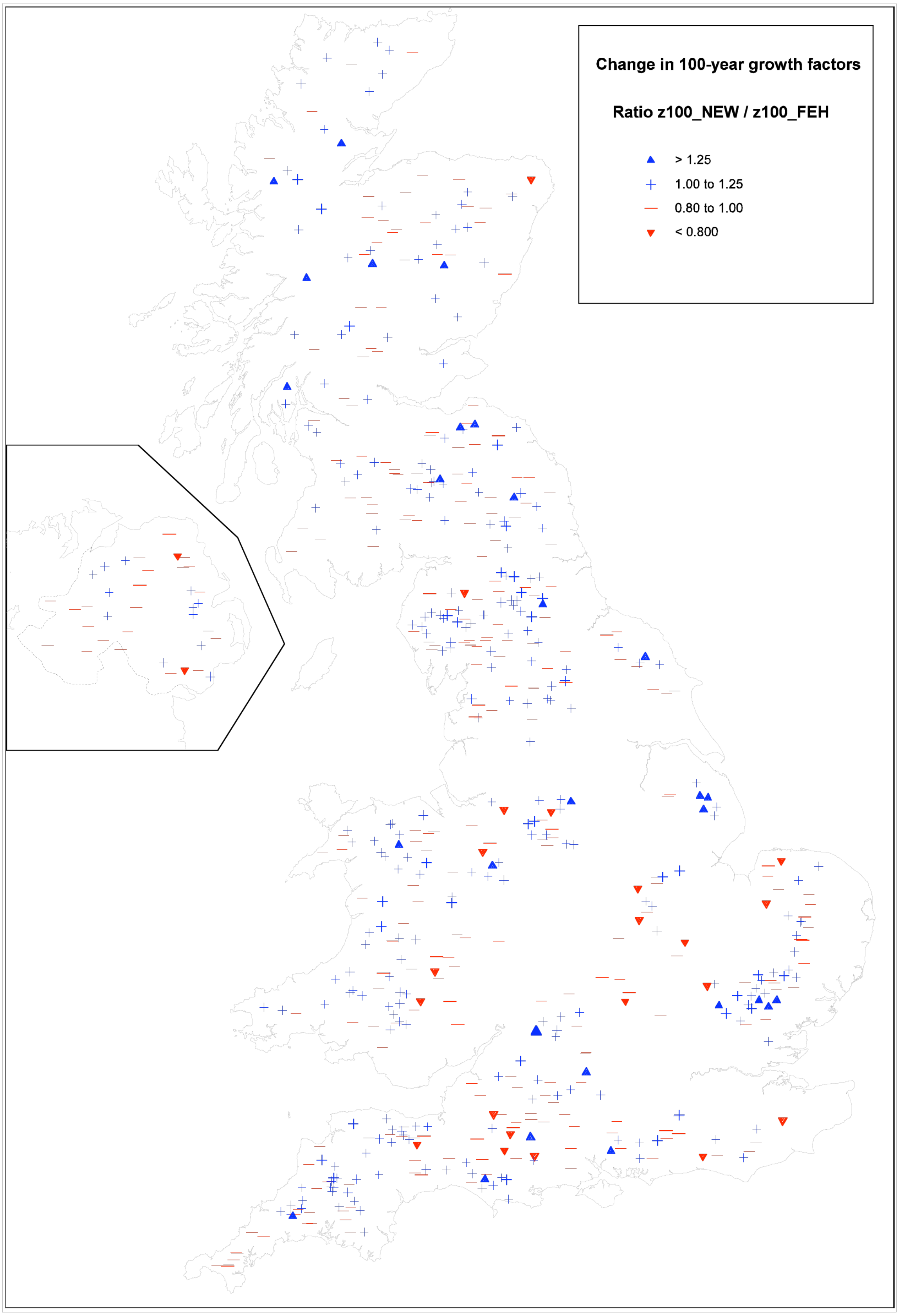

Figure 3 Comparison of growth curve estimates, z_{100}, for ungauged catchments using i) the new pooling method and ii) the FEH pooling method.

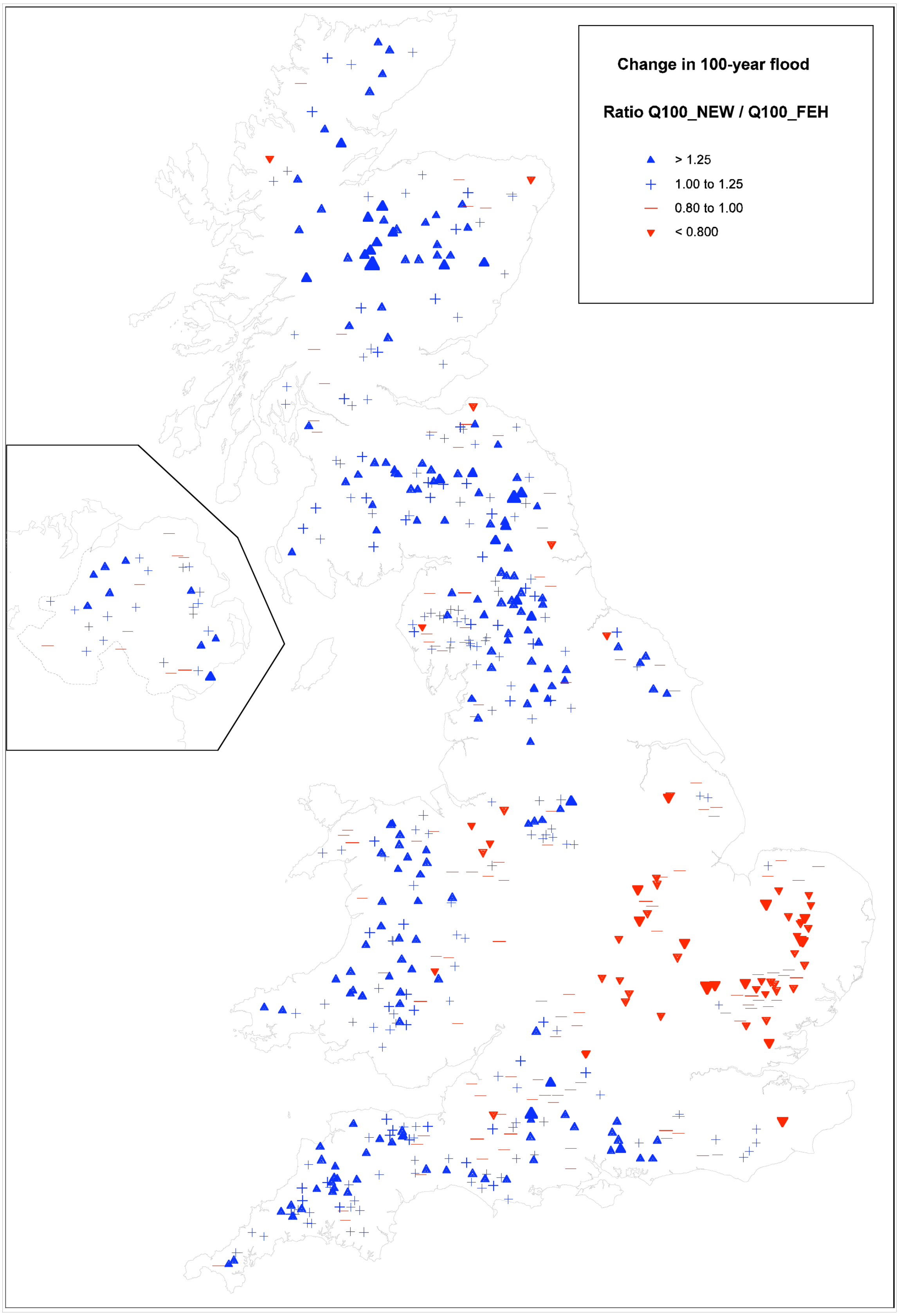

Figure 4 Comparison of Q_{100} estimated for ungauged catchments as the final estimates from i) the new recommendations and ii) the FEH procedure.

IMPLEMENTATION
The Environment Agency intends to issue an Operational Instructional containing guidance on how these new methods will affect how people working in or for the Environment Agency use the FEH methods. Without pre-judging that instruction, it is clear that parts of the new method could be adopted immediately, but others will require further software development. The new QMED equation from catchment descriptors could be applied by hand using existing data. The revised procedure for donor catchments could also be implemented; although the procedure requires the locations of catchment centroids, which are not generally used at present, values are available by querying the selected catchment in version 2 of the FEH CD-ROM (catchment centroids may need to be specified under options>preferences). However, the new procedures to estimate pooled growth curves cannot be applied at present. The method requires values of the floodplain extent descriptor, FPEXT, which are not currently available, and the procedures themselves will need to be implemented in a future update of WINFAP-FEH.

CONCLUSIONS
The research presented in the Science Report for project SC050050 (EA, 2008) and summarised in this paper constitutes an improvement to the existing FEH statistical procedures for flood frequency estimation. The improvements are a result of both i) new modelling techniques and ii) an updated data set (HiFlows-UK). The statistical procedures outlined in the FEH made the region-of-influence approach operational in the UK, which was considered a major achievement and a benchmark for research and development both nationally and internationally. As a result, the new developments introduced in this project build on the foundations laid by the FEH and further improve the reliability of flood frequency estimation in the UK.

However, the scope of this project did not encompass all aspects of the FEH methodology. Also, during the course of the project, particular parts of the methodology were identified where further research and development would be beneficial and potentially provide further improvements of flood frequency estimation in the UK.

HiFlows-UK
Besides simply extending the records at the existing set of catchments included in HiFlows-UK, it is important to consider whether these catchments are sufficiently representative of catchments where flood estimation problems arise in practice. In particular, a view has been expressed by users of FEH methodology that they are often concerned with catchments that are rather smaller than those included in the HiFlows-UK data-set. Future research should pay particular attention to collection of hydrometric data and the performance of FEH methodologies on small catchments.

Urban catchments
While the qualitative effect of increased urbanisation on flood response from a catchment is well understood (increase in the percentage runoff and decrease in the response time) the challenge in applied hydrology is to quantify these effects and to make a generally applicable model. There is an urgent need for more research and collection of data on the hydrological behaviour of urbanised catchments.

Permeable catchments
The new QMED model was found to provide a better description of *QMED* on permeable catchments than the FEH model. For the pooled analysis, the soil term (*BFIHOST*) was removed from the measure of hydrological similarity used for the creation of pooling groups as it was found not to have any significant influence on the growth curves. Thus, the new procedure does not explicitly consider permeable catchments and note that the effects in FEH might have been largely illusory when considering the growth curves. It is recommended that further research should be undertaken to investigate the existence of such multiple mechanisms and, if confirmed, to determine how to incorporate such effects into the current procedures.

Use of donor catchments
A very important improvement to the FEH procedure presented in this report is the revised procedure for estimation of *QMED* using data transfer from a gauged donor catchment to an ungauged subject catchment. While the new donor procedure is an important improvement, it is currently limited by allowing only one potential donor site to be used. Further model development and testing is necessary

to allow more than one donor catchment to be used and to assess the effect of such a methodological extension.
Within the new framework for using donor catchments, project SC050050 initiated work on distinguishing between a donor catchment located on the same river network as the subject catchment and other donor catchments. More work is needed to further classify donor catchments according to location relative to donor catchments before such a system could be made operational. However, it would be an intuitive extension to the framework and could potentially add further improvements to the method.

Acknowledgements
This project has benefited from the improved datasets made available through the Environment Agency Hiflows-UK project.

REFERENCES

Defra and Environment Agency (2005) Revitalisation of the FSR/FEH rainfall-runoff method, Technical Report - FD1913/ TR, Available at http://www.defra.gov.uk/environ/fcd/research/default.htm

Environment Agency (2007) Dissemination of the FSR/FEH rainfall-runoff method, Science Report SC040029 / SR, Available at http://www.defra.gov.uk/environ/fcd/research/default.htm

Environment Agency (2008) Improving the FEH Statistical Index Flood Method and Software, Science Report *SC050050*/SR, Available at http://www.defra.gov.uk/environ/fcd/research/default.htm

Centre for Ecology & Hydrology (2007) FEH CD-ROM Version 2.00. Centre for Ecology & Hydrology, Wallingford, UK.

Institute of Hydrology (1999) Flood Estimation Handbook, 5 Volumes, Institute of Hydrology, Wallingford, UK.

Kjeldsen TR and Jones DA (2007) Estimation of an index flood using data transfer in the UK. Hydrological Sciences Journal, 52(1), 86-98

Natural Environment Research Council (1975) Flood Studies Report, 5 Volumes, Natural Environment Research Council, London, UK.