HIFLOWS-UK

www.environment-agency.gov.uk/hiflows-uk

Peter Spencer (Environment Agency, Warrington), Noel Higginson (Rivers Agency, Belfast),
Tim Palmer (Black & Veatch, Chester), Paul Wass (JBA Consulting, Newcastle)

INTRODUCTION

The flood hydrologist relies heavily on high flow data. Ideally, all data would be of good quality and readily applicable. In practice, it is vital that hydrologists understand the quality and limitations of the data they use. This is particularly important now because the standard method of flood estimation for natural catchments in the UK, the Flood Estimation Handbook (FEH), relies heavily on data. The quality of the data itself and the quality of the information which is used to make decisions about its application are both important.

1. PROJECT OBJECTIVES

This paper describes the HiFlows-UK project to extend and update the flow data and supplementary station information used with the Flood Estimation Handbook. The objectives were to:

- Add recent data and additional gauging stations
- Review data to produce a single authoritative dataset
- Improve background information to assist users of the FEH to make valid decisions
- Make data widely available via the internet
- Establish structures and procedures for future updating and dissemination

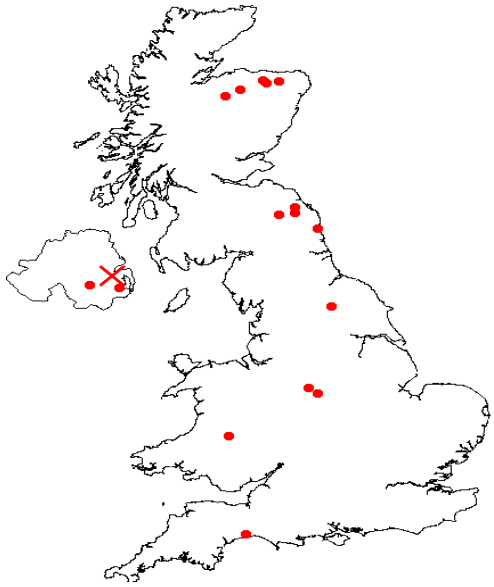

Figure 1 Lagan @ Newforge - location of pooling group stations

2. BACKGROUND

The FEH (1999) provides the primary flood estimation methods in the UK. Although the FEH presents both statistical and rainfall-runoff methods, the statistical method is the main method. The rainfall-runoff method was little changed form the Flood Studies Report (1975) and a "re-vitalised" rainfall-runoff method is the subject of a current R&D project. The statistical method is based on an Index Flood, which is the median annual flood, QMED, and the preferred method for estimating QMED at ungauged sites is by data transfer from a similar gauged catchment. Peak flows of the required probabilities (return periods) are obtained by applying growth factors to the index flood. Growth factors are generally based on statistical analysis of data from a number of similar catchments, termed a "pooling group". A pooling group is composed of similar catchments which can be anywhere in the UK, and Figure 1 illustrates this by showing a pooling group for the 100 year (0.1% annual exceedance probability) for the gauging station on the Lagan at Newforge. Thus both the data itself and users' understanding of the catchments and the gauging stations are important. The FEH

stresses the importance of observed data in arriving at a flood estimate, and five of the six maxims which summarise the FEH's viewpoint on flood estimation are about the use of data.

The FEH and software which implements the statistical method, WINFAP-FEH, provide annual maxima (AMAX) and peaks-over-threshold (POT) flows and catchment descriptors for around 1000 gauging stations on CD. Although the FEH project checked data quality and rejected dubious data, timescales did not allow for extensive data review and updating. Hence, the FEH dataset extends no later than water year 1994/95, and at some currently-open stations only to the early 1980's. At some stations, FEH and gauging authority flows are now quite different because of recent rating reviews by the gauging authorities. The FEH recognised the importance of updating flow data, and suggested a networked database of both level and flow data. FEH users also felt the need for updated data, particularly as the Easter floods of 1998 and the autumn floods of 2000 are not included in FEH dataset.

The HiFlows-UK project was therefore setup, with direct Treasury funding, as a joint project between the three UK gauging authorities - the Environment Agency, the Scottish Environment Protection Agency (SEPA), and the Rivers Agency of the Department of Agriculture and Rural Development, Northern Ireland (DARD). The Department of the Environment and Rural Affairs (DEFRA) and the Centre for Ecology and Hydrology (CEH) Wallingford have supported the project. It has been managed by the Environment Agency.

In order to shorten the timescale and to make good use of available expertise, many Consultants have been involved in the project. JBA Consulting have had a key central technical role, with Black and Veatch providing a project co-ordinator and additional support. Data capture was carried out by Kirk McClure Morton (Northern Ireland), Ba[illegible] (Scotland), JBA (England - North West and North East), Posford Haskoning (Wales and England - Anglian), Mott MacDonald (England - Thames and Midlands), and Symonds (England - Southern and South West). At a later stage these consultants, together with Halcrow, Bullens, and Atkins (supported by Peter Brett Associates and Jacobs) reviewed stations throughout the UK based on their existing knowledge.

3. COLLECTION OF DATA AND INFORMATION

3.1 Setting up the data collection

The project agreed with the gauging authorities the stations to be included, issued procedures and held training workshops. The data capture then took place between December 2002 and August 2003. This was greatly assisted in England and Wales by the Environment Agency making available its interim full digital archive from its Hydrometric Archive Replacement Project (HARP).

The data capture consultants' brief was to collect data and relevant information from gauging authority staff, to obtain available photographs, review and consolidate existing ratings (but not derive new relationships), and write catchment and station descriptions, building upon what was already in the FEH dataset. Digital FEH catchment descriptors were extracted from the 1999 FEH CD-ROM where these did not already exist.

In addition, Martin Lees of CEH Wallingford supplied the project with a scanned copy of all the station files in the National River Flow Archive (NRFA), and Andrew Black of Dundee University supplied a copy of their AMAX and POT series for Scotland.

3.2 Data quality

Ensuring data quality has been challenging. The challenges arose from a number of factors, including the many gauging stations, the number of different offices and their different data archiving systems, the several different sources of data (including the FEH dataset itself, digital, written, and microfiche), the complexities of some rating histories, and the often limited information about past events, particularly before the start of digital records. A frequent source of uncertainty was how some stations really behave at the highest flows, especially where they have been designed and operated primarily

for the measurement of low flows. The data quality procedures have been carried out at two stages. Firstly, by setting up clear and consistent data quality procedures for the data capture phase, and secondly by subsequent review.
Consistent data quality procedures were established for the data capture phase by JBA preparing a database for the Consultants capturing the data to complete, and this also provided tools to visualise and check the data, including:

- Comparison of rating relationships with flow gaugings
- Checking changes in ratings against the history of physical changes
- Time series graphs of the AMAX and POT series
- Trend analysis
- Comparing the HiFlows-UK data with that in the FEH dataset
- Comparing peak flows with other stations on the same river
- Visually examining the flow hydrographs for the largest digitally available 5 peaks

The data collected and ratings applied were reviewed and approved by gauging authority staff. The Consultants then prepared a summary report. All this work was done under the Consultants' own quality systems (including review and approval) before supply to the project.

The next, and very important stage, was a further review of the initial data and information. This was done by releasing the database to selected Consultants and then by making a complete pilot website available for external review. This gave users an opportunity to contribute their knowledge. The project received over 500 "station-person" comments, all which were followed up. In addition, JBA and project staff have carried out further extensive reviews of the data and consultations with hydrometric staff.

4. RESULTS - DATA

4.1 Number of stations and station-years

Although the project included additional stations to those in the FEH, around 200 stations from the original FEH dataset were eventually rejected on data quality grounds. The final total of a little under 1000 stations is similar, though a little less, than the number to the FEH. The project has listed the necessary work (often small compared with the years of effort already put in) for rejected stations to be included in HiFlows-UK in future. The project outputs should help to prioritise future collection and calibration of high flow data.
Overall, the HiFlows-UK website significantly increases the flood peak data available. The mean record length has increased from 23 to 32 years, and the number of station-years of AMAX and POT data has increased as follows:

Number of station-years of record

Data type	FEH dataset	HiFlows-UK	% increase
AMAX	23,000	32,000	40
POT	18,000	26,000	45

4.2 'Indicative Suitability' of stations based on data quality

Although the authors of the FEH were very aware of the problems of data quality, it is not stated as an explicit factor in the selection of stations for QMED and pooling groups, although the effects of poor quality may subsequently lead to a station being removed from a pooling group on statistical grounds.

This project provided a good opportunity to present information on whether the station data are sufficiently reliable to derive QMED, or for the station to be included in a pooling group. For example, a station might be reliably gauged to bankfull and considered suitable for QMED, but out-of-bank flows might be ungauged, the applicability of the rating unknown, and hence the station considered unsuitable for a pooling group. The term 'Indicative Suitability' is used because the decision can be more complex than a simple Yes or No. The choice as to whether to use a station may well be a trade-off between its data quality and other characteristics suitable for a particular

application. For example, a FEH user might chose to estimate QMED using an analogue station where the data are known to have some problems, but where the catchment is highly suitable in terms of catchment descriptors and geographical proximity. Also, stations classed as simply 'Not Suitable' will include borderline stations little worse than some classed as Suitable. An important feature is that the website gives the reasons for the indicative category, allowing users to make their own decisions.

The 'Indicative Suitability' refers only to the data quality. It does not include other requirements in the FEH such as the catchment area being greater than 0.5km^2, the record length greater than seven years, or aspects covered by catchment descriptors such as reservoirs, lakes, or urbanisation. It is important to appreciate this point because although WINFAP-FEH excludes urban catchments, it will include those affected by lakes and reservoirs, and this aspect (as measured by the FARL catchment descriptor) should be checked by users when deriving a pooling group.

Around 85% of the stations in the dataset are indicated as being suitable to derive QMED, and around 60% as being suitable for pooling groups. Only 40% of stations with a catchment area less than 30km^2 are indicated as being suitable for pooling groups, so that small catchments are under-represented.

The FEH rule of thumb is that the stations in a pooling group should collectively supply five times as many years of record as the target return period. Hence, a typical pooling group for the 100 year flood would have 22 stations using the FEH dataset but 16 stations with HiFlows-UK. It is hoped that pooling groups will now be more homogeneous and the results more reliable because the record at each station is longer (so fewer outliers), there are fewer stations in the pooling group (so they should be more similar to the subject site), and the data quality is improved. With improved data and further research, it may be that the number of station-years in pooling groups could be less than five times the design return period.

4.3 HiFlows-UK and Gauging Authority data

Although much has been done to produce a single authoritative dataset, flows within HiFlows-UK are not always the same as those currently held internally by the three gauging authorities. In some cases differences occur because HiFlows-UK has adopted a single rating applicable only to high flows over a period of years where there is no reason to expect a change in the rating, whereas the gauging authorities hold a series of ratings over the same period, which better reflect changes in low flows, without supporting high-flow gaugings. Where differences are large, stations will probably not be indicated as suitable for QMED or pooling groups. The project has identified these differences, and it is intended to reduce them in future years based on further data collection and analyses.

4.4 Examples of data issues

Reviewing a gauging station can be like opening Pandora's box. Perhaps it is best not to look, but once you have looked, many questions can arise.

The project has also had to adopt a pragmatic approach to some broad questions, which are easier to handle when studying one particular station in detail, rather than inn the standardised approach which has had to be adopted by HiFlows-UK. Such questions include (for example):

- Are the criteria for independence of POT data set out in the FEH rigid rules or guidelines?
- For AMAX, when do you not include the highest flow in a year with an incomplete record (and what should you do about it)?
- How to handle duplicate stations (when one station replaces another)?
- How to handle combined stations (where there are separate low and high flows records)?

4.5 Northern Ireland data

The FEH dataset contains peak flow data for 47 gauging stations, selected on the grounds of quality and usefulness, out of about 64 stations which record river flow. HiFlows-UK has used 45 of these stations, removing two because the high flows gauged at the sites are affected by bypassing or diversion. Seven of these 45 stations are now closed. The Indicative Suitability of 38 of the stations is

suitable for QMED, and 21 are suitable for pooling groups. The data capture was carried out by Kirk McClure Morton, with River's Agency staff extracting independent peaks for the POT series.

At twelve of the stations, flows have changed significantly from those in the FEH, mainly caused by rating reviews which have been carried out since the data was originally supplied to the FEH team about ten years ago. The ongoing processes of additional flow gauging and rating review are likely to produce further changes in the future. This is particularly so because at 16 stations existing gaugings do not extend above QMED.

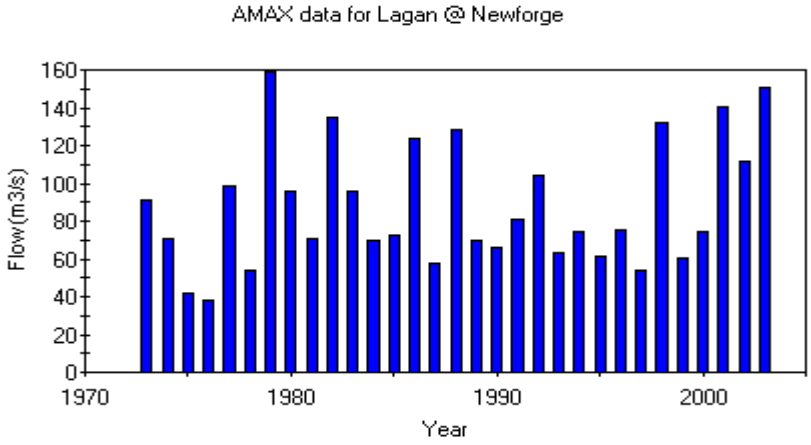

Figure 2

As an example, consider the gauging station on the Lagan at Newforge. The FEH has 21 years of AMAX data from water year 1972/73 to 1992/93, with a QMED of 75.4 m³/s. HiFlows-UK has 31 years of record from 1972/73 to 2002/03, shown in Figure 2, with a QMED of 74.7 m³/s, although individual AMAX values are around 6% less in HiFlows-UK because a different rating equation has been used.

5. THE WEBSITE

www.environment-agency.gov.uk/hiflows-uk

The website is hosted by the Environment Agency with HiFlows-UK branding. The station pages and data are effectively "published" as static pages (rather than being generated on demand). Atkins Management Consultants were responsible for development of the functionality of the website, with the content being extracted from the project database and written in published form by JBA. When the pilot website was issued to interested parties, they were asked to comment on the functionality of the website, as well as on the data and information as previously described. These comments were then acted upon as far as possible.

HiFlows-UK Home Page

The Home Page provides brief welcome text and access to the following pages:

- About -Background information on the HiFlows-UK project –objectives, funding, history, future updating and acknowledgements
- Disclaimer - Notes about data quality etc
- Download - HiFlows-UK data in WINFAP-FEH and CSV formats
- Maps - Maps of the UK showing gauging stations in HiFlows-UK
- Station search - Search for stations by a range of criteria
- FEH Differences - Describes the differences between the data in HiFlows-UK and the FEH
- Links - Links to related websites
- FAQ - Answers Frequently Asked Questions (FAQs)

- Glossary - Glossary of terms used in the website
- Feedback form - For users to feed comments, additional information, etc, to help updates

Download: The website provides a replacement dataset for use with the FEH by download in either WINFAP-FEH or .csv formats. The data includes both flows and levels. The file size is around 7MB, so that download times are relatively short.

Maps: The initial map shows the whole of the UK with the internal and external boundaries of the gauging authorities. On clicking one of the regions, users are then presented with a regional map which marks the location of all stations within the region, and on hovering over a mark the station and watercourse name are shown. Clicking on this mark brings up the Station Home Page.

Search page: The search page enables users to search for a station by name, watercourse, hydrometric area, NRFA number, text in the catchment description, etc. Part-completion of a name will bring up all matching stations. Users can also search within a range of catchment descriptors. Thus a search for all stations with catchment area between 5 and 10km^2, Average Annual Rainfall (SAAR) greater than 1000mm, and "Millstone Grit" in the Catchment Description finds Burbage Brook@Burbage Moor. Where several stations fit the criteria, selecting one of them brings up its Station Home Page.

Links page: The links page includes, amongst others, the Rivers Agency of Northern Ireland and hydrometric data for the Republic of Ireland presented on the website of the Office of Public Works.

FAQ and Glossary pages: Although the website assumes some knowledge of the FEH, these pages provide information about FEH and other terms used in the website, including technical aspects such as hydrometric calendars and rating curves.

Station Home Pages
These pages provide station details for FEH users and are reached through the search or map pages. The central panel contains core information such as station type, station and catchment descriptions, catchment area, start and end dates, and indicative suitability for QMED and pooling. Child pages provide photographs, AMAX and POT data, ratings, and summaries of station history and missing data. The AMAX and POT data are in both tabular form and a time series plot, which gives an instant visual display of the range of peak flows and years. The rating curve plot also shows gauged values and other information such as banktop level.

6. THE FUTURE

An important aspect of the project was to set-up both structures and procedures for future updating. The data will be updated annually for the preceding water year to the end of September. CEH Wallingford will be responsible for the long-term maintenance of the data on the site, in a similar role to that at present under the Surface and Groundwater Archive (SAGA). The gauging authorities will provide annual updates to CEH Wallingford who will then review this additional data for consistency with earlier data, and agree changes with the gauging authorities where necessary. Procedures have been agreed to add or remove stations.

7. BENEFITS

The advantages of the data itself may be summarised:
- Increased quantity of flood data, giving better representativeness & confidence
- Increased quality of flow data, giving more accurate flood estimates
- More meta data, increasing the confidence of hydrologists in the data they using, and hopefully leading to a more intelligent use of data

- Consistency among datasets, ensuring more compatible flood estimates between different hydrologists and organisations
- Efficiency savings in supplying and using data
- Better flood risk decisions (e.g. on flood alleviation schemes)

Although the dataset has been collated solely for the purpose of flood estimation, it may prove to be useful in other applications. The general information, photographs, catchment descriptors, etc, which are provided about each station may themselves be useful.

One possible example, which draws on both the data and the information, is trend analysis. Once a trend has been identified it is necessary to consider whether the trend is caused by external factors such as change in land use, climate change, reservoir construction, etc, or by physical changes at the station or in the data processing (e.g. rating curves), and the station and catchment information in HiFlows-UK may help to clarify this. Figure 3 shows an apparent (though low confidence) trend at two gauging stations on the River Duddon in Cumbria.

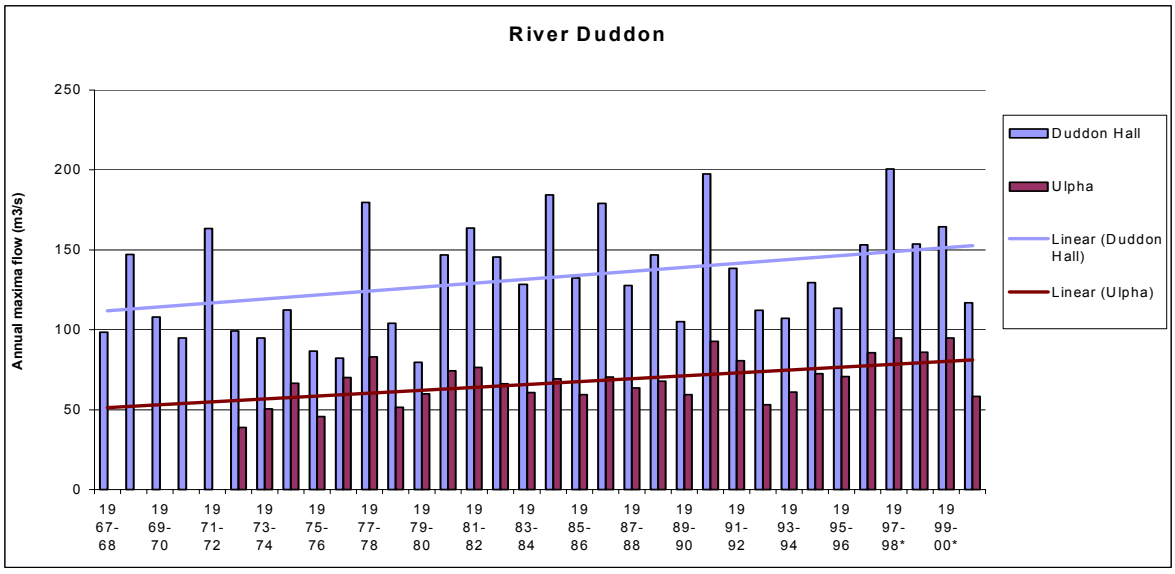

Figure 3

Much has been said in recent years about taking a catchment or basin--wide approach, as in the Environment Agency's programme of Catchment Flood Management Plans (CFMPs), and the Water Framework Directive's river basin characterisation. The HiFlows-UK website should be directly applicable to the CFMPs, and may also prove useful in other studies, particularly where it contains a number of gauging stations in the river basin being studied.

The authors hope that the HiFlows-UK website will lead to more robust flood estimates, with resultant improvements in flood alleviation schemes, flood maps, and other applications. We also hope that such a readily available data repository will lead to a shared understanding of high flow data and its importance between data collectors and data users. It is important that there is continued commitment to future updating of the HiFlows-UK dataset, and to increasing the number catchment types which are poorly represented..

Summary of the most important new features

HiFlows-UK will provide users with:

- AMAX & POT as tables of levels as well as flows, with a time series graph
- An audit trail provided by station datums and rating histories, levels and rating curves

- Core information provided as station and catchment descriptions as text, with station photographs, and indicative suitabilities.
- The ability to search for a station not only by name, number, river, or catchment, but also by user-defined features

Implications for FEH users

- the website should be as much a part of FEH user's toolkit as the Handbook or the software
- new data will give different results
- results should be more robust, and users should have more confidence in them

Acknowledgement
The authors would also like to acknowledge the efforts of, in Northern Ireland, Derrick Pinkerton, Data Manager, Rivers Agency, and Kirk McClure Morton (who were responsible for the data capture in Northern Ireland) and in the rest of the UK, the Environment Agency, SEPA and the consultants' staff who participated in the data capture and other aspects of the project. We are also grateful for the feedback received.