River Gauging Station Data Quality Classification (GSDQ)

R&D Technical Report W6-058/TR

River Gauging Station Data Quality Classification (GSDQ)

R&D Technical Report W6-058/TR

Lamb, R., Zaidman, M.D., Archer, D.R.[1]
Marsh, T.J., Lees, M.L.[2]

Research Contractors:
JBA Consulting – Engineers & Scientists[1]
Centre for Ecology & Hydrology (CEH) Wallingford[2]

Publishing Organisation
Environment Agency, Rio House, Waterside Drive, Aztec West, Almondsbury
Bristol BS32 4UD
Tel: 01454 624400 Fax: 01454 624409
Website: www.environment-agency.gov.uk

ISBN 1 844 32256 4

© Environment Agency 2003

All rights reserved. No part of this document may be produced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without the prior permission of the Environment Agency.

The views expressed in this document are not necessarily those of the Environment Agency. Its officers, servants or agents accept no liability whatsoever for any loss or damage arising from the interpretation or use of the information, or reliance upon the views contained herein.

Dissemination Status
Internal: Released to Regions
External: Publicly Available

Statement of Use
Environment Agency hydrometry/hydrology staff will be required to implement this R&D after receiving training in the use of the Gauging Station Data Classification tool. Guidance will be provided by Policy regarding prioritisation of this work. The tool may be used to set gauging station data targets and to provide data users with information on data quality.

It is anticipated that other agencies responsible for hydrometric networks may also use the classification system as described above.

Research Contractor
This document was produced under R&D Project W6-058 by :
JBA Consulting – Engineers & Scientists, South Barn, Broughton Hall, Skipton, North Yorkshire, BD23 3AE UK
Tel : +44 (0)1756 799919 Fax : +44 (0)1756 799449 Website:www.jbaconsulting.co.uk

Environment Agency Project Manager
The Environment Agency's Project Manager for R&D Project W6-058 was
Dave Stewart, Phoenix House, Global Avenue, Leeds, W. Yorkshire, LS11 8PG.

Further copies of this report are available from:
Environment Agency R&D Dissemination Centre
WRc, Frankland Road, Swindon, Wilts. SN5 8YF
Tel: 01793 865000 Fax: 01793 514562 E-mail: publications@wrcplc.co.uk

FOREWORD

The authors would like to thank the Environment Agency Project Board for their valuable management and technical input throughout this R&D project. We gratefully acknowledge the Environment Agency staff who provided data and carried out software tests and benchmarking of the Gauging Station Data Quality Classification.

We would also like to thank the individuals and organisations listed in the main report who kindly gave their time to respond to consultation during the project.

EXECUTIVE SUMMARY

One of the most commonly asked questions about river flow (or level) data is ‘How good are the data?’. This generic question is asked by almost all data users, whether consultants working on flood or water resources studies, scientists undertaking research, ‘strategic’ data users, such as government departments, the public or media, and environmental managers. It is also a question asked by hydrometrists themselves. There are published sources of information on data quality, such as the station notes in the UK Hydrometric Register produced by CEH Wallingford, yet few formal schemes exist in the UK or elsewhere to provide comparative measures of data quality, or to serve as performance targets for hydrometric agencies.

Perhaps the main reason for this is that there is a very wide range of factors that can influence the quality of data at a flow or level gauging station. Not all factors are equally important, and the influences on quality will vary according to the type of gauging station and flow regime. Consequently, reports of data quality tend to be either rather subjective, or to comprise basic facts about the characteristics of a station from which quality could be assessed, given knowledge of the hydrometric methods. Neither approach allows for an easy, broad-brush comparison between stations.

The Environment Agency therefore commissioned R&D project W6-058 to fill this gap by developing a new method for representing the quality of gauging station data. A consortium of JBA Consulting Engineers & Scientists and CEH Wallingford were appointed to carry out the work.

The project comprised the following elements –

- An extensive consultation (both in the UK and overseas) with hydrometrists, data analysts, hydrologists and water managers to determine the requirement for a repeatable, empirical scheme for representing data quality.
- A review of existing approaches, both in the UK and overseas.
- The identification of factors that influence gauging station data quality.
- The development of a Gauging Station Data Quality (GSDQ) classification scheme based on attribute scoring.
- Implementation of the GSDQ classification in a software tool.
- Provision of an R&D Technical Report, software user guide and training materials.

This is the main R&D Technical Report for the project.

The GSDQ classification is described in detail in this report, which aims to explain the concepts underlying the scheme and the reasons why the attribute scoring approach was adopted. Attributes (specific factors that influence data quality) are defined in detail, and a software implementation of the data quality classification is described. The GSDQ classification encompasses the following main types of gauging station:

- Rated sections
- Structures (built and maintained to British Standard)
- Structures (non-standard)
- Ultrasonic (transit time)

- Electromagnetic (buried coil)
- Level-only

The classification provides a quantitative, repeatable and objective measure of data quality that still aims to be flexible and general enough to cope with the very wide range of circumstances that can occur at gauging stations operated by the Environment Agency. It includes statistically-based estimates of uncertainty in flow measurement, derived from current British/International Standards where possible, quantitative attributes, such as the number and deviations of check gaugings, and categorical attributes such as assessments of the significance of by-passing or weed growth.

Basic station information (including ratings, flow gaugings, station dimensions etc.) can be entered and stored in the GSDQ software. The software itself is a customised Microsoft Excel spreadsheet application. The GSDQ spreadsheet calculates all the required attribute values from basic inputs and returns a classification score. This is a number between 0.0 and 1.0, where 1.0 indicates best quality. The numerical score is also sub-divided in three classes, CAUTION, FAIR and GOOD.

As part of this R&D project, a benchmarking exercise was carried out to test whether the GSDQ classification met with the expectations of hydrometry officers and data users. The results indicate broad support for the classification.

CONTENTS

LIST OF TABLES

page:

LIST OF FIGURES

1 INTRODUCTION

1.1 Background

The measurement of river flows and levels is an essential part of the business of the Environment Agency (EA). The data sets are an important resource, both within the EA and externally, and are used in part to fulfil statutory requirements under the Water Resources and Environment Acts. The EA operates an extensive network of river gauging stations to monitor flows and levels and has developed consistent operational standards for hydrometry. Despite the high standards that are set, it is acknowledged that the quality of gauging station data can vary. Users and suppliers of the data need information about its quality, but there has not been a complete uptake of earlier approaches to classify data quality.

The Environment Agency therefore established R&D Project W6-058 *Identification Of A Method For Representing The Quality Of Gauging Station Data* to address this need. The overall aim of the project was:

> *'to provide the Agency with a review of the current gauging station classification procedure and to provide a revised procedure which is both statistically robust and easily understood'.*

The EA appointed a consortium of JBA Consulting - Engineers & Scientists and The Centre for Ecology and Hydrology (CEH) Wallingford to carry out the work. The project commenced in late July 2002.

1.2 Quality of Gauging Station Data in the UK

Gauging station data is used for a wide variety of purposes, including flood forecasting, water resources planning and design, hydrological research, operation of hydraulic structures and setting of abstraction consents. Equally, gauging station data is used (both directly and indirectly) by a wide variety of organisations, including water utilities, government departments, district councils and other public organisations, private consultants, insurance companies, academia and research consortia, as well as the Environment Agency itself.

It is important that users are aware of any limitations and uncertainties associated with hydrometric data. Influences on the quality of gauging station data can include reliability and accuracy of the stage recorder, errors associated with the stage-discharge relationship used at a site, and truncated or misleading records caused by out-of-bank flows or weed growth, for example. The method of gauging generally determines which sources of uncertainty are likely to be influential at particular sites.

Provision of information regarding the quality of hydrometric data is within the remit of the EA's Hydrometric Service. The EA is usually aware of the main factors influencing data quality at particular sites through routine site visits, maintenance visits and calibration surveys, and through their efforts to maintain good practice and quality assurance standards (particularly ISO 9002). Some information about gauging station characteristics and data quality (particularly related to the flow record) is also held by the National Water Archive (NWA), maintained by CEH Wallingford, and is summarised in the UK Hydrometric Register[1], which is presently published every fifth year.

It is often very difficult to reconcile the disparate sources of uncertainty and to translate them into a tangible and quantifiable understanding of data quality. As a consequence there can be a tendency for users either to accept hydrometric data at face value or to ignore the impact of data uncertainty. It is also difficult to make broad comparative assessments of data quality. One solution is to employ a formal classification scheme to categorise stations according to their overall data quality and to present this to the user in summarised form.

A classification was developed for this purpose by the National Hydrometric Group of the then National Rivers Authority (NRA) in 1995[2]. In this scheme, stations were classified primarily according to statistics derived from flow gaugings where flows were computed from a rating equation, on the theoretical error where a structure was used, or on the accuracy of measurement where a level recorder was used. In each case the 'performance of the gauge' was assessed for the high flow range (mean annual flood), medium flow range (average daily flow) and low flows range (Q95 flow). However, for various reasons, the scheme has not been applied rigorously across all regions of the Agency, and as a result the classification results have not been widely disseminated to end-users within the scientific and water communities. Experience gained from the '1995 NRA classification' has highlighted the difficulty in describing the overall quality of data in a succinct and objective manner and emphasised the need for an objective and repeatable method for representing gauging station data quality.

1.3 Project Objectives

The lessons learnt from the 1995 NRA classification provided a starting point for the development of a revised scheme. The 1995 classification was perceived to have a probable mathematical bias towards current meter gauging. An important issue was therefore how to include aspects of gauging performance other than statistics derived from flow gaugings in the new quality classification procedure. The 1995 classification was also perceived to be overly complex. This is not really the case (further discussion is presented in Section 2, and the new scheme developed in this project is, in fact, necessarily more complex). However, the perception illustrates the need for the principles of the classification to be transparent to users, for the classification procedures to be easy to implement and for the results (i.e. classes assigned) to be simple to interpret by end-users. The overarching aim of this project was therefore that development of the revised classification should employ straightforward, practical procedures, based on sound analysis and backed up by software and training.

The terms of reference for the project specified that the new method should be appropriate for the following different types of gauging station:

---

[1] CEH Wallingford, 2003, *Hydrological Data UK: Hydrometric register and statistics 1996-2000*, ISBN 1 903741 06 8.

[2] National Rivers Authority. 1995. Gauging Station Classification. Guidance on the method and application of river gauging station classification system. Report of the National Hydrometric Group. August 1995. 20pp.

- Rated-section (including open-channel and non-standard structures),
- Structures (i.e. weirs and flumes) meeting British Standard specification,
- Transit-time ultrasonic river flow gauges,
- Electromagnetic river flow gauges,
- Sites monitored using a level-recorder only.

Further specific requirements for the revised classification were as follows:

- The data quality classification should be straightforward to understand and use,
- The procedure should have a sound statistical basis,
- The R&D outputs will include an Excel software tool, but the methods should also be capable of being implemented efficiently in other software environments (such as the new WISKI hydrometric database currently being implemented within the Agency through the HARP programme),
- The R&D should link appropriately into a new gauging station rating training course being developed as an addition to existing national hydrometry training courses.

Whilst it seeks to lead best environmental practice in the UK, the Agency, as a public body, must also be able to demonstrate value for money in the services it procures. The issues raised concerning measures for gauge data quality therefore had to be balanced with the need for a procedure that was cost-effective to develop and that could be delivered entirely within the scope of this R&D Project. It was therefore intended that the new scheme should help demonstrate the delivery of the Agency's hydrometric service to users of the data.

1.4 Project Management and Programme

The Agency's project manager for this contract was Dave Stewart (Head Office), based at the Ridings Area Office, Phoenix House, Leeds. The project board was Dave Stewart, Luci Allen (Midlands), David Brown (Southern, now North West), Alison Hanson (North West) and Will Lidbetter (Thames).

The project programme was arranged into eight main tasks, which are set out in Table 1.1.

Table 1.1: Project Tasks

Task 1	To undertake a review and investigation of methods used currently by the Agency and by other hydrometric agencies worldwide.
Task 2	To undertake a survey of Agency customers, internal and external, to assess what information is required about data quality.
Task 3	To provide a revised procedure which will: o provide a statistically sound method for measuring accuracy at all types of surface water flow and level sites. o ensure that such methods are consistent between different types of sites. o provide a classification descriptor, which includes a measure of reliability at a site. o provide a classification in an easily understood and transparent format.
Task 4	To present worked examples using the new procedure from a wide range of types of hydrometric sites.
Task 5	To produce a manual describing the application of the new procedure. This will include flow charts or other suitable means of allowing the procedure to be easily programmable for future inclusion into Agency hydrometric software.
Task 6	To produce an Excel spreadsheet application in order to automate the procedure.
Task 7	To provide a training input for station classification to be included in an existing gauging station rating development course (currently under development).
Task 8	To provide specific training in the new classification procedure to any Agency staff who may have undertaken a gauging station rating training course without the inclusion of this R&D output.

1.5 Report Structure

The review (Task 1) is presented in Section 2 of this report. This includes an appraisal of the 1995 classification currently in use in the Agency, and a review of the approaches adopted in other countries. Advantages and disadvantages of possible methods are considered.

User requirements (Task 2) were assessed during a workshop held in York on 26 September 2002 with 11 attendees from the Agency, academia and consultancy. The views of 22 users or providers of hydrometric data were also canvassed via questionnaire and telephone interviews. A number of initial options for the gauging station data quality classification were discussed at the workshop. Taking account of the strengths, weaknesses and general comments made about each option, the participants agreed that a classification system based on an attribute scoring method should be adopted. The review of user requirements is discussed in Sections 3 and 4 of this report.

Tasks 3 to 6 were conducted in parallel during the period from October 2002 to June 2003. Following the York workshop, a prototype classification was designed and implemented as an Excel spreadsheet tool. Whilst this prototype version did not include full automation of the procedure, it illustrated the key features of the scheme, and was endorsed by the project board. Further development of the classification and software tool was carried out, with guidance from the project board and reference to worked examples. A fully automated Excel spreadsheet tool was developed by April 2003 and further modifications have taken place following initial testing.

The general principles and features of the finalised classification scheme are described in Section 5 of this report, whilst implementation of the classification for different gauging station types is discussed in Section 6.

Section 7 describes in detail the individual attributes considered within the classification and the benchmarking procedure respectively.

Section 8 describes the design and implementation of the Excel spreadsheet tool. There is also a software user guide, produced as a separate report, which includes step-by-step instructions for use of the Excel tool and a number of worked examples.

A 'benchmarking' process was added to the project to test and fine-tune the data quality scheme using data from real stations. This was conducted during the period June to August 2003 and is reported in Section 9.

Some guidance on using and maintaining the classification is given in Section 10. Finally, a number of recommendations for future research and actions to promote uptake of the classification are discussed in Section 11.

2 REVIEW OF EXISTING DATA QUALITY MEASURES

2.1 Introduction

This section reviews approaches to measuring or reporting the quality of gauging station data that have been used in the UK and elsewhere. It is based on a review of relevant literature and direct communication with hydrometric agencies and national archiving services. The review begins with a discussion of particular features of the England and Wales hydrometric network, and their implications for data quality.

2.2 The England and Wales Gauging Station Network

In global terms England and Wales has a very dense gauging station network, currently comprising around 900 primary stations augmented by a substantial number of secondary and temporary monitoring sites. The number and disposition of stations is a necessary response to the drainage network (a multiplicity of mostly small basins) and to the diversity of England and Wales in terms of its climate, topography, geology, land use and patterns of water utilisation.

The England and Wales network is also very distinctive with regard to the variety of gauging stations deployed. Simple river sections, by far the dominant category globally, comprise less 30% of the overall network; this is a unique distinction at the national scale. The network includes over 600 gauging structures (embracing many different designs and configurations) reflecting the modest size of most rivers and grant-aid provision in the 1960s and 1970s. Most purpose-built gauging structures have proved robust and reliable, and often capable of successful operation outside their design ranges (normally based on tank tests). 'New technology' stations have been increasingly deployed, particularly over the last 15 years. There are currently more than 60 ultrasonic gauging stations in operation, with seven on the Thames alone. The majority of gauging stations are 'hybrid' (exploiting different measurement techniques for different flow ranges) and a significant minority are multi-site (e.g. multiple channel, high and low flow component sites), requiring more complex level-to-flow data processing arrangements.

This great diversity and complexity virtually precludes a comprehensive, objective and yet simple data quality measure from being developed. However, some general comments can probably be made without risking too much controversy. Accuracy bands that characterise the medium flow ranges can seldom be approached in the extreme flow ranges. Under low flow conditions, limited water depth places a premium on reliable water level sensing and recording, whilst at bankfull and above, stage-discharge relations are often uncertain.

Very broad guides to the accuracy considered attainable in the average flow range using different flow measurement techniques have been published (see Table 2.1). These may assume gauging station operation to BS/ISO standards but are, in reality, a basis for discussion and general comparisons only and do not reflect issues of reliability and stability that affect continuous measurements. Much depends on local circumstances.

Procedures exist to develop site-specific assessments of gauging station accuracy but their complexity (and data demands) preclude general application.

Internationally, and within the UK, the Standard error of the mean relation (SMR) is widely used to index the accuracy of river flow data (see Section 7 for the definition and discussion of SMR). This statistic is most valuable at gauging stations where all flows are contained, hydraulic conditions are relatively stable and sufficient gaugings have been completed to fully characterise the rating.

Table 2.1: Attainable uncertainties in a single measurement of discharge

Method	**Percentage uncertainty (at 95% confidence level)** *plus or minus*
Current meter measurement	5
Floats	10-20
Slope-area	10-20
Fall-discharge	10-20
Dilution techniques	5
Thin plate weir	2
Thin plate V-notch	2
Triangular profile (Crump) weir	5
Flat V weir	5
Rectangular profile weir	5
Round nosed weir	5
Flumes	5
Moving boat	5
Ultrasonic	5
Electromagnetic	5-10

Notes:
Source: R.W. Herschy (1995) *Streamflow Measurement*, p. 486. Values may not apply to extremes of flow.

The performance characteristics of most gauging structures used in England and Wales have been thoroughly investigated, both in the laboratory and the field. However, many factors can combine to introduce variations in the accuracy of measured flows (e.g. survey or datum errors, algae on weir crests, accretion on upstream aprons, drowning of structures). As a consequence, current-metering at structures has been established practice over many years, generally to confirm the continuing applicability of the appropriate theoretical/laboratory based rating.

There has been a tendency to undertake many more current meter gaugings at structures in the recent past, partly stimulated by a perceived need to quantify accuracy bands more explicitly. The potential dangers of this approach have been illustrated at a number of stations where gauging-based calibrations have superseded the laboratory-based rating, notwithstanding the inherently greater accuracy of the structure compared to gaugings (particularly those based on single-depth metering).

Generally of most significance in relation to the ability of a gauging station to furnish accurate river flow data is the limited water depth in UK streams and rivers; stage values corresponding to low flows are commonly less than 100 mm, often much less. The conventional 15-minute recording interval implies that random errors in computed mean daily flows tend to be very low. By contrast, systematic bias in measured river levels (caused, for example, by algal growth on weir crests or datum errors) can be substantial and difficult to eliminate.

The most vulnerable gauging stations are those where small head changes correspond to substantial changes in flow, i.e. those stations with insensitive controls. Figure 2.1 shows the change in river flow associated with a 5 mm change in stage at the Q95 flow, based on over 1000 gauging stations throughout the UK. A systematic error of 5 mm translates into a 10% flow error for more than 35% of the gauging station network. Stations in the English Lowlands are disproportionately represented in the higher error bands; a 5 mm stage error corresponding to a flow change of 15% or more at almost a third of the gauging stations.

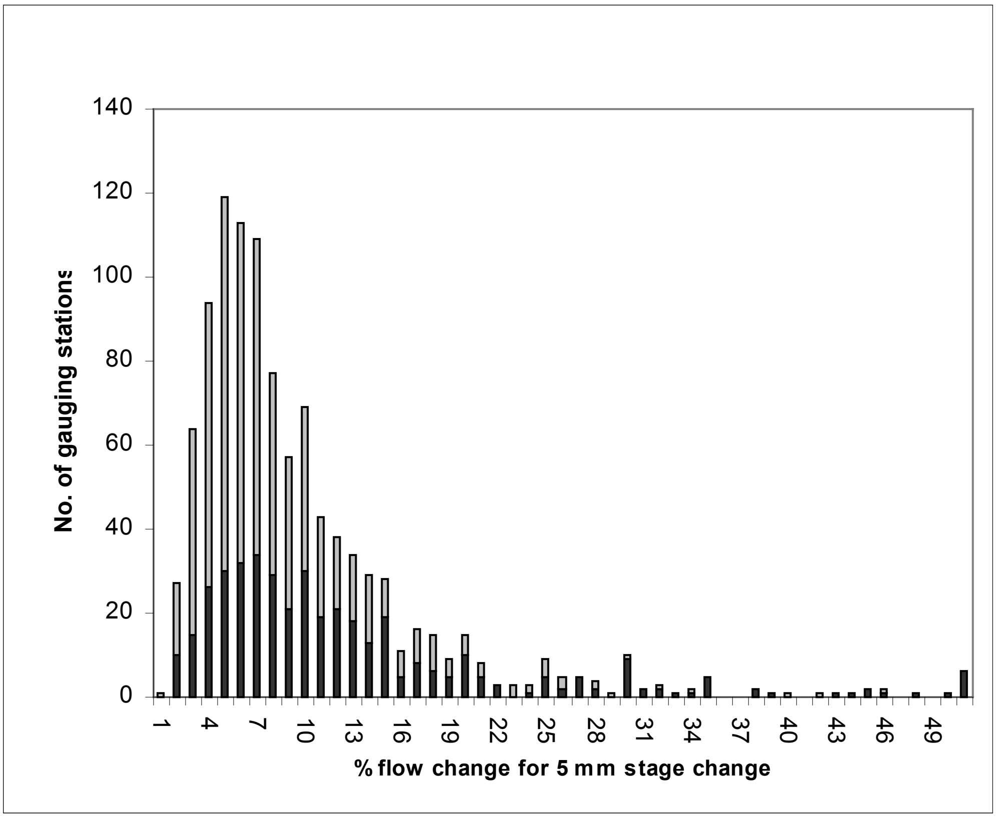


Figure 2.1: The sensitivity of UK gauging stations at Q95

Notes: Darker shading indicates stations in the English lowlands. (After Marsh, T.J. 2002. Capitalising on river flow data to meet changing national needs – a UK perspective. *Flow Measurement and Instrumentation*, 13, 291-298).

A salutary exercise is to compute 'Associated Accuracy', i.e. the precision to which stage needs to be measured to ensure that the change in flow (due to this cause alone) is less than 5%. Whilst systematic errors are most significant at low flows, they can be influential across the flow range e.g. due to a lack of adjustment to computed flows to account for weir operation in the non-modular range or through the excessive extrapolation of stage-discharge relations. Instrumentation and data-processing procedures exist to address most of these problems but, with many competing demands on the time of hydrometric personnel, their application is patchy both spatially and through time; this can significantly impact on the homogeneity of river flow time series, high flows in particular.

2.3 The National Rivers Authority 1995 Classification

A classification for data quality was developed by the National Hydrometric Group of the then National Rivers Authority (NRA) in 1995[3]. This scheme was based on calculating the value of a performance descriptor for 'low', 'medium' and 'high' categories of flow/levels (corresponding to Q95, average daily flow and mean annual flood flows respectively for flow gauges, and 90%, 50% and 10% of the range respectively for level gauges). This was coupled with a measure of gauge reliability (i.e. percentage data capture).

Several methods of determining the performance descriptor were used in order to reflect the inherent differences in data quality for different types of gauging station. Where flows were computed from a rating equation, the performance descriptor was derived empirically based on statistics derived from flow gaugings. The theoretical 'as built' error was used for BS/ISO compliant structures, and also for non-compliant structures having no confirmatory check gaugings. The accuracy of measurement was used where a level recorder was used. Quality codes were then assigned based upon the value of the performance descriptor and the type of gauged considered. For example if measurement accuracy of a level gauge was within 2mm it would be assigned an 'S1' quality code, if between 2-5mm it would be assigned a quality code of 'S2', if between 5-10mm a quality code of 'S3' and so on.

Quality codes for each flow range were combined to provide a classification for each station in the following format

Date from - Date to, Lx(y), Mx(y), Hx(y), Rw.

where L, M and H represented the chosen indicators for low, medium and high flow/levels recorded at a gauging station, x represented the performance quality code, and y was a code representing the method of flow measurement.

For example, a code of the form

0389-0492, LF1(OC), MF3(OC), HS1(LV), R1

indicated a classification applicable from March 1989 to April 1992, for a station comprising an open channel (OC) rated section at low and medium flows, and reverting to a level only site for high flows. The data quality is good at low flows (class F1), fairly poor at medium flows (class F3), whilst level measurement is very accurate (class S1). The reliability of the gauge is class 'R1' which indicates a 98% data capture rate in this case.

Classifications were time banded so that a series of classifications might exist over the length of the record. For example

[3] National Rivers Authority. 1995. Gauging Station Classification. Guidance on the method and application of river gauging station classification system. Report of the National Hydrometric Group. August 1995. 20pp.

0389-0492, LF1(OC), MF3(OC), HS1(LV), R1

0592-0602, LF1(OC), MF1(OC), HF3(OC), R1

indicates an improvement in data quality after May 1992 (in this case due to the installation of a cableway at the gauging site enabling the rating curve to be extended over the high flows range).

2.4 UK Gauging Station Appraisal Schemes

The practical (and conceptual) difficulties of ascribing specific error bands to river flow data provided a stimulus for a number of national assessments of gauging station performance (including appraisals of the quality/representativeness of the associated datasets). The most notable grading exercises have been undertaken as part of major national research programmes. The aim has normally been to identify broad categories of stations designed to match the needs of the project – in most cases these were similar to those of a much wider user community. SMR was used in the grading exercises, but as part of a wider appraisal programme incorporating station, river or catchment characteristics that may be expected to impinge on the quality or utility of the river flow data. Within the National River Flow Archive at CEH Wallingford such information (e.g. bankfull flow, station sensitivity, and Factors Affecting Runoff) is augmented by concise descriptive material relating to the gauging station's hydrometric performance, flow record, and catchment. It is important to recognise that assessments of accuracy, important though they are, are but one element in the mix of information required by the users of river flow data.

Flood Studies Report Gauging station categorisation

The Flood Studies Report (FSR)[4] review of stations was comprehensive; virtually all the 1150 stations in the United Kingdom were visited. Personnel were either Principal or Senior Hydrologist/Engineer level. There was significant emphasis on discussion with local personnel regarding flood characteristics and out of bank inundation. The rating history was scrutinised and an optimum condensed sequence of ratings established with their dates of currency. Although the rating quality was characterised by a simple letter code, the full station appraisal had four components:

A gauging station form summarising the significant elements for floods interest, typically:

1. A gauging station form summarising the significant elements for floods interest, typically:

    - Basic reference material,
    - Highest peak recorded,
    - Type of station,
    - Current metering in high flow range,
    - 100% gauged? If not bypassing details,
    - Estimation techniques outside highest gauging/structure full,

[4] Natural Environment Research Council (NERC), 1975, *Flood Studies Report*, London

- Plans and channel sections,
- Recorder details and limitations (if any),
- Backwater or other relevant control variation,
- Other levels or flows at or near the site - including nested or neighbouring catchments.

2. A stage discharge rating quality assessment (see Table 2.2).

3. Rating curve log-log plot with relevant details, such as:

   - Plotted segment(s) with associated equation(s),
   - Highest or higher group of recorded gaugings,
   - Estimate of the mean annual flood,
   - Maximum recorded flood,
   - Structure full and/or bankfull stage.

4. Written description of the station, highlighting hydrometric characteristics such as:

   - Quality and features of the chart record,
   - General hydrograph character,
   - Catchment features (including predominant geology, significant storages or diversions),
   - Peaks-over-Threshold (PoT) threshold stage and time to peak.

No such nationwide survey has been carried out since the FSR. Other contractors may have carried out regional surveys (e.g. the Anglian Region Asset Survey by Hydraulics Research Ltd that included appraisal of the flood ratings) and the EA have local arrangements for asset survey and station appraisal, but results may not be available in a standardised form. The National River Flow Archive over many years has visited stations and discussed their features with measuring authority staff. Findings are summarised in the thumbnail station and catchment descriptions available with retrievals, in NRFA publications and from the Website (http://www.nwl.ac.uk/ih/nrfa/index.htm).

Table 2.2: The Flood Studies Report station stage-discharge relation grading criteria

Grade	River section	Gauging structure	Other structures
A1	Rating well defined by current meter	Rating in modular range and within design limits and specifications	Weir in good condition and rated by current meter or careful modelling
A2	Rating less well defined	Rating in non-modular range using two recorders	Weir in good condition rated by credible formula
B	Valid extrapolation of a valid A grade rating to level where cross section geometry and flow conditions change	Non-modular range with one recorder. Extrapolation as for river section	Weir in poor condition. Excessive silting in the channel. Weir submerged. Extrapolation as for river section
C	Further extrapolation of B grade rating beyond channel conditions characteristic of base rating. Limited to an increase in width equal to main channel width. Upgrade to B if indirect measurements in this range have been taken.	Extrapolation of structure rating beyond structure capacity. Limit and upgrading as for river section	As for river section
D	As for C, but width of flood plain greater than width of the main channel. Upgrade to C if indirect measurements in this range have been taken	As for river section	As for river section
E	Rejection grade - Low flow rating only; rating relation not unique owing to tidal influence or persistent backwater		
Z	Rejection based on facts other than rating - Levels only, excessive truncation, persistent malfunction of installation, very short record, reservoir discharge, spring flow		

The FSR material exists in hard copy form and is retained at CEH Wallingford. However, scanned versions of the material will shortly be available on the World Wide Web via the HIFLOWS-UK project (http://www.hiflows-uk.info).

Low Flow Studies (1980)

The Low Flow Studies (1980) appraisal[5] was concerned with the reliability of naturalised flows. It addressed 1467 gauged catchments and had three broad criteria: first, the accuracy of flow measurement, second, the extent of artificial influences and third, the length of record. The determination of flow accuracy was not stringent; the Low Flows team were more concerned with consistent bias, perhaps associated with

[5] Institute of Hydrology, 1980, Low Flow Studies. Institute of Hydrology, Wallingford, Oxon, OX10 8BB. In four volumes.

errors of stage, e.g. persistent weed growth, than the absolute accuracy of gauging. Typically, accuracy was assessed at 25% of average discharge as an index to low flow performance. The measure used was the factorial standard error (fse), which is the antilog of the standard error of the log-deviations of gaugings, and is a common form for expressing errors about a power law relationship.

The assessment of the impact of artificial influences was to sum the flow paths within and across catchment boundaries and arrive at a net loss or gain and/or redistribution in time. The estimate of fse was assessed as a function of the size of this loss, gain or redistribution relative to average flows. The final fse was arrived at by combining the error from the artificial influences with the rating appraisal to arrive at a single value for each station.

The final grading was to assess the time period for which the data was considered acceptable, coupled with the criterion that the total fse must be below 10%. So, to use monthly flow data, the error in the estimation of monthly flows and the periodicity of the artificial influences had to be within 10%. The best stations were graded ‘A-daily’, i.e. having precise low flow ratings and few artificial influences. One, two, five and twenty years were the minimum length of records suitable for daily, weekly, monthly or annually graded stations. Of the 1467 gauges investigated, 632 were thought unsuitable and of the remainder, the numbers acceptable in the daily to annual categories were: 396 daily, 121 weekly, 163 monthly and 7 annual.

Low Flow Estimation in the United Kingdom (1992)[6]

The 1992 Low Flow Estimation study (often referred to as IH Report 108) used a similar technique to the Low Flows Studies (1980) accuracy assessment, except that the flow point used to determine the accuracy was the one day duration Q95 flow, Q95(1). A difference with the structures appraisal was that the fse was based upon an estimate of the probable error in deriving a gauged flow from the head (principally, the accuracy of measuring head). From these structures it was assumed, unless more detailed evidence was available, that the fse was 1.02. A second major component was related to the sensitivity of the gauge at Q95(1), i.e. how great or little was the effect upon flow of a small stage increment. The sensitivity was described by the percentage change of Q95(1) represented by a +10mm increment of head above the Q95(1) stage, that is

$$\text{Sensitivity Index} = (Q(Q95(1)\ \text{stage} + 10\text{mm}) - Q95(1)) / Q95$$

expressed as a percentage.

This was a particularly telling measure as, in many cases, it related to stages where the uncertainties in flow measurement were largely conditioned by the accuracy of measuring head (so, in some cases, implying a contradiction with the assumed value for fse for good structures). With the accumulation of algae on the weir crest possibly exceeding 15mm, the uncertainties at Q95(1) could be large. The overall hydrometric quality grade for a station was then evaluated based on the rules set out in Table 2.3.

[6] Gustard, A., Bullock, A. & Dixon, J.M., 1992. Low Flow Estimation in the United Kingdom (IH Report No. 108). Institute of Hydrology, Wallingford, Oxon, OX10 8BB. 88pp.

Table 2.3: Hydrometric quality grading for IH Report 108

Grade	Deterioration*	Sensitivity	Factorial standard error† at Q95(1)
A	No obvious deterioration	<20%	<1.1
B	Observed periodic	20% - 50%	1.1 - 1.2
C	Observed sustained	> 50%	>1.2

Notes:
* Due to siltation, weed growth, vandalism etc.
† Based on scatter of spot gaugings about rating the curve or assigned.

The wavy borders between the gradings in Table 2.3 are drawn to emphasise that some judgement was exercised in the combination of factors. For example, high sensitivity might be allowed to ameliorate poorer rating performance or weed growth susceptibility. A grade U (unclassified) was applied if insufficient information was available.

The assessment of the impact of artificial influences was again different to the earlier measure. This was achieved by estimating the bias due to artificial influences upon the ratio of Q95(1) and the natural mean flow. This index was used as when estimating low flows at ungauged sites; the low flow statistics were standardised by the mean flow. Thus if there was no change in this ratio once artificial influences had been assessed, the catchment was regarded as natural. Values of Q95(1) affected by net loss (abstraction) from the catchment, or gain (imports or effluent returns) would cause a change in the ratio.

The flow statistics Q95(1) and Mean Flow were calculated from the National River Flow Archive, which principally contains gauged flows that can be subject to artificial influences. To calculate the 'natural' ratio, an attempt was made to estimate the two flow statistics under natural conditions. Extensive liaison with measuring authorities[7], using the licensing details, and reservoir yields and compensation flows provided the basic data for the naturalising method (fully described in IH Report 108).

The magnitude of the bias remains small where abstractions and returns are principally within the catchment boundary and become larger with increasing imports or exports of water. The initial grading is illustrated in Table 2.4.

Table 2.4: Grading of artificial influences for IH Report 108

Q95 / Mean Flow ratio	<0.5	0.5-0.79	**0.8 - 1.2**	1.21- 1.5	>1.5
Grade	C	B	**A**	B	C

A grade U (unclassified) was applied if insufficient data were available to estimate bias. A total of 1643 gauges were reviewed, of which 1366 were classified. Gauging stations graded as AA were defined as pristine and those graded AB, BA, BB were defined as usable. 490 stations were graded as pristine and 865 were usable.

[7] Gustard et al, 1987 - A study of compensation flows in the UK

The Review of the Northern Ireland hydrometric network
This review, carried out in 1996 by the Institute of Hydrology, took a comprehensive look at the Northern Irish network, with five complementary studies. Only that related to the hydrometric data quality review is summarised here.

In order to index the performance and data quality of stations in the network a broadly based Data Utility Score (DUS) was developed to establish the relative value of the time series of gauged flows associated with each individual monitoring site. One objective was to allow comparisons to be made between the user-perceived value of a station and the actual quality of hydrometric data it may be expected to provide. Obvious mismatches between the station value survey and the DUS could inform decisions regarding network evolution. There were seven elements within the DUS, as set out in Table 2.5.

The Data Utility Score was derived arithmetically using the following formula:

$$2L + S + 2H + Q + R + C + F + A$$

The weightings of the components may obviously be altered; no attempt was made to give different weights to the high or low ranges, for example. Of the 52 stations considered, the range of DUS was between 36.5 and 1.0; only 5 were below 10 and 15 scored 30 or above. Primary network gauges might be expected to score above 25.

Table 2.5: Elements of the Northern Ireland Data Utility Score

Element	Description
Low Flow score, *L*	A four division appraisal carried out by DANI staff involved in maintaining the gauging programme, taking into consideration the hydraulic characteristics of the gauging section its temporal stability, susceptibility to weed growth, scour and accretion effects. An approach similar to the IH 1992 study was not feasible owing to rating relationship practice at DANI.
Sensitivity score and Associated Accuracy, *S*	Sensitivity was defined identically to the IH 1992 study. Associated accuracy was defined as the precision, in mms, that the stage needed to be measured at Q95 to restrict errors (from this source) to less than 5%. Not used in the scoring, it was nonetheless a valuable indicator as to the hydrometric standards that may be necessary to ensure sound data quality.
High Flow score, *H*	This was based on the grading procedure adopted for the FSR and focussed on the existence of confirmatory gaugings in the high flow range and the degree of containment in the flood range. A four-category assessment was carried out with DANI personnel.
QBAR/Highest gauging, *Q*	A mechanistic ratio derived from a schedule of mean annual flood values provided by DANI divided by the flow at the highest gauging. A fivefold index score (1 - 5) was awarded.
Length of record, *R*	Scores were ascribed in proportion to the number of years of data held on the NRFA.
Completeness of record, *C*	An evaluation period of 1983-90 was adopted. A score of 5 implied that >98% of monthly records were complete, a 1 indicates that at least 12% of the months were incomplete. Adjustments were made if the start or end of the record fell within the 1983-90 period.
Factors affecting runoff, *F*	Natural catchments comprise about 80% of the NI network (corresponding figures in E&W are ~15%). The factors were those utilised by the NRFA and published in the Hydrometric Register and Statistics volume and on the Web site. Natural catchments were awarded a 'natural flow increment' of 2.
Adjacency with primary WQ monitoring sites, *A*	An adjacency score was a fourfold ranking (1 - 4); a score of 4 indicated that the sites were sensibly coincident, and 1 a departure of 2km or more.

2.5 Measures used by other Hydrometric Agencies

New Zealand

From a practical point of view, the nature of the New Zealand environment presents rather more severe hydrometric challenges than the United Kingdom. The key to good data quality is regular stream gauging to monitor the stability of rating curves in the face of sediment movement (which is very significant in New Zealand's rivers) and vegetation growth (aquatic and terrestrial).

Data in New Zealand are collected by several agencies, which means that some of the organisations are quite small. However, technical oversight is maintained by the National Institute of Water & Atmospheric Research, and the maintenance of records has been facilitated by their TIDEDA data processing package. Mosley and McKerchar (1989)[8] describe the objectives of a National Hydrometric Reference Network, and the

[8] Mosley, M. P. & A. I. McKerchar (1989) Quality assurance programme for hydrometric data in New Zealand, Hydrological Sciences Journal 34, 185-202.

necessary standards for data collection are defined. This, and later work (Mosley & McKerchar, 1993)[9] describe how they have adopted ISO recommendations throughout.

McKerchar has also advised that

> "Most streamgauges in NZ are natural channel reaches: pervasive sediment movement tends to mitigate against the use of flumes and weirs that are relatively common in the UK. Our practice for rating curves differs from the UK practice. We plot the gaugings on natural scale paper and use hand-drawn rating curves, often following a quadratic form. In the data archive (TIDEDA) the recorded water levels are stored as times series, and stage/discharge ratings are stored as x, y, coordinates with a quadratic interpolation routine for intermediate values. The ratings are applied as required. (In contrast to packages that archive mean daily flows, the advantages are that you can extract peak flows or water levels, and adjust extrapolations of ratings as new gaugings of extreme flows come to hand). We don't work with linear regression in log/log space.
>
> To assess data quality for a streamgauge, we do time series plots for each streamgauge of the difference between the measured stage at the time of gauging, and the stage read from the rating curve for the gauged flow. This is the so-called bedplot because systematic shifts for a series of gaugings can point to a shift in the hydraulic conditions, and indicate that a change in the rating is needed.
>
> Our field teams do regular reports on data quality and give summaries of differences between gauged and rated flows and compare these with relevant ISO standards."

The most recent NZ list of sites is published in Walter (2000)[10]. For most regions of the country, plots of recent flow data for number of streamgauges are available on line: see for example Auckland Regional Council, Otago Regional Council, and Canterbury Regional Council - River Flows. However, the quality of data is not provided on services such as these. Mosley and McKerchar (1993) stated that, among the standards adopted by the Water Resources Survey of New Zealand, "flow gaugings and revised rating curves shall be available on the Water Resources Archive within a maximum of six months of the date of the gaugings".

Around the time that that was written, however, New Zealand generally, and that organisation (the former Department of Scientific and Industrial research) in particular, went through massive restructuring, with much disruption to programmes. NIWA states[11], a description of the Water Resources Archive, that "the goal of this programme is to provide comprehensive and accessible data as a basis for improved knowledge on New Zealand's climate and freshwater resources", with "a key aspect of the programme is application of stringent quality control procedures ensuring national consistency and providing assurance that data can be confidently used for scientific and planning purposes". However, this archive does not yet seem to be in operation.

---

[9] Mosley, M. P. & A. I. McKerchar (1993) Streamflow, Chapter 8 in Handbook of Hydrology, ed. D. R. Maidment, McGraw-Hill.
[10] Walter, K.M. (2000) Index to recording sites in New Zealand, Technical Report 73, National Institute of Water & Atmospheric Research, Wellington, N. Z.
[11] See http://www.niwa.co.nz/rc/prog/database/

To conclude, despite a high level of compliance with ISO Standards in New Zealand, and the provision of quality data, there is no objective classification of river gauging stations provided on an operational basis.

Australia

The physical environment of Australia, as already noted for New Zealand, is such that quality of data can be variable, primarily due to sediment movement in this case. The overall situation for data management is somewhat similar to New Zealand, in that data are collected by many agencies, and made available by those agencies, while oversight is maintained by the Bureau of Meteorology. The maintenance of uniform data protocols has been facilitated by the widespread use of the "HYDSYS" data processing package.

Stream Gauging Information, Australia[12] is a WWW-based compilation of the water quantity monitoring stations operated by the State and Territory Water Agencies and made available by the Bureau of Meteorology. Individual stations may be searched for, and their details determined. Various options include:

- Drainage Divisions and River Basin Boundaries
- Water Agency Contact Details
- Chronology of data base updates
- How to include details for additional stations
- Download the Catalogue (the listing of some 7000 stations plus details)

The observed data are not available from the Bureau's web site and must be obtained from the agency operating the station. A catalogue of the State agencies that provide data has been compiled by the State and Territory water agencies and the Bureau of Meteorology under the auspices of the Agricultural and Resource Management Council of Australia and New Zealand, and can be obtained electronically on http://www.bom.gov.au/hydro/wr/sgc/agencies.shtml. There are some 18 different agencies in that list.

The catalogue of gauging stations, on http://www.bom.gov.au/hydro/wr/sgc/sgc_database.zip, gives a listing of some 7000 stations. There are details provided for each station and guidance on interpreting the information can be downloaded from the Bureau of Meteorology site[13]. In the context of this report, the section on Quality of Data is the most important, and is quoted here:

> "The quality of data depends on many factors which include methods and frequency of recording stream height, stability and sensitivity of station control, frequency and range of measurement used for calibration and the method of computation. These factors can all vary with time and because of the complex interaction of these factors, it is impossible to provide a quantitative measure of data quality.
>
> In providing information, operating authorities subjectively qualify the data quality as GOOD, FAIR or POOR as judged by their current standard."

The comment closing the first paragraph seems to state an Australian point of view, that "it is impossible to provide a quantitative measure of data quality", and that no such

[12] http://www.bom.gov.au/hydro/wr/sgc/

[13] http://www.bom.gov.au/hydro/wr/sgc/help.shtml

process is to be attempted. The second quoted paragraph shows how far a classification system goes, concerning the data provided centrally by the Bureau of Meteorology.

The individual States that provide the data may or may not provide metadata associated with the gauging station and the quality of its data. Some sites have been examined in carrying out the review for this project. New South Wales provides convenient data, graphically and numerically[14]. The Victorian Water Resources Data Warehouse[15] provides data in an uncomfortable format. Queensland[16], on the other hand, provides a convenient way of obtaining data; a wealth of information and history is provided, however, once again there is no evidence of any objective classification.

2.6 Fitness for Purpose

The utility or 'fitness for purpose' of river flow data reflects its accuracy but is also strongly influenced by a number of other factors. Even a small proportion of missing data can greatly reduce the ability to derive meaningful summary statistics (e.g. annual runoff totals or 30-day minima). As importantly, the nature of hydrometric measurement determines that missing data tends to cluster disproportionately in the extreme flow ranges. In the UK there has been an increase in the proportion of missing data submitted to the NRFA over recent years. In part, this reflects the very unusual flow conditions experienced, but also underlines the need for effective procedures to derive estimates of missing flows. Judgement needs to be exercised in applying such procedures to avoid archiving misleading flow estimates. In most circumstances however the inclusion of an auditable and flagged estimate rather than leaving a gap in the record will produce significant benefits in relation to the overall utility of the time series.

In terms of data utility it should be recognised that 'inaccurate' data can be of great value. There will be many circumstances (e.g. during extreme flood or exceptional drought conditions) where data of low but indeterminate accuracy represent a major benefit in terms of time series data utility. The recent increase in the proportion of missing or truncated flows in datasets submitted to the National River Flow Archive can be traced, in part, to a perceived need to avoid the derivation of 'inaccurate' flows. This caution is understandable but it has real potential to degrade the information content of river flow time series.

It is suggested that in any further development of procedures to determine data quality, consideration should also be given to the user requirement for data to be 'fit for purpose'. This could usefully encompass both objective and descriptive measures of station performance and dataset utility. It is worth noting that in recent years a requirement for an annual sequence of daily flows (say) to be 'signed-of' by a competent engineer/hydrologist, as part of a rigorous data auditing procedure, would have considerably reduced the frustration in the user-community confronted by river flow data of very variable quality.

[14] http://waterinfo.dlwc.nsw.gov.au/
[15] http://www.vicwaterdata.net/
[16] http://www.nrm.qld.gov.au/watershed/

2.7 Summary

This review has illustrated the complexity of the England and Wales gauging station network, and shown that any measure of data quality has to be able to account for the different factors that affect a range of types of gauging station. The fact that rated sections do not dominate the hydrometric network means that data quality cannot necessarily be assessed solely on the basis of the scatter in gaugings about a fitted rating curve. A review of international practices has not revealed any established, quantitative, objective measure of data quality in widespread use.

Other agencies do adopt systems based on rigorous quality audits, and (generally subjective) comments on data quality may be linked to the degree of compliance with the target standards. Interestingly, we have not found any system that attempts to measure data quality by referring directly and quantitatively to uncertainties calculated according to the ISO standards for open channel flow measurement.

Previous studies for floods and low flows in the UK have included national assessments of data or gauging station quality and fitness-for-purpose. These assessments have been based on assigning scores or grades based on criteria including statistical measures of fit to flow gaugings, but also broader factors that may affect data quality and utility.

3 CONSULTATION – (I) QUESTIONNAIRE SURVEY

3.1 Introduction

A questionnaire survey was used to seek views on the existing 1995 NRA classification of gauging station data quality, and also on features that may be useful in a new scheme. The questionnaire was compiled by the JBA and CEH project team and approved by the Agency Project Board. This report section describes the questionnaire survey and discusses the results and conclusions drawn from it.

3.2 Questionnaire Design

The questionnaire survey was designed to provide information on the use of the 1995 NRA classification, its strengths and weaknesses and on actual requirements for data quality information, both from hydrometry data providers and users of data. A copy of the questionnaire is included as Appendix A of this report. There was a total of 18 questions, some based on a 'tick-box' format and others requiring the respondent to rank options in order of preferences. Some questions also required a more detailed response. Questions were divided into five categories (A to E) as follows:

A. Use of gauging station data (1 question),
B. Use of existing classification (6 questions),
C. Qualitative/categorical information with which to judge accuracy and reliability (6 questions),
D. Additional information on which to judge accuracy and reliability (3 questions),
E. Form of classification (2 questions).

A list of potential respondents was agreed between JBA and the Agency's project board, and the respondents were contacted to confirm their willingness to participate. The list included a mix of Agency staff and external parties covering both data providers and data users. Data users included those with a specific interest in flood risk, low flows and other aspects of water resources. In total, JBA issued 27 questionnaires during September 2002. Telephone interviews were conducted with each of the respondents to work through the questionnaire and take down their responses and comments.

In total 22 responses were collated. The responses include two members of the Project Team, who filled in the questionnaire at an early stage prior to the development of options for the methodology development. For the purpose of the survey, respondents were asked to identify themselves as 'suppliers' or 'users' of data, or both. Five replied as data suppliers, seven as users and 10 as 'both'. The respondents were as follows:

Luci Allen	Midlands	Peter Spencer	North West
Dave Brown	Southern (now NW)	Dave Stewart	North East
Gordon Davies	Midlands	Mike Vaughan	Southern
Alison Hanson	North West	David Archer	JBA Consulting
Richard Iredale	Midlands	Stewart Child	Hydro-Logic Ltd
Anne Kemlo	South West	Richard Cole	DARD-NI
Michael Law	Consultant for Thames	Andrew Grime	Weetwood Services
Will Lidbetter	Thames	Malcolm Macconnachie	SEPA
David Lindsay	North East	Terry Marsh	CEH Wallingford
Russell Long	Southern		
Stephen Marks	South West		
Sue Morris	Thames		
Ann Ruane	North East		

3.3 Use of the 1995 NRA Classification

A summary of responses regarding use of the 1995 NRA classification is given in Table 3.1. Respondents have been divided into suppliers, users and those who carry both responsibilities in varying proportions. Most users were unaware of the existence of the classification and had never been offered it with data supplied. Of the suppliers who use the system at least sometimes, one developed the system and one is responsible for classifying regional data. Further conclusions are:

- suppliers were also generally dissatisfied (distrust of SMR, the standard error of the mean statistic);
- the existing system was thought over-complex but ‘doesn’t tell you what you need to know’;
- there was thought to be disparity in its use between different station types;
- there was a preference to use CEH Wallingford or Region-specific station summary sheets, or to refer directly to hydrometric staff.

Table 3.1: Use of existing system

			Category		
	Always	**Sometimes**	**Rarely**	**Never**	TOTAL
Suppliers	1	1	1	2	5
Users				8	8
Both			4	5	9
TOTAL	**1**	**1**	**5**	**13**	*22*

3.4 Quantitative or Categorical Information with which to Judge Reliability

Box C1 of the questionnaire presented a list of 24 criteria that might be used to judge the reliability of a record for stations where the rating relationship is determined primarily on the basis of current meter gaugings. Respondents chose those they thought useful, and were then asked to pick out and rank the four most important. A summary is presented in Table 3.2. Column 3 shows the number of times a criterion was selected and column 4 the number of times it appeared in the four most important criteria. In column 5 criteria in Ranks 1 to 4 were given scores (4 for Rank 1 to 1 for Rank 4). The scores were totalled over all respondents. In columns 6 and 7 scores were subdivided between data suppliers and users/both categories.

Broadly speaking, the Standard Error (SE) appeared to be the statistical measure of error preferred by both suppliers and users, although the distrust of SMR may be related to it being a slightly more complicated statistic, although it is in fact not a difficult concept to explain (SE estimates the uncertainty of a sample, SMR estimates the uncertainty of a rating curve).

The survey also revealed a strong demand for:

- A measure of ratio of maximum/minimum gauged to observed,
- A sensitivity measure,
- The number of gaugings.

Furthermore, it would appear that data users generally want more information than data suppliers think necessary. Suppliers chose an average of six measures, whereas users chose an average of 11.

Table 3.2: C1 Numerical/categorical information to judge stations based on current meter gauging

	Criterion	No. of times selected	No of times Rank 1-4	Score for Rank 1-4	Suppliers Score 1-4	Users/Both Score 1-4
A	Total No. gaugings	11	5	25	9	16
B	No. Gaugings since 1990	9	2	6		6
C	Total No. rating changes	8	2	6	2	4
D	No. rating changes since 1990	7	1	1		1
E	No. of structure changes	6		7		7
F	No. structure changes 1990-	5				
G	Freqency of datum surveys	5				
H	Standard error of gaugings	14	7	35	12	23
I	SMR	8	3	15	6	9
J1	Max. gauged : max obs. flow	18	5	23	3	20
J2	--- " --- rank 2 highest gauging	5				
J3	--- " --- rank 3 highest gauging	5	1	4		4
K1	Max gauged flow : QMED	12	3	14		14
K2	--- " --- rank 2 highest gauging	4				
K3	--- " --- rank 3 highest gauging	3				
L1	Min. gauged : min obs. flow	14	2	10	4	6
L2	--- " --- rank 2 lowest gauging	4				
L3	--- " --- rank 3 lowest gauging	4	1	3		3
M1	Min. gauged : Q95	10	2	6		6
M2	--- " --- rank 2 lowest gauging	3				
M3	--- " --- rank 3 lowest gauging	4				
N	% dev. 3 highest gauged	10	3	13		13
O	% dev. 3 lowest gauged	10	3	8		8
P	Index of sensitivity	19	9	21	9	12

Notes:
A full listing of the selection criterion is given in Appendix A. Blanks indicate zero values.

A smaller number of criteria were listed with respect to gauging stations with structures, ultrasonic or electromagnetic gauges. Respondents were asked again to tick those considered useful and to rank in order of usefulness. Several respondents selected but did not rank. Table 3.3 shows, for each main type of gauging station, the number ticked for gauging structures and the number ranked 1 and 2.

The following broad conclusions were drawn from this analysis:

- Respondents value check gaugings for all gauging stations (including standard structures)
- Knowledge of modular limit is a key issue for structures
- An index of sensitivity is also essential, such as the percentage change in Q95 caused by a 10 mm change in stage, as adopted by the National River Flow ArchiveMore important is some measure of comparison between recorder and check gauge (number, average difference etc)

[17] Gustard A., Bullock, A. & Dixon, J.M. 1992. Low Flow Estimation in the UK (Report 108), Institute of Hydrology.

Table 3.3: C2 Information selected to judge the accuracy of gauging stations

	Criterion	Number ticked	Number Rank 1	Number Rank 2
	Structures			
A	Sensitivity index	20	6	3
B	Type of station	19	6	2
C	Modular limit	20	5	6
D	Use of reduction factors	12		4
E	Vulnerability to accretion/weed	12	1	1
F	Date of last survey	11		1
	Time of travel ultrasonic stations			
A	No of bed surveys per year	12	1	5
B	Check gaugings per year	17	13	
C	Procedure for > bankfull flow	14	1	4
D	Index of flight path failures	10		3
E	Checks for sediment and aeration	13		3
	Electromagnetic stations			
A	No of bed surveys per year	18	11	2
B	Bankfull flow /limit of measurement	14	3	7
	Level-only stations			
A	No. manual checks of level	21	12	5
B	% capture rate of water level	14	2	8
C	Type of instrument	16	5	3

3.5 Additional Information to Judge the Accuracy of Gauging Station Data

In addition to asking respondents about the above quantitative or categorical information, the questionnaire also asked for views on useful additional information that might be less easily quantified, or less ‘crisp’. Such information might be expressed as additional codes or description. The responses are summarised in Table 3.4.

Table 3.4: D2/D3 Additional information to judge the accuracy of gauging station data

	Criterion	Number of times ticked
	Requirement for a validation code	
A	Hydraulic modelling	17
B	Crest tapping and d/s gauges	19
C	Extrapolation of Velocity and Area	13
D	Volume checks with neighbours	16
	Requirement for a station attribute code	
A	Weed growth	17
B	Sediment accretion	15
C	Tidal influence	13
D	Other backwater effects	18
E	By-passing	17
F	Structural limitations on high Q	14
G	Physical capability to gauge	15

The following conclusions were drawn from the responses:

- There is strong support for information based on local knowledge (e.g. weed growth and sediment accretion) or on informed judgement (e.g. backwater effects and bypassing) as well as on quantitative information.
- Users felt that they would be better able to judge station reliability, particularly in extrapolated ranges, if one or several validation checks confirmed the rating and they were informed of such validation.
- There were conflicting opinions on usefulness of validation by hydraulic modelling.

3.6 Form of the Classification

Respondents were asked whether the classification should be purely numerical, purely descriptive or both, and whether an abbreviated form should be provided:

- 19 respondents thought the classification should be both numerical and descriptive,
- no respondents thought it should be only descriptive,
- one respondent thought it should be only numerical,
- 16 respondents felt the need for an abbreviated classification as a first-stop measure for all users.

3.7 Discussion of the Survey Findings

The questionnaire survey findings were presented at a workshop held in York on the 26 September, 2002. The workshop proceedings are described in the next section of this report, but a number of points were raised in discussion following the presentation of survey findings.

The main finding of the survey was that he present system is considered unsatisfactory where known. Users were often unaware of the existing classification, and it is clear that any future system needs to be promoted and offered to users. Feedback from the recent Hydrometry Good Practice study by Hydro-Logic and HR Wallingford was that quality codes were only valued if applied consistently.

Respondents valued the existence of check gaugings at all station types as a means of validation. There was a strong demand for measures of the ratio of maximum and minimum gauged to maximum and minimum observed to indicate the extent to which the whole flow range had been confirmed by gaugings. Standard error was better understood than SMR.

There is a continuing need for data quality to be considered in the three ranges of high, medium and low flow. For high flows, it was felt that knowledge of modular limit and by-pass flow are essential for high flow reliability, although this is likely to be based on observation or judgement.

In terms of general presentation, users appeared to want more information than suppliers thought necessary. There was strong support for a classification that is both numerical and descriptive, and some users thought a graphical display of the rating curve should be provided. There was also strong support for an abbreviated expression of classification.

There was some suggestion that the questionnaire did not indicate very strong support for a low flow sensitivity measure, but this could be attributed to there being some bias in the interests of the respondents, despite the best efforts to ensure a representative poll.

A comment was made that the theoretical ratings for well-maintained BS/ISO structures, especially when operating in the modular range, should take preference over ratings derived from current meter gaugings. This is a point of view that derives from field and laboratory determinations of uncertainty in flow measurement at structures (typically within ±10% at the 95% confidence level), compared with the uncertainties about rating curves based on gaugings. The uncertainty about a rating curve fitted to gaugings should be less than that in a single gauging, as individual measurements can be regarded as samples subject to random error, which the fitted curve seeks to minimise. But a single gauging, or a small sample of gaugings, may have a larger associated error, which should be considered when seeking to ‘confirm’ a theoretical relationship at a structure. Furthermore, the number of gaugings may not in itself be meaningful as a measure of quality – the implication for data quality depends on the history and reasons for the gaugings.

The value of knowing about the influence of ‘field-to-office’ procedures was noted, as was the difference between potentially useful information and ‘at-a-glance’ useful information. It was noted that the Agency’s WISKI database could be programmed to include a field or fields to record the values of a quality classification, although complex forms of representing data quality would be demanding to implement. WISKI includes facilities to interrogate rating data, and to include and identify points derived from hydraulic modelling. Every flow value will have a quality flag attached to it.

3.8 Summary

The questionnaire survey, and subsequent open discussion, led to the following features being identified as desirable in the new measure of data quality.

- The quality assigned to gauge data should be objectively ‘provable’, based on common criteria. It should therefore be possible to make, with confidence, the same statements about gauges that have the same quality measure values.
- The performance of some gauging stations is inevitably known in more detail than others. Procedures should be as robust as possible to avoid assigning an overly-good quality to ‘less familiar’ sites simply because details that would raise quality concerns are not widely known. The quality measure should have an empirical basis. Where the information to confirm good quality is unknown, it may be appropriate to assign data a somewhat lower status, simply as a reflection of that lack of specific knowledge.
- It is valuable to retain discrimination between different flow regimes, ideally including information on critical limits of gauge performance. This will be useful to dovetail with use categories to assess fitness for purpose. It will help

indicate the balance between ‘reasonable quality over a wide range of flows’ and ‘excellent performance over a more limited range’.

- The quality classification should not attempt to represent every detail of gauging performance.

4 CONSULTATION – (II) WORKSHOP

4.1 Introduction

Following the initial literature review and questionnaire survey, a workshop was held at the Environment Agency offices at Coverdale House, York on 26 September 2002. Seventeen participants were invited to the workshop (as agreed with the Agency Project Manager), of which thirteen attended. The workshop participants were as follows:

Luci Allen	Midlands	David Archer	JBA Consulting
Dave Brown	Southern	Stewart Child	Hydro-Logic Ltd.
Alison Hanson	NW	Rob Lamb	JBA Consulting
Richard Iredale	Midlands	Martin Lees	CEH Wallingford
Will Lidbetter	Thames		
David Lindsay	NE		
Ann Ruane	NE		
Dave Stewart	NE		
Simon Wood	Anglian		

The proceedings of the workshop are reported in what follows, and we also summarise its main conclusions and decisions.

4.2 Aims of the Workshop

The main aims of the workshop were to consult with representatives of the hydrometric and hydrology communities to decide on the broad concept to be developed as a method for representing data quality. The workshop agenda was to report the initial findings of the review and questionnaire survey and to present initial proposals for five distinct options for the gauging station data quality classification based on the findings of the questionnaire survey. More general issues relating to data quality and the review of the current classification were also discussed. The strengths and weaknesses, and implications for data users, data systems and for the Agency's hydrometric staff were considered for each option. It was also intended that the workshop participants would agree on an option (or combination of options) to progress into a full classification system.

4.3 Background Considerations

There were a number of relevant considerations borne in mind when developing the five initial options. The project brief was to identify a method for representing the quality of gauging station data, but the precise form of representation was left open, with options including real-number statistical measures (i.e. on the interval scale), scoring systems (i.e. on the nominal scale), qualitative description, visual representation or combinations of the above. The main issues considered are described below.

Performance target or measure of fitness?
The project brief was not specific as to whether the 'Method of representing the quality of gauging station data' should primarily meet the needs of data suppliers or data users. The 1995 NRA classification has been seen to be both a performance target as well as a retrospective record of performance. The responses to the survey questionnaire indicate that most data users had not heard of it or had not been offered it along with the data. This suggests that it has not been effective as a measure of performance suitable for data users.

In discussion with the Agency project board, the main requirement for this project was has been interpreted as a method of representing the quality of gauging station data to enable users to assess whether data are fit for the purposes for which they wish to use it. The focus was therefore on the data user rather than on the supplier. However, it is expected that measures of representing data quality for the benefit of users may also be appropriate as performance targets.

Simplicity/Complexity
The project specification expressed concern at the complexity of the outputs of the 1995 NRA classification system, which can potentially contain as many as 25 descriptor characters. This view has been supported by users' responses to the questionnaire survey, although many of them were seeing an example of the 1995 classification method for the first time. Nevertheless it is clear that the assessment of gauging station reliability is indeed a complex issue. Uncertainty takes many forms. The main challenge here is to condense the complexity into something simple.

The view of most of the questionnaire respondents (including all the users), was that a considerable body of information should be provided to enable users to judge data reliability. However, the most critical elements should be combined to create a condensed representation that would be provided not just to 'non-technical' users, but also to others. Users would then go on to look more closely at the full quantitative and qualitative information to make a judgement, should the 'quality measure' fall below a critical value for their purposes.

'Representation' or 'classification'?
In developing initial options for data quality representation, it was not assumed that the chosen representation of gauging station data quality should necessarily take the form of a classification (i.e. an approach by which a number of gauges could easily be compared on a standardised scale). However, many workshop and survey participants tended to reject options that were not suitable for classification, and it therefore became apparent that a classification scheme was most likely to be chosen.

Management of the classification
The context of the quality classification and its management were considered as part of the development of the five initial options. Figure 4.1 illustrates the management of the quality classification. Key features are the hydrometric database and a corresponding database of 'quality attributes' (which may or may not be intersect with the hydrometric database), and the procedures and software tools that encode the classification. It was proposed that the classification should be led by the requirements of data users (this is taken to be an aspect of the project brief), however Figure 2.1 illustrates that the provider of data is central to the maintenance of the quality classification, and that it must therefore be manageable.

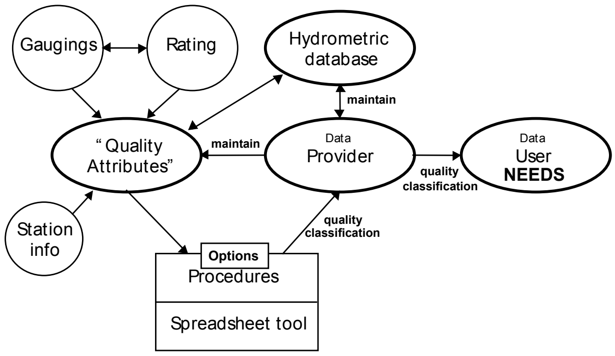


Figure 4.1: Context of the classification system

4.4 Options for Representing Gauging Station Data Quality

Five options were proposed at the workshop. The options are not entirely exclusive; aspects of several options could potentially be combined into a final scheme. Each option presented was followed by a discussion session. The options are described below, along with summaries of the workshop discussions.

4.5 Option 1 – Attribute Scoring

Methodology

This proposed option was based on a matrix of 'quality attributes', likely to be drawn from the established attributes included within the survey questionnaire, for example SMR, bankfull flow and indicators of whether weed growth is present or managed. The principles of the option were to assign scores to different values of the chosen attributes and to aggregate the scores into a more succinct form by using a weighting function. The scoring rules would be common for all stations. The weighting function would be adjusted for different station types, with care being taken to avoid, as far a possible, any inconsistencies between station types. A simple and effective weighting function would be a weighted sum (or average) of scores. Weights might also potentially be adjusted to reflect different user priorities.

Scores would need to be assigned to each quality attribute to translate the attributes to a common scale. For example, a scoring scale of (0,5) may be adopted, where 0 indicates lowest quality and 5 indicates best quality. An example scoring rule may be of the form given in Table 4.3 for a quality attribute defined as the ratio between the highest check-gauged flow and the highest recorded flow (i.e. the highest flow determined from the station rating). The rule shown here has been specified arbitrarily and is illustrative only, although there will inevitably be a degree of judgement exercised in practice in determining the scoring rules. The weighting function used within attribute scoring would be determined on the basis of gauge type, with weights being chosen so as to reconcile the quality codes obtained for different gauge types.

Table 4.3: Example quality attribute scoring rule

Ratio of highest check-gauged flow to highest recorded flow ($R = \max\{ Q' \} : \max\{ Q_{obs} \}$)	**Score**
$0.9 \le R$	5
$0.8 \le R < 0.9$	4
$0.7 \le R < 0.8$	3
$0.6 \le R < 0.7$	2
$0.5 \le R < 0.6$	1

It was proposed that a spreadsheet tool would support the attribute scoring option. This would contain the scoring rules and weights, and encode these in a fixed form, visible to the user. As an initial proposal, the spreadsheet tool may comprise two sheets, the first being a space for managing quality attribute data and the second containing the scoring. Dummy layouts for the tool were presented to the workshop, as illustrated in Figure 4.2 (quality attributes sheet) and Figure 4.3 (scoring).

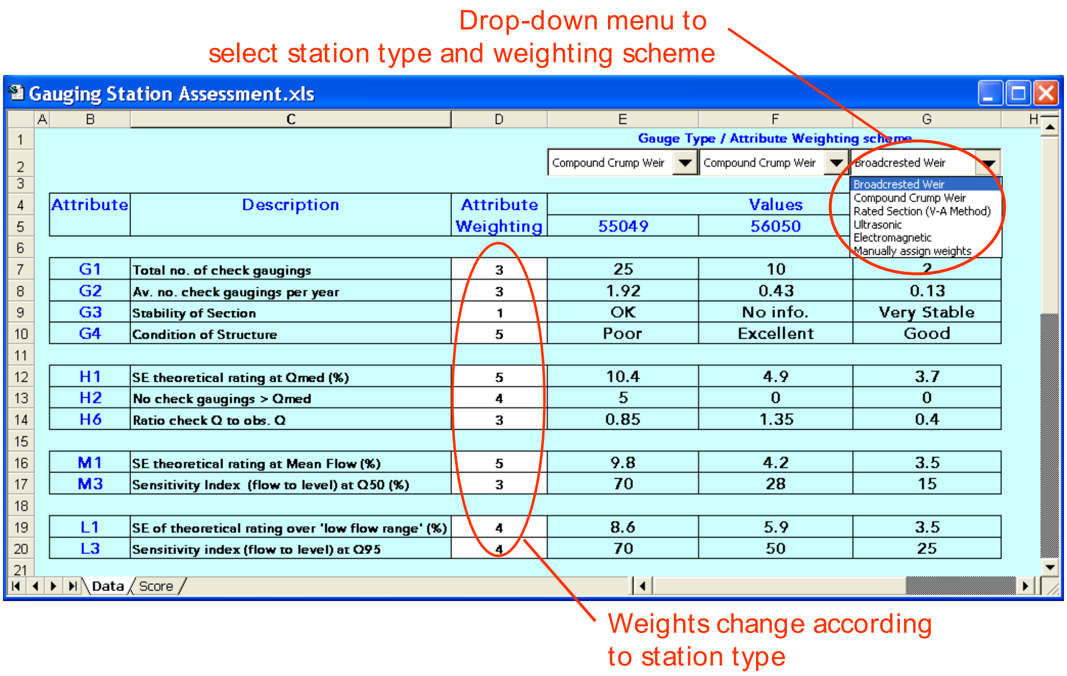
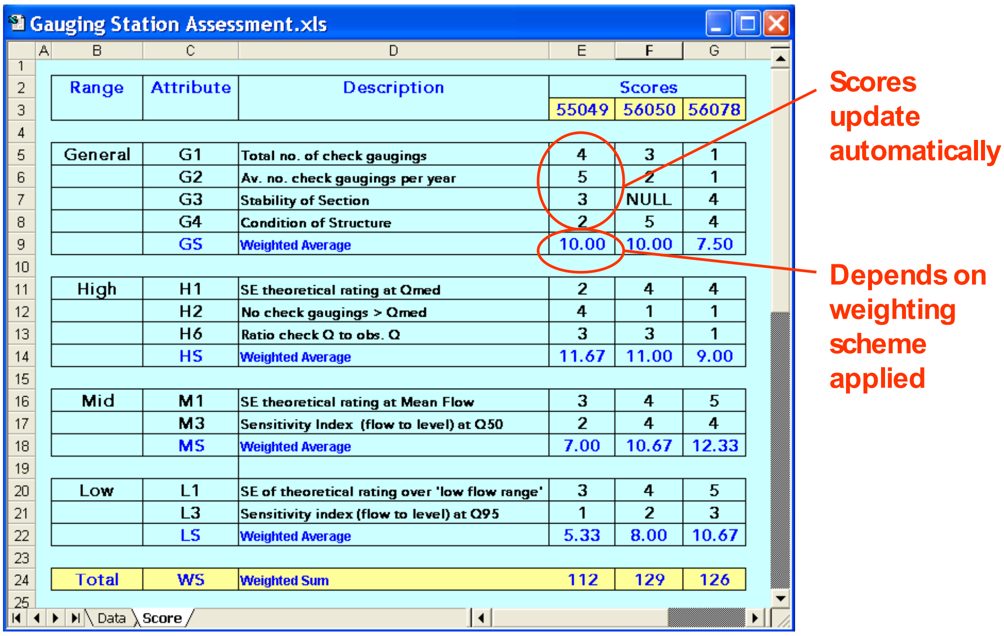


Figure 4.2: Option 1 – Spreadsheet tool (quality attributes)



Range	Attribute	Description	Scores		
			55049	56050	56078
General	G1	Total no. of check gaugings	4	3	1
	G2	Av. no. check gaugings per year	5	2	1
	G3	Stability of Section	3	NULL	4
	G4	Condition of Structure	2	5	4
	GS	Weighted Average	10.00	10.00	7.50
High	H1	SE theoretical rating at Qmed	2	4	4
	H2	No check gaugings > Qmed	4	1	1
	H6	Ratio check Q to obs. Q	3	3	1
	HS	Weighted Average	11.67	11.00	9.00
Mid	M1	SE theoretical rating at Mean Flow	3	4	5
	M3	Sensitivity Index (flow to level) at Q50	2	4	4
	MS	Weighted Average	7.00	10.67	12.33
Low	L1	SE of theoretical rating over 'low flow range'	3	4	5
	L3	Sensitivity index (flow to level) at Q95	1	2	3
	LS	Weighted Average	5.33	8.00	10.67
Total	WS	Weighted Sum	112	129	126

Figure 4.3: Option 1 – Spreadsheet tool (scoring)

Workshop discussion (Option 1 - Attribute scoring)
This option was well received by the workshop participants. It was felt to have parallels with a system of prioritisation for Thames region gauging stations[18]. It was noted, in general, that data users are not always well-informed about what makes data 'good' in quality terms; Option 1 would provide a framework for providing this information. Option 1 provides for objective classification. It was suggested that this will be an important feature. It was noted that the attributes and scoring rules could be aligned with the recent Hydrometry Good Practice R&D.

It was noted that the attribute scoring option is similar in principle to the approach used by CEH Wallingford in their 1996 review of the Northern Ireland hydrometric network.

A key issue that was discussed was obtaining a suitable balance between scores and weights so as to avoid inconsistencies between gauges, and also in terms of complexity versus simplicity. The discussion then focussed on identifying strengths and weaknesses. The comments made are summarised in Table 4.4.

Attributes would have to be chosen to ensure that the method could be resourced. The final presentation of quality as a score, or set or scores, would help to promote integration with other data systems. Overall, Option 1 was thought to be a flexible approach that offered the potential to combine disparate information in a consistent way.

[18] Hydro-Logic Ltd and HR Wallingford. 2002. *Hydrometry Good Practice*, R&D Report W6-055/TR.

Table 4.4: Option 1 – Attribute scoring (Discussion)

Strengths	Weaknesses
The approach is objective.	It could become too complex.
It can be designed to be thorough.	It could be opaque to users without full training and explanation.
The method can accommodate differences between station types.	Some attribute information could be difficult to obtain.
The method can be tuned.	There may be difficulty in achieving consistency between stations.
The method discriminates between different flow ranges.	The weights require careful choice.

4.6 Option 2 – Descriptive

Procedure

The second option considered, Option 2, was a purely descriptive representation of data quality, based on a comprehensive set of the information affecting station and data quality. The format would be as station summary sheets, modelled perhaps on the National River Flow Archive summaries, or those used in some Agency regions. Quantitative information could form a part of the descriptive approach, for example summary sheets may include information about bankfull flow, rating curves etc. Expert judgment of hydrometric staff might also be included.

The descriptive option would therefore consist of a re-working of some current practice, with an emphasis on establishing a consistent format and layout. The option would provide classifications for individual aspects of station data quality in that all values of a particular attribute, say the sensitivity of flow to level at Q95, could be extracted and compared. The option would not, however, attempt to combine the different factors that determine station data quality into an overall measure (or suite of measures). As such, conclusions drawn from the descriptive option would depend entirely on interpretation by the user.

Workshop discussion

During the Workshop discussion it was noted that the primary disadvantage of this option is that it is not a classification, and that it cannot therefore be used to compare different stations directly, especially if a large number of stations were to be considered. The descriptive option was felt to be in many ways similar to the CEH station summary sheets, and there was some general discussion of the potential for incorporating additional information into the CEH summary sheets, if this were provided by the Agency.

One other aspect of Option 2 that was noted was that it might formalise and standardise a set of relevant information to be collected and recorded throughout the Agency. Strengths and weaknesses are summarised in Table 4.5.

Table 4.5: Option 2 – Descriptive (Discussion)

Strengths	Weaknesses
Simple.	Not a classification.
Easily understood – requires no specific explanation for a technically informed user.	Can't be used to compare stations directly.
Would get key messages across.	Not capable of being used as a performance target.
Would satisfy user needs (for individual stations)	Risk of inconsistency.
	Potential resourcing problem.

Some of the information that would form part of the descriptive option will be available from HARP, whilst other information could potentially be added, though the resource implications for doing this would need to be considered carefully.

Overall, it was felt that Option 2, whilst useful, would not be suitable as a stand-alone method. It would not deliver a classification. It would therefore be more difficult than with other options to make a comparison over a large number of stations (for example selecting nationally all stations providing 'good' high flows data, as required for the current HIFLOWS-UK initiative). Another important drawback of the descriptive approach is that it would not be useable as a performance target, something that the 1995 classification does support. As a final issue, it was identified that there could be substantial resource implications in collating and recording all descriptive information about a station and its data quality.

The value of descriptive information available in a separate form, but additional to a quality classification, was discussed and noted.

4.7 Option 3 – Standard Error 'Plus'

Methodology

The starting point for the third initial option is the current (1995) gauging station classification. Option 3 as proposed would look similar to the current system, but might have a different 'feel' as a result of support from a spreadsheet tool, a revised tabular format, revised treatment of gauging statistics and the addition of some new features.

It was proposed to review the statistical measures used to assess the rating relationship. The current classification makes use of two related statistics. The first is the standard error of estimate (*SEE*, which is defined in detail in Section 7). The standard error of estimate is a measure used to assess the degree of goodness-of-fit of a regression line to a sample of data. The statistic *SEE* has been referred to in some documents as simply the 'standard error', *Se*. The standard error of estimate can be derived separately for different ranges of discharges. The second related statistic is the 'Standard error of the Mean Relationship' or (*SMR*, see also Section 7), which expresses uncertainty about the regression line itself (the 'mean relationship' being the regression line in this case).

For practical purposes there are two important differences between the two statistics. The first is that *SMR* increases as one moves away from the mean of the sample data towards its exterior. This is to be expected, for it indicates that uncertainty about the regression line increases as it is used to extrapolate beyond the sample data. The second difference is that SMR increases in value more steeply as the sample size (i.e. number of spot gaugings) becomes smaller, reflecting the reduction in the precision with which the rating curve can be fitted as the number of spot gaugings reduces.

One proposal for presenting data within Option 3 was to display actual uncertainties at specified flow points, rather than derived scores. A suitable statistic might be the width of the 95% confidence interval, displayed at, say Q95, Q50 and QMED/2 as a percentage of the rated flow. A proposed enhancement to the standard error option was to ascribe a theoretical value of *SEE* to BS/ISO structures rather than allowing check gaugings to prevail. This would avoid the risk of data from good structures being given an artificially poor quality for cases where current meter gaugings may be inherently less accurate than the structure itself. It was also proposed that bankfull flow or the modular limit and the existence of gaugings above bankfull or application of corrections for non-modular flow should be indicated.

The final element of the classification would be an index into a structured lookup table that would be made widely available and would attempt to summarise most of the key information that could affect data quality. For example, a 'Class 1' station might be a structure and 'Class 1' stations might be sub-divided amongst other things into those where the modular limit was below QMED ('Class 12', say) and those where reduction factors had been applied consistently ('Class 13', say). Figure 4.4 is an example of the presentation of Option 3, including a concept sketch for the structured lookup table.

Date		95% confidence band			Bank	No. > bank	Class
		Low	Med	High			
01-Jan-85	15-Oct-91	12%	9%	21%	120	3	12
16-Oct-91	03-Jun-01	10%	9%	14%	200	2	13

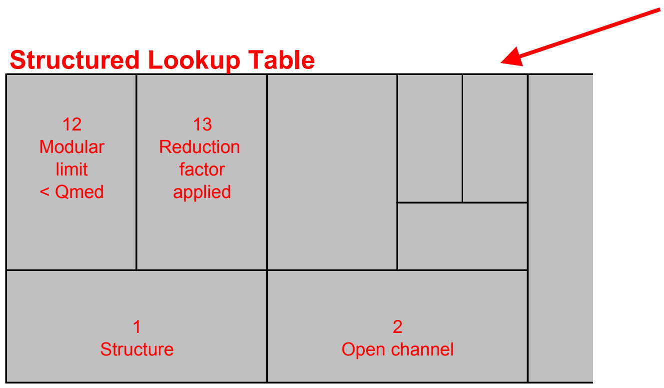


Figure 4.4: Option 3 – Concept sketch

Workshop discussion of Option 3

The statistical basis of the method was discussed in some detail, and it was noted that the approach used to define error in the stage discharge relationship, including the implications of different station types, would be reviewed within the development of the option. Strengths and weaknesses are summarised in Table 4.6.

Some of the information needed to compute the statistics for Option 3 will be readily available from within WISKI, including the rating curve and check gaugings. However, it was felt that there may still be a resourcing issue to keep track of the information needed to derive the quality codes.

Table 4.6: Option 3 – Standard Error 'Plus' (Discussion)

Strengths	**Weaknesses**
Quantifiable.	Thought to look too much like the current system.
Empirical.	May be difficult to get full information for the statistical analysis.
Repeatable.	Would require different information for different types of site.
Provides a classification – i.e. can be used for comparisons	The lookup table could become very complicated.
Auditable	Requires explanation to be interpreted correctly.
The lookup table could be effective.	

4.8 Option 4 – Time Series Audit

Methodology

Option 4 is as an entirely visual representation of data quality. It would be based on data quality codes assigned in continuous time (i.e. to every value in a series, or at least on a daily basis). It was proposed that the option should be based on a subset of quality attributes. Making use of the statistical analysis of the rating relationship, the time series audit would plot confidence intervals alongside the data and would also plot specified items such as the minimum and maximum gauged flows, bankfull and/or modular limit and missing, truncated or in-filled values.

The time series audit would be supported by a simple data browser tool that could be made available as a public download. A model for this concept is the data browser in the CEH Low Flows 2000 software. A mock up of the browser for Option 4 was presented in Figure 4.5.

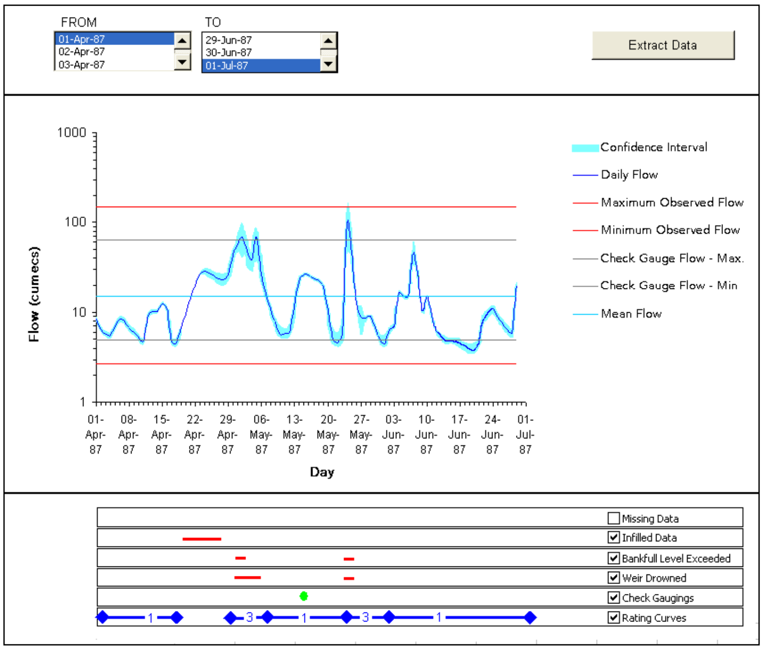


Figure 4.5: Option 4 – Mock-up data quality browser

Workshop discussion

Option 4 represents data quality in a visual format. This option was thought to be attractive and useful, but, as can be seen from the strengths and weaknesses (Table 4.7), it was essentially ruled out on three grounds. Firstly, it does not provide a classification. Secondly, it would duplicate some of the functionality of the WISKI database. Finally, it would risk creating a duplicate of a core Agency data set.

Table 4.7: Option 4 – Time Series Audit (Discussion)

Strengths	Weaknesses
Gives direct view of the data.	Not a classification.
Immediate and intuitive.	Duplication of HARP functionality.
Useful as a data 'validation' tool.	Duplication of core Agency data set.
	Does not provide a compact summary.

It was noted that WISKI will provide some 'high-level' statistics, for example numbers of missing or incomplete days of record, as well as visualisation functionality. The provision of this visualisation for the external data user was not discussed.

4.9 Option 5 – Abbreviated Presentation

Methodology

The final option presented to the workshop was an abbreviated and highly simplified form of classification. This would divide the flow regime into 'low', 'mid' and 'high' flows at specified thresholds. It would then classify the quality of data within each flow range as ‘poor’, ‘fair’ or ‘good’. The classification would be based either on subjective expert judgement or on an attribute scoring approach, as presented in Option 1. The abbreviated classification is therefore not a ‘stand alone’ option, but was proposed as an alternative means of presentation. It is illustrated for a single date span in Figure 4.6, though it would be likely to be defined over a number of date ranges for any given station.

	Good	Fair	Poor
High		✓	
Mid	✓		
Low			✓

Figure 4.6: Option 5 – Abbreviated classification

Workshop discussion

The ‘abbreviated’ option was strongly supported for its simplicity and ease of interpretation. It was felt, however, that this option would exist only in conjunction with an underlying method to determine the quality classes. The strengths and weaknesses were straightforward to identify, and are shown in Table 4.8.

Table 4.8: Option 5 – Abbreviated (Discussion)

Strengths	**Weaknesses**
Verbal and intuitive.	It is necessary to dig down to obtain detailed information.
Provides a classification.	

It was suggested in the discussion that Option 5 need not be intended solely for the ‘non-expert’ data user. It would provide a convenient starting point for all data users and providers.

4.10 Workshop Conclusions

The aim of the discussion session held during the workshop was for the project team to gain an overall steer that would indicate which option should be taken forward and developed into a new procedure. The discussion did not aim to fix all of the details of a classification, prescribe any particular statistic or design a finished product. The discussion was structured so as to identify, for each option in turn, strengths, weaknesses and implications (compatibility with data systems, resourcing implications and practicality of uptake). This section summarises the key points raised during the discussion and presents the analysis of strengths and weaknesses. It does not therefore attempt to record every comment made during the discussion that lasted nearly three hours.

Categories of use
Users are often interested in a particular range of flow. Within Options 1, 3 and 5 it is suggested to provide a quality classification for different parts of the flow regime. Attention tends to focus on high and low flows, and there was some debate as to whether many users are specifically interested in a class for medium flows, centred round accuracy at mean flow level. However there are certain aspects of 'general' data quality that will be relevant to the full range of flows, including those between the extremes. It is suggested to organise classification around three indicative flow ranges, as follow:

High Flows
These data will mainly be used for FEH analysis. Since much of the flood frequency analysis depends on pooling of flood growth curves, some indication of reliability is essential for flows well above QMED, which is the base level for growth curves.

Low flows
The low flows category will be centred on use for assessing abstractions and impacts of polluting discharges as well as on hydroecology. Although much of this analysis is focused around Q95, the reliability of low flow extremes well below this level is also important.

General
This category would apply mainly to water resources uses. The interest is in the reliability of the mean daily or monthly flow series, which will contain data over the full spectrum. In addition to reliability in the middle of the range, the reliability of low flows are likely to be of more concern than high flows because of their persistence and implication for drought studies.

4.11 Summary

Taking account of the strengths, weaknesses and general comments made about each option, the discussion concluded with an attempt to identify the overall concept that should be taken forward. The conclusion from this discussion is that the workshop endorsed the following approach:

- A classification system,
- Based on the attribute scoring method,
- Incorporating an abbreviated presentation.

5 PRINCIPLES OF GAUGING STATION DATA QUALITY CLASSIFICATION

5.1 Introduction

This section of the report describes the principle elements of the gauging station data quality classification. It sets out the types of attributes chosen to represent factors that influence data quality and the overall framework of the scoring schemes used to combine attributes. Further details about the component parts of the classification are given in subsequent sections.

5.2 Background and Main Considerations

Workshop

As we have discussed in Section 4 of this report, the main outcome of the York workshop was to adopt attribute scoring as the method for representing gauging station data quality. This choice was subsequently endorsed by the project board. The attribute scoring approach assumes that factors having some influence on flow measurement (attributes) can be identified, assessed objectively, as far as possible, and the results can be combined to provide an overall picture of the quality of gauged data.

There was also a general consensus that, in order to ensure that uptake of the scheme is as wide as possible, a relatively simple form of classification should be developed. Classification results would be summarised into a 'poor/fair/good' style format.

The main challenge in developing a classification based on these approaches was to capitalise on the flexibility offered by the attribute scoring methodology, whilst keeping scoring procedures straightforward and transparent to the user. Key considerations included:

- The need to maintain an objective and representative classification,
- That the classification should be consistent for different type of gauging station, but at the same time should reflect the intrinsic strengths and weaknesses of different gauging methods and whether stations operate to British Standard,
- The need to choose representative and appropriate attributes,
- The need to ensure that any analyses used in deriving attributes are statistically sound,
- The need for a scoring procedure that will work equally for attributes based a variety of disparate sources,
- The need to align attribute values with the Hydrometry Good Practice Report[19],
- The need to combine attribute scores in a simple, yet statistically valid, manner,
- The need to assign appropriate abbreviations to (ranges of) attribute scores.

[19] R&D Technical Report W6-055/TR. Review of Good Practices for Hydrometry, Hydrologic and HR Wallingford consortium. 2001. S. Child, B. Woods-Ballard, A. Clare-Dalgleish, P. Sayers.

Features retained from the 1995 NRA classification

The 1995 NRA classification was a basic starting point for the new scheme. In particular the method of discriminating data quality for different parts of the flow regime (High, Low and General) was highlighted by participants of the York workshop as one of the strengths, and was carried through to the new classification.

The uncertainty associated with each measurement, as used in the 1995 classification, was considered to be an important factor and carried through as an attribute in the new scheme. However the statistics used were reconsidered - the way SMR is used in the 1995 procedure underestimates the error at extremes (see Section 7). The statistical procedures used to calculate uncertainty for rated-sections were therefore revised as appropriate whilst uncertainty calculation for other types of station was aligned to BS 3680 Part 4 (ISO 6416 and 9213) where possible.

Attributes

Attributes are those factors considered to influence the quality of the data from the station. Much consideration was given to how knowledge of a wide range of influences on gauging performance could be built into the quality classification procedure as attributes. It was felt that the attributes used should encompass properties of the flow record, features of the gauging station situation and features of the flow regime.

Attributes were identified on the basis of the literature review and the review of user requirements. They were chosen carefully to reflect data quality issues, rather than broader 'fitness for purpose' (such as the degree of artificial influence upstream and so on), the latter being beyond the scope of the project. The primary issue was to select attributes that meet users' requirements and which can either be derived from data readily available within the Agency's Hydrolog or WISKI databases or that are routinely recorded in the station files.

In particular, ways of incorporating local knowledge were considered. By 'local knowledge' we mean specific factors that may affect data quality from a station that are known to hydrometric staff, but might not be explicitly represented by an arbitrary sample of flow gaugings, for example

- Does the cross section or control change at high flows (say above the highest gauging)?
- Is the structure or cableway bypassed at high flows?
- Is the station affected by variable backwater?
- Is the structure modular at extreme flows and is the modular limit reliably known?
- Are low flows affected by weed growth (thus with different seasonal accuracy)?

However local knowledge must not introduce bias for or against those gauges for which there is less information available.

The role of 'field-to-archive' factors such as rounding up of data, transcription errors was also considered. It was decided that these could not be represented explicitly as attributes, but would, however, be represented implicitly by attributes that compared gauged flows with continuously recorded data held on archive.

A further difficulty was to reconcile the need to include enough attributes to ensure a representative and meaningful classification, against the need to keep things as simple as possible. Consideration was also given to how best to deal with attributes that influence data quality differently for different types of gauging station, and how to reflect the relative influence of different attributes on data quality. Weighting of attributes or a classification table were considered as likely options for bringing this information into the quality measure.

Use of 'check gaugings' to 'confirm' theoretical ratings
In early discussions, and at the York workshop in September 2002, the view was expressed that structures built to British Standard were to be regarded as inherently more accurate than rating curves established by gauging. There has since been concern expressed that it is nonetheless useful to know whether gaugings at a station support a theoretically calculated flow, or a flow in the non-modular range for which corrections have been applied. These represent two conflicting views on the value of gaugings compared to the inherent capability of a structure

Much consideration was therefore given to ways of incorporating the level of agreement to current meter check gaugings, whilst recognising that the level of random error or 'noise' within a small sample of current meter gaugings is likely to be high

Level-only stations
The classification was also to include scoring procedures for sites where only level is measured. It was recognised that for such sites, the accuracy of level measurement would be key to data quality and that the attributes to be included would necessarily be more focussed on accuracy than when considering gauging stations.

Software tool development
The requirement to incorporate the classification into a spreadsheet tool was an important consideration in the approach adopted. A certain level of sophistication was considered appropriate for the software tool, allowing the classification to be slightly more complex than if it was to be implemented by hand. For example the ability to automate procedures in the software tool would allow the inclusion of attributes requiring more complex, lengthy or repetitive calculations, which would not otherwise have been suitable. At the same time, a pragmatic approach to software development was preferred. Thus simpler solutions were generally favoured over the complex, so that the software tool would not become too unwieldy or difficult to use.

5.3 Approach Adopted in the New Classification

Overall framework
The framework of the new classification was developed at the outset, although modifications and improvements were made as and when required, and tested on a trial and error basis using example stations. The framework is illustrated in Figure 5.1. There are three main stages.

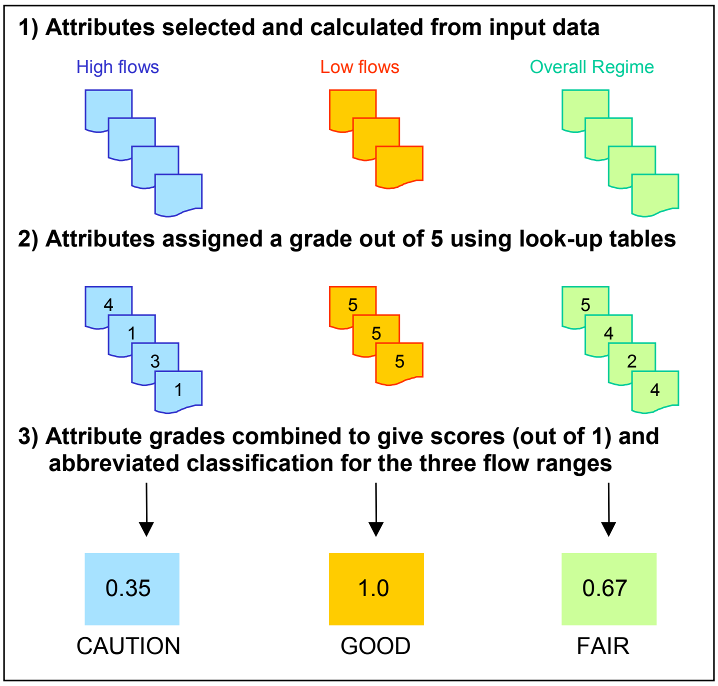


Figure 5.1: Basic framework of the data quality classification

The first stage involves defining a set of attributes appropriate to the method of gauging (between 16 and 20 attributes are considered in each case), which are then divided into three flow ranges:

- High Flows
- Low Flows
- General Flow Regime

Some attributes are evaluated in each of the three ranges, whilst other attributes are relevant only for a single flow range.

For High Flows, QMED (the median annual maximum flow) is used as an 'index' value against which certain attributes can be defined (for example rating curve confidence intervals). For Low Flows, Q95 is used as the index flow. The mean daily flow is used as an index for the General class, although greater emphasis is placed on overall measures of accuracy (such as the standard error of estimation of a rating curve) and a mixture of High/Low attributes are included. The relative importance of each attribute (within its flow range) is represented using a weighting factor. The arrangement of attributes and weights is called the scoring scheme. For level-only sites it is not considered appropriate to divide attributes in to the three flow ranges, rather a single overall-range, incorporating seven separate attributes, is used.

In the second stage each attribute is evaluated 'objectively' using look-up tables to assign a nominal grade between 1 and 5, where 5 is the best grade, depending on its value.

In the third stage the weighting is applied to the grades for each attribute, and the attributes within each flow range are combined into a numeric score. This combination of grades and weights is calculated using the geometric mean. The combined numeric score is presented as a decimal fraction (i.e. a real number between 0 and 1). For each site, the quality of flow data is then classified as 'Caution', 'Fair' or 'Good depending on the score achieved.

The classification represents an assessment of quality of gauging station data observed between two discrete points in time, i.e. the results of the scoring scheme applied over a particular period of time in the flow record. The choice of cut-off points between classification periods will be a matter of judgement; further guidance is given in Section 10 of this report.

Terminology
The terminology used in the classification is intended to be generally self-evident, however a definitive list of terms used is given in Table 5.1.

Table 5.1: Terminology used in the classification

Term	Description
Classification	The assessment of quality of gauging data observed between two discrete points in time based on applying the appropriate scheme given the type of gauging station used.
Attribute	A factor having a strong influence (either negative or positive) on gauging station quality, generally one of the following: • A physical feature of the gauging station / recorder • A statistical / numerical property the flow or stage record • A statistic relating to check gaugings made at the site.
Attribute value	The numerical or, where the attribute cannot be described numerically, categorical value of the attribute.
Grade	An integer between 1 and 5 representing the 'quality' of each attribute. A grade of 5 indicates that an attribute is of a fully acceptable standard and has a neutral effect on quality. Any reduction from 5 has a detrimental effect, so that a grade of 1 has a very detrimental effect on quality.
Look-up table	A table from which the grade associated with a particular attribute value is determined.
Attribute Score	The grade expressed as a fraction of the maximum grade (i.e. out of 5).
Weight	A weighting factor used to adjust the degree of influence that each attribute has on the final data quality score for a particular type of gauging station.
Scoring scheme	The arrangement of attributes and weights for a particular gauging station type.
High flows range	The part of the flow duration curve above the flow percentile equivalent to 0.5 x the median annual flood (QMED) in m^3s^{-1}.
Low flows range	The part of the flow duration curve below the 5th flow percentile (Q95) in m^3s^{-1}.
General flow regime	The whole of the flow duration curve, but specifically describing flows not described as high or low. Index flow event is the average daily flow (ADF) m^3s^{-1}.
(Combined) Score	The weighted geometric mean of the attribute grades for each category. Presented as a decimal fraction between 0.0 and 1.0.
Abbreviated score	The abbreviated classification of data quality expressed as 'CAUTION', 'FAIR' or 'GOOD'.

5.4 Types of Gauging Station Represented

Rather than use one set of attributes for all types of gauging station, it was considered more appropriate to use particular combinations of attributes for different types. Six combinations, referred to as scoring schemes, were therefore developed, as described in Table 5.2. Different combinations of attributes are used for each scheme, although many attributes are common to all.

One generic category is used for all types of weir or flume (providing these operate to British Standard specification). The associated scoring scheme does allow for degrees of non-compliance with BS 3680 (Part 4), however non-standard structures that are operated solely as a rated section for gauging purposes should be scored using the rated-section scheme. A further scoring scheme is used for structures formally supported by use of a rating curve at high flows (e.g. those structures with a limited modular range). This scheme is equivalent to the BS/ISO structures scheme for overall regime / low flows and the Rated Section scheme for high flows.

Scoring schemes are described in greater detail in Section 6.

Table 5.2: Scoring schemes

Scheme Name	Gauging Station Types
Level only	Level recorder only
Rated section	1) Open channel cross-section with natural control where the stage-discharge relationship (rating) is derived from current meter (or other) flow gaugings. 2) Cross-section with artificial control (structure) where the stage-discharge relationship (rating) is derived from flow gaugings. Includes stations where the rating is derived from gaugings made using ADCP or intermittent ultrasonic gauge readings as well as current metering using hand held devices or cableways
BS Structure	1) All structures designed and operated to British Standard/ISO e.g. o Weir meeting BS 3680 (Part 4) having a theoretical stage-discharge relationship o Flume meeting BS 3680 (Part 4) having theoretical stage-discharge relationship o Compound structure meeting BS 3680 (Part 4) having theoretical stage-discharge relationship. 2) Structures deviating from ISO/British Standard but where the theoretical stage-discharge relationship is used regardless.
Ultrasonic	Permanent gauging station based on transit-time ultrasonics, installed and operated to BS 3680 (Part 3E).
Electromagnetic	Permanent electromagnetic (EM) gauging station with either a suspended or buried induction coil, installed and operated to BS 3680 (Part 3H).
BS Structure with rated section at high flows.	Structures formally supported by use of a rating curve at high flows, but operating to ISO/BS at other times.

5.5 Grading of Attributes

Selection of attributes

The attributes adopted in the classification have been chosen following consultation with data users and providers, discussion at the York workshop and further discussion with the Project Board. The chosen attributes are an attempt to provide a reasonably complete description of factors that may affect data quality. This necessarily involves seeking a balance between excessive detail, which would render the classification unworkable from a practical point of view, and not capturing enough information to provide a realistic assessment of data quality.

The classification uses 25 core attributes. These are described in more detail in Section 7. However, considering those attributes that may be evaluated for two or more flow ranges, or for two or more different gauging station types, brings the total number of attributes to 77. On a broad level these can grouped as follows:

- those relating to the uncertainty of the flow measurement (or stage-discharge relationship),
- those relating to the agreement of check gaugings,
- those relating to local factors which may influence the measurement such as out-of-bank flows, weed growth, or deterioration of the gauge.
- those relating to the reliability of the gauge or recorder and the number of missing data,
- and those relating to the accuracy of the stage measurement.

The attributes are, as far as possible, quantitative and objective. However, some of the factors affecting data quality have proven difficult to define quantitatively, especially those relating to weed management, by-passing and missing data. In such cases, it has been concluded that hydrological judgement is more important and a more robust way of determining the effect on data quality. In such cases, qualitative assessments are therefore used.

Not more than 20 attributes are considered for any one scoring scheme, of which no more than 8 apply to one particular flow range (as described in Section 6). If larger numbers of attributes were to be considered, it would be far more difficult to assign each one an appropriate weight and, perhaps more importantly, a much larger amount of input data would be required in order to calculate attribute values. In the proposed scheme input data is kept to a minimum and in many cases the same basic information may be used to derive several different attributes. Appendix C summarises the input fields required.

Look-up tables for attribute grades
Each attribute is assigned an integer grade between 1 and 5, with 5 being the best grade.

For numeric attributes, grades have been assigned by dividing the range of possible values into five intervals. These need not necessarily be equal, and interval widths are weighted if appropriate (e.g. the interval width for a grade of 3 may be wider than that for a grade of 5). This information is stored in a look-up table for the attribute.

Nominal grades are also given for descriptive attributes. Owing to the rather less precise definition of the descriptive attributes, broader nominal grades of 1, 3 or 5 have generally been assigned (with the grades of 2 and 4 not being used). Alternatively (and only where appropriate) the attribute is compartmentalised into five categories assigned grades of 1 to 5. Descriptive attributes are subjective to the views of the individual entering data into the scheme. To minimise the impact of this, users of the scheme are directed to follow more detailed guideance on selecting an appropriate categories, as explained in Section 7 and in the Software User Guide. Although it is subjective, this categorical approach allows for a degree of flexibility and judgement that is often needed. As with numeric attributes, the rules to convert attribute values into grades are stored in a look-up table.

A grade of 5 does not imply perfection. Rather grades have been set realistically, with a grade of 5 aligned to good practice where appropriate, or to other reasonable targets. Values used in look-up tables were refined by reference to worked examples.

In the scoring procedure the attribute grades are assigned automatically with reference to the appropriate look-up tables. Figure 5.2 illustrates this procedure.

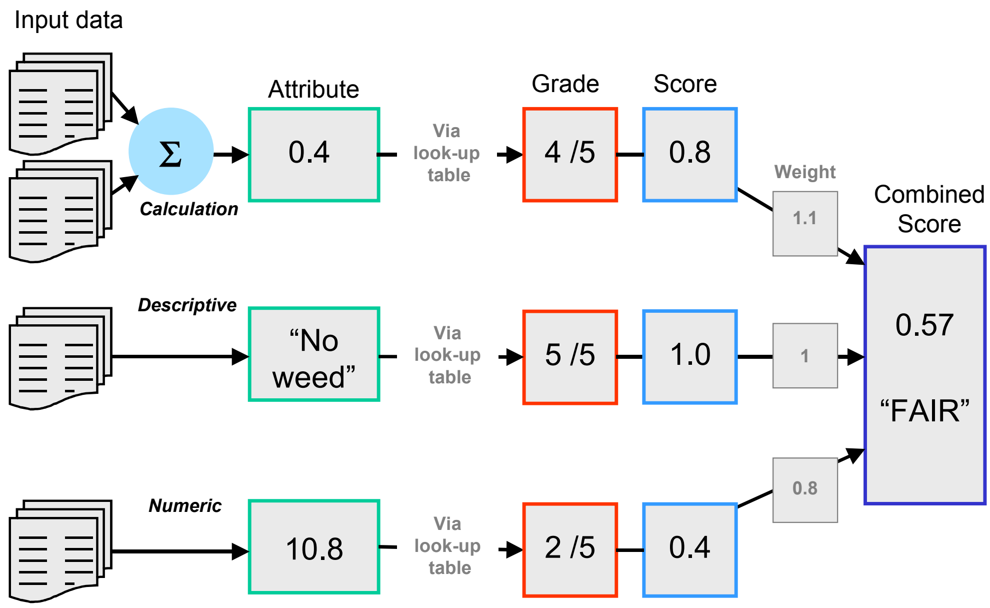


Figure 5.2: Procedure for grading and combining attributes

Attribute Weights

Attribute weights determine the influence of a particular attribute (whatever grade) on the score achieved for the category. All weights are adjusted from an initial value of 1. The weight is reduced for attributes that are considered less important, and increased for those that are considered more important. The uncertainty associated with the flow measurement is the key attribute in the schemes. This importance is reflected by the relatively high weights applied (1.4-1.6). In the final classification weights were refined on the results of the benchmarking exercise.

5.6 Combination of Attributes

Two different methods were considered for combining grades for individual attributes to create a final score. The scoring scheme was initially developed using weighted linear combinations of attribute grades ('additive' scoring). However a multiplicative scheme based upon the geometric mean of attribute grades was adopted after the prototype stage. The two alternatives are discussed further below. Although the additive scheme was initially proposed because of its simplicity, it will be seen that the multiplicative scheme later adopted has certain advantages.

Additive scoring scheme
The additive scoring effectively supposes that each attribute cumulatively 'adds quality' to the overall score. Weights are used to adjust the amount added in each case (having a set of weights that always sum to a fixed value can be used to ensure even-handedness between different station types and flow categories). The combined weighted score, S_A, achieved under the additive scheme for a category containing *n* attributes can be written as

$$S_A = (g_1 w_1) + (g_2 w_2) + ...(g_n w_n) \quad (1)$$

where r_i is the grade for attribute *i* and w_i is the corresponding weight.

The additive scoring method is conceptually straightforward, but is not in fact a realistic approach, as attributes that have a negative effect on the data quality should really 'subtract' from the combined score, if the scores are not to be biased. For example, a station might be good at high flows in respect of several attributes (e.g. uncertainty in rating, accuracy of level instrument, no long gaps between gaugings), but if the station was, for the sake of argument, subject to significant by-passing then this one factor should strongly reduce the combined score, rather than adding to it.

In the additive approach, this kind of reduction was difficult to achieve without assigning a large weight to the single significant low-scoring attribute, which would then also tend to dominate the combined score if it had a good grading of 4 or 5. There is also some difficulty in arranging the weights for the case where two or more attributes should be able to exert negative control on the combined score. Yet it is arguable, for most of the attributes chosen for the quality classification, that the combined score should be reduced strongly if there are low scores for individual attributes.

Multiplicative scoring
The above arguments point towards a multiplicative scheme for combining attribute grades. A convenient way to calculate a weighted combination of *n* individual attribute grades is to use the weighted geometric mean, S_G, which can be written as:

$$S_G = \sqrt[n]{r_1^{w1} r_2^{w2} r_3^{w3} \ldots r_n^{wn}} \quad (2)$$

where r_i is the <u>score</u> for attribute *i* and w_i is the corresponding weight. A feature of the approach is that the multiplicative nature of S_G means that the final combined score will fall within the range of 0.0 and 1.0 and thus is more consistent with the way individual attributes scores are presented with a lower limit of 0.2 and a maximum of 1.0.

Comparison of scoring methods
Table 5.3 illustrates the original arithmetic scoring scheme in comparison with the multiplicative scheme. The example shows a simple case, where three attributes are all graded as 3 out of 5. The weights have all been set to equal 1.0 and the final combined scores have been standardised as a percentage, to allow an even comparison of the schemes. It can be seen that the schemes produce the same total score.

Table 5.4 shows an example where the grading (but not the weights) varies between attributes. Two attributes have been given a very good grade of 5, while one has a very poor grade of 1. There is now a difference between the scoring schemes; the multiplicative scheme generates a lower overall score, reflecting more strongly the influence of the one very poor attribute.

Table 5.3: Comparison of weighted scores (1) – uniform grades

	Grade, g	Score, r	Weight, w	g x w	r^w
Attribute 1	3	0.6	1	*3*	*0.6*
Attribute 2	3	0.6	1	*3*	*0.6*
Attribute 3	3	0.6	1	*3*	*0.6*
Additive scheme combined score, S_A				9	
S_A as % of maximum achievable				**60%**	
Geometric mean combined score, S_G					0.216
S_G as % of maximum achievable					**60%**

Table 5.4: Comparison of weighted scores (2) – varying grades

	Grade, g	Score, r	Weight, w	g x w	r^w
Attribute 1	5	1.0	1	*5*	*1*
Attribute 2	5	1.0	1	*5*	*1*
Attribute 3	1	0.1	1	*1*	*0.2*
Additive scheme combined score, S_A				11	
S_A as % of maximum achievable				**73%**	
Geometric mean combined score, S_G					0.2
S_G as % of maximum achievable					**58%**

In the geometric approach, weights applied to individual attribute scores have a non-linear effect, so that the range of weights representing important and not so important attributes is smaller than in the additive approach. The total (sum) of weights must also be equal to the number of attributes. Table 5.5 shows an illustration of the geometric mean scoring for real quality attributes (in this case for a standard structure within the low flow range).

Table 5.5: Calculating scores

Attribute	**Weight**	**Grade**	**Weighted Score**
Width of 95% confidence interval based on BS3680 (% Q95)	1.6	1	0.08
Significance of missing data	0.5	3	0.77
Accuracy of level measurement (mm)	0.5	4	0.89
Weed growth management	1.4	3	0.49
Percentage archived flows within ±15% of gauged (flows below Q95)	1	5	1.00
Combined score			**0.481**
Abbreviated classification			**FAIR**

5.7 Abbreviated Score

For many users of the scheme the numeric score will have little meaning. An abbreviated label is therefore applied to each scheme. Descriptive labels 'CAUTION', 'FAIR' and 'GOOD' are assigned to bands within the range of possible combined scores as shown in Table 5.6.

Boundaries between CAUTION, FAIR and GOOD were not selected arbitrarily. Rather the boundary between CAUTION and FAIR was determined by calculating the typical score when all attributes were assigned common values of {2,2,2,…,2}, whilst an appropriate value for the boundary between FAIR and GOOD was determined using common values of {4,4,4,…,4} for all attributes.

Table 5.6: Abbreviated Scores

Combined score	Label
score < 0.55†	CAUTION: Data should be treated with caution, although it may still be of use.
0.55† ≤ score < 0.7	FAIR: The overall quality of the data is reasonable, but in some aspects the data- quality will be poor. The user should investigate in further detail.
0.7 ≤ score < 1.0	GOOD: The quality of data is generally of a good standard, aligning to best practice.

Notes
Threshold originally set to equal 0.4, but revised following benchmarking tests (see Section 9)

As noted in Table 5.6, the threshold between CAUTION and FAIR was originally set in this way to equal 0.4, but revised to 0.55 following a benchmarking exercise, described in section 9 of this report.

6 SCORING SCHEMES

6.1 Introduction

This section of the report describes the scoring schemes developed for each of the types of gauging station within the data quality classification. The scoring schemes lie at the heart of the classification. They define the way in which quality grades are assigned to each attribute, and the weighting applied to the attributes. The attributes are introduced here with their basic definitions and weights. More precise definitions, details of the calculation and the grading of the attributes are given in the following Section 7.

6.2 Scoring scheme 1 – Rated sections

The Rated Section Scheme applies primarily to open channel sections (with either a natural or artificial control) where discharge is routinely derived using a stage-discharge relationship of the form

$$Q = c\,(h + a)^{b} \qquad (3)$$

where Q is the discharge through the cross section (m^3s^{-1}), h is the stage (m above datum), and c, a and b are constants.

It is assumed in principle that the relationship is derived via least squares regression through a set of discharge-stage pairs (although this assumption need not hold for the classification to be valid). It is assumed that the gaugings are independent single measurements of discharge, probably based on the use of hand-held propeller or EM flow meters in the conventional manner, or are derived using techniques such as ADCP or other portable ultrasonics. Table 6.1 lists the attributes used in the scoring scheme and the weights applied to each attribute.

Table 6.1: Rated Section Scheme

Code	Attribute description	Weight	Type of attribute
R-H1	Width of 95% confidence interval at QMED (as a % of QMED)	1.6	Numeric
R-H2	Significance of missing data	0.5	Descriptive
R-H3	Effective accuracy of level measurement (mm)	0.7	Numeric
R-H4	Occurrence of unmeasured bypass flow	1	Descriptive
R-H5	Average annual number of gaugings at flows over 0.5 x QMED	1	Numeric
R-H6	Maximum gauged flow ÷ maximum archived flow	1.4	Numeric
R-H7	Longest gap length between gaugings at flows over 0.5 x QMED (years)	0.8	Numeric
R-L1	Width of 95% confidence interval at Q95 (as a % of Q95)	1.6	Numeric
R-L2	Significance of missing data	0.5	Descriptive
R-L3	Effective accuracy of level measurement (mm)	0.7	Numeric
R-L4	Sensitivity (%)	1.2	Numeric
R-L5	Average annual number of gaugings at flows below Q95	1	Numeric
R-L6	(Q95-minimum gauged flow) ÷ (Q95 - minimum archived flow)	1	Numeric
R-L7	Longest gap length between gaugings at flows below Q95 (years)	0.8	Numeric
R-L8	Weed growth management	1.2	Descriptive
R-G1	Standard error of estimate (as a % of mean daily flow)	1.2	Numeric
R-G2	Average annual number of missing daily flows	0.8	Numeric
R-G3	Effective accuracy of level measurement (mm)	1	Numeric
R-G4	Average annual number of check gaugings	1	Numeric
R-G5	Gauged flow range ÷ archived flow range	1	Numeric

The key attribute is the width of the 95% confidence interval at specified index flow rates (QMED, Q95 and mean daily flow). The confidence interval is a measure of uncertainty about the fitted rating curve, based on the information available in a set of flow gaugings. Its statistical basis is discussed in detail in Section 7 of this report. The main points to note here are that the confidence interval width takes into account the goodness of the fit between the rating curve and the flow gaugings, the number of gaugings, their scatter, and the range of flows or stage over which gaugings have been taken. For the General flow score, the scheme uses the standard error of estimate across the whole flow range (expressed as a percentage of the mean daily flow), rather than the confidence interval (which changes over the flow range).

Attributes relating to the gaugings used to derive the rating curve also figure highly in the scheme. A suite of attributes are used to evaluate whether the number, coverage (of the flow range) and frequency of the gaugings are sufficient, complementing the confidence interval attribute to build up a picture of the overall uncertainty associated with the rating. These include the average annual number of gaugings (in the index range), the longest gap length between gaugings and a comparison between the highest gauging and highest observed flow (representing the extension of the rating curve beyond the observed range).

The remaining attributes mainly deal with local and environmental factors such as missing data, the management of weed growth and whether the cross-section is by-passed at high flows. With the exception of weed growth, these factors are assigned relatively low weights.

One complication at rated sections is the use of stage-discharge adjustment (shift) procedures. These are used to adjust the rating curve incrementally to account for the effects of vegetation growth during the summer season. Typically, a winter base curve will be fitted to flow gaugings taken in the absence of vegetation and the ‘shift parameter’ (a in Eqn (3) above) is adjusted at, say, weekly or monthly intervals to update the rating curve. This adjustment should be based on the deviation between the base curve and one or more flow subsequent gaugings. The advantage of the approach is that it provides a quick and relatively easy way to maintain an applicable rating curve under changing conditions. The main disadvantage is that the shift is very much reliant on a small number of gaugings, which may be subject to error.

Shift procedures are difficult to incorporate within the statistical analysis of the rating curve. This is in part because there are effectively multiple rating curves in operation that could have different statistics and, perhaps more significantly, because there are possible variations in the exact methods used (for example, whether or not shifts are based on averaged or interpolated deviations from a sequence of gaugings or on the deviation at the end of a period of weed growth). Within the gauging station data quality classification, the use of shift procedures is therefore treated as a form of weed growth ‘management’. Where shift procedures are applied, the scoring scheme interprets this as indicating worst-possible uncertainty in the rating curve at low flows combined with partial management of weed growth.

6.3 Scoring scheme 2 – Structures

The BS Structures Scheme applies to both weirs and flumes. These should in principle have been designed and operated to British/ISO Standard, although the scheme in fact allows for penalties to be applied where there are deviations from the standards. The scheme is appropriate for the following types of structures

- Thin plate weirs
- Broad-crested weirs
- Triangular profile weirs
- Compound weirs
- Flumes

For structures that have a non-standard design or that have been severely damaged it is assumed that the theoretical weir equations no longer apply, and that the station would be treated as a rated section, albeit one with an artificial control. Table 6.2 lists the attributes used in the BS Structures Scoring Scheme. The scheme is necessarily more complex than for rated sections, mainly to try to account for the possible correction procedures that can be applied to compensate for the effects of non-modularity at high flows.

Table 6.2: BS Structures Scheme

Code	Attribute description	Weight	Type
S-H1	Width of 95% confidence interval based on BS3680 (as a % of QMED)	1.6	Numeric
S-H2	Significance of missing data	0.5	Descriptive
S-H3	Effective accuracy of upstream level measurement (mm)	0.7	Numeric
S-H4	Occurrence of unmeasured bypass flow	1.2	Descriptive
S-H5	Deviation from BS / Percentage archived flows within ±10% of gauged (flows over 0.5 x QMED)	1.0	Descriptive / Numeric
S-H6	Corrections applied for non-modular flows	1.0	Descriptive
	• *Modular limit, if non-modular flows occur within high flows range*	*0.15*	Numeric
	• *Average annual number of gaugings at flows over 0.5 x QMED*	*0.10*	Numeric
	• *Maximum gauged flow ÷ maximum archived flow*	*0.10*	Numeric
	• *Longest gap length between gaugings (years)*	*0.05*	Numeric
	• *Effective accuracy of tailwater stage measurement (mm)*	*0.05*	Numeric
S-L1	Width of 95% confidence interval based on BS3680 (as a % of Q95)	1.6	Numeric
S-L2	Significance of missing data	0.5	Descriptive
S-L3	Effective accuracy of (upstream) level measurement (mm)	0.5	Numeric
S-L4	Sensitivity (%)	1.2	Numeric
S-L5	Weed growth management	1.2	Descriptive
S-L6	Deviation from BS / Percentage archived flows within ±15% of gauged (flows under Q95)	1.0	Descriptive / Numeric
S-G1	Width of 95% confidence interval based on BS3680 (% of mean daily flow)	1.5	Numeric
S-G2	Average annual number of missing daily flows	0.8	Numeric
S-G3	Effective accuracy of level measurement (mm)	0.7	Numeric
S-G4	Modular Range	1.0	Numeric
	• *Approx. average annual number days in which non - modular flow occurs*	*0.2*	Numeric
S-G5	Deviation from BS / Percentage archived flows within ±15% of gauged (full flow range)	1.0	Descriptive / Numeric

For structures that have a non-standard design or that have suffered severe deterioration over time it is assumed that theoretical weir equations no longer apply. The BS Structures Scoring Scheme would then not be strictly appropriate; instead the station would be regarded as a rated section, albeit one with an artificial control. However it is recognised that non-standard structures sometimes continue to be operated using theoretical weir equations. In such cases check gaugings should highlight the imprecise flow measurement and a low score will be obtained if the site is classified under the BS

Structure scheme. In the unlikely event that no check gauging has been carried out at the site, the perceived deviation from BS/ISO design is considered under the SH6, SL6 and SG5 attributes.

Some structures are managed as hybrid stations with a formal rating curve applied over parts of the flow range. This approach is used in particular where the modular limit of the structure is reached at a relatively low point on the flow duration curve resulting in uncertain measurement of discharge during periods of high flow, or if there are significant differences in the stage-discharge relationship between rising and falling limbs. For the purposes of the classification gauges managed in this way may be treated under either the BS Structure Scoring Scheme or the Structure with Formal Rating at High Flows Scheme. There is no hard-and-fast rule to say when to use one scheme or the other, but as a guide, the latter scheme was intended for use only where a rating is applied exclusively to flows higher than about 0.5 x QMED, and is not appropriate if a rating is applied only during the low flow part of the flow duration curve. Also it should not be used if the gaugings corresponding to the rated control are not readily available (as these are required in order to calculate vital error statistics used in the scoring).

In common with the other schemes, the width of the 95% confidence interval is the key attribute in each category (e.g. SH1, SL1, and SG1). However as current meter gaugings may be regarded as less accurate than a well-maintained structure operating within its modular range, it is not considered appropriate in this case to calculate confidence interval on the basis of a sample of check gaugings. Instead confidence interval widths are derived from theoretical uncertainties discussed in BS 3680 *'Measurement of Liquid Flow in Open Channels, Part 4: Weirs and Flumes*', and quoted in terms of a 95% confidence interval, which is consistent with the approach adopted for ratings derived from flow gaugings. The derivation of these uncertainties is discussed further in Section 7.

Acknowledging that check gauging can be a useful way of independently confirming the performance of a structure, especially during the non-modular range, check gaugings have been incorporated into the scheme as an optional measure. Where check gaugings are used, the degree of agreement with archived flow has been determined based on the proportion of the gaugings that lie within either ±10% (for high flows) or ±15% (low flows and general categories) of the corresponding archived flow (e.g. attributes SH5, SL6, and SG6). Where no check gaugings are available to substantiate the performance of the gauge, a score is assigned based on a descriptive attribute that reflects local knowledge about whether there are deviations from BS3680 standards at the station.

Non-modularity

The scheme includes attributes that account for any correction procedures applied to compensate for the effects of non-modularity on the flow record (SH6). Three types of correction procedure are considered, namely use of a rating curve, use of a tailwater stage recorder to determine head loss over the structure and crest-tapping to measure the pressure loss over the structure. The flow at which the modular limit of the structure is reached is also an attribute in the General category of the scheme (SG4).

Often knowledge of the modular range of a structure itself is rather vague, and it is difficult to produce a definitive assessment of the uncertainty associated with corrected flows. For the purpose of assessing data quality, descriptive attributes were therefore chosen to represent the modular limit and the broad effects of non-modularity, and whether correction factors, tailwater measurement or crest tapping have been applied. Discussion amongst the project team and Agency project board led to a view that the following factors would most likely improve the quality of data, ranked from worst-case to best-case:

- No correction applied for non-modular flows,
- Rating used over non-modular range,
- Correction based on tailwater stage measurements,
- Correction based on crest tapping.

However, it is clear that the quality of data will depend not just on the type of correction applied, but also the efficacy of the application. For example, use of a tailwater measurement may not greatly improve data quality if the tailwater gauge is itself unreliable or inaccurate. The attribute for non-modular corrections therefore includes a set of secondary attributes that modify the grade given to the basic descriptor. The rules adopted to modify the attribute grade are explained in Section 7.

6.4 Scoring scheme 3 – Structures with Rated Section at High Flows

This scheme is appropriate where structures are managed as hybrid gauges with a rating curve applied during periods of high flow (for guidance, roughly where flows are equal to or greater than 0.5 x QMED) but treated as conforming the BS/ISO Standard (i.e. theoretical weir equations apply) at low flows. The high flows rating must be of the same functional form assumed for the rated section. The scheme is an amalgamation of the Rated Section Scheme applied within the High Flows range and the BS Structures Scheme applied for Low Flows and General categories. The attributes used in the scheme (shown in Table 6.3) are therefore discussed further in the relevant sections of the report.

Table 6.3: BS Structure operating a rating at high flows

Code	Attribute description	Weight	Type
R-H1	Width of 95% confidence interval at QMED (as a % of QMED)	1.6	Numeric
R-H2	Significance of missing data	0.5	Descriptive
R-H3	Effective accuracy of level measurement (mm)	0.7	Numeric
R-H4	Occurrence of unmeasured bypass flow	1.0	Descriptive
R-H5	Average annual number of gaugings at flows over 0.5 x QMED	1.0	Numeric
R-H6	Maximum gauged flow ÷ maximum archived flow	1.4	Numeric
R-H7	Longest gap length between gaugings at flows over 0.5 x QMED (years)	0.8	Numeric
S-L1	Width of 95% confidence interval based on BS3680 (as a % of Q95)	1.6	Numeric
S-L2	Significance of missing data	0.5	Descriptive
S-L3	Effective accuracy of (upstream) level measurement (mm)	0.5	Numeric
S-L4	Sensitivity (%)	1.2	Numeric
S-L5	Weed growth management	1.2	Descriptive
S-L6	Deviation from BS / Percentage archived flows within ±15% of gauged (flows under Q95)	1.0	Descriptive / Numeric
S-G1	Width of 95% confidence interval based on BS3680 (as a % of mean daily flow)	1.5	Numeric
S-G2	Average annual number of missing daily flows	0.8	Numeric
S-G3	Effective accuracy of level measurement (mm)	0.7	Numeric
S-G4	Modular Range	1.0	Numeric
	Approx. average annual number days in which non - modular flow occurs	0.2	Numeric
S-G5	Deviation from BS / Percentage archived flows within ±15% of gauged (full flow range)	1.0	Descriptive / Numeric

6.5 Scoring scheme 4 – Ultrasonic Stations

The Ultrasonic Scoring Scheme refers specifically to those stations where transit-time ultrasonics (acoustics) are used to gauge discharge through the channel. The scheme is not appropriate for sites gauged using other acoustic devices e.g. velocity-area stations with ratings developed using acoustic Doppler profiling methods should be classified using the rated sections scheme. Whilst acoustic Doppler systems are likely to be adopted for continuous flow measurement in the near future, at present they do not fall within the scope of the gauging station data quality classification.

Table 6.4 describes the attributes used in the scoring scheme for ultrasonic gauges. Unlike the other gauging methods, the ultrasonic method does not depend on a stage-discharge relationship. Rather the transit time of the ultrasonic pulse is used to determine the velocity of water movement, and stage is considered explicitly only in the way this is integrated across the column. The scheme is therefore dominated by attributes describing the configuration of the equipment including the number of paths, and the arrangement of transducers (e.g. UH5, UL6). Knowledge of the bed-profile is also more important for the ultrasonic scheme, as represented by the inclusion of the UG4 attribute describing the frequency of bed-level surveys.

Table 6.4: Ultrasonic Scoring Scheme

Code	Attribute description	Weight	Type
U-H1	Width of 95% confidence interval based on BS3680 (as a % of QMED)	1.6	Numeric
U-H2	Significance of missing data	0.5	Descriptive
U-H3	Effective accuracy of level measurement (mm)	0.5	Numeric
U-H4	Occurrence of unmeasured bypass flow	1.0	Descriptive
U-H5	Height of uppermost path ÷ max. archived stage	1.4	Numeric
U-H6	Deviance from BS / Percentage archived flows within ±10% of gauged (flows over 0.5 x QMED)	1.0	Numeric
U-L1	Width of 95% confidence interval based on BS3680 (as a % of Q95)	1.6	Numeric
U-L2	Significance of missing data	0.5	Descriptive
U-L3	Effective accuracy of level measurement (mm)	0.5	Numeric
U-L4	Weed growth management	1.4	Descriptive
U-L5	(H95 - height of lowermost path) ÷ (H95 - mean bed level) (%)	1.0	Numeric
U-L6	Deviation from BS or % archived flows within ±15% of gauged (flows below Q95)	1.0	Numeric
U-G1	Width of 95% confidence interval based on BS3680 (as a % of mean daily flow)	1.2	Numeric
U-G2	Average annual no. of missing daily flows	0.8	Numeric
U-G3	Effective accuracy of level measurement (mm)	0.5	Numeric
U-G4	Average annual number of bed-level surveys	1.2	Numeric
U-G5	Deviation from BS or % archived flows within ±15% of gauged (full range)	1.3	Numeric

In this case the computation of uncertainty for ultrasonic stations follows the calculation method described in BS3680, Part 3E[20], and expressed, for consistency with the other schemes, as a 95% confidence interval (attributes UH1, UL1, UG1).

Check gaugings are incorporated within the scoring scheme in the same way as for BS/ISO structures. If entered (optional) the percentage gaugings within 10% or 15% of the archived flows are calculated (e.g. UH6, UL6, UG5). Where no check gaugings were available to substantiate the performance of the gauge, local effects are taken into consideration.

[20] BS3680 Measurement of Liquid Flow in Open Channels. Part 3E: Stream Flow Measurement of discharge by the ultra-sonic (acoustic) method.

6.6 Scoring scheme 5 – Electromagnetic Stations

The Electromagnetic Scoring Scheme is appropriate for buried permanent electromagnetic (EM) gauging stations with either a suspended or buried induction coil, installed and operated to BS 3680 (Part 3H). The attributes considered in the scheme are shown in Table 6.5.

Table 6.5: Electromagnetic Scoring Scheme

Code	Attribute description	Weight	Type
E-H1	Standard error of deviations for flows over 0.5 x QMED (as a % of QMED)	1.6	Numeric
E-H2	Significance of missing data	0.6	Descriptive
E-H3	Effective accuracy of level measurement (mm)	0.8	Numeric
E-H4	Occurrence of unmeasured bypass flow	1.2	Descriptive
E-H5	Average annual number of gaugings at flows over 0.5 x QMED	0.8	Numeric
E-H6	Maximum gauged flow ÷ maximum archived flow	0.8	Numeric
E-L1	Standard error of deviations for flows below Q95 (as a % Q95)	1.6	Numeric
E-L2	Significance of missing data	0.5	Descriptive
E-L3	Effective accuracy of level measurement (mm)	0.7	Numeric
E-L4	Average annual number of gaugings at flows below Q95	1	Numeric
E-L5	(Q95-minimum gauged flow) ÷ (Q95 – minimum archived flow)	1	Numeric
E-G1	Standard error of deviations for full range of flows (as a % of mean daily flow)	1.2	Numeric
E-G2	Average annual number of missing daily flows	0.7	Numeric
E-G3	Effective accuracy of level measurement (mm)	0.7	Numeric
E-G4	Integrity of insulating membrane around coil	1.2	Numeric
E-G5	Average annual number of check gaugings	1	Numeric

It was felt that the input data required to calculate uncertainty associated with an EM gauge using the method reported in British Standard BS3680, Part H, would not be readily available and unlikely to be held on the WISKI database. Rather an approach similar to that used for rated sections (based on check gaugings) was adopted, although the attributes are represented in terms of standard error of deviations rather than width of 95% confidence interval (EH1, EL1, EG1).

The integrity of the coil insulation/membrane is an important influence on data quality and is included as attribute EG5. The only other significant feature of the scheme is that weed growth is not included as an attribute as EM gauges are generally tolerant of both in-stream and bank vegetation.

6.7 Scoring Scheme 6 – Level-only Stations

The Level-only Scoring Scheme is applicable for sites where a level recorder is used. Unlike the other scoring schemes it is only evaluated for the General category, and includes just eight attributes. These attributes are listed in Table 6.6.

Table 6.6: Attributes in the Level-only Scoring Scheme

Code	Attribute description	Weight	Type
L-G1	Type of instrument	2	Numeric
L-G2	Non-capture rate (percentage missing data)	0.8	Descriptive
L-G3	Accuracy of level measurement (mm)	1.2	Numeric
L-G4	Average annual number of manual checks for level	1.2	Numeric
L-G5	Truncation (of measured level) at high flows	0.4	Descriptive
L-G6	Truncation (of measured level) at low flows	0.4	Descriptive
L-G7	Weed growth management	1.2	Descriptive
L-G8	Siltation management (of intake pipe & stilling well)	0.8	Descriptive

This scheme is focussed on the effective accuracy of stage measurement. The instrument type (LG1) and precision of measurement (LG3) are considered, in addition to factors that influence the accuracy of measurement such as the possibility of the measurement being truncated (LG4, LG5) and the effect of siltation (LG8) and weed growth (LG7). The reliability of the gauge is also considered (LG2).

6.8 Setting Weights and Grades

A great strength of the attribute scoring approach is its flexibility. This flexibility does, however mean that there are many possible (and possibly equivalent) ways in which weights can be assigned to attributes, and in which the attributes themselves can be graded (i.e. assigned their individual scores, which we express for convenience as a grade 'out of five'). The process by which weights and grades are set can be thought of as the tuning of the GSDQ classification.

The first decision in building the classification was the weight that should be assigned to each attribute. The weights, which have been tabulated in this section of the report, represent expert judgements about the relative importance of the attributes in determining data quality. As a general principle, greatest weight was given to statistical measures of uncertainty in the measurement of discharge and to the primary physical factors such as by-pass flow, weed growth or the elevations of the limiting paths for ultrasonic stations. Many data users and providers value the confirmation provided by check gaugings, and attributes relating to check gaugings were therefore given relatively heavy weights. The weights were agreed during discussions between the R&D team and the Environment Agency project board during the period November 2002 to June 2003 and adjusted following a benchmarking exercise reported in Section 9.

Grades have been assigned to different ranges of values (for quantitative attributes) or categories (for categorical attributes) by a similar process of judgement, based on experience of what is achievable, what is required within British Standards and values for parameters suggested in recent R&D on hydrometric good practice. The grades assigned to attributes are set out in look-up tables in Section 7. Where the same attribute is used for more than one station type, the grading scheme is kept the same, as far as possible, to ensure even-handedness across station types. The ranges of attribute values assigned to each grade do, however, vary between flow ranges, reflecting judgements about the achievable or necessary quality parameters at different flows.

Weights and grades were set during trials of the GSDQ procedures on 14 stations, spread over all represented station types. The setting of weights and grades is not, however, a calibration of the classification because there is no absolute measure of data quality, and hence no single independent standard against which to calibrate. Instead, it is best to regard the classification an ‘expert system’ and it is possible that the optimum grading schemes and weights will only become apparent once a large number of station classifications have been completed.

7 ATTRIBUTE DESCRIPTIONS

7.1 Introduction

The attributes used in the classification encompass influences on data quality that are specific to one or more gauging techniques, as well as those that are more generic in nature such as properties of the flow record, flow regime and gauging station site. For each station type and scoring scheme, the relevant attributes are selected and evaluated. This section defines the attributes and sets out the look-up tables used to assign grades (and hence scores) to each attribute for the different flow ranges.

We have broadly grouped attributes into the following categories.

Quantification of the uncertainty associated with the determination of discharge
These attributes consider the uncertainty associated with the determination of discharge from the basic variables measured at the station. For a rated section or BS-compliant structure this will be a stage/discharge relationship, or rating curve. Ultrasonic and electromagnetic gauges are essentially velocity-area stations, and so do not rely on a stage-discharge relationship, theoretical or otherwise. The specific attributes are:

1. Standard error of estimate (for the general flow regime).
2. Width of 95% confidence interval at QMED, Q95 and mean daily flow (DMF).
3. Standard error of deviations for check gaugings in high, low and general flow ranges.

For rated sections a statistical analysis of the rating curve can be used. For structures, simplified estimates of uncertainty have been derived from BS 3680 (Part 4). For ultrasonics, the uncertainty of each flow measurement is associated with the estimation of transit time and distance of the ultrasonic pulse, and depends on the configuration of the gauge, number of paths used, path angles and path heights. It can be determined using methods described in BS 3680 (Part 3E, ISO 6416).

Similarly for EM gauges, the configuration of the EM coil has a large influence on the calculation of the uncertainty using BS/ISO methods. However in this case it is felt that the BS/ISO methods are too complex to adopt, and that it is more appropriate (and also a realistic reflection of the view of operators of EM stations) to evaluate uncertainty based on flow gaugings used to calibrate the instrument.

Applicability of theoretical stage-discharge relationship.
The following attributes consider the likelihood and significance of the gauging structure being operated beyond its theoretical range.

4. Method of correcting for non-modular flow over a structure (including effective accuracy of tailwater gauge or crest-tapping, if used)
5. Modular limit of structure
6. Proportion of time modular limit exceeded

Effective accuracy of stage measurement
This group of attributes describes the confidence assigned to stage measurement at the site, incorporating both the inherent accuracy associated with instruments used, and site effects.

7. Effective accuracy of stage measurement (high flows, low flows and general ranges).

Confirmation by independent check gaugings
The use of these attributes is twofold. For the rated section they provide an additional evaluation of the gaugings used in the stage-discharge relationship, whilst for the other gauge types they provide a means to evaluate the degree of deviation from BS/ISO.

8. Average annual number of check gaugings in high flows, low flows and general ranges.
9. Ratio of gauged range / station (archive) range for high flows, low flows and general regime.
10. Gap length between successive gaugings in high flows, low flows and general ranges.
11. Percentage of gaugings where station flows deviate from gaugings by less than 15% (low and general flow ranges) or 10% (high flow range).

Gauging station configuration
These attributes are used as additional checks on data quality related to the disposition and operation of the station.

12. Average annual number of bed-level surveys.
13. Ratio between ultrasonic path height range (high flows and low flows).
14. Sensitivity of weir.
15. Integrity of membrane around EM coil.

Reliability of gauging station and significance of missing data
These attributes are used to evaluate the amount and impact of missing data.

16. Average annual of missing entries on mean daily flow archive
17. Significance of missing data at high flows and at low flows

Local factors relating to siting of gauging station
These attributes are used to ensure local knowledge regarding the site is incorporated into the schemes.

18. Effect of weed growth and management practices adopted
19. Occurrence of unmeasured bypass flows

Level-only sites
These attributes refer specifically to level-only sites.

20. Type of instrument
21. Non-capture rate (percentage missing data)

22. Accuracy of instrument
23. Average annual number of manual checks on level
24. Truncation of measured level (at high flows and low flows)
25. Siltation of stilling well & management practice.

Of these core attributes, the uncertainty measures (attributes 1 to 3) play a key role in the classification, being assigned the highest weights in the scoring schemes. The statistical theory behind these attributes is discussed in detail in the next few sections, followed by definitions of the way in which the statistics have been used to formulate attributes of the classification scheme. The other attributes are then described, by category, in subsequent sections.

7.2 Derivation of Uncertainty (1) – Stage/discharge Relationship at a Rated Section

Confidence intervals are used to represent the degree of uncertainty about a flow value calculated on the basis of the rating relationship between stage and flow. It is general practice to express the relationship between flow q and stage h as a power law relationship that is linearised using a log-log transformation. The parameters are then estimated by fitting this relationship by least squares to a set of independent flow gaugings. As the procedure is analogous to linear regression (with the rating relationship being the regression line), the theory of regression analysis has been used to provide measures of confidence in the rating relationship.

Standard Error of Estimate
The Standard Error of Estimate *(SEE)* is a statistic used to measure the goodness-of-fit between a regression line (the stage discharge relationship) and a sample of N flow gaugings. For flow rating, it is appropriate to compute *SEE* using the log-transformed data, in which case

$$SEE = \sqrt{\frac{\sum_i \left(q_i^* - \hat{q}^*\right)^2}{N-2}}\ , \qquad (5)$$

where q_i is a gauged flow, $\hat{q}$ is the corresponding estimate from the rating curve at a given stage and we use the notation $q^* = \ln q$.

Standard Error of the Mean
The statistic *SEE* is a total measure of the goodness-of-fit of the rating relationship to a sample of gaugings. There will also be a range of uncertainty about the fitted line, which can be interpreted as an estimate of the mean value of flow for any given stage, in the logarithmic space. A suitable expression for this uncertainty is the Standard Error of the Mean Relationship,

$$SMR_{(\ln h)} = SEE\sqrt{\frac{1}{N} + \frac{\left(h^* - E(h_i^*)\right)^2}{\sum\left(h_i^* - E(h_i^*)\right)^2}} \qquad (6)$$

where $h^* = \ln h$, subscript i denotes the i^{th} gauging and $E(h_i^*)$ is the expected value, or mean, of log-stage for the gaugings. It will be noted that *SMR* varies according to the position on the rating curve relative to the mean stage of the sample of gaugings. The effect of this variation is that *SMR* is at a minimum at the point $E(h_i^*)$, and increases towards the lower and upper end of the rating, reflecting the greater uncertainty that exists as one moves away from the centre of the range of observed data.

Substituting $h^* = E(h_i^*)$, it can be seen that

$$SMR_{(\ln h = E(\ln h))} = \frac{SEE}{\sqrt{N}} \, . \tag{7}$$

Equation (6) has the form of the general expression for the sampling error in estimates of the mean, that is, the random variation that can be expected in estimates of the mean when calculated from samples of size *N*. Although the special case (7) is more straightforward to calculate, the information needed to compute *SMR* at any given stage should also be available if *SEE* has been computed.

The full expression in (6) has the advantage that it reflects both the degree of fit between the rating curve and the flow gaugings, and the degree to which a stage or flow falls within the range of stage/flow values at which gaugings have been made.

Calculation of 95% Confidence Intervals

The 100(1 - α)% confidence intervals for the flow calculated from the rating curve at a given stage *h* can then be computed from

$$\hat{q}^* \pm t_\alpha \left(SMR_{(\ln h)} \right) \tag{8}$$

where t_α is the critical value of the Student *t* distribution corresponding to a total probability of α under both tails of the distribution for $N - 2$ degrees of freedom. (The probability under each tail is $\alpha/2$, and $t_{\alpha/2}$ should be used for a one-tailed table of *t* values). For the gauging station data classification, we use 95% ($\alpha = 0.05$) confidence intervals, in line with earlier analysis. The confidence intervals, when plotted in log space around the rating relationship, are a pair of curved lines. The interval width is narrower at the centroid of the gauging data, and becomes wider towards each end of the range.

Strictly speaking, the preceding analysis requires an assumption that the rating equations used are a reasonable model for the variation of flow with stage, and that the differences between flow gaugings and rated flows conform to the distributional assumptions for linear regression. If QMED or Q95 are within a portion of the rating curve that has been extrapolated beyond the range of the highest or lowest flow gauging, then the confidence intervals will expand to reflect the lack of observational data in the extrapolated range. They will not incorporate errors in the extrapolated curve that could occur if the extrapolated relationship is physically unrealistic, for example because of a significant change in cross section for out of bank flows.

The confidence interval width has been chosen for use in the High and Low flow categories, evaluated at the points corresponding to QMED and Q95, respectively. For the Overall category, the simpler standard error *SEE* is used.

The calculation of confidence intervals for the classification system can be rather complicated, depending on the complexity of the rating relationship (number of limbs, number of changes in rating, number of gaugings). These calculations are carried out within the Excel spreadsheet tool to assist in calculating the 95% intervals. The Excel spreadsheet tool calculates *SEE* and *SMR* given a set of flow gaugings (as {flow, stage} data pairs) and rating curve segments, which are defined by the parameters *a*, *b* and *c* in the power law

$$Q = c(h + a)^b . \quad (9)$$

Where the rating curve is made up of multiple segments, we have currently treated these as a single function for calculating uncertainties. In equations (5) and (6) we have therefore taken the sums of squares in aggregate over all rating limbs, rather than evaluating separate piecewise statistics for each segment. This approach avoids inconsistencies at the boundaries between the segments.

Comparison with the 1995 NRA Gauging Station Classification

The treatment of uncertainty proposed here builds upon the 1995 classification scheme. In the 1995 scheme, uncertainty was expressed as a range ±(2 x SMR^{1995}). The statistic SMR^{1995} is in principle the same as the statistic defined by equation (7) above, but was written in terms of percentage deviations between gauged flows and the rating curve, without logarithmic transformation. The rating curve was split into three segments – those containing Q95, mean daily flow and QBAR (mean annual maximum flow). For each segment, a single value of SMR^{1995} was calculated, equivalent to the value for the mean gauged stage within the segment.

The two-tailed value for $t_{0.05}$ is approximately 2.0 for samples of size ~30 or more, and ±(2 x SMR^{1995}) therefore approximated a 95% confidence interval at the mid-point of the sample of gaugings within a rating segment.

Within the 1995 classification, quality codes are then based on the width of the interval, using what is in effect a standardised scoring rule. For example, a flow gauge is assigned to class '1' (the best score) if 2 x $SMR_{(0)}$ for the mid point of a segment of the rating curve is less than 6% of the rated flow at that point.

7.3 Derivation of Uncertainty (2) – British Standard Structures

For weirs and flumes operating to British Standard, the confidence interval width is derived from uncertainties associated with the theoretical weir equations. The uncertainty will therefore vary primarily depending on the type of structure, and design used. Theoretical uncertainties for different gauging structures are discussed in BS 3680 *'Measurement of Liquid Flow in Open Channels, Part 4: Weirs and Flumes*', specifically in the following sections:

- Part 4A: Thin plate weirs (1981) + Annex
- Part 4B: Triangular profile weirs
- Part 4D: Compound gauging structures (1989) + Appendix A, B, C
- Part 4E: Rectangular broad-crested weirs
- Part 4C: Flumes

For the purposes of the classification, fixed values have been adopted for each structure type, based on conservative estimates quoted in BS 3680 (Part 4). The standard values are quoted in terms of a 95% confidence interval, which is consistent with the approach adopted for ratings derived from flow gaugings. The values are shown in Table 7.1.

Table 7.1: Theoretical uncertainties for structures, after BS 3680 (Part 4)

Thin plate weirs	8%
Broad-crested weirs	10%
Triangular profile weirs	10%
Flumes	10%
Free overfall	20%

Notes
Figures are widths of the 95% confidence interval. For example, a conservative estimate of uncertainty at the 95% level is ±4% for a thin plate weir.

For compound weirs the total uncertainty can generally be considered as the root mean square uncertainty of the component weir types, although this requires an assumption that the components contribute equally to the overall uncertainty and are uncorrelated.

7.4 Derivation of Uncertainty (3) – Transit Time Ultrasonic Stations

The computation of uncertainty for ultrasonic stations follows the calculation method described in BS3680[21], part 3E. This method mainly takes account of uncertainty arising during the computation of cross-sectional area and mean velocity, and uses the first order (root mean square) method to determine the overall uncertainty as follows:

$$X_Q = \pm (X_{VL}^2 + X_d^2 + X_b^2 + X_p^2)^{1/2} \qquad (10)$$

where X_Q is the percentage uncertainty of the determination of discharge, X_{VL} is the percentage uncertainty in the determination of line velocity, X_d is the percentage uncertainty in the determination of depth of flow, X_b is the percentage uncertainty in the determination of channel breadth, and X_p is the percentage uncertainty that results from the limited number of paths used.

Taking the component factors in turn, the uncertainty arising during determination of line velocity, $X_{VL,}$ depends on the accuracy to which path angle, path length and time of transit of the ultrasonic wave are measured. Typical accuracy of path angle is ±0.2^o, typical accuracy of path length measurement is ±30mm and typical accuracy of time measurement is ±30ns. These may be converted to percentage uncertainties based on the appropriate path angle and path length values.

The uncertainty arising from determination of depth, X_d, can be taken as equivalent to the accuracy of stage measurement at the site. The uncertainty arising from channel breath can be assumed to be approximately 0.1%. For uncertainty arising from the number of paths, a value of 1% can be assumed if a cross-configuration, multi-path

[21] BS3680 Measurement of Liquid Flow in Open Channels. Part 3E: Stream Flow Measurement of discharge by the ultra-sonic (acoustic) method.

system is used, a value of 5% can be assumed if a multi-path system is used, whilst a value of 10% can be assumed if only a single-pathway is operated at the site.

7.5 Derivation of Uncertainty (4) – Electromagnetic Stations

British Standard BS3680 includes a section on electromagnetic flow gauges[22] which includes calculations for uncertainties. It is stated that 'generally for an insulated channel, the random uncertainty at the 95% confidence level in the value predicted from the calibration relation may be of the order of +/-2%'.

The calculation of uncertainty in BS3680 is for a single determination of discharge requires an understanding of the errors associated with the generation of the electromagnetic field. Component uncertainties include:

- The dimensions of the electromagnetic coil, and its position relative to the bed,
- The variability of the water velocity profile,
- The measurement of the coil current,
- The measurement of depth, relative to the channel bed,
- The measurement of electrode potential.

Errors also arise during calibration of the gauge against current meter measurements.

It is felt that the input data required to calculate uncertainty associated with an EM gauge using the method reported in British Standard BS3680, Part H, would not be readily available in most cases. Therefore the BS3680 method is not implemented in the scoring scheme. Rather an approach similar to that used for rated sections was adopted, in that error was based on check gaugings.

Here check gaugings (preferably those used during the calibration of the gauge) were used to determine the standard error of deviation. The goodness-of-fit between *archived* flows (i.e. those computed using the EM device) and a sample of *N* flow gaugings was measured. This method was applied to the general regime (where all suitable gaugings were used in the sample of *N* flow gaugings), to the high flows range where only gaugings made at flows higher than 0.5 x QMED were used in the sample *N,* and to the low flows range, where only gaugings made at flows below Q95 were used.

7.6 Attributes Quantifying Uncertainty of Measurement

The statistics described in the preceding sections are used in a number of ways to define attributes for different station types and flow ranges. The single most important such attribute is the width of the 95% confidence interval, which has been used as a standardised measure of uncertainty.

[22] BS3680 Measurement of Liquid Flow in Open Channels. Part 3H: Stream Flow Measurement: Electromagnetic method using a full-channel-width coil (1993).

Width of 95% confidence interval
The total width of the 95% confidence interval can vary with flow, and therefore has been expressed either as a percentage of the QMED flow (e.g. for High flow attributes R-H1, S-H1 and U-H1), or as a percentage of the Q95 flow (e.g. for low flow attributes R-L1, S-L1 and U-L1). As has been discussed earlier, the 95% confidence interval is based on the computation of uncertainty about the stage-discharge relationship fitted to flow gaugings for rated sections, or on methods and data published BS3680, for both structures and ultrasonic gauges. The 95% confidence interval is not used in the level-only or electromagnetic scoring schemes.

Within the BS structures and ultrasonic scoring schemes, the total width of the 95% confidence interval is also calculated as a percentage of the mean daily flow (MDF) and used for attributes S-G1 and U-G1.

Table 7.2 shows how the confidence interval width is discretised into five grade classes. The width of the 95% confidence interval at Q95 is calculated using the same procedures as for QMED. However to reflect the inherently better performance that might be expected at low flows, the attribute is assessed using a more stringent grading scheme. Note that this scheme implies that it is considered reasonable for there to be a greater proportional error in flow measurement at high flows owing to greater uncertainty in the stage/discharge relationship or in components of the measurement at structures and velocity-area stations. For a rating curve, the main reason for this greater uncertainty is likely to be the smaller number of gaugings that are generally available at very high flows. (This is a different issue to the proportional error in an individual flow gauging, which has been judged to be greater at low flows – see Section 9).

Table 7.2: Grade look-up table for the width of the 95% confidence interval for discharge

Grade	Total interval width W at QMED (as % of QMED)	Total interval width W at Q95 (as % of Q95)	Total interval width W at DMF (as % of DMF)
1	W > 25	W > 16	W > 16
2	20 < W ≤ 25	16 < W ≤ 12	16 < W ≤ 12
3	15 < W ≤ 20	12 < W ≤ 8	12 < W ≤ 8
4	10 < W ≤ 15	8 < W ≤ 4	8 < W ≤ 4
5	W ≤ 10	W ≤ 4	W ≤ 4

Standard error of estimate as a percentage of mean daily flow
The standard error of estimate (expressed as a percentage of mean daily flow) is used to represent the uncertainty of flow measurement over the flow duration curve as a whole. It is used in the rated section scheme (Table 7.3).

Table 7.3: Grade look-up table for standard error of estimate as % of mean daily flow

Grade	SEE (as %DMF)
1	SEE > 8
2	6 < SEE ≤ 8
3	4 < W ≤ 6
4	2 < W ≤ 4
5	W ≤ 2

Standard error of gauging deviations
The standard error of deviations (SED) is used to assess the uncertainty associated with flow measurement at EM flow gauges and is calculated separately for the high flows, low flows and overall regime ranges. It is the same form of statistic as SEE, but is computed by comparing the deviations between check gaugings and corresponding flows measured by the gauging station (as opposed to derived from a rating curve). The attribute is derived from the check gaugings as follows

- For high flows – from check gaugings made at flows above 0.5 x QMED.
- For low flows – from check gaugings made at flows below Q95.
- For the general performance of the station– from all check gaugings entered.

A minimum of three gaugings is required in each case. The grading scheme is shown in Table 7.4.

Table 7.4: Grade look-up table for standard error of deviations from check gaugings

Grade	SED for high flows	SED for low flows	SED overall regime
1	SED > 12.5	SED > 8	SED > 8
2	10 < SED ≤ 12.5	6 < SED ≤ 8	6 < SED ≤ 8
3	7.5 < SED ≤ 10	4 < SED ≤ 6	4 < SED ≤ 6
4	5 < SED ≤ 7.5	2 < SED ≤ 4	2 < SED ≤ 4
5	SED ≤ 5	SED ≤ 2	SED ≤ 2

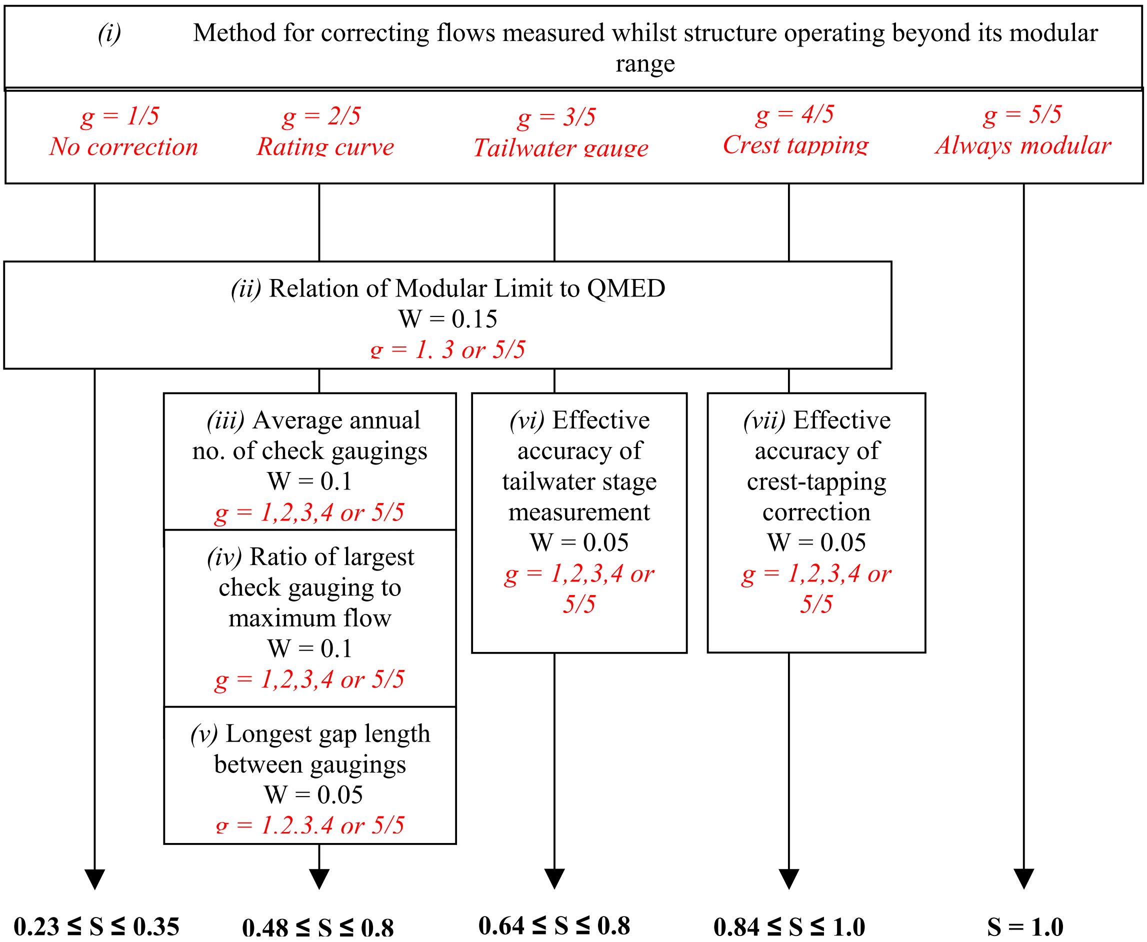
7.7 Attributes Describing Applicability of Theoretical Stage-Discharge Relationship

When a weir is operated beyond its modular range, it is reasonable to expect a reduction in data quality. Ensuring that effects of non-modularity are accounted for is an important element in the scoring scheme for structures.

For the majority of structures non-modularity becomes an issue in the high flow range. During non-modular conditions the stage exceeds the maximum stage for which the stage-flow relationship obeys the theoretical weir equation - the structure is said to be ‘drowned’ out. Typically the highest 10-30% of flows can occur in the non-modular range, although this depends on the type of structure, and the flow regime at the site. Occasionally non-modularity becomes an issue for low flows (for instance where the flow/stage is insufficient to maintain an adequately aerated nappe over the weir crest).

Correction applied for non-modular flow over a structure
The method of correcting for non-modularity is incorporated as an attribute within the high flow range of the BS structures scoring scheme. The attribute is designed to reflect the overall impact of non-modularity on data quality, taking account of how often the modular limit of the weir is exceeded and any correction procedures that might be applied to measured flow values.

In order to incorporate all the issues contributing to data quality when such correction procedures are applied, a three-tier approach has been used to directly determine an appropriate score for the attribute. This involves evaluating seven component attributes (indexed as components *i* to *vii*) and combining their grades to calculate a single attribute score. The procedure is best illustrated by Figure 7.1.



Notes
Figures given in *red italics* represent the grade (*g*) that can be returned for each component factor, those given in **bold** represent the range of scores, **S**, that are achieved in each case, whilst W is the weighting applied to each factor.

Figure 7.1: Factors assessed as part of the non-modularity attribute

In the first tier the method of correcting non-modular flows (*i*) is considered. Three types of correction procedure routinely used for computing flows in the non-modular range of structures are considered. Situations where no correction procedures are applied or where no correction procedures are required (e.g. where high flows never exceed the modular range of the instrument) are also considered. Grades are assigned as follows:

- No correction procedures applied: grade = 1
- A rating derived from check gaugings is applied during the non-modular range: grade = 2
- Correction procedure based on a downstream level where tailwater stage measurements are used to determine the head drop across the weir: grade = 3

- Correction based on tapping of the weir crest to provide information on the head drop across the weir; grade = 4
- High flows never exceed the modular range of gauge, therefore correction procedures not required: grade = 5

In the second tier (Figure 7.1) the position of the modular limit is considered. This component (*ii*) is assigned a grade of either 1,3 or 5 as follows:

- Limit of modular range of structure is between 0.5 x QMED and QMED (grade = 1)
- Limit of modular range of structure is between QMED and 1.5 x QMED (grade = 3)
- Limit of modular range of structure is greater than 1.5 x QMED (grade = 5)

Five additional components are considered in the third tier, as shown in Figure 7.1. If 'rating based on gaugings' is selected in tier 1, components *iii*, *iv*, and *v* (average annual number, ratio to maximum observed flow and longest gap length) are evaluated. These do not explicitly look at the rating used, but consider the number, range and frequency of gaugings. The user is not therefore obliged to enter any gaugings, although this will influence the score achieved (if a formal rating is used for gauging high flows at structures, it is intended that the 'Structure with formal rating at high flows' scoring scheme (scheme 3) would be followed instead). Likewise if correction procedures are based on downstream level are selected in tier 1, component *vi* (accuracy of tailwater stage measurement) is considered and similarly where correction procedures are based on crest-tapping, component *vii* (the accuracy of crest-tapping) is evaluated.

In order to determine a score for the attribute, the grades achieved by the seven 'components' are combined in an additive manner (this should not be confused with the multiplicative approach used elsewhere in the classification) as follows:

$$r = \left[g_i + \left(g_{ii}W_{ii}\right) + \left(g_{iii}W_{iii}\right) + \left(g_{iv}W_{iv}\right) + \left(g_vW_v\right) + \left(g_{vi}W_{vi}\right) + \left(g_{vii}W_{vii}\right)\right] \qquad (11)$$

where r is the attribute score, g_i is the grade for component *i*, W_i is the weight factor for component *i* (W_i = 1 always), and so on.

The above procedure can be thought of as adjusting the main component, (*i*), depending on the values of components (*ii*) to (*vii*). Where components are not evaluated they receive a grade of '0' and drop out of the equation. Weights are used so that the level of adjustment depends on the relevance of the component. For example, as it is an important factor, an adjustment factor of $W = 0.15$ is applied to the modular limit (*ii*) whilst an adjustment of $W = 0.05$ is considered appropriate for longest gap length (*v*) which has much less influence. This approach ensures that components *ii* to *vii* have appropriate influence on the attribute score and within the high flows scoring scheme as a whole. It is effective in situations where particular component attributes have little relevance, as they do not necessarily have to contribute to the attribute score in such cases.

Modular Range
The modular range is an attribute used to help determine quality for the general flow range at structures. It combines two elements, the modular limit and the proportion of time that this limit is exceeded. Each element is graded independently, but the two

grades are then combined. It is expected that in most cases the modular limit will be known only approximately. It is therefore evaluated in relation to the QMED flow. The options for modular limit are described below:

- **Always within modular range**
  - The operating range of the structure is conditions always modular and the theoretical rating is applicable across the full range of measured flows

- **1.5 x QMED < Modular limit**
  - The operating range of the structure is such that the theoretical rating is applicable up to flows equivalent to 1.5 x QMED flow or higher.

- **QMED < Modular limit < 1.5 x QMED**
  - The operating range of the structure is such that the theoretical rating is applicable up or over the QMED flow, but the structure is thought to become non-modular before the 1.5 x QMED flow is reached

- **0.5 x QMED < Modular limit < QMED**
  - The operating range of the structure is such that the theoretical rating breaks down before the QMED flow is reached, but is still applicable when the flow is equal to 0.5 x QMED. It is assumed that if the modular limit is below 0.5xQMED, a formalised rating curve would be used to rate all flows in the high flow range, and in this case scoring should follow the procedure described in the 'BS structure with formal rated-section (high flows)' scoring scheme.

- **Non-modular at low flows**
  - Non-modularity is primarily a problem during periods of lower flows.

The proportion of time for which the modular limit is exceeded is expressed in terms of the number of days during the classification period for which the weir/structure is known/thought to have been operating outside its modular range. Grades are assigned to each element as shown in Table 7.5.

Table 7.5: Grade look up table for Modular Range

Grade assigned g_i or g_{ii}	Modular limit	Number of days non-modular, N_d
1	Non-modular at low flows only	$N_d > 14$
2	Modular limit between 0.5x QMED and QMED.	$14 \geq N_d > 7$
3	Modular limit between QMED and 1.5xQMED	$7 \geq N_d > 3$
4	Modular limit greater than 1.5xQMED	$3 \geq N_d > 1$
5	Never exceeds modular range	$1 \leq N_d$

The two components are combined to determine a score for the modular range attribute as follows,

$$r = g_i + 0.2 g_{ii} \qquad (12)$$

where r is the attribute score, g_i is the grade for component i (modular limit), and g_{ii} is score for component ii (number of days non-modular).

7.8 Attributes Relating to Effective Accuracy of Stage Measurement

General concepts
There are two distinct interpretations of the accuracy of stage measurement within the gauging station data quality classification. One is pure instrument accuracy – this is really the absolute precision of a correctly installed and well maintained instrument and is often quoted by the manufacturer. Typically the range quoted for level measurement is of the order of a few millimetres. Pure instrument accuracy is needed for the Level Only station type.

In other cases, what is needed for the GSDQ classification is instead an *effective accuracy* of stage measurement. The concept of effective accuracy recognises that the stage used to calculate flows, especially at structures, is an idealised hydraulic variable and that the water level recorded by a sensor and then on an archive may not quite correspond to the desired hydraulic variable.

Instrument precision
Where there is no available information regarding instrument precision at a particular site, Table 4-1 may be used to estimate typical value for a variety of instrument types. Where one or more types of level recorder are used at a gauging station, the attribute should be scored on the least accurate.

Effective accuracy of stage measurement
The *effective accuracy of stage measurement* is defined as the resultant accuracy of a measurement taking into account the effects of the combination of instrument and sensor accuracy and resolution, site effects and any other impacts such as analogue to digital signal conversion resolution. Site effects that might introduce an additional error to a stage measurement include:

- incorrect installation or calibration of instrument,
- instrument drift,
- instrument reliability,
- instrument datum being inconsistent with that of flow gauge,
- inappropriate range of instrument (e.g. stage board poorly located, wrong choice of pressure sensor),
- draw-down effects,
- superelevation,
- siltation within stilling well,
- channel turbulence, especially during periods of high flow.

Effects caused by weed growth are not included because weed growth is a separate quality attribute. Table 7.6 provides some guidance as to typical values that might be expected for effective accuracy of stage.

Table 7.6: Effective accuracy of stage measurement

Sensor Type	**Recording medium**	**Effective Accuracy in mm** Good conditions		**Effective Accuracy in mm** Poor conditions	
Shaft encoder	Chart	In stilling well, steady conditions, high resolution chart	**±2**	Rapidly changing stage, difficult to read gauge board, poor chart resolution.	**±20**
Shaft encoder	Logger / Outstation	In stilling well, steady conditions. At least 12 bit A/D conversion, use of internal well dip.	**±1**	Rapidly changing stage, difficult to read gauge board.	**±10**
Pressure transducer	Logger / Outstation	Level range small, sensor calibrated to range, high quality transducer.	**±2**	Large level range, sensor not calibrated to range, poor quality transducer.	**±25**
Upward looking ultrasonic	Logger / Outstation	Steady conditions, small range.	**±3**	Choppy surface or rapidly changing stage. Moderate stage range.	**±10**
Downward looking ultrasonic	Logger / Outstation	Steady conditions, small range.	**±3**	Choppy surface or rapidly changing stage. Moderate stage range.	**±10**

Table 7.6 provides a guide for a range of conditions from good to poor. It is possible that effective accuracy may lie outside the above limits where better or worse conditions apply. For example, using a logger or outstation with only an 8 bit A/D conversion attached to a pressure transducer could provide a resolution of 25mm. The effective accuracy under these conditions would therefore be no better than ±12.5mm. Another example is a site which incurs say, 50mm draw-down of stilling well level during flood flows, and no compensation for this is allowed. This site will have at best an effective accuracy of ±50mm at high flows.

The effective accuracy of stage measurement is considered in all scoring schemes with the exception of the level-only (a different approach is used for level-only sites where the instrument accuracy and truncation of measurements are considered separately). It is also evaluated for all three flow ranges (high, low and general). For BS/ISO structures, effective accuracy of stage measurement at downstream level recorder or effective accuracy of crest-tapping may also be considered as part of the correction for non-modular flows attributes.

Table 7.7 shows the grade look-up table for the effective accuracy attributes. The grading scale for effective accuracy at low flows and for the overall regime is much narrower to reflect the difficulties in measuring level at higher flows.

Table 7.7: Grade look-up table for effective accuracy attributes

Grade assigned	High flows	Low flows	General	Tailwater Stage	Crest-tapping
1	A > ± 30	A > ± 15	A > ± 15	A > ± 30	A > ± 30
2	± 30 ≥ A > ± 20	± 15 ≥ A > ± 10	± 15 ≥ A > ± 10	± 30 ≥ A > ± 20	± 30 ≥ A > ± 20
3	± 20 ≥ A > ± 10	± 10 ≥ A > ± 5	± 10 ≥ A > ± 5	± 20 ≥ A > ± 10	± 20 ≥ A > ± 10
4	± 10 ≥ A > ± 6	± 5 ≥ A > ± 3	± 5 ≥ A > ± 3	± 10 ≥ A > ± 6	± 10 ≥ A > ± 6
5	A ≤ ± 6	A ≤ ± 3	A ≤ ± 3	A ≤ ± 6	A ≤ ± 6

7.9 Attributes Relating to the Level of Confirmation Provided by Independent Check Gaugings

General concepts

Check gaugings are considered in all the scoring schemes, with the exception of the level-only. For the rated-section and electromagnetic schemes, gaugings are used in the estimation of the uncertainty attribute, as discussed above. The following three additional attributes are also used in the rated section and EM schemes as measures of how representative the set of gaugings is:

- Average annual number of check gaugings
- Ratio between the range of flows included in the set of gaugings with the full range of flows observed at the gauging station
- Gap length between successive gaugings.

There are a two opposing opinions on the use of check gaugings to validate flow data from structures and ultrasonic gauges. Check gaugings are inherently ‘noisy’ compared with the hydraulic relationships at a well-maintained structure or with velocity-area calculations where good measurements are available. At structures, this noise is manifested in random scatter amongst gaugings when plotted alongside a rating curve. (Note that for rated sections, flows are calculated using a relationship that has been fitted to the gaugings in a way that minimises the scatter about the curve). Whilst this random scatter could be a reflection of the greater uncertainty in gauging methods rather than of poor performance of the structure, it may also be the case that a series of gaugings indicate systematic error in the flow data at a structure. This will be seen as bias in the data. In this case, the average size of the differences between check gaugings and the archived flow will be significantly greater than zero.

Note that multi-depth gaugings are less likely to exhibit the systematic bias which is commonly associated with single-point methods especially when water depth is limited.

A simple measure of this gross deviation is provided by counting the proportion of check gaugings at a station that deviate from the theoretically-measured flow by more than a set percentage. This threshold is set at ±15% for high flows and ±10% for the low flows and general ranges. Where insufficient gaugings are available, an attribute describing the degree of deviation at the station from British Standard is used instead.

Average annual number gaugings in flow category

This attribute is used as an assessment of the frequency of flow gaugings. Given the errors that are inherently associated with gauging of flows using current meters, a larger number of gaugings helps to ensure data quality. The attribute is calculated as follows:

$$A_G = \frac{N_G}{N_y} \tag{13}$$

where A_G is the average annual number of gaugings during the classification period, N_G is the total number of gaugings during the classification period and N_y is the number of years in the classfication period.

For the high flows range N_G is taken as the total number of gaugings taken where the flow is greater than half the QMED flow, whilst for the low flows range N_G is taken as the total number of gaugings made during flows lower than the Q95 flow. The grading scheme for the attribute is shown in Table 7.8.

Table 7.8: Look-up table for annual average number of gaugings

Grade assigned	Average annual number of gaugings	Average annual number of gaugings > 0.5 x QMED	Average annual number of gaugings < Q95
1	$A_G \le 1$	$A_G \le 0.1$	$A_G \le 0.1$
2	$1 < A_G \le 5$	$0.1 < A_G \le 0.5$	$0.1 < A_G \le 0.5$
3	$5 < A_G \le 10$	$0.5 < A_G \le 1$	$0.5 < A_G \le 1$
4	$10 < A_G \le 30$	$1 < A_G \le 2$	$1 < A_G \le 2$
5	$A_G > 30$	$A_G > 2$	$A_G > 2$

Longest gap length between gaugings in flow category

This attribute is used as an assessment of how good the coverage of flow gaugings is over the period of record. Although a lack of gaugings could well indicate that a site has a stable control (which would promote good data quality), there is a perception that gaugings should be evenly distributed over the classification period and that any long gaps reduce the confidence in data because the performance of the station has not been confirmed.

The following procedure is adopted to calculate the 'longest gap length' attribute:

- Gaugings are arranged in chronological order.
- The length of time, *T*, (in years) between each gauging and the next (the gap length) is calculated.
- In the event that the start date of the classification precedes the earliest gauging, the period between the two is included as a 'gap length'.
- In the event that the last gauging precedes the end of the classification, the period between the two is included as a 'gap length'.
- The longest gap length is determined.

For the High Flows range, *T* is taken as the longest gap in years between gaugings taken where the flow is greater than half the QMED flow, whilst for the Low Flows range *T* is taken as the longest gap in years between gaugings made during flows lower than the Q95 flow. The grading scheme for the attribute is shown in Table 7.9.

Table 7.9: Grade look-up table for longest gap length between gaugings

Grade assigned	Longest gap length between gaugings > 0.5 x QMED T (years)	Longest gap length between gaugings < Q95 T (years)
1	T > 5	T > 5
2	2 < T ≤ 5	2 < T ≤ 5
3	1 < T ≤ 2	1 < T ≤ 2
4	0.5 < T ≤ 1	0.5 < T ≤ 1
5	T ≤ 0.5	T ≤ 0.5

Ratio of gauged flow to archived flow

This attribute is used as an assessment of whether the gaugings are able to represent the range of flows observed at the site. The calculation of this attribute varies slightly depending on the gauge type and the flow range for which it is being evaluated.

For the High Flows range the maximum archived flow is used as an index level. The attribute used is based on the ratio of the largest gauging to the maximum archived value as follows:

$$R = \frac{QG_{\max}}{Q_{\max}} \qquad (14)$$

where QG_{max} represents the largest flow measured during check gauging and Q_{max} is the largest flow held on archive for the site, during the classification period, and can either be an 'instantaneous flow' if flows are determined on a continuous (15 minute) basis or the largest mean daily flow if continuous data are not held.

For Low Flows the index gauging is the minimum gauged flow, and is compared to the lowest flow observed at the site. To avoid problems where zero flows are recorded, the flows are expressed as deviation from Q95 as follows:

$$R = \frac{Q95 - QG_{\min}}{Q95 - Q_{\min}} \qquad (15)$$

where QG_{min} represents the smallest flow measured during check gauging, and Qmin represents the smallest flow observed at the site. For the General category, the comparison between the range of gauged flows and range of observed flows is considered. The attribute is calculated as follows:

$$R = \frac{QG_{\max} - QG_{\min}}{Q_{\max} - Q_{\min}} \qquad (16)$$

where QG_{max} and QG_{min} are defined as before. The grading scheme for the attribute is shown in Table 7.10.

Table 7.10: Look-up table for ratio of gauged to archived flows, R.

Grade assigned	High flows	Low flows	Overall regime
1	$R \le 0.5$	$R \le 0.5$	$R \le 0.5$
2	$0.5 < R \le 0.7$	$0.5 < R \le 0.7$	$0.5 < R \le 0.7$
3	$0.7 < R \le 0.8$	$0.7 < R \le 0.8$	$0.7 < R \le 0.8$
4	$0.8 < R \le 0.9$	$0.8 < R \le 0.9$	$0.8 < R \le 0.9$
5	$0.9 < R$	$0.9 < R$	$0.9 < R$

Percentage of Check gaugings within 15% or 10% of archived value

This attribute is used to evaluate the gross deviation between check gaugings and the corresponding theoretically-measured flow (referred to as the station flow or archived flow). Check gaugings are separated into appropriate ranges:

- High flow gaugings are considered as those where the flow calculated from the check gauging exceeds the 0.5 x QMED value.
- Low flow gaugings are considered as those where the flow calculated from the check gauging is smaller than the Q95 value.
- All gaugings are considered in order to evaluate the attribute in the General category.

The bias, B, between each gauging and its corresponding archived flow is

$$B = 100 \times \frac{Q_G - Q_S}{Q_G} \tag{17}$$

where Q_G is the gauged flow and Q_S is the corresponding station flow. The percentage of gaugings is therefore determined as follows:

$$P = 100 \times \frac{N(B \le 15\%)}{N_G} \tag{18}$$

where N_G is the number of gaugings in the flow range and $N(B \le 15\%)$ is the number of these with bias values not exceeding ±15%. The grade look-up table is shown in Table 7.11.

Table 7.11: Look-up table for percentage of gaugings, P, with bias not exceeding 15%

Grade assigned	High flows	Low flows	Overall regime
1	$P \le 30$	$P \le 30$	$P \le 30$
2	$30 < P \le 45$	$30 < P \le 45$	$30 < P \le 45$
3	$45 < P \le 60$	$45 < P \le 60$	$45 < P \le 60$
4	$60 < P \le 75$	$60 < P \le 75$	$60 < P \le 75$
5	$75 < P \le 100$	$75 < P \le 100$	$75 < P \le 100$

Deviation from BS design

The preceding attribute is replaced by an evaluation of deviation from British/International Standard **only where no check gaugings are available** to verify flow measurements for BS structures and ultrasonic scoring schemes. This refers to the compliance of structure with the BS/ISO standard and provides an opportunity for the user to enter local knowledge about the condition and performance of the structure/gauge (whether or not this has been quantified by detailed review of the structure or represents a general perception). In the absence of any other information it is assumed that best judgement based on local experience or anecdotal evidence will be used to select the most appropriate option.

The grading scheme is based upon three options as follows and graded as shown in Table 7.12 (note there are no grades 2 or 4 for this attribute):

- Strong deviation from BS
  - The stage-discharge relationship is known to deviate strongly from the theoretical or other features of the gauge deviate severely from BS/ISO specification.
  - Structures may strongly deviate from BS/ISO specification if there are defects such as geometry of the weir not to specification, incorrect or over-design of structure, strong influence of upstream/downstream conditions or turbulence in channel.

- Little deviation from BS
  - The stage-discharge relationship is known to deviate moderately from the theoretical or other features of the gauge deviate moderately from BS/ISO specification.
  - Minor deviation includes corrosion/ poor maintenance of structure, wrongly positioned level device, grit/gravel deposition, re-circulating flows, poor condition of weir crest, bowing of flume cheeks and so on. Where such issues are sever then the station may best be treated as deviating strongly from BS.

- No deviation from BS
  - The stage-discharge relationship does not deviate from the theoretical and/or the structure is built and maintained to BS/ISO specification.

Table 7.12: Grade look-up table for deviation from BS

Grade assigned	Deviation from BS
1	Strong deviation from BS
3	Little deviation from BS
5	No deviation from BS

7.10 Attributes Related to Configuration of Gauging Station

Average annual number of bed level surveys

The average annual number of bed level surveys is calculated as follows:

$$\overline{N}_B = \frac{N_B}{n_{ys}} \tag{19}$$

where $\overline{N_B}$ is the attribute value, N_B is the total number of bed surveys conducted between the start and end dates of the classification period and N_{ys} is the number of years between the start and end dates of the classification period. The attribute is therefore a measure of the general performance of an ultrasonic gauging station and contributes to the scoring scheme for the General range. A grading scheme in line with good-practice guidelines is adopted (Table 7.13).

Table 7.13: Grading scheme for average annual number of bed surveys

Grade assigned	Range
1	$0.25 \leq \overline{N}_B$
2	$0.5 \geq \overline{N}_B > 0.25$
3	$0.75 \geq \overline{N}_B > 0.5$
4	$1 \geq \overline{N}_B > 0.75$
5	$\overline{N}_B > 1$

Ratio of stage to ultrasonic path height

Within the ultrasonic scoring scheme two attributes are used to consider whether path heights are appropriate. For high flows the ratio between the height of the uppermost ultrasonic flight path and the maximum archived stage is considered. The ratio is determined as follows:

$$R = \frac{h_{up}}{H_{\max}} \tag{20}$$

where R is the attribute value, h_{up} is the height of the upper ultrasonic flight path specified in m above datum and H_{max} is the maximum stage observed at the gauging station (during the period of record) in m above datum.

For low flows the ratio of height of lowermost path to mean bed level below lowermost path is considered as follows:

$$R = \frac{H95 - h_{lo}}{H95 - MBL} \tag{21}$$

where R is the attribute value, $H95$ is the stage associated with the Q95 flow (or equivalently the stage level exceeded or equalled for 95% of the time), h_{lo} is the height of the lowermost ultrasonic flight path specified in m above datum and MBL is the mean level of the channel bed below the path specified in m above datum. The look-up tables for the attributes are shown in Table 7.14.

Table 7.14: Look-up table ratio, R, of stage to ultrasonic path height

Grade assigned	High flows	Low flows
1	$R > 0.9$	$R > 80$
2	$0.8 < R \leq 0.9$	$60 < R \leq 80$
3	$0.7 < R \leq 0.8$	$40 < R \leq 60$
4	$0.5 < R \leq 0.7$	$20 < R \leq 40$
5	$R \leq 0.50$	$R \leq 20$

Sensitivity Index

The 'gauge sensitivity' is an attribute in the Low Flow range of the scoring schemes for rated-sections and structures. It defined as the change in flow associated with a 10mm increase in stage at the Q95 flow. It is as one of the key hydrometric statistics in the Hydrometric Register, where it is calculated based on the period of record Q95. The sensitivity index provides a means of quantifying the overall precision of flow measurement during low flow periods when errors in stage measurement have a large influence, proportionally, on flow measurement. It is expressed as a percentage.

For the rated section the rating equation is used to calculate the sensitivity based on the following steps:

- Rating equation used to calculate stage associated with Q95 flow.
- The stage is increased by 10mm.
- Rating equation is used to calculate flow associated with new stage value.
- The increase in flow is expressed as a percentage of the Q95 flow.

For structures the stage associated with Q95 flow and flow associated with a stage of 10mm higher than this are input by the user (having for example been read from the rating table). The difference in flow is then calculated and expressed as a percentage of the Q95 flow. The look-up table is shown in Table 7.15.

Table 7.15: Grading scheme for sensitivity, S

Grade assigned	Sensitivity, S (%)
1	$S > 40$
2	$40 \geq S > 30$
3	$30 \geq S > 20$
4	$20 \geq S > 10$
5	$S \leq 10$

Membrane Condition

The integrity of the coil insulation / membrane is an important control on data quality of EM gauges. The membrane is usually made of thick polythene which, having a very high electrical resistivity, prevents any electrical leakage between the coil and the water in the channel that would lead to incorrect potentials being recorded. The insulation membrane may be damaged by local scour. Integrity of the membrane was therefore included as an attribute. Where check gaugings show good agreement with measured flow, and in the absence of other information, the membrane may be considered to be in fair condition. The grade look-up table is shown in Table 7.16.

Table 7.16: Membrane condition

Grade assigned	Membrane condition
1	Poor condition
3	Condition unknown
5	Good condition

7.11 Attributes Characterising Gauge Reliability and Significance of Missing Data

The term 'missing data' describes a scheduled flow or level measurement where no value was recorded, or where a value recorded was not carried through to archive. Instrument failure, failure of the logging device, flood events, zero/low flow events and so on, may all result in missing data. Data may also be lost during the archival process (e.g. human error or rejection following quality control). Missing data are usually archived using a –9999 identifier. Zero flows should not, in general, be counted as missing.

Whilst it is obvious that the presence of missing data does not have a direct affect, per se, on quality of flow measurements, missing data may have a significant influence on the quality of the flow record as a whole. For example the user may have little confidence in POT data if large portions of the flow record are missing.

Missing data cannot by definition be quantified according to the flow range. Accordingly, it is not appropriate to attempt to categorise missing measurements into high flow or low flow categories. For gauging stations, the quantity of missing data is therefore evaluated in terms of the average annual number of missing daily flows during the classification period. (For level-only stations, where the continual measurement of stage is desirable (e.g. for Flood Warning), missing data is represented in terms of percentage non-capture rate as described in Section 7.10)

The significance of missing data can, however, vary depending on when it occurs. The user must therefore be able to qualify the importance of missing data separately for periods of low flows and high flows. This is achieved in the classification by use of significance of missing data at high flows and significance of missing data at low flows attributes.

Average annual number of missing daily flows
The average annual number of missing mean daily flows is the number of days during the classification period for which there was insufficient flow data to determine the mean daily flow. It is calculated as follows:

$$\overline{N}_M = \frac{N_M}{n_{ys}} \qquad (22)$$

where $\overline{N}_M$ is the attribute value, N_M is the total number of entries on the daily flow archive for the gauging station that have been entered as 'missing' values between the start and end dates of the classification period and n_{yrs} is the number of years between the start and end dates of the classification period.

Mean daily flow is generally derived either as the weighted mean of all flows measured on one particular calendar day, or from the mean daily stage. For example, suppose that flow is determined on a 15-minute basis for a particular gauging station. In this case there would be 96 individual measurements of flow per day.

The loss of, say, 12 flow measurements is unlikely to preclude calculation of mean daily flow for that day. However the mean daily flow is likely to be archived as 'missing' in the event that, say, 80 flow measurements are lost. The attribute is therefore a measure of the general performance of the gauging station, and contributes to the scoring scheme for the General category. Table 7.17 shows the grading scheme.

Table 7.17: Grading scheme for average annual number of missing check gaugings

Grade assigned	Range
1	$\overline{N}_M > 21$
2	$21 \geq \overline{N}_M > 14$
3	$14 \geq \overline{N}_M > 7$
4	$7 \geq \overline{N}_M > 3$
5	$3 \leq \overline{N}_M$

Significance of missing data at Low and High flows

This attribute is based on the significance of missing data and encompasses truncated measurements, whether or not such events are recorded as –9999 (note that this attribute applies only to flow-gauges, truncation of level measurement is under a separate attribute).

The attribute is assessed qualitatively and is assigned to one of the following three categories, with grades assigned according to Table 7.18:

- Significant missing data

  This designation should be selected if missing data occurs frequently during the classification period.

  For example if flood peaks are consistently missed this would be 'significant'.

- Some missing data

  Missing data are observed at the site, but are not a regular occurrence.

- Insignificant missing data

  The gauging station is generally reliable, and little or no missing data have been observed during the classification period. For example, if missing data arises only rarely due to vandalism at the site.

Table 7.18: Grading scheme for significance of missing data

Grade assigned	Significance for high flows	Significance at low flows
1	Significant	Significant
3	Some	Some
5	Insignificant	Insignificant

7.12 Attributes Related to Local Conditions at Site of Gauging Station

Weed Growth

Weed growth is detrimental to gauging station data quality. It usually has greatest impact during periods of low flow, when the stage is low and waters are slow moving, particularly as the lowest flows are often occur during the summer months when weeds grow most vigorously.

For open-channel rated-sections the presence of weeds increase stage for a given flow whilst weed and algal growth can also affect the performance of structures, especially if along the weir crest. For ultrasonic stations weed growth along the banks can inhibit signal receipt. Electromagnetic gauging stations are often commissioned in part as a solution for sites with vegetation growth. A number of different management practices are employed to minimise the impact of weed growth. These include clearance of weed and the use of 'shift procedures' where the rating-curve is continually adjusted by the use of check gaugings to account for changes in stage.

An attribute combining both the significance of weed growth at a station (i.e. the requirement to manage weed growth) and the degree to which it has been managed is applied in the classification. The attribute is based on four categories, which have been designed to be unambiguous as far as possible, and are intended to reflect the actual history of management at a station.

The four categories are:

- Shift procedures applied
  - Only appropriate where rating curve shift procedures are applied at a rated-section gauging station

- Not managed
  - Weed growth is a problem but has not been managed.

- Partially managed

- Some action has been taken to manage weed growth, but this may fall short of the ideal level of management
    - This might encompass situations such as control of weed growth on an infrequent basis.

- No weed / well managed
    - No significant weed growth, or significant weed growth would occur, but is managed such that it has negligible impact on flows.
    - This may encompass situations where weed growth is controlled on a frequent basis, relative to the vigour of the growth.

The shift procedures option is applicable only for rated-section gauging stations. The use of shift procedures implies that there will be many time-dependent rating curves, rather than a single unique rating equation, and that the standard error (SE) associated with flows estimates will be large. It is therefore appropriate that a low grade should be assigned to attributes related to standard error or confidence intervals. Where shift procedures have been applied, the following approach is taken in the data quality classification:

- The weed growth attribute is assigned a grade of 3.
- Only those check gaugings taken outside the period in which shift procedures are operating should be considered in the classification.
- Only the rating equation describing the winter base curve should be considered in the classification.
- The relevant standard error of estimate (SEE) attribute for the high flows range is graded in the normal manner (i.e. based on gaugings).
- The relevant SEE attributes for low flows and general ranges are assigned a grade of 1.

Table 7.19: Grading scheme for Weed growth

Grade assigned	Weed growth and management
1	Not managed
3	Partially managed
5	No weed / well managed
3	Shift procedures applied

Unmeasured By-Pass Flow

Unmeasured bypass flow is here defined as that part of the flow conveyed past a gauging station that is not actually captured by a flow measurement. It is considered as a High Flow attribute and contributes towards the score for the High Flow range. The attribute is therefore intended to encompass situations such as out-of-bank flow on the floodplain around a gauging station or unmeasured flow in a secondary channel.

Strictly speaking unmeasured flow through sediments on the river bed and so on or leakage under the gauging structure (where flow is measured using a weir, flume or EM gauge) also represent unmeasured bypass flow, but are unlikely to have much significance during the periods of high flow, and can be essentially ignored for the purposes of the classification.

It is not intended that this attribute be used to describe the degree to which a gauging station provides a complete closure of catchment water balance. For example, in some permeable catchments there may be a significant proportion of the water balance that is exported as subsurface flow and therefore, in a sense, 'by-passes' any gauging. This water would not be counted as 'un-measured bypass flow' for the gauging station quality classification.

By definition, 'unmeasured bypass flow' can only ever be an estimate. A qualitative assessment of the significance of impact of by-passing at the station is therefore used in the classification, with one of the following three categories being selected as appropriate:

- Frequent or significant bypass flow
  Unmeasured bypass flow occurs frequently during the classification period, or if it occurs less frequently, represents a significant proportion of flow at the site.

- Infrequent or insignificant bypass flow
  Unmeasured bypass flow occurs infrequently during the classification period or, if it occurs more frequently represents a small proportion of the flow at the site.

- No or negligible bypass flow
  There is no record of bypass flow at the site, or bypass flow has occurred rarely during the classification period.

Some judgement will therefore be required to provide a realistic assessment of by-passing that is appropriate for the classification period. Although this approach is subjective, it avoids the need to produce a numeric estimate of bypass flow. A number of methods are suggested by which the significance of bypass flow can be evaluated including:

- considering truncated peaks within the flow record,
- comparing peak flows to those at upstream/downstream gauging stations,
- reviewing other evidence regarding the peak stage during flood events, e.g. observations by members of the public / Agency staff, photographs and wrack marks.

The following look-up table is appropriate:

Table 7.20: Grading scheme for bypass flow

Grade assigned	Significance of bypass flow
1	Frequent or significant
3	Infrequent or insignificant
5	No or negligible bypass flow

7.13 Attributes Relating to Level-only Sites

A series of attributes are used to characterise the accuracy of stage measurement, and a different approach is used for level-only sites, where the instrument accuracy and truncation of measurements are considered separately, compared to flow gauges, where the effective accuracy is the dominant measure.

Type of instrument
The type of instrument used is a key attribute in the scoring scheme for level-only sites. The attribute is graded using the following look-up table.

Table 7.21: Instrument types

Grade	Instrument
1	Stage board
2	Chart recorder or Punched tape recorder (PTR)
3	Up-looking or down-looking ultrasonic water level gauge
4	Pressure transducer with diaphragm / pneumatic sensor
5	Shaft encoder

Sensor accuracy
The sensor accuracy represents the operational accuracy of the level gauge, and is entered as a numeric value in mm. For example if level can be assumed to be measured to within ± 1mm the instrument accuracy is 1mm. Manufacturers will normally provide details of accuracy. The accuracy of stage measurement is graded according to Table 7.22.

Table 7.22: Grading scheme for instrument accuracy, A_I

Grade assigned	Range of instrument Accuracy
1	$A_I > \pm 15$
2	$\pm 15 \geq A_I > \pm 10$
3	$\pm 10 \geq A_I > \pm 5$
4	$\pm 5 \geq A_I > \pm 3$
5	$A_I \leq \pm 3$

Truncation of measured level
Whilst many level-recording instruments allow stage to be measured very precisely, the value recorded might not necessarily be an accurate one. For example stage measurements are often truncated or are affected by large systematic errors. The degree to which measured level is truncated is assessed for both high and for low flow ranges. For each of these ranges the user is required to assign one of three categories, using best judgement to evaluate the frequency and/or severity of truncation of level measurement, as described below. The grading scheme is shown in Table 7.23.

- Frequent

  This designation should be selected if truncation / errors occurs frequently during the classification period.
  For example if flood peaks are consistently missed at high flows, or if stage measurement is insensitive once levels drop below L95 at low flows.

- Occasional

  Truncated data or large errors are observed at the site, but are not a regular occurrence.

  For example if levels are truncated due to the formation of ice within the channel during severe winters, or if levels below L99 cannot be measured.

- Rare

  The level recorder is generally reliable, and little or no truncation of data is thought to have occurred during the classification period.

Table 7.23: Look up table for truncation of level

Grade assigned	Truncation (high levels)	Truncation (low levels)
1	Frequent	Frequent
3	Occasional	Occasional
5	Rare	Rare

Average annual number of manual checks for level

The average annual number of manual checks for level is used as an additional check on the accuracy of level measurement for the Level-only Scoring Scheme. It is assumed that frequent independent checking of level will lead to better quality of data, as sources of error will be identified more quickly. It is calculated as follows:

$$\overline{N}_L = \frac{N_L}{n_{ys}} \qquad (23)$$

where $\overline{N}_L$ is the attribute value, N_L is the total number of independent checks on level made between the start and end dates of the classification period and n_{yrs} is the number of years between the start and end dates of the classification period. Table 7.24 is used to determine grades for the attribute:

Table 7.24: Look up table for average annual number of manual checks for level, N_L

Grade assigned	Truncation (high flows)
1	$\overline{N}_L \leq 6$
2	$6 \leq \overline{N}_L \leq 12$
3	$12 \leq \overline{N}_L \leq 20$
4	$20 \leq \overline{N}_L \leq 40$
5	$\overline{N}_L > 40$

Non-capture rate (level measurements)
The percentage non-capture rate, NCR, is calculated as follows:

$$NCR = 100\ N_{MS}\ \frac{F_L}{n_{hrs}} \qquad (24)$$

where N_{MS} is the number of missing stage measurements during the classification period, F_L is the frequency of level measurement in hours (e.g. 15-minute data would have a frequency of 0.25 hours) and n_{hrs} is the length of the classification in hours. The total number of missing stage measurements can usually be determined by counting the number of entries having a -9999 identifier appearing on the level record between the start and end dates (inclusively) of the classification. Table 7.25 shows the look-up table for the non-capture rate attribute.

Table 7.25: Look-up table for non-capture rate

Grade assigned	NCR (%)
1	NCR > 10.0
2	5.0 < NCR ≤ 10.0
3	1.0 < NCR ≤ 5.0
4	0.5 < NCR ≤ 1.0
5	NCR ≤ 0.5

Management of siltation of stilling well
The siltation attribute is considered only in the level-only scheme. It addresses the severity and management of siltation that might occur around the level gauge, but refers to silt affecting the stilling well or access/feeder pipes rather than accretion in the main channel. The user is required to make some judgement as to the balance between the severity of the problem and the success of any management practices that are adopted in selecting one of three options using the drop-down list box, which are graded as shown in Table 7.26:

Table 7.26: Siltation management

Grade	Instrument
1	Severe, or not managed
3	Minor, or partially managed
5	None, or well-managed

8 EXCEL SPREADSHEET TOOL

8.1 Introduction

This report section gives an overview of the software tools that have been developed to implement the gauging station data quality classification. For logistical reasons, the software has been developed within the Microsoft Excel spreadsheet application, using Visual Basic for Applications (VBA) to provide automation and user interface functions. We describe here how the software tool was developed and its basic design. A separate user guide, which accompanies this report, provides guidance on how to enter data in the tool.

8.2 Specification and Requirements

The original project specification required a customized Excel spreadsheet for implementing the classification, but no further particulars were given. Discussions at the York workshop and elsewhere demonstrated a demand for a user-friendly interface for entering input data and attribute information. Users also wanted the methods for implementing the full (numerical) and abbreviated (descriptive) classification to be, at least, semi-automated.

It also became clear early on that, in addition to a tool for implementing the scoring procedures, a spreadsheet tool for storing the results from different gauging stations was needed. This would essentially be used as a 'register' of gauging station data quality, allowing classifications for different sites, or for different time periods to be accessed and compared easily.

The requirement to develop the software within Excel places certain limitations on the programmer. Although the many built-in functions of Excel are available to the user, there is less flexibility when designing the user interface than in a full application. Two separate tools therefore had to be developed, as follows:

- the **Classification Tool** – which automates procedure for calculating and grading attributes, for calculating final scores at individual sites and for determining the abbreviated classifications for high, low, and overall categories for a particular gauging station.

- the **Register Tool** – which stores the results of classifications for different sites in tabular format.

There will be one copy of the classification spreadsheet for every individual gauging station classification. There will then be a single copy of the register spreadsheet for any group of classification results stored in one folder.

The Classification Tool was designed as a blank template, with the intention that classifications for individual stations should use copies of it, and be saved under unique filenames. The Register tool was designed so that it would access all such files, read the classification results stored within and summarise these in tabular format. Figure 8.1 illustrates this framework.

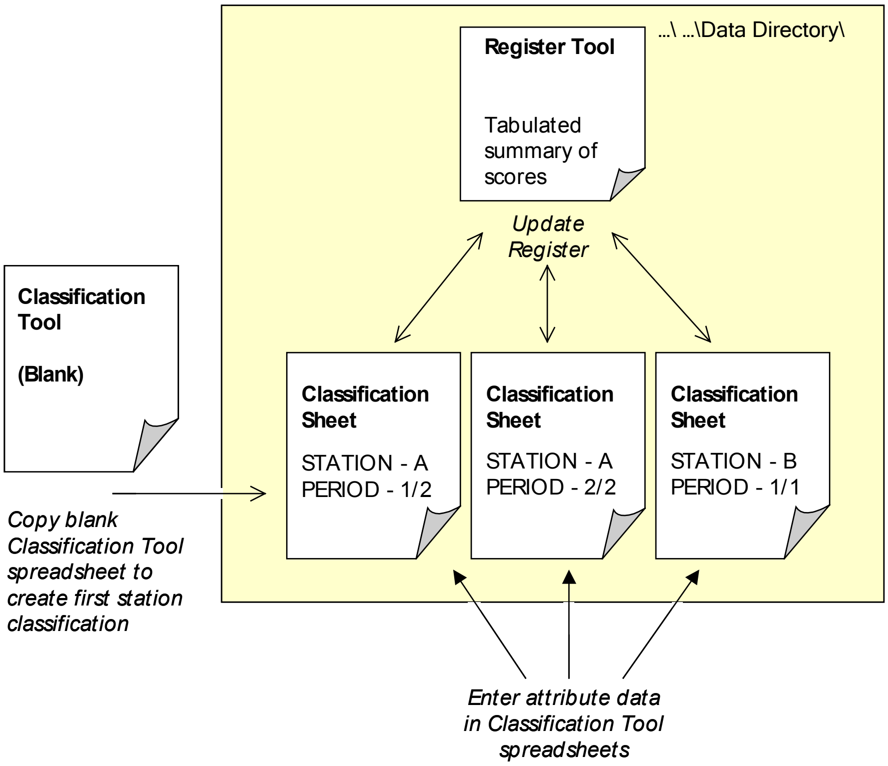


Figure 8.1: Schematic overview of Excel tool v1.1

The main principles used within this framework are described in the remainder of this chapter. More detail with regard to the programming, layout, and operation of the spreadsheet tools is provided in the software user guide.

8.3 Development Issues

It was a requirement of the project to avoid any software solution that would create installation or other system maintenance and security issues. This precluded the development of a stand-alone executable application, or of dynamically linked code libraries. An implication of this is that the code required for the software tool has to reside within the spreadsheet file for each classification of a gauging station, which is inefficient in terms of disk usage, but does mean that any single copy of the software tool is entirely self-contained. Furthermore, Excel stores every cell formatting instruction within a spreadsheet, which can lead to relatively large (over 1 Mb) file sizes even for a 'blank' scoring scheme that contains no real data. There is unfortunately no practical way of avoiding this inflation of spreadsheet file sizes, although we have sought to minimise it as far as possible.

The spreadsheet tools were built using Microsoft Excel 97, and were tested under both the Microsoft Windows 2000 and XP Professional operating systems. The Visual Basic for Applications (VBA) functionality of Microsoft Excel was used to automate some of the calculations and functions required. This means that the spreadsheet tools run Visual Basic macros in order to implement some parts of the scoring procedure and that the 'Enable Macros' option should be selected when using the files. All the macros used were written by JBA, and none require the user to install or reference non-standard software components such as Dynamic Link Libraries (DLLs), or to make changes to the operating system.

In order to prevent accidental changes, a password protection was applied to parts of both the Classification Tool and Register Tool workbooks. The VBA code was also protected in the same way.

8.4 GSDQ Classification Tool

The Classification Tool was designed as a modular system, where required functions are combined as appropriate (depending on the gauging station type) to calculate attribute grades and populate the scoring scheme. Although the scoring schemes include some complex attributes, the user is asked only to input basic data relating to the characteristics of the gauge and the flow record.

The user interface of the Classification Tool is based on a number of worksheets. There are three worksheets for data entry. However these have been designed so that where the required station information is readily available, the spreadsheet takes only a few minutes to complete. Of course accessing the station data and deciding which rating revisions to classify or which gaugings to include may take somewhat longer.

Results are shown on separate worksheets that appear to the user only when all input data has been entered correctly and all calculations and procedures used in the classification fully completed. The tool is designed so that the scoring procedure will not be implemented unless all required data have been entered. The tool will return an error message if the user attempts to run the procedure without having completed the input worksheet(s) and the status with be 'Unclassified'. Figure 8.2 shows the basic procedure and arrangement of modules.

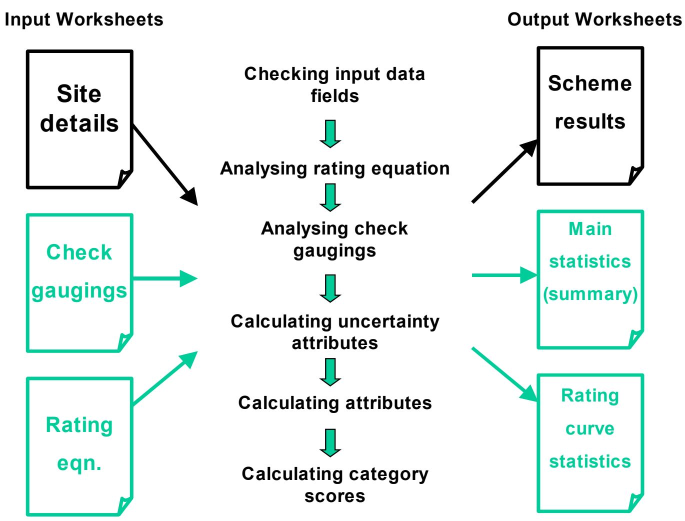


Figure 8.2: Modular structure used in the GSDQ classification tool

Table 8.1 summarises the worksheets within the classification tool. The Input Station Info worksheet is the key input sheet and is activated as default on opening the Classification Tool. It also allows the user to select the Scoring Scheme to be used, on the basis of the gauge type. This in turn determines which of the input and output worksheets are displayed. The data entry fields shown on the worksheets will also vary automatically according to the type of gauging station being considered. (It should be noted that parts of the workbook are password protected in order to prevent accidental changes to cell formats and formulae).

Table 8.1: Worksheet descriptions

Worksheet Name	Schemes applicable	Use
Input Station Info	All schemes	*Entry of*: o Classification details o Site details (gauge type, and flow regime) o Missing data o Effective accuracy o Modularity of structure o Configuration (for ultrasonic gauge)
Input Rating Equation	R, SR	*Entry of*: o Parameters of the rating equation used at rated sections o Range of applicability of the rating equation
Input Check Gaugings	R, SR, E S, U (optional)	*Entry of*: o Gaugings used to derive the rating curve if used, o Gauging used to check, validate or calibrate flow measured at Electromagnetic gauges, ultrasonic gauges or structures.
Guidance (Station Info)	All Schemes	Guidance on entry of station info parameters
Guidance (Rating Equation)	R, SR	Guidance on entry of the rating equation
Guidance (Check Gaugings)	R, SR, E S, U (optional)	Guidance on the entry of check gaugings
Level-only Scoring Scheme	L	Classification results in tabular format
Rated Section Scoring Scheme	R	
BS Structure Scoring Scheme	S	
Structure (Rating at high flows)	SR	
Electromagnetic Scoring Scheme	E	
Ultrasonic Scoring Scheme	U	
Guidance (Scoring scheme results)	All Schemes	Guidance on the results tables
Rating Curve	R, SR	*Report of*: o Plot of the rating curve (if used) with 95% confidence intervals generated.
Further Details	R, SR, S, U, E	*Report of*: o Intermediate calculations generated during the scoring procedures. o Notes and warnings issued

Notes: R –Rated section Scoring Scheme, SR – Structure with Rating at high flows Scoring Scheme, U – Ultrasonic Scoring Scheme, E – Electromagnetic Scoring Scheme, S- BS Structure Scoring Scheme, L –Level-only Scoring Scheme.

The key results worksheet is the 'Scoring Scheme' worksheet, which shows the classification results, including attribute grades, weights and scores, in tabular format. Most of the worksheets have an accompanying Guidance Note. These give further details about input fields, report any special considerations and give suggested values or ranges.

Input Worksheets

The 'Input Station Info' worksheet (Figure 8.3) is used for entry of data concerning the physical characteristics of the gauging station and statistical characteristics of the flow record and is the default worksheet when the tool is opened. Drop-down lists are used for data entry. However the user is also required to enter numeric data into blank cells. The sheets have been designed so that data can be entered with the standard 'Copy and Paste' facility of Excel.

The key field is the **Gauge type** drop-down box, which enables the user to select the gauging station type. This selection determines which scoring scheme will be implemented and which worksheets are to be displayed to the user. For example if a Rated Section is selected the 'Input Rating Equation', and 'Input Check Gaugings' worksheets automatically appear automatically.

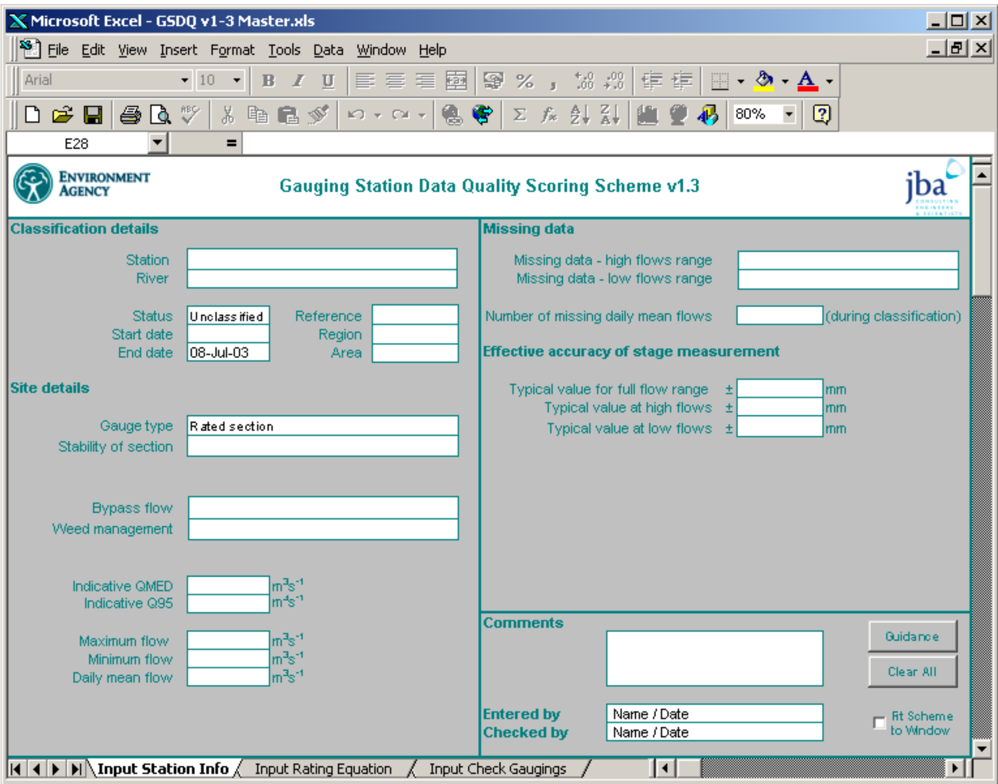


Figure 8.3: GSDQ classification tool – 'Input station info' worksheet

When the user is satisfied that he/she has completed all data entry, a button ('**Calculate scores**') is used to run the scoring procedures. Provided all input fields are filled with suitable values (e.g. a number is entered for a numeric field) this facility can be used to re-calculate the classification scores at any stage.

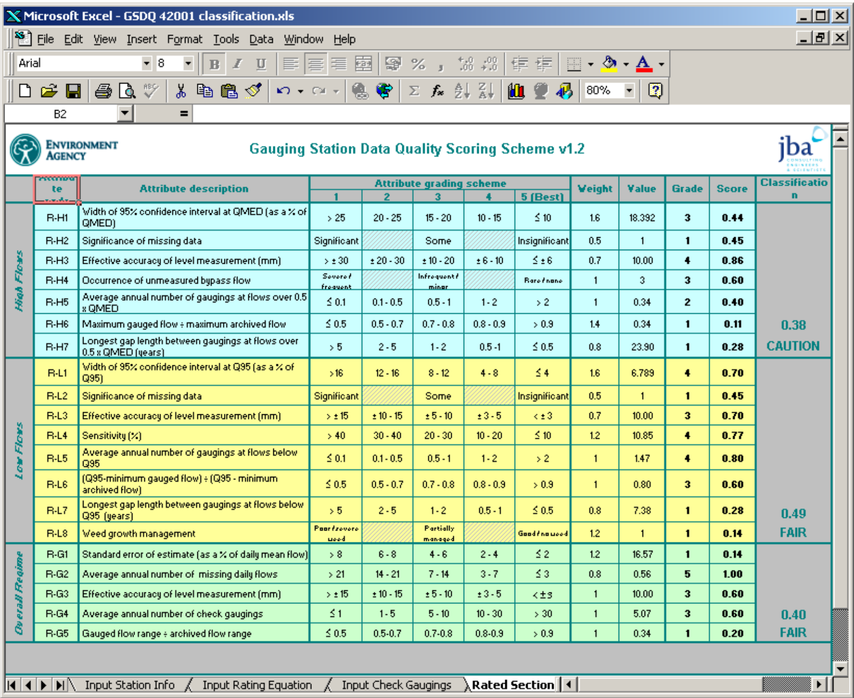
The ‘Input Check Gaugings’ worksheet must be filled in for the rated section and EM schemes, as gaugings are used to assess the standard error of estimate/deviation associated with the rating equation. Typically the set of gaugings that were used to derive the parameters of the rating equation will be input. The date of gauging, observed stage and observed flow are required for each gauging entered. The rating equation is then used to determine the rated flow corresponding to each observed stage.

A ‘suitability’ field is also included, which allows gaugings that have been entered on the sheet to be disregarded when implementing the scoring scheme procedure. A maximum of 1000 gaugings may be entered. These do not need to be entered in chronological order, but if the date field is not completed the gauging will not be counted or included. On processing, the gaugings are re-ordered by date.

For the other types of gauging station (Ultrasonic / Weir / Flume) input of check gaugings is optional. For these the archived flow (in m^3s^{-1}) must also be entered for each gauging. The archived flow is the corresponding flow measured at the gauging structure. Note that the observed and archived flows should be taken at concurrent times as far as possible. The archive flow should also be entered for a weir that is gauged by rated section at high flows. The rating equation (entered in the 'Input Rating Equation' tab) is used to estimate rated flows for gaugings over 0.5 x QMED.

Output Worksheets

The scoring scheme results are shown in tabular format on separate worksheets (Figure 8-4). These tables are not shown to the user until the classification is completed. Attribute grades, weights and scores are shown in each case. The combined attribute scores in the High, Low and General ranges are also shown.



Attribute code	Attribute description	Attribute grading scheme 1	2	3	4	5 (Best)	Weight	Value	Grade	Score	Classification
R-H1	Width of 95% confidence interval at QMED (as a % of QMED)	> 25	20 - 25	15 - 20	10 - 15	≤ 10	1.6	18.392	3	0.44	
R-H2	Significance of missing data	Significant		Some		Insignificant	0.5	1	1	0.45	
R-H3	Effective accuracy of level measurement (mm)	> ± 30	± 20 - 30	± 10 - 20	± 6 - 10	≤ ± 6	0.7	10.00	4	0.86	
R-H4	Occurrence of unmeasured bypass flow	Severe / frequent		Infrequent / minor		Rare / none	1	3	3	0.60	
R-H5	Average annual number of gaugings at flows over 0.5 x QMED	≤ 0.1	0.1 - 0.5	0.5 - 1	1 - 2	> 2	1	0.34	2	0.40	
R-H6	Maximum gauged flow ÷ maximum archived flow	≤ 0.5	0.5 - 0.7	0.7 - 0.8	0.8 - 0.9	> 0.9	1.4	0.34	1	0.11	
R-H7	Longest gap length between gaugings at flows over 0.5 x QMED (years)	> 5	2 - 5	1 - 2	0.5 -1	≤ 0.5	0.8	23.90	1	0.28	0.38 CAUTION
R-L1	Width of 95% confidence interval at Q95 (as a % of Q95)	>16	12 - 16	8 - 12	4 - 8	≤ 4	1.6	6.789	4	0.70	
R-L2	Significance of missing data	Significant		Some		Insignificant	0.5	1	1	0.45	
R-L3	Effective accuracy of level measurement (mm)	> ± 15	± 10 - 15	± 5 - 10	± 3 - 5	< ± 3	0.7	10.00	3	0.70	
R-L4	Sensitivity (%)	> 40	30 - 40	20 - 30	10 - 20	≤ 10	1.2	10.85	4	0.77	
R-L5	Average annual number of gaugings at flows below Q95	≤ 0.1	0.1 - 0.5	0.5 - 1	1 - 2	> 2	1	1.47	4	0.80	
R-L6	(Q95-minimum gauged flow) ÷ (Q95 - minimum archived flow)	≤ 0.5	0.5 - 0.7	0.7 - 0.8	0.8 - 0.9	> 0.9	1	0.80	3	0.60	
R-L7	Longest gap length between gaugings at flows below Q95 (years)	> 5	2 - 5	1 - 2	0.5 - 1	≤ 0.5	0.8	7.38	1	0.28	
R-L8	Weed growth management	Poor / severe weed		Partially managed		Good / no weed	1.2	1	1	0.14	0.49 FAIR
R-G1	Standard error of estimate (as a % of daily mean flow)	> 8	6 - 8	4 - 6	2 - 4	≤ 2	1.2	16.57	1	0.14	
R-G2	Average annual number of missing daily flows	> 21	14 - 21	7 - 14	3 - 7	≤ 3	0.8	0.56	5	1.00	
R-G3	Effective accuracy of level measurement (mm)	> ± 15	± 10 - 15	± 5 - 10	± 3 - 5	< ± 3	1	10.00	3	0.60	
R-G4	Average annual number of check gaugings	≤ 1	1 - 5	5 - 10	10 - 30	> 30	1	5.07	3	0.60	
R-G5	Gauged flow range ÷ archived flow range	≤ 0.5	0.5-0.7	0.7-0.8	0.8-0.9	> 0.9	1	0.34	1	0.20	0.40 FAIR

Figure 8.4: GSDQ classification tool – Output of classification results

Two further worksheets reporting additional information are also produced. The 'Rating Curve' worksheet is generated as part of the scoring scheme for rated sections. It allows the influence of individual gaugings on the standard error of estimate and standard error of the mean relationship to be examined. The variation in SMR with flow is illustrated as a chart. The 'Further details' worksheet is generated for all schemes (except the level-only) and summarises data fields generated as intermediate steps in the calculation of attribute values

Scoring Procedures

Figure 8.5 illustrates the scoring procedures, which are initiated by the user by pressing a button labelled 'Calculate Scores'. The tool is designed so that error messages are generated if the user has not provided suitable values for all input fields. Problem fields are then highlighted in red on the input worksheets. Similarly if any errors occur when calculating attribute values (this is usually related to problems with input data, such as a decimal place entered in wrong place or the wrong set of check gaugings entered) then the scoring procedures will also be aborted.

Warnings are also issued to inform the user of inconsistencies in the input data that would not prevent the scoring procedures from being completed, but may cause the classification results to be wrong. In this case the scoring procedures will still be completed as normal.

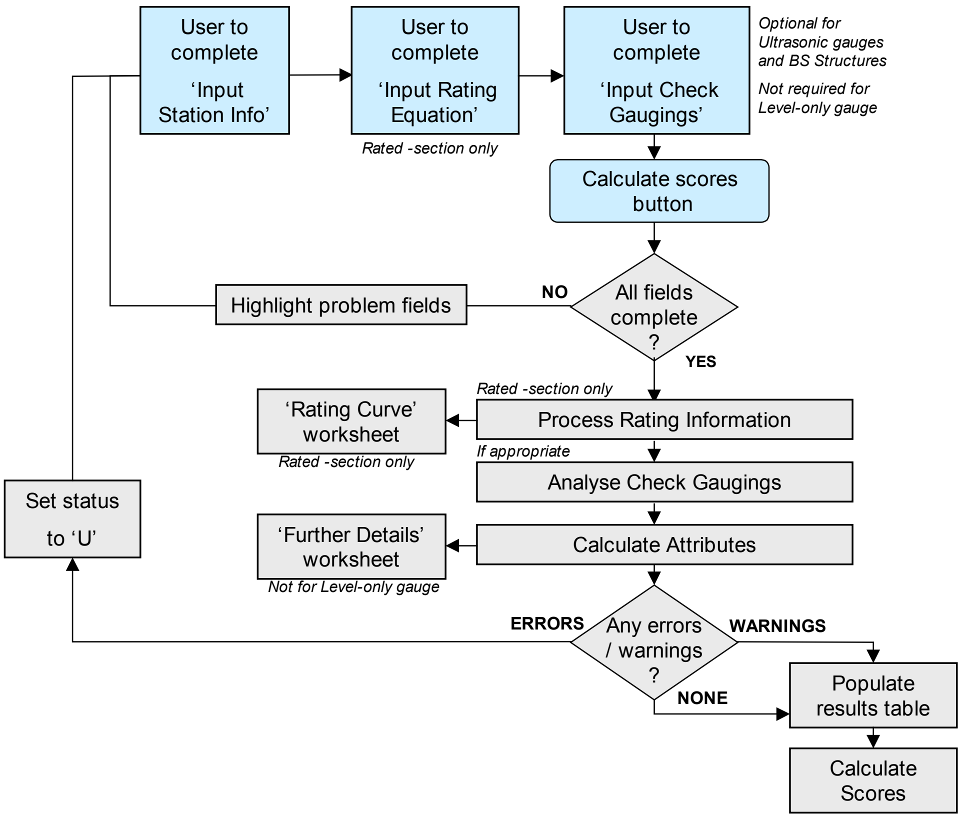


Figure 8.5: Flow chart for scoring procedures

Notes: Stages shown in blue indicate activity by the user, those in grey indicate activity automated in the Classification Tool

The tool is designed so that error messages are generated if the user has not provided suitable values for all input fields. Problem fields are then highlighted in red on the input worksheets. Similarly if any errors occur when calculating attribute values (this is usually related to problems with input data, such as a decimal place entered in wrong place or the wrong set of check gaugings entered) then the scoring procedures will also be aborted.

Warnings are also issued to inform the user of inconsistencies in the input data that would not prevent the scoring procedures from being completed, but may cause the classification results to be wrong. In this case the scoring procedures will still be completed as normal.

8.5 GSDQ Register Tool

The Register Tool is a single sheet Excel workbook. Figure 8.6 shows the sheet, which is essentially a table showing details of completed classifications. The only input field required is the path name of the directory in which the scoring sheets are located. The register reads all files named GSDQ*.xls in this directory. Up to 100 files may be accessed by a single register.

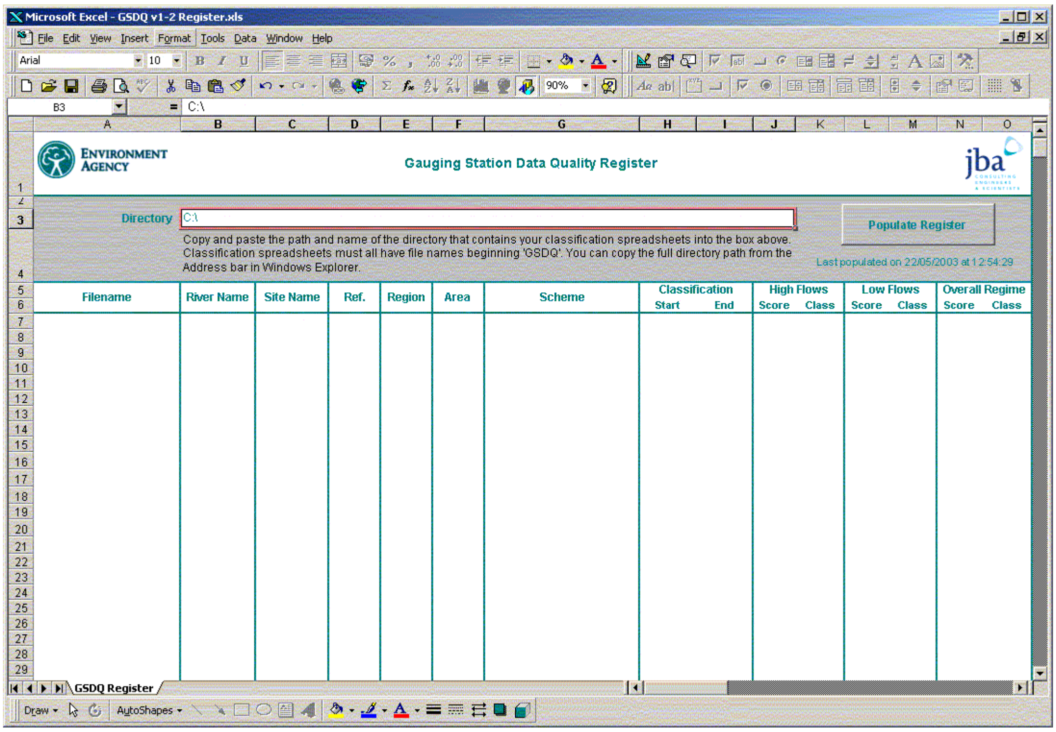


Figure 8.6: GSDQ Register tool

The register is operated using the **Populate Register** button. The running time depends on the number of files from which data must be retrieved. Typically about 30 seconds are required to access 10 files. The GSDQ site files do not need to be opened by the user. The logic of the register is shown in Figure 8.7.

The register reads the following fields from the scoring sheets:

- River Name
- Site Name
- Site Reference
- EA Region
- EA Area
- Start Date (of classification period)
- End Date (of classification period)
- Scheme type
- High Flows Score (if applicable)
- Low Flows Score (if applicable)
- General Category Score

The date at which the register was taken is saved to the sheet. When run again, the tool will overwrite existing entries.

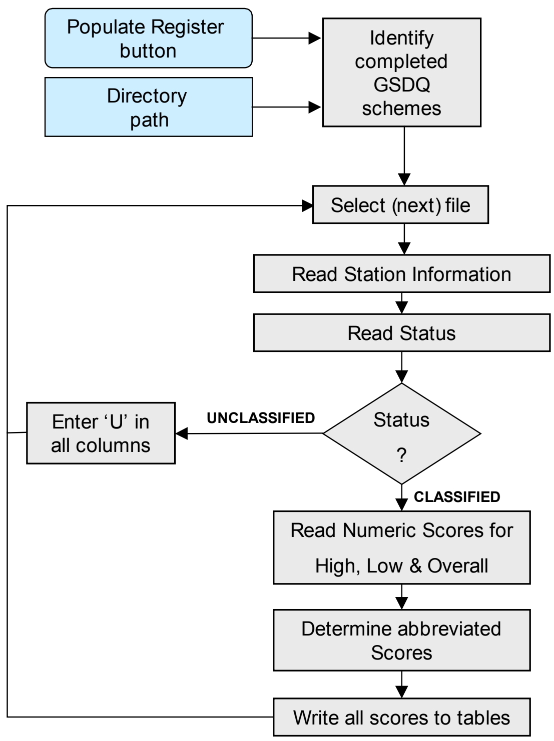


Figure 8.7: Flow Chart for Register Tool

Notes: Stages shown in blue indicate activity by the user, those in grey indicate activity automated in the RegisterTool

9 BENCHMARKING THE GSDQ CLASSIFICATION

9.1 Introduction

There is no single, absolute measure of gauging station data quality. The GSDQ classification attempts to provide a set of conventions and a repeatable method for assessing data quality over discrete periods of record. It has always been recognised that the classification may need to be 'tuned' to provide the best results, and this can be done through the grade look-up tables (see section 7), the weights applied to each attribute or the setting of thresholds between the CAUTION, FAIR and GOOD classes.

To test the classification, a benchmarking exercise was carried out using a test version of the GSDQ software tools. Environment Agency staff were asked to complete the classification for a selection of stations, and to comment on whether the results met with expectations, and also on the design and implementation of the classification. Analysis of the benchmarking results was used to refine the scoring schemes within the classification.

9.2 Test Sites

The gauging stations selected for benchmarking are listed in Table 9.1. It will be seen that a number of stations were treated as more than one type, or run with alternative data entry, for the purpose of experimentation. In total 30 stations were used, but 31 classifications evaluated because one station was split into two classification periods.

Table 9.1: Stations selected for benchmarking

River	Station	Region	GSDQ scheme
Henmore Brook	Ashbourne	Midlands	BS/ISO structure
Sow	Great Bridgford	Midlands	Rated section
Severn	Montford	Midlands	Rated section
Severn	Montford	Midlands	Ultrasonic
Severn	Buildwas	Midlands	Ultrasonic
Strine	Crudgington	Midlands	Electromagnetic
Soar	Littlethorpe	Midlands	Electromagnetic
Aire	Armley	North East	Rated section
Broughton Beck	Broughton hall	North East	Level only
Calder	Methley	North East	Ultrasonic
Whitting	Sheepbridge	North East	Standard (BS/ISO) weir
Doe Lea	Staveley	North East	Standard flume
West Beck	Snakeholme lock	North East	Electromagnetic
Rother	Whittington	North East	BS/ISO weir (rated section at high flows)
Thames	Cricklade	Thames	BS/ISO weir (compound structure)
Ampney Brook	Sheppen bridge	Thames	Standard (BS/ISO) weir
Blackwater	Farnborough	Thames	Electromagnetic
Hart	Bramshill	Thames	Standard (BS/ISO) weir
Loddon	Twyford	Thames	Ultrasonic
Colne	Watford (Berrygrove)	Thames	Non-standard weir treated as BS/ISO weir
Bourne	Addlestone	Thames	Rated section

Table 9.1 (cont): Stations selected for benchmarking

Wey	Tilford	Thames	BS/ISO weir (rated section at high flows)
Mole	Leatherhead	Thames	Electromagnetic
Thames	Teddington	Thames	Non-standard weir treated as BS/ISO weir
Thames	Kingston	Thames	Ultrasonic
Eden	Vexour	Southern	Rated section
Anton	Fullerton 4.75	Southern	Structure
Itchen	Riverside Park	Southern	Ultrasonic
Test	Chilbolton	Southern	Electromagnetic
Test	Longbridge	Southern	Electromagnetic

9.3 Does the GSDQ Agree with Users' Expectations?

The main conclusion of the benchmarking exercise was that the GSDQ classification does reflect the overall perception of knowledgeable local staff about data quality at most gauging stations. It has to be remarked that benchmarking can only be carried out against a subjective view of the station, which may itself be changed as a result of using the GSDQ. Furthermore, no general scheme will ever work optimally in every specific case. However, there was a good agreement with users' expectations for all types of station and over different flow ranges.

Each regional team was asked to assess whether the GSDQ classification agreed with their expectation for each station and for each flow range (High, Low, General). The responses have been classed as either 'agree', 'not agree' or 'unsure'. Table 9.2 summarises the outcome of the benchmarking, broken down by station type. It shows that the benchmarking teams agreed with the classification in the majority of cases, and were more likely to be unsure of the 'true' classification than to disagree.

Table 9.2: Benchmarking results by station type

	Rated section	**BS Structure**	**BS structure with rated section**	**Ultrasonic**	**Electro-magnetic**	**Level only**	**ALL TYPES OF STATION**
Agree	75%	63%	83%	50%	50%	100%	62%
Not agree	25%	13%	-	11%	33%	-	17%
Unsure	-	23%	17%	39%	17%	-	21%
No. classified	*15*	*30*	*6*	*18*	*21*	*3*	93

It is clear that the GSDQ classification works best for rated sections, then for structures, but that there is more debate about its results for ultrasonic and electromagnetic stations. One surprising finding is that there would seem to have been greater debate about whether the classification works well for ultrasonic stations then for electromagnetics, although the sample sizes are not large.

The greatest criticism of the classification was for electromagnetic stations. Here, there was concern that accuracy of stage was given too much weight. It was notable that the disagreement with the classification at EM stations was particularly for the low flows score, whereas no such emphasis was evident for other station types. Comments expressed in relation to the low flow scores at EM stations drew attention both to possible marking down by the classification, but also to possible unreliability of check gaugings. This is perhaps a reflection of the inherent difficulty of making good, accurate check gaugings at low flows, especially at sites where an electromagnetic station has been installed, quite likely because of vegetation growth that makes gauging in shallow, low velocity flows very difficult.

9.4 Modifications Arising from the Benchmarking Exercise

The benchmarking results were reviewed by the JBA team and Agency project manager. As a result, a number of modifications were made to fine-tune the classification and to improve the GSDQ software interface. The sole attribute change was to the margin used for counting deviations of check gauging at structures, which was reduced from ±15% to ±10% at high flows, reflecting the greater *proportional* errors likely in low flow gaugings.

The most significant modification was to revise the threshold for the abbreviated classification results between CAUTION and FAIR. A number of stations had been benchmarked as FAIR when comments suggested they might be better classed as CAUTION. The threshold was therefore raised from 0.4 to 0.55. This may inevitably result in some stations being classified as CAUTION when local opinion would regard then as FAIR. It is important to recognise that any choice of threshold between quality classes will always lead to some stations seeming to be in the wrong category, given the lack of any absolute scale for data quality. Setting the threshold between CAUTION and FAIR at 0.55 is essentially a conservative decision that reflects a preference, *on aggregate*, to 'underestimate quality' rather than to give the benefit of the doubt to stations.

It should also be emphasised that the three descriptive quality classes were intended only as a secondary output of the gauging station data quality scoring scheme. The primary measure of data quality is the numerical score (between 0.0 and 1.0), and stations will fall into the same relative positions on this scale regardless of the thresholds adopted for the descriptive classes.

Where Agency staff do not agree intuitively with a CAUTION, FAIR or GOOD assessment, an alternative may be to report not only the numerical GSDQ score for the data in question, but also to compare it with the scores from another station classification that is agreed to be poor, and one that is agreed to be good.

10 MANAGEMENT OF THE CLASSIFICATION

10.1 Organisation of Excel Tool Files

Each Classification Tool spreadsheet can be used to calculate the data quality classification for a given station and for a given period of record. To create a classification record for a different station, it is only necessary to make a copy of the Master spreadsheet.

To create a classification for a gauge where a spreadsheet already exists for a different period of record, copy the existing sheet to a new file and edit the date ranges, any changed attributes and gauging data accordingly.

10.2 Classification Date Ranges

The gauging station data quality classification has been designed to represent data quality on the basis of attributes that may vary over time. Following consultation and discussion at the September 2002 Workshop, it was, however, decided not to represent quality as a continuous time series (in which the data quality code could vary with each value in a flow data series). Instead, the classification is based on discrete 'blocks' or time periods in which the data quality attributes are thought to be relatively static.

The delineation of classification periods will therefore be linked to changes in gauging station disposition that will in turn cause discrete changes in some or all of the station quality attributes. It will sometimes be a matter of judgment as to what constitutes a 'discrete change' in data quality attributes. Some likely situations are listed below:

- Change of station type (e.g. replacement of rated section with ultrasonic),
- Significant revision of the rating curve,
- Re-engineering of the station (e.g. widening to reduce flow by-passing),
- Change in weed growth management practices,
- Change of the approach taken to correct for drowned flow,
- Change in management of siltation,
- Replacement of instruments of different tolerances or reliability.

In some of these situations, much of the existing quality attribute data may be carried over straightforwardly between classification periods. The specific issue of choice of flow gaugings is discussed below.

10.3 Selection of Flow Gaugings for the Classification Period

In the simplest case, a station could have a single rating curve or calibration, and a single set of flow gaugings which would be used to calculate uncertainty statistics. In reality, multiple rating curves or calibration curves and gaugings often exist, relating to different periods of time. It will be necessary to judge which gaugings to use in the spreadsheet tool, and this is perhaps best left open to the knowledge and expertise of hydrometry staff using the classification. Flow gaugings used to assess data quality for a given period of the record should be accurate independent measurements of flow, relevant to the hydraulic control or measurement instruments operating during that period for the flow/stage range.

For example, if reliable gaugings have been carried out at a new rated section for 5 years and the rating equation is then updated, but the control at the station is not thought to have changed, then we would suggest that early gaugings should continue to be used to calculate uncertainty about the new rating. If, however, the rating has been changed because it is thought that the control has in fact changed, then the old ratings are in principle not a 'fair' independent check on the new rating and should not be used. Judgement may be needed to decide, if it is thought that the control has shifted slowly, whether to allow some of the older gaugings to be included notwithstanding.

10.4 Setting Indicator Flows

There are several 'fixed' points in the flow range that the gauging station data quality scheme uses to calculate quality attributes. These are:

- Median annual maximum flood (QMED);
    - QMED and 0.5 x QMED are used to define the 'High' flow range in the classification

- 95th percentile of flow duration curve (Q95);
    - used to define the 'Low' flow range in the classification
    - used to calculate the sensitivity attribute

- Minimum and Maximum recorded flows
    - Used to calculate attributes to indicate the degree to which the full range of flows can be adequately measured

- Mean daily flow
    - Used to scale standard error statistics for the 'General' flow category

These 'indicator' flows are just that – *indicative*. For the data quality classification it is not expected, or necessary, that the Q95, QMED, minimum and maximum flows are exact, provided that they are reasonable estimates. These indicator flows have been chosen because they are familiar quantities, and, in the case of Q95 and QMED, generalised calculation methods exist to derive them.

The indicator flow values should be estimated for the entire period of record at the station, rather than the individual sub-periods over which classification is calculated. This avoids introducing any inconsistencies in the classification as a result of any differences in the length of classification periods.

10.5 Overlap Between Classification Periods and Flow Ranges

It may be that there are some cases where a change at a station does not impact equally on the quality classification at low and high flows. In such cases, it may be necessary to create a new classification spreadsheet, but to change only the attribute data corresponding to one of the flow ranges. An example might be a station where weed growth management changes (to improve low flow measurement) but high flow measurement is not affected.

10.6 Updating the Classification

Initial retrospective application of the data quality classification may require some care in making suitable judgments about the sub-division of records into separate classification periods, if appropriate. However, once established, the classification should require little maintenance. If any significant changes are made to the operation or fabric of the station, including changes in ratings, then it would be advisable to update the quality classification accordingly. Otherwise, it is suggested that a routine annual check should be carried out to update classification spreadsheets, adding any new check gaugings.

Although estimates for the indicator flows might change as more data are added to the record at a station, it is recommended that these values are not adjusted within the classification spreadsheets unless the changes are substantial, say greater than 15%. If indicator flow estimates are adjusted for a particular station, then the adjustment should also be carried out retrospectively to classification spreadsheets for earlier periods or record, if any exist.

After any change to classification spreadsheets, the Register spreadsheet should be updated to ensure that the tabulated summary of classification scores is kept up-to-date.

11 RECOMMENDATIONS AND ISSUES ARISING

11.1 Fitness for Purpose of Gauging Station Data

Throughout this project there has been considerable discussion about whether the Gauging Station Data Quality classification should address the fitness for purpose of processed data sets in addition to the more tightly defined question of hydrometric data quality. The focus has been fixed on the latter (although the division of the classification in Low and High flow ranges was motivated in the main by the needs of data users). This is in part a reflection of the difficulty in encapsulating factors that determine fitness for purpose in a general format for a suitably wide variety of purposes. In some cases, say the construction of pooling groups for flood estimation or the analysis of water balance for recharge estimation, the debate about the acceptability of a gauged record may be quite subtle, may require a detailed knowledge of modelling or analysis methods and may also be influenced very strongly by context.

There are some pointers to fitness for purpose in the GSDQ classification, including individual attributes for weed growth, by-passing and other important factors. However, wider catchment effects are not included. For example, non-closure of the water balance in an impermeable catchment would not be evident from the classification. Likewise, secular trends in a series of peak flows would not be acknowledged if the cause was urban development within the catchment during the period of record. Even non-stationarity attributable to the station itself (for example deterioration of a weir crest) would only be evident from the GSDQ classification if hydrometric staff were able to distinguish periods of the record having different quality attribute values. This may be difficult if changes in station condition have been gradual and the period of record is not very long.

Yet there is a clear need to address fitness-for-purpose in presenting data to users. This has been demonstrated by the consultation carried out for this project, and, more strongly, by the resources invested in assessments of gauging station networks for projects including those discussed in Section 2 of this report.

It is our hope that users will be assisted in judging fitness for purpose by the detailed results of the GSDQ classification (that is, the actual attributes and attribute values collated for each station). But for wider issues there remains a need for visualisation of data and descriptive commentary to be available to the user. Data visualisation permits users to recognise patterns that may confirm the fitness of flow data or may call it into question. An obvious example is to inspect the time series of data for two stations on the same river on a time series plot. Descriptive commentary is a second, vital aspect in communicating issues that may affect fitness for purpose. The primary sources of such information are local Agency hydrometry staff and, more formally, station summaries as produced by the National River Flow Archive (see Figure 11.1), which also provides visual summaries of river flow time series for its data holdings.

We would therefore recommend that the GSDQ classification is seen not as a total measure of data quality, but as a classification of gauging station data to be used in conjunction with data visualisation and station summaries to present a balanced picture to data users. It is important that the GSDQ classification scores are not used as a way of avoiding the detective work involved in checking the fitness for purpose of supplied data (most users will probably want more information than the GSDQ score can provide on its own). It is therefore recommended that the Agency consider providing both GSDQ scores and Classification Tool results sheets to users, perhaps along with a disclaimer drawn from the above arguments.

National River Flow Archive
Data Retrieval Service

NRFA Gauging Station Information Sheet

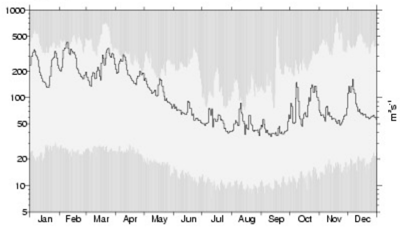
Thames at Kingston

Measuring Authority: Environment Agency
Grid Reference: 51 (TQ) 177 698
Station Type: Ultrasonic

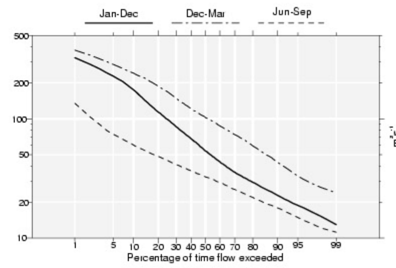
Gauged Flows and Rainfall: 1883-2003
NRFA Station Number: 39001
Local Number: 3400

Daily Flow Hydrograph*

Max. and min. daily mean flows from 1883 to 2003 excluding those for the featured year (2001; mean flow: 125.00 m^3s^{-1})



Flow Duration Curve*



Flow Statistics*

(Units: m^3s^{-1} unless otherwise stated)

Mean flow	78.50	
Mean flow (ls^{-1}/km^2)	7.89	
Mean flow ($10^6m^3/yr$)	2480.0	
Peak flow / date		
Highest daily mean / date	806.0	18 Nov 1894
Lowest daily mean / date	7.370	9 Jul 1934
10 day minimum / end date	9.544	19 Aug 1944
60 day minimum / end date	11.288	30 Aug 1921
240 day minimum / end date	16.336	15 Jan 1922
10% exceedance (Q10)	174.400	
50% exceedance (Q50)	53.810	
95% exceedance (Q95)	18.700	
Mean annual flood		
IH Baseflow index	0.64	

Rainfall and Runoff*

	Rainfall (1883-2001) mm					Runoff (1883-2003) mm				
	Mean	Max/Yr		Min/Yr		Mean	Max/Yr		Min/Yr	
Jan	66	137	1948	13	1997	38	90	1915	8	1997
Feb	50	127	1900	3	1959	33	88	1904	6	1905
Mar	52	142	1947	3	1961	31	100	1947	7	1944
Apr	50	147	2000	3	1938	23	53	2001	7	1976
May	54	137	1932	7	1990	18	49	1932	5	1944
Jun	52	137	1971	3	1925	13	47	1903	4	1944
Jul	57	130	1920	8	1921	10	24	1968	3	1921
Aug	63	147	1912	3	1940	9	24	1931	3	1976
Sep	59	157	1896	3	1959	9	36	1968	3	1898
Oct	74	188	1891	5	1978	14	50	1903	4	1934
Nov	72	188	1940	8	1945	22	84	1894	5	1921
Dec	73	185	1914	13	1933	31	93	1929	6	1921
Year	722	1001	1903	414	1921	249	418	1951	98	1934

Station and Catchment Characteristics

Station level	(mOD)	4.7
Sensitivity	(%)	
Bankfull flow		450.0
Catchment area	(km^2)	9948.
Maximum altitude	(mOD)	330
FSR slope (S1085)	(m/km)	0.32
1961-90 rainfall (SAAR)	(mm)	706
FSR stream frequency (STMFRQ)	($junctions/km^2$)	
Urban extent	(0-1)	0.0429
Flood Attenuation Index	(0-1)	0.9129

Station and Catchment Description

Ultrasonic station commissioned in 1974; multi-path operation from 1986. Full range. No peak flows pre-1974 when dmfs derived from Teddington weir (a 70m wide complex of gates, sluices, weirs and locks); significant structural improvements since 1883 but high hydrometric accuracy not achievable. Tailwater rating used for flows >85 cumecs. Leakage and lockages result in underestimation of pre-1951 low flows; mill operation also evident on early hydrographs. 1894 peak gdf re-assessed (800 cumecs). Baseflow sustained mainly from the Chalk and the Oolites. Runoff substantially decreased by major PWS abstractions - naturalised flows available.

Diverse topography, geology and land use which - together with the pattern of water utilisation - has undergone important historical changes.

Factors Affecting Runoff

- Reservoir(s) in catchment affect runoff.
- Runoff influenced by groundwater abstraction and/or recharge.
- Runoff reduced by public water supply abstraction.
- Runoff reduced by industrial and/or agricultural abstraction.
- Regulation from surface water and/or ground water.
- Runoff increased by effluent returns.

* Featuring, or based on, naturalised flows

Summary of Archived Data

Gauged Flows and Rainfall

Key:

	All rainfall	Some or no rainfall
All daily, all peaks	A	a
All daily, some peaks	B	b
All daily, no peaks	C	c
Some daily, all peaks	D	d
Some daily, some peaks	E	e
Some daily, no peaks	F	f
No gauged flow data	=	-

	0 1 2 3 4	5 6 7 8 9
1880s	- - - c C	C C C C C
1890s	C C C C C	C C C C C
1900s	C C C C C	C C C C C
1910s	C C C C C	C C C C C
1920s	C C C C C	C C C C C
1930s	C C C C C	C C C C C
1940s	C C C C C	C C C C C
1950s	C C C C C	C C C C C
1960s	C C C C C	C C C C C
1970s	C C C C C	B A A A A
1980s	B B A A A	A A A A A
1990s	A A A A A	A A A A A
2000s	A A a e	

Naturalised Flows

Key:

All daily, all monthly	A
Some daily, all monthly	B
Some daily, some monthly	C
Some daily, no monthly	D
No daily, all monthly	E
No daily, some monthly	F
No naturalised flow data	=

	0 1 2 3 4	5 6 7 8 9
1880s	- - - A A	A A A A A
1890s	A A A A A	A A A A A
1900s	A A A A A	A A A A A
1910s	A A A A A	A A A A A
1920s	A A A A A	A A A A A
1930s	A A A A A	A A A A A
1940s	A A A A A	A A A A A
1950s	A A A A A	A A A A A
1960s	A A A A A	A A A A A
1970s	A A A A A	A A A A A
1980s	A A A A A	A A A A A
1990s	A A A A A	A A A A A
2000s	A A A C	

Centre for Ecology and Hydrology, Wallingford, Oxon OX10 8BB, UK.Tel. (01491) 838800. 6th October 2003

Figure 11.1: National River Flow Archive (NRFA) Station Summary Sheet

11.2 Linkage with the National River Flow Archive

The National Water Archive is maintained by the Centre for Ecology and Hydrology (CEH) at Wallingford and is one of the Natural Environment Research Council's eight Designated Data Centres. The aim of the Designated Data Centres is to provide a focus for NERC's environmental data holdings and provide information and advisory services to a wide range of users.

The National River Flow Archive and National Groundwater Level Archive form the core of the National Water Archive. A broad range of hydrological (and related) data are being assimilated into the coordinated management provided by the NWA The NRFA maintains a national database of river flow data and a publicly accessible catalogue of the UK primary gauging station network.

There is potential to harmonise the GSDQ classification with the information available from the NRFA, particularly through the Gauging Station Information Sheets, which are widely used and were shown to be valued during the consultation phase of this project. There is clearly scope for the GSDQ classification results to be incorporated within the NRFA information sheets, and CEH Wallingford envisage doing this, subject to agreement with the Environment Agency.

Conversely, the text summaries held by the NRFA would usefully complement the GSDQ and would help to address situations where significant issues affect data quality or utility that cannot easily be reflected in the GSDQ scheme. There can be an important synergy between the classification and more subjective material compiled by experienced hydrologists. This synergy would help bridge the gap between the GSDQ as a classification of hydrometric quality and the wider issues of data utility for a user community primarily interested in the fitness for purpose of time series data. We would therefore recommend that NRFA descriptive summaries are added to the comments fields of the GSDQ Classification Tool when first compiled. CEH are able to make this information available, indexed by NRFA station number.

11.3 Applications beyond England and Wales

The GSDQ has been developed in close consultation with the Environment Agency (England and Wales) and is designed to reflect the nature of the England and Wales gauging station network, as discussed in Section 2 of this report. Both the Scottish Environment Protection Agency (SEPA) and Rivers Agency (Northern Ireland) were consulted during the R&D project and the Environment Agency is keen to promote uptake of the GSDQ classification by these national agencies, should they wish.

One of the findings of our background review was that there are few, if any, formal data quality classifications used elsewhere, at least in English-speaking countries. The GSDQ would not necessarily translate immediately to use in other countries. In the USA, for example, there are very few formal structures compared with the proportion deployed in the UK, and many rated sections are operated using shifting controls. The GSDQ approach for rated sections treats shift procedures in a manner judged appropriate for the UK situation, but further development might be suitable elsewhere.

Attributes relating to issues not common in the UK, such as ice cover, could easily be introduced to the system.

However, the principles of the attribute scoring approach should translate easily to other hydrometric networks. In particular, the approach for rated sections provides a method of combining statistical measures with more subjective factors in a consistent and repeatable framework. Although the statistical theory within the GSDQ is geared towards the common UK practice of fitting rating curves by least squares, the approach could be adapted to work even when fitting has been carried out in other ways (including 'eyeball' visual curve fitting).

11.4 Further Research

The GSDQ system was based on attribute scoring in part to allow for flexibility in the setting of quality attributes. The choice of attributes, the weights applied to each attribute and the look-up tables that divide the attributes into 'grades' have all been determined by a heuristic process. These components, which together form the scoring schemes for each station type, cannot be set on entirely objective grounds because there is no absolute total measure of data quality. The benchmarking carried out within this study has helped to confirm, and, in some cases, refine the scoring schemes. However the benchmarking could only be carried out on a limited selection of stations and by a relatively small number of hydrometric staff. Whilst there is no reason to believe that the selection of stations was unrepresentative of the broader network, there will be an opportunity, once the GSDQ has been in use for some time, to repeat the benchmarking using a much larger selection of stations.

We would therefore suggest that a repeat benchmarking study be considered after the GSDQ classification has been in use for about a year. This study would collate results from a larger selection of sites, along with the assessments of local hydrometric staff about their validity and, importantly, the component attribute values themselves. The classification information could easily be extracted from GSDQ Classification spreadsheets using automated procedures.

Two useful tasks could then be completed. Firstly, the 'extended benchmarking' could be used either to confirm that the current GSDQ scoring schemes are suitable for long term use or to adjust the schemes to bring greater consistency with the assessments of hydrometric staff . This task would be aided by having used a wide sample of stations to determine objectively the attainable ranges and distributions of attribute values.

The second, related, task would be to provide a clear, empirical summary of the ranges and distributions of values taken by GSDQ attributes and classification scores for England and Wales. This would provide a 'snapshot' of the current quality of gauging station data over the hydrometric network, and could serve as a sound empirical basis for setting targets and measuring improvement.

11.5 Linking GSDQ and WISKI

The GSDQ classification tools have been produced to the Environment Agency's specification as stand-alone customised spreadsheets. This approach is suitable for a classification applied to discrete time periods. However, the WISKI hydrometric database now being rolled out within the Environment Agency would provide a natural home for data quality information. Whilst WISKI includes quality codes relating to specific factors, such as instrument failure or missing data, it could also, in principle, be used to store GSDQ classification scores (and even, potentially, the component attributes).

There are a number of alternative approaches that could be taken. The simplest would be to calculate GSDQ classifications using the current spreadsheet tools, but to transcribe the results for storage in WISKI. A more sophisticated approach would be to add functionality to the GSDQ spreadsheet tools to write GSDQ outputs as files that could be picked up directly by WISKI. The most direct link would be to program the GSDQ algorithms directly in WISKI, making use of the description of the attributes, scoring schemes and software algorithms in this report and the companion GSDQ Software User Guide.

The WISKI system is a time series database and this would open opportunities to refine the way in which the GSDQ classification is managed. Some attributes could be recoded as time series on different scales (for example by-passing or level truncation might be recorded on a daily basis whereas the frequency of check gauging could be calculated for an *N*-year moving window). The GSDQ classification might then become a more gradually varying quality indicator. In any case, integration of GSDQ with WISKI would need to preserve the important comment information currently accommodated within the spreadsheet tools.

11.6 Summary of Recommendations

1) The GSDQ classification should be promoted as one element of a three-part process of flow data quality assessment comprising:

   - the GSDQ classification,
   - data visualisation,
   - descriptive summaries of the station and catchment.

2) The Environment Agency and CEH Wallingford should plan to incorporate GSDQ classification results within the information held and disseminated through the National River Flow Archive.

3) GSDQ classification comments should include text drawn from the NRFA station summary sheets.

4) There should be a complete review of the GSDQ classification results after approximately one year of use to extend the benchmarking carried out within this project and to fine-tune the scoring schemes based on the results.

The benefit of this review would be to strengthen and confirm the classification with a solid empirical assessment.

5) The review proposed in item (4) could easily be combined with analysis both of GSDQ classification results and of the values of attributes across the gauging station network. Such a summary would provide a comprehensive view of gauging station parameters for the UK network and serve as an empirical basis for confirming standards of hydrometric performance.

APPENDIX A

Survey Questionnaire

R&D W6-058 Identification of a method for representing the quality of gauging station data

ENVIRONMENT AGENCY

Customer Survey

jba CONSULTING

Client Environment Agency - North East Region
Date 27 August 2002
Revision 4

Your time taken in completing this questionnaire is greatly appreciated and will be of value to this study. Please note that the content of this form will be treated in the strictest confidence and in accordance with the Data Protection Act. The information you provide will be not be attributable to you The contact details will not be used for any other purpose or divulged to any other party.

RESPONDENT DETAILS	
Name:	
Job Title:	
Organisation:	
Address:	
Country:	
Tel No.:	
Fax No.:	
E-mail:	

Please tick if you are a		
Data supplier	Data user	Both

JBA Consulting, South Barn, Broughton Hall, Skipton, North Yorkshire, BD23 3AE, UK
V:\2002s375 (RL) - Environment Agency - North East Region - R&D W6-058 - Quality of Gauging Station Data\Survey Questionnaire\2002s375 - W6-058 - questionnaire - 4.doc

Page 2 of 10

A . USE OF GAUGING STATION DATA

A1. How frequently do you use/supply gauging station data? *(tick one or more boxes)*					
Frequency of use:		**Daily**	**Weekly**	**Monthly**	**>Monthly**
Categories of use	15 min flows				
	15 min level				
	Daily mean flow				
	Monthly flow stats				
	AM and POT flow				
	Min. flow data				
	Long term data series				
	Other (specify)				

B . USE OF EXISTING GAUGING ACCURACY CLASSIFICATION

The existing Environment Agency gauging station quality classification is based on a statistical comparison between current meter gauging data and rating curve (or accuracy of level measurement, for level-only stations). It also uses BS 3680 for expressing theoretical performance for ungauged BS structures.

The quality code incorporates:

- period of record,
- indicators of quality at low, medium, and high flow ranges
- station type
- an indication of the number of missing data in a record.

For example, a gauge classification code may look like this:

0389-0492, LF1(BS), MF3(BS) HS1(LV), R1

This gauge classification refers to the period from March 1989 to April 1992. The gauge is a British Standard Structure with flow data of Class F1 for Low Flows and Class F3 for Medium Flows. At High Flows however, the structure is drowned out and the gauge reverts to being a level only of Class S1. The reliability of the gauge is R1, eg at least 98% data has been recovered.

B1. Do you use the existing accuracy classification?

Always	Sometimes	Rarely	Never

B2. If not, why not?	*(tick box – more than one if appropriate)*	
A	Don't know it exists	
B	Don't know where the classification record is stored	
C	Have never been offered classification statement with the data	
D	Classification system is too complex	
E	Classification system is too simple	
F	Classification provides no useful information for my uses	
G	Other	

Further comment

B3. If used, what flow range(s) of the classification do you refer to?				
All flows	Low lows	Medium flows	High flows	River level

B4. If used, for what categories of gauging station do you use the classification?						
Level	Ultrasonic	Electromagnetic	BS structure	Non-BS structure	Open Channel	All

B5. Do you reject/accept use of stations on the basis of the classification?	Y	N
(please tick)		
For what purposes?		

B6. Are you satisfied with the existing classification system?	Y	N
(please tick)		
If not, what limitations do you think the system has?		

C. QUANTITATIVE OR CATEGORICAL INFORMATION ON WHICH TO JUDGE THE ACCURACY/RELIABILITY OF A GAUGING STATION

What at-a-glance information would enable you to judge the accuracy/reliability of a gauging station?

C1. Stations where rating accuracy is based on current meter gauging.

(Tick as many boxes as appropriate)

A	Number of gaugings per year since inception of station	
B	Number of gaugings per year since 1990	
C	Number of changes of rating since inception of station	
D	Number of changes or rating since 1990	
E	Number of changes in structure since inception	
F	Number of changes in structure since 1990	
G	Frequency of datum surveys	

H	A value of the standard error (Se) based on gaugings over a given range of flow.	
I	A value of the standard error of the mean relationship (Smr) at a given level.	

J	Ratio of highest current meter gauged Q to highest observed Q (based on whatever rating used, see figure)	
	--- " --- 2nd highest current meter gauged Q (Rank 2, see figure)	
	--- " --- 3rd highest current meter gauged Q (Rank 3, see figure)	

K	Ratio of highest current meter gauged Q to QMED	
	--- " --- 2nd highest current meter gauged Q (Rank 2, see figure)	
	--- " --- 3rd highest current meter gauged Q (Rank 2, see figure)	

L	Ratio of lowest current meter gauged Q to lowest observed Q (based on whatever rating used)	
	--- " --- 2nd lowest current meter gauged Q (Rank N-1, see figure)	
	--- " --- 3rd lowest current meter gauged Q (Rank N-2, see figure)	

M	Ratio of lowest current meter gauged Q to Q_{95}	
	--- " --- 2nd lowest current meter gauged Q (Rank N-1, see figure)	
	--- " --- 3rd lowest current meter gauged Q (Rank N-2, see figure)	

N	Greatest percentage deviation of the three **highest** gauged flows from flows derived from rating curve for given stage	
O	Greatest percentage deviation of the three **lowest** gauged flows from flows derived from rating curve for given stage	
P	Index of sensitivity of flow to level	

Q	Would you prefer a measure in Questions (I) to (L) on the basis of stage rather than flow?	Stage	
		Flow	

What are the four most important measures from the list (A) to (O)? Please rank your selections, in order of 1 (most important) to 4, of those you feel provide the best measures of accuracy/reliability

1. ☐ 2. ☐ 3. ☐ 4. ☐

What at-a-glance information would enable you to judge the accuracy/reliability of a gauging station?

C2. Stations where accuracy is based on a standard structural rating.

(Tick as many boxes as appropriate and add rank where 1 = most important)

		(tick)	(rank)
A	An index of sensitivity of flow to level		
B	Type of station		
C	Modular limit		
D	Whether reduction factors are applied		
E	Vulnerability to accretion		
F	Date of last survey		

What at-a-glance information would enable you to judge the accuracy/reliability of a gauging station?

C3. Stations where accuracy is based on 'transit time' ultrasonics.

(Tick as many boxes as appropriate and rank)

A	Number of bed level surveys per year		
B	Number of current meter check gaugings per year		
C	Whether a procedure has been applied to assess flows exceeding bankfull		
D	Index of 'flight path' failures		
E	Checks for possible suspended solids or aeration problems		

What at-a-glance information would enable you to judge the accuracy/reliability of a gauging station?

C4. Stations where accuracy is based on electromagnetic flow gauging.

(Tick as many boxes as appropriate and rank)

A	Check gaugings on record		
B	Good knowledge of bankfull flow		

What at-a-glance information would enable you to judge the accuracy/reliability of a gauging station?

C5. Level measurement (either for level only or for flow stations).

(Tick as many boxes as appropriate and rank)

A	Number of field manual checks of level per year		
B	Percentage capture rate of water level data		
C	Type of instrument (float / pressure transducer)		

C6. What additional numerical measures would you consider helpful in defining gauging station accuracy/reliability?
A. For stations whose rating reliability is based on current meter gauging:
B. For stations whose rating reliability is based on a standard structural rating:
C. For stations where discharge is based on 'transit time' ultrasonics:\s
D. For stations where discharge is based on electromagnetic flow gauges:
E. For level measurement (either for level only or for flow stations):

D . ADDITIONAL INFORMATION ON WHICH TO JUDGE THE ACCURACY OR RELIABILITY OF A GAUGING STATION

	Y	N
D1. Should the station classification include an index of confidence where there are no measurements to establish accuracy empirically? An example could be the subjective degree of confidence in modularity at high flows based on the opinion of hydrometrics officers.		

		Y	N
D2. Should the station classification include a code to indicate whether validation checks have been carried out where there are no measurements to establish accuracy empirically?			
	(please tick)		
A	Confirmation of high flow rating by hydraulic modelling		
B	Confirmation of non modular rating by u/s and d/s gauges or crest tapping.		
C	Separate extrapolation of Area and Velocity at high flows		
D	Hydrological validation by volume check with u/s and d/s stations		

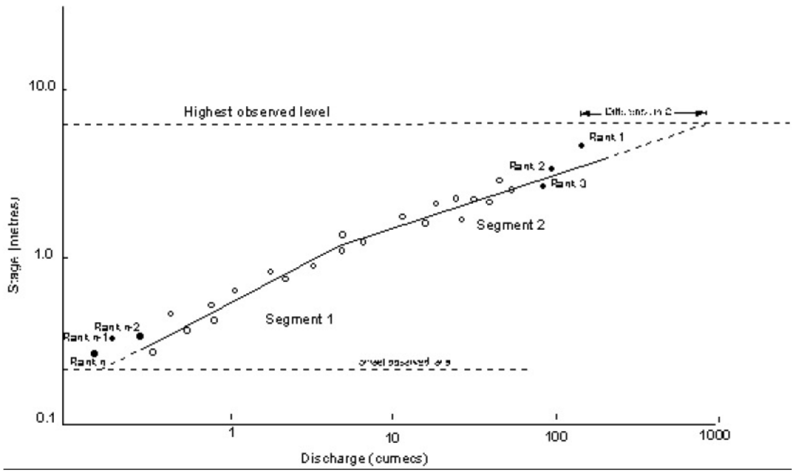
		Y	N
D3. Should the station classification include a code to indicate the existence of attributes of the station which may affect the accuracy or reliability of the station?			
	(please tick)		
A	Weed growth (and level of severity)		
B	Sediment accretion		
C	Tidal influence		
D	Other backwater effects		
E	Known by-passing of flow a high levels		
F	Structural limitations on level recording (e.g. limits on float travel)		
G	Physical capability to gauge full flow range (e.g. cableway at site)		

E. FORM OF THE CLASSIFICATION

E1. Should an abbreviated classification be provided for certain limited users?	Y	N
	☐	☐

E2. Should the classification be…		*(please tick)*
A	… purely numerical?	
B	… purely descriptive?	
C	… both?	

Figure 1. Sample rating curve (see Question C1.)



APPENDIX B

List of Attributes

Table B1.1. List of numeric attributes

Attribute	Level only	Rated section	Structure	US	EM
Width of 95% confidence intervals on flow data	N/A	Confidence intervals derived from gaugings and rating curve. Width of interval increases away from centre of range, therefore evaluated at Q95 and QMED flows.	Derived from BS3680 analysis of theoretical uncertainties. Depends on structure type. Constant throughout range, but evaluated as percentage values of QMED, Q95 and MDF.	Derived from BS3680 analysis of theoretical uncertainties. Depends on path arrangements. Constant throughout range, but evaluated as percentage values of QMED, Q95 and MDF.	N/A
Standard error of deviations on flow data	N/A	N/A	N/A	N/A	Standard deviation between observed flows and check gaugings in high flows, lows flows and entire flow ranges.
Standard error of estimate	N/A	Overall measure of fit between rating curve and gaugings. Expressed as a percentage of the mean daily flow.	N/A	N/A	N/A
Average annual number of missing daily flows	N/A	The typical number of days, per year, for which no daily flow value can be obtained from flow measurements made at the site	The typical number of days, per year, for which no daily flow value can be obtained from flow measurements made at the site	The typical number of days, per year, for which no daily flow value can be obtained from flow measurements made at the site	The typical number of days, per year, for which no daily flow value can be obtained from flow measurements made at the site
Effective accuracy of level measurement (in mm) / truncation	N/A	Estimate of likely errors in measurement of stage, including truncation effects.	Estimate of likely errors in measurement of stage, including truncation effects.	Estimate of likely errors in measurement of stage, including truncation effects.	Estimate of likely errors in measurement of stage, including truncation effects.
Accuracy of level measurement (in mm)	Instrumentation accuracy under ideal operating conditions	N/A	N/A	N/A	N/A
Average annual number of gaugings at flows above 0.5 x QMED	N/A	Provides a measure of how many check/calibration gaugings have been carried out, standardised by record length.	N/A	N/A	Provides a measure of how many check/calibration gaugings have been carried out, standardised by record length.

Attribute	Level only	Rated section	Structure	US	EM
Average annual number of gaugings at flows below Q95	N/A	Provides a measure of how many check/calibration gaugings have been carried out, standardised by record length.	N/A	N/A	Provides a measure of how many check/calibration gaugings have been carried out, standardised by record length.
Average annual number of check gaugings	N/A	Provides a measure of how many check/calibration gaugings have been carried out, standardised by record length.	N/A	N/A	Provides a measure of how many check/calibration gaugings have been carried out, standardised by record length
Longest gap length (in years) between gaugings	N/A	Provides a measure of the distribution of gaugings throughout the lifetime of the gauging station. Long periods without a gauging are taken are considered to be negative.	N/A	N/A	N/A
Sensitivity: % change on flow for a 5mm change in stage at Q95	N/A	Evaluates rating relationship at low flows, based on slope of rating curve at Q95	Evaluates rating relationship at low flows, based on theoretical weir equations	N/A	N/A
Ratio of highest gauged flow to highest recorded flow	N/A	Provides an indication of the degree to which high flow measurements have been ‘confirmed’ by check-gauging	N/A	N/A	Provides an indication of the degree to which high flow measurements have been ‘confirmed’ by check-gauging
Ratio of (Q95-Minimum gauged flow) to (Q95-Minimum recorded flow)	N/A	Provides an indication of the degree to which high flows have been ‘confirmed’ by check-gauging	N/A	N/A	Provides an indication of the degree to which high flows have been ‘confirmed’ by check-gauging
Ratio of gauged flow range to archived flow range	N/A	Provides an indication of the range of applicability of the rating curve	N/A	N/A	N/A

Attribute	Level only	Rated section	Structure	US	EM
Ratio of highest path to highest observed level	N/A	N/A	N/A	Provides an indication of the errors to which high flows can be captured by the path configuration.	N/A
(Elevation at Q95 – elevation of lowest US path) ÷ (Elevation at Q95- Minimum recorded flow)	N/A	N/A	N/A	Provides an indication of the errors to which low flows can be captured by the path configuration.	N/A
Average annual number of bed level surveys	N/A	N/A	N/A	To confirm cross-sectional area does not change over time	N/A
Average annual number of manual checks for level	Provides an indication of whether inaccuracies of level measurement are detected	N/A	N/A	N/A	N/A
Percentage of flows with bias values within 10% in high flow range and 15% otherwise	**N/A**	N/A	Based on observed gaugings in desired flow range, if available	Based on observed gaugings in desired flow range, if available	N/A

Table B1.2. List of descriptive attributes

Attribute	Level only	Rated section	Structure	US	EM
Significance of missing data	N/A	Evaluated for high and low flow ranges	Evaluated for high and low flow ranges	Evaluated for high and low flow ranges	Evaluated for high and low flow ranges
Non capture rate	Approximate percentage of fifteen-minute stage measurements that are not recorded.	N/A	N/A	N/A	N/A
Truncation of measured level	Significance of truncation of level measurements evaluated at high and low flow ranges	N/A	N/A	N/A	N/A
Occurrence of unmeasured bypass flow	N/A	Significance and frequency of unmeasured bypass flow occurring during high flow range	Significance and frequency of unmeasured bypass flow occurring during high flow range	Significance and frequency of unmeasured bypass flow occurring during high flow range	Significance and frequency of unmeasured bypass flow occurring during high flow range
Weed growth management	Provides an indication of the degree by which management practices have mitigated problems associated with measuring flows during periods of weed growth	Provides an indication of the degree by which management practices have mitigated problems associated with measuring flows during periods of weed growth	Provides an indication of the degree by which management practices have mitigated problems associated with measuring flows during periods of weed growth	Provides an indication of the degree by which management practices have mitigated problems associated with measuring flows during periods of weed growth	N/A
Integrity of insulating membrane around coil	N/A	N/A	N/A	N/A	To confirm that gauge is operating correctly
Modular range (incl.) Average number of days per year in which non - modular flow occurs.	N/A	N/A	Extent to which measured flows are affected by non-modular flow over structure	N/A	N/A
Corrections applied for non-modular flows	N/A	N/A	Used to determine corrections applied to flows measured during non-modular conditions	N/A	N/A

Attribute	Level only	Rated section	Structure	US	EM
Type of instrument	Operational accuracy of method	N/A	N/A	N/A	N/A
Siltation management	Indicates level of mitigation of problems associated with siltation	N/A	N/A	N/A	N/A
Local effects / deviation from BS	N/A	N/A	To determine 'bias' associated with flow measurement when no check gaugings are available	To determine 'bias' associated with flow measurement when no check gaugings are available	N/A

APPENDIX C

List of Input Fields

C1. Drop-down lists

The **Gauge type** field refers to the type of gauging station.

The **Instrument type** field refers to the device used to measure stage at a level-only site.

The **Siltation** field addresses the severity and management of any siltation that might occur around the level gauge. It refers to silt affecting the stilling well or access/feeder pipes rather than any siltation in the main channel. A qualitative response is required, therefore the user is required to make some judgement. It is considered only in the level-only scheme.

The **Stability of section** field refers to the channel stability for a rated-section gauge. The user is required to make some judgement regarding the channel stability; as a general guide a concrete or artificial channel may be considered to have good stability, whilst a gravel bed can be considered to have poor stability. It is considered only in the rated-section scheme, but is not used to derive any attribute grades.

The **Weir type / Flume type** field is required for the BS structure scoring scheme. For the purposes of the classification only eight types of structure are considered. For unusual structures the most similar category should be selected (NB many non-standard weirs are treated as rated section). For compound weirs, up to two component weir types may be specified. All flumes are classed as one generic type.

The **Configuration** field refers to the number and arrangement of flight paths used in an Ultrasonic gauge, and is considered only in the ultrasonic scoring scheme. Where two or more flight paths (at different heights in the water column) are used, the gauge is said to have a multi-path configuration. Where two symetrical flightpaths are used to measure the velocity at a particular height in the water column, a cross-configuration is in use.

The **Membrane condition** field refers to the condition of the protective membrane that insulates the EM coil, and is considered only in the electromagnetic scoring scheme. The coil may be located above or below the bed. The user is required to make some judgement regarding the membrane; as a general guide condition will be poor if the membrane is ripped, leaks or shows general deterioration.

The **Bypass flow** field refers to the degree/importance of unmeasured bypass flow around the gauge. Again the user must use some judgement when setting this field (see Section 2 for further guidance).

The **Weed management** field addresses both the severity of weed growth at the site and any management practices that are used to reduce it weed. Removal of weed from the channel and application of a shift/correction procedure are considered as methods of weed management (if correction procedures are applied the accuracy of stage measurement field should be set accordingly).

The **Local effects** field describes the known condition of the gauge (structures and ultrasonic gauges only). Expressed in deviation from the 'British Standard'.

The **Missing data - high flows range** represents the significance / importance of missing data. For example if flood peaks are consistently missed this would be 'significant'.

The **Truncation of stage** field combines the frequency at and degree to which stage measurements are truncated. Truncation is considered separately for high and low flows.

The **Type of correction field** refers to the type procedure applied to correct flow measurements during periods in which the weir/structure is known/thought to have been operating outside its modular range. If no correction procedure is applied the user should select ' uncorrected' from the drop-down menu.

The **Modular range** field describes the approximate position of the modular limit of the weir/structure. The user is able to specify whether flows over the structure are always within the modular range, or otherwise. In the latter case the user must indicate whether the modular limit is above or below the QMED flow.

C2. Numeric fields

The **Start date** field refers to the first day included in the classification.

The **End date** field refers to the last day included in the classification. As a default the end date field is set at today's date.

The **Indicative QMED** field refers to the median annual maximum flow. A numeric value in units of m^3s^{-1}should be entered. QMED will typically be estimated (e.g. using FEH procedures) or calculated from the flow record, if this is of sufficient length. The value entered should be indicative rather than accurate.

The **Indicative Q95** field refers to the flow equalled or exceeded 95% of the time. A numeric value in m^3s^{-1} should be entered. Q95 will typically be calculated from the flow record, or estimated if this is of insufficient length. The value entered should be indicative, but as accurate as possible. For sites where the Q95 is zero, a small value such as 0.01 m^3s^{-1} should be entered.

The **Maximum flow** represents the maximum flow (in m^3s^{-1}) recorded at the site during the entire period of record (not just during the classification period), or if the record period is very short an estimate of the maximum flow may be entered.

The **Minimum flow** represents the minimum flow (in m^3s^{-1}) recorded at the site during the entire period of record (not just during the classification period), or if the record period is very short an estimate of the maximum flow may be entered

The **Mean daily flow** represents the approximate mean value (in m^3s^{-1}) of the daily flows on archive for the gauge during the entire period of record (not just during the classification period).

The **Stage at QMED** field represents the stage value corresponding to the QMED flow. It will usually be determined from the rating table for BS structures or from calibration ratings for EM and US gauges. For sites with hysteresis in the stage-discharge relationship, the largest of the stage values at QMED should be entered. Stage should be given in metres above datum.

The **Stage at Q95** field represents the stage value corresponding to the Q95 flow. It will usually be determined from the rating table for BS structures or from calibration ratings for EM and US gauges. For sites with hysteresis in the stage-discharge relationship, the smallest of the stage values at QMED should be entered. Stage should be given in metres above datum.

The **Maximum stage** field applies specifically for the Ultrasonic gauge type and is the maximum stage recorded at the site during the entire period of record (not just during the classification). Stage should be given in metres above datum.

The Mean bed level represents the typical or average elevation of the river or stream bed. Where the bed surface is very irregular the minimum bed level should be used. Elevation should be given in metres above datum

The **Flow at Q95 stage + 10mm** field applies specifically for the BS structure scoring scheme. It is the flow corresponding to a stage 10mm higher than the stage at Q95 flow, and should be determined from the rating table for the weir/flume.

The **Frequency of stage measurement** field represents the frequency at which stage is recorded, assuming stage is continuously logged. For example if stage was measured at 15 minute intervals the frequency would be 0.25 hours, if recorded daily a frequency of 24 hours should be entered.

The **Number of missing stage measurements** field refers to the number of stage measurements (during the classification period), that are recorded as null or zero values. E.g. if frequency of measurement was 0.25 hours, and the gauge was out of operation for one hour, four measurements would be missing.

The **Number of manual checks on level** field refers to the number of confirmatory manual measurements of stage taken during the classification period.

The **Number of missing daily flow values** field represents the total number of days during the classification period that that have null or zero values on the mean daily flow archive.

The **Typical effective accuracy of stage measurement** field represents the effective accuracy to which stage may be measured, in general, throughout the entire flow range.

The **Effective accuracy of stage measurement - high flows** field represents the effective accuracy to which stage may be measured during periods of high flows (flows between 0.5 x QMED and QMED).

The **Effective accuracy of stage measurement - low flows** field represents the accuracy to which stage may be measured during periods of low flows (flows at or below Q95 flow).

The **Accuracy of tailwater stage measurement** field represents the typical accuracy of tailwater level gauge if operated at the site, and applies specifically to the BS structures scoring scheme.

The **Number of daily flows in non-modular range** field represents the number of days during the classification period for which the weir/structure is known/thought to have been operating outside its modular range. It applies specifically to the BS structures scoring scheme.

The **Highest flight path** is the height of the uppermost flight path operated at an Ultrasonic gauging station. Elevation should be given in m above datum.

The **Lowest flight path** is the height of the lowermost flight path operated at an Ultrasonic gauging station. If the gauge is a single-path type, the lowest and highest path fields will be equal. Elevation should be given in m above datum.

The **Path angle** is the angle (in degrees) between the direction of the flight path and the direction of flow in the channel, for an ultrasonic gauging station. For a multi-path system the mean or typical path angle should be entered.

The **Path length** is the length (in m) of the flight path (i.e. distance between transmitter and receiver) for an ultrasonic gauging station. For a multi-path system the mean / typical path length should be entered.

The **Number of bed surveys** per year is the typical number of surveys of the channel bed (cross-section) per year, and applies specific to the ultrasonic scheme.

C3. Use of input fields for attribute derivation

Table B3-1 details how each of the input data fields is used to calculate attributes. In the table, the following abbreviations are used for each of the gauging station types:

- L Level only
- R Rated section
- S Standard structure
- E Electromagnetic
- U Ultrasonic

The attribute codes are those quoted in the scoring spreadsheets.

Table C.1. Use of input fields in attribute derivation

			Attributes in which variable used				
Input variable	**Schemes**	**Options**	**Level-only scheme**	**Rated-section scheme**	**BS Structure scheme**	**US scheme**	**EM scheme**
Gauge type	L, R, S, E, U	Level only Rated section Rated section with cableway Standard (BS) weir BS weir with rated section at high flows BS weir - compound structure Non-standard weir treated as rated section Non-standard weir treated as BS weir Standard flume Ultrasonic Electromagnetic None	*Used to select appropriate scoring scheme*				
Instrument type	L	Stage board Pressure transducer - pneumatic sensor Pressure transducer - diaphgram sensor Uplooking ultrasonic water level gauge Downlooking ultrasonic water level gauge Shaft encoder Chart recorder Punched tape recorder (PTR)	LG1				
Weir type / Flume type (also compound structure)	S	Triangular profile (Crump) 1:2, 1:5 Triangular profile (Crump) 1:2, 1:2 Triangular profile flat-vee Rectangular thin plate (sharp-crested) Triangular thin plate ('V-notch') Round-nose broad-crested Triangular broad-crested Broad-crested rectangular profile weir			S-H1 S-L1 S-G1		
Configuration	U	Single - path Multi - path Multi-path, cross configuration				U-H1 U-L1 U-G1	
Membrane	E	Poor Condition					E-G4

Table C.1. Use of input fields in attribute derivation

			Attributes in which variable used				
Input variable	**Schemes**	**Options**	**Level-only scheme**	**Rated-section scheme**	**BS Structure scheme**	**US scheme**	**EM scheme**
condition		Condition of membrane unknown Good Condition					
Stability	R	Poor stability Fair Stability (e.g. bedrock) Good Stabililty (e.g. concrete)					
Siltation of stilling well/pipes	L	Severe, or not managed Minor, or partially managed None, or well managed	L-G8				
Bypass flow	L, R, S, E, U	Frequent or significant bypass flow Infrequent or insignificant bypass flow No or negligible bypass flow		R-H4	S-H4	U-H4	E-H4
Weed management	L, R, S, E, U	Weed growth not managed Weed growth partially managed No weed or weed well managed Shift procedures applied	L-G7	R-L8	S-L5	U-L4	
Truncation of stage - high flows	L	Frequent Occasional Rare	L-G5				
Truncation of stage - low flows	L	Frequent Occasional Rare	L-G6				
Local effects	S, U	Strong deviation from BS Some deviation from BS No deviation from BS				U-H6 U-L6 U-G5	
Missing data - high flows range	R, S, E, U	Significant missing data Some missing data Insignificant or no missing data		R-H2	S-H2	U-H2	E-H2
Missing data - low flows range	R, S, E, U	Significant missing data Some missing data Insignificant or no missing data		R-L2	S-L2	U-L2	E-L2
Type of correction	S	Uncorrected High flow rating curve Tailwater measurement Crest-tapping Modular range not exceeded			S-H6 S-G4		
Modular limit	S	Non modular in low flows range Between 0.5 x QMED & QMED Between QMED & 1.5 QMED Above 1.5 x QMED Modular range not exceeded			S-H6		
Site Name	L, R, S, E, U	Text string					
River Name	L, R, S, E, U	Text string					
Reference	L, R, S, E, U	Text string					
Start date	L, R, S, E, U	Date	L-G2 L-G4	R-H5 R-L5 R-G2 R-G3	S-H6 S-G2 S-G4	U-G2 U-G4	E-H5 E-L4 E-G2 E-G6 E-G5
End date	L, R, S, E, U	Date					
QMED	R, S, E, U	Numeric value in m^3s^{-1}		R-H1 R-H5 R-H7	S-H6 S-L4	U-H1	E-H1 E-H5

Table C.1. Use of input fields in attribute derivation

			Attributes in which variable used				
Input variable	**Schemes**	**Options**	**Level-only scheme**	**Rated-section scheme**	**BS Structure scheme**	**US scheme**	**EM scheme**
Q95	R, S, E, U	Numeric value in m^3s^{-1}		R-L1 R-L4 R-L5 R-L6 R-L7		U-L1	E-L1 E-L5 E-L4
Maximum Flow	R, S, E, U	Numeric value in m^3s^{-1}		R-H6 R-G4			E-H6
Minimum Flow	R, S, E, U	Numeric value in m^3s^{-1}		R-L6 R-G4	S-H6		E-L5
Mean daily flow	R, S, E, U	Numeric value in m^3s^{-1}				U-G1	
Maximum stage	L	Numeric value in m				U-H5	
Stage at QMED	S	Numeric value in m					
Stage at Q95	S	Numeric value in m			S-L4		
Flow at Q95 stage + 10mm	S	Numeric value in m^3s^{-1}			S-L4		
Frequency of stage measurement	L	Numeric value in hours	L-G2				
Number of missing stage measurements	L	Integer	L-G2				
Number of manual checks on level	L	Integer	L-G4				
Number missing daily flow values	R, S, E, U	Integer		R-G2	S-G2	U-G2	
Number daily flows in non-modular range	S	Integer			S-H6 S-G4	E-G2	
Typical accuracy of stage measurement	L, R, S, E, U	Numeric value in mm	L-G3		S-G3	U-G1 U-G3	E-G3
Accuracy of stage measurement - high flows	R, S, E, U	Numeric value in mm		R-H3	S-H3	U-H1 U-H3	E-H3
Accuracy of stage measurement - low flows	R, S, E, U	Numeric value in mm		R-L3	S-L3	U-L1 U-L3	E-L3
Accuracy of tailwater stage measurement	S	Numeric value in mm			S-H6		
Highest flight path	U	Numeric value in m				U-H5	
Lowest flight path	U	Numeric value in m				U-L5	
Path angle	U	Numeric values in decimal degrees				U-H1 U-L1 U-G1	
Path length	U	Numeric value in m				U-H1 U-L1 U-G1	
Total number of bed surveys	U	Integer				U-G4	
Mean bed level	U	Numeric value in m				U-L5	
Ratings	R			R-H1 R-L1 R-L4 RG1			

Table C.1. Use of input fields in attribute derivation

			Attributes in which variable used				
Input variable	**Schemes**	**Options**	**Level-only scheme**	**Rated-section scheme**	**BS Structure scheme**	**US scheme**	**EM scheme**
Check Gaugings	R, S, E, U			R-H1 R-H5 R-H6 R-H7 R-L1 R-L5 R-L6 R-L7 R-G1 R-G2 R-G4	S-H5 S-H6 S-L6 S-G5	U-H6 U-L5 U-G5	E-H1 E-H5 E-H6 E-H7 E-L1 E-L6 E-G1 E-G5 E-G6

APPENDIX D

Scoring Scheme Reference Tables

Table D.1. Rated Section Scoring Scheme

	Attribute code	Attribute description	Look-up table					Weight
			1 (Worst)	2	3	4	5 (Best)	
High Flows	R-H1	Width of 95% confidence interval at QMED (as a % of QMED)	> 25	20 - 25	15 - 20	10 - 15	≤ 10	1.6
	R-H2	Significance of missing data	Significant		Some		Insignificant	0.5
	R-H3	Effective accuracy of level measurement (mm)	> ± 30	± 20 - 30	± 10 - 20	± 6 - 10	≤ ± 6	0.7
	R-H4	Occurrence of unmeasured bypass flow	Severe / frequent		Infrequent / minor		Rare / none	1
	R-H5	Average annual number of gaugings at flows over 0.5 x QMED	≤ 0.1	0.1 - 0.5	0.5 - 1	1 - 2	> 2	1
	R-H6	Maximum gauged flow ÷ maximum archived flow	≤ 0.5	0.5 - 0.7	0.7 - 0.8	0.8 - 0.9	> 0.9	1.4
	R-H7	Longest gap length between gaugings at flows over 0.5 x QMED (years)	> 5	2 - 5	1 - 2	0.5 -1	≤ 0.5	0.8
Low Flows	R-L1	Width of 95% confidence interval at Q95 (as a % of Q95)	>16	12 - 16	8 - 12	4 - 8	≤ 4	1.6
	R-L2	Significance of missing data	Significant		Some		Insignificant	0.5
	R-L3	Effective accuracy of level measurement (mm)	> ± 15	± 10 - 15	± 5 - 10	± 3 - 5	< ± 3	0.7
	R-L4	Sensitivity (%)	> 40	30 - 40	20 - 30	10 - 20	≤ 10	1.2
	R-L5	Average annual number of gaugings at flows below Q95	≤ 0.1	0.1 - 0.5	0.5 - 1	1 - 2	> 2	1
	R-L6	(Q95-minimum gauged flow) ÷ (Q95 - minimum archived flow)	≤ 0.5	0.5 - 0.7	0.7 - 0.8	0.8 - 0.9	> 0.9	1
	R-L7	Longest gap length between gaugings at flows below Q95 (years)	> 5	2 - 5	1 - 2	0.5 - 1	≤ 0.5	0.8
	R-L8	Weed growth management	Poor / severe weed		Partially managed		Good / no weed	1.2
General	R-G1	Standard error of estimate (as a % of daily mean flow)	> 8	6 - 8	4 - 6	2 - 4	≤ 2	1.2
	R-G2	Average annual number of missing daily flows	> 21	14 - 21	7 - 14	3 - 7	≤ 3	0.8
	R-G3	Effective accuracy of level measurement (mm)	> ± 15	± 10 - 15	± 5 - 10	± 3 - 5	< ± 3	1
	R-G4	Average annual number of check gaugings	≤ 1	1 - 5	5 - 10	10 - 30	> 30	1
	R-G5	Gauged flow range ÷ archived flow range	≤ 0.5	0.5-0.7	0.7-0.8	0.8-0.9	> 0.9	1

Table D.2. BS Structure Scoring Scheme

	Attribute code	Attribute description	Look-up table					Weight
			1 (Worst)	2	3	4	5 (Best)	
High Flows	S-H1	Width of 95% confidence interval based on BS3680 (as a % of QMED)	> 25	20 - 25	15 - 20	10 - 15	≤ 10	1.6
	S-H2	Significance of missing data	Significant		Some		Insignificant	0.5
	S-H3	Effective accuracy of upstream level measurement (mm)	> ± 30	± 20 - 30	± 10 - 20	± 6 - 10	≤ ± 6	0.7
	S-H4	Occurrence of unmeasured bypass flow	Severe / frequent		Infrequent / minor		Rare / none	1.2
	S-H5	Percentage archived flows (over 0.5x QMED) within ±10% of gauged **OR** Deviation from BS	Severe Deviation		Moderate Deviation		No Deviation	1
	S-H6	Corrections applied for non-modular flows	Uncorrected theoretical rating used for non-modular flows	Rating based on flow gaugings used for non-modular flows	Correction factor based on d/s level applied for non modular flows	Correction factor based on crest-tapping applied for non modular flows	High flows "always" within modular range	*1*
		Modular limit, if non-modular flows occur within some/all of high-flows range	Between 0.5 x QMED and QMED		Between QMED and 1.5 x QMED		Greater than 1.5 x QMED	*0.15*
		Average annual number of gaugings at flows over 0.5 x QMED	'≤ 0.1	0.1 - 0.5	0.5 - 1	1 - 2	> 2	*0.1*
		Maximum gauged flow ÷ maximum archived flow	'≤ 0.5	0.5 - 0.7	0.7 - 0.8	0.8 - 0.9	> 0.9	*0.1*
		Longest gap length between gaugings (years)	> 5	2 - 5	1 - 2	0.5 - 1	≤ 0.5	*0.05*
		Effective accuracy of tailwater stage measurement or crest-tapping correction (mm)	> ± 30	± 20 - 30	± 10 - 20	± 6 - 10	≤ ± 6	*0.05*

	Attribute code	Attribute description	Look-up table					Weight
			1	2	3	4	5	
Low Flows	S-L1	Width of 95% confidence interval based on BS3680 (as a % of Q95)	> 16	12 - 16	8 - 12	4 - 8	≤ 4	1.6
	S-L2	Significance of missing data	Significant		Some		Insignificant	0.5
	S-L3	Effective accuracy of (upstream) level measurement (mm)	> ± 15	± 10 - 15	± 5 - 10	± 3 - 5	≤ ± 3	0.5
	S-L4	Sensitivity (%)	> 40	30 - 40	20 - 30	10 -20	≤ 10	1.2
	S-L5	Weed growth management	Poor / severe weed		Partially managed		Good / no weed	1.2
	S-L6	Percentage archived flows (< Q95) within ±15% of gauged **OR** Deviation from BS	Severe deviation		Moderate deviation		No deviation	1
General	S-G1	Width of 95% confidence interval based on BS3680 (as a % of daily mean flow)	> 16	12 - 16	8 - 12	4 - 8	≤ 4	1.5
	S-G2	Average annual number of missing daily flows	> 21	14 - 21	7 - 14	3 - 7	≤ 3	0.8
	S-G3	Effective accuracy of level measurement (mm)	> ± 15	± 10 – 15	± 5 - 10	± 3 - 5	≤ ± 3	0.7
	S-G4	Modular Range	Non-modular flows within low flow range	Between 0.5 x QMED and QMED	Between QMED and 1.5 x QMED	Greater than 1.5 x QMED	Flows 'always' within modular range	*1*
		Approx. average annual number days in which non - modular flow occurs	> 14	7 - 14	3 - 7	1 - 3	0 - 1	*0.2*
	S-G5	Percentage archived flows (full flow range) within ±15% of gauged **OR** Deviation from BS	Severe deviation		Moderate deviation		No deviation	1

Table D.3. BS Structure operating a rating at high flows

	Attribute code	Attribute description	Look-up table					Weight
			1 (Worst)	2	3	4	5 (Best)	
High Flows	R-H1	Width of 95% confidence interval at QMED (as a % of QMED)	> 25	20 - 25	15 - 20	10 - 15	≤ 10	1.6
	R-H2	Significance of missing data	Significant		Some		Insignificant	0.5
	R-H3	Effective accuracy of level measurement (mm)	> ± 30	± 20 - 30	± 10 - 20	± 6 - 10	≤ ± 6	0.7
	R-H4	Occurrence of unmeasured bypass flow	Severe / frequent		Infrequent / minor		Rare / none	1
	R-H5	Average annual number of gaugings at flows over 0.5 x QMED	≤ 0.1	0.1 - 0.5	0.5 - 1	1 - 2	> 2	1
	R-H6	Maximum gauged flow ÷ maximum archived flow	≤ 0.5	0.5 - 0.7	0.7 - 0.8	0.8 - 0.9	> 0.9	1.4
	R-H7	Longest gap length between gaugings (years)	> 5	2 - 5	1 - 2	0.5 -1	≤ 0.5	0.8
Low Flows	S-L1	Width of 95% confidence interval based on BS3680 (as a % of Q95)	>16	12 - 16	8 - 12	4 - 8	≤ 4	1.6
	S-L2	Significance of missing data	Significant		Some		Insignificant	0.5
	S-L3	Effective accuracy of (upstream) level measurement (mm)	> ± 15	± 10 - 15	± 5 - 10	± 3 - 5	≤ ± 3	0.5
	S-L4	Sensitivity (%)	> 40	30 - 40	20 - 30	Oct-20	≤ 10	1.2
	S-L5	Weed growth management	Poor / severe weed		Partially managed		Good / no weed	1.2
	S-L6	Percentage archived flows (< Q95) within ±15% of gauged **OR** Deviation from BS	0 - 30	30 - 45	45 - 60	60 - 75	75 - 100	1
General	S-G1	Width of 95% confidence interval based on BS3680 (as a % of daily mean flow)	> 16	12 - 16	8 - 12	4 - 8	≤ 4	1.5
	S-G2	Average annual number of missing daily flows	> 21	14 - 21	7 - 14	3 - 7	≤ 3	0.8
	S-G3	Effective accuracy of level measurement (mm)	> ± 15	± 10 - 15	± 5 - 10	± 3 - 5	≤ ± 3	0.7
	S-G4	Modular Range	Non-modular at 'low flows'	Modular limit below QMED	Modular limit between QMEd and 1.5 x QMED	Modular limit above 1.5 x QMED	Flows "always" in modular range	1
		Average annual number of days in which non-modular flow occurs	> 14	7 - 14	3 - 7	1 - 3	≤ 1	0.2
	S-G5	Percentage archived flows (full flow range) within ±15% of gauged **OR** Deviation from BS	0 - 30	30 - 45	45 - 60	60 - 75	75 - 100	1

Table D.4. Ultrasonic Scoring Scheme

	Attribute code	Attribute description	Look-up table					Weight
			1 (Worst)	2	3	4	5 (Best)	
High Flows	U-H1	Width of 95% confidence interval based on BS3680 (as a % of QMED)	> 25	20 - 25	15 - 20	10 - 15	≤ 10	1.6
	U-H2	Significance of missing data	Significant		Some		Insignificant	0.5
	U-H3	Effective accuracy of level measurement (mm)	± > 15	± 10 - 15	± 5 - 10	± 3 - 5	± ≤ 3	0.5
	U-H4	Occurrence of unmeasured bypass flow	Severe / frequent		Infrequent / minor		Rare / none	1
	U-H5	Height of uppermost path ÷ max. archived stage	≤ 0.5	0.5 - 0.7	0.7 - 0.8	0.8 – 0.9	> 0.9	1.4
	U-H6	Percentage archived flows (over 0.5xQMED) within ±15% of gauged **OR** Deviation from BS	Severe deviation		Moderate deviation		No deviation	1
Low Flows	U-L1	Width of 95% confidence interval based on BS3680 (as a % of Q95)	> 16	12 - 16	8 - 12	4 - 8	≤ 4	1.6
	U-L2	Significance of missing data	Significant		Some		Insignificant	0.5
	U-L3	Effective accuracy of level measurement (mm)	± > 15	± 10 - 15	± 5 - 10	± 3 - 5	± ≤ 3	0.5
	U-L4	Weed growth management	Poor / severe weed		Partially managed		Good / no weed	1.4
	U-L5	(H95 - height of lowermost path) ÷ (H95 - mean bed level below lowest path) (%)	≤ 20	20 - 40	40 - 60	60 - 80	> 80	1
	U-L6	Percentage archived flows (< Q95) within ±15% of gauged **OR** Deviation from BS	Severe deviation		Moderate deviation		No deviation	1
Overall Regime	U-G1	Width of 95% confidence interval based on BS3680 (as a % of daily mean flow)	> 16	12 - 16	8 - 12	4 - 8	≤ 4	1.2
	U-G2	Average annual no. of missing daily flows	> 21	14 - 21	7 - 14	3 - 7	≤ 3	0.8
	U-G3	Effective accuracy of level measurement (mm)	> ± 15	± 10 - 15	± 5 - 10	± 3 - 5	± ≤ 3	0.5
	U-G4	Average annual number of bed-level surveys	≤ 0.25	0.25 - 0.5	0.5 - 0.75	0.75 - 1	> 1	1.2
	U-G5	Percentage archived flows (full flow range) within ±15% of gauged **OR** Deviation from BS	Severe deviation		Moderate deviation		No deviation	1.3

Table D.5. Electromagnetic Scoring Scheme

	Attribute code	Attribute description	Look-up table					Weight
			1 (Worst)	2	3	4	5 (Best)	
High Flows	E-H1	Standard error of deviations for flows over 0.5 x QMED (as a % of QMED)	> 12.5	10 - 12.5	7.5 - 10	5 - 7.5	≤ 5	1.6
	E-H2	Significance of missing data	Significant		Some		Insignificant	0.6
	E-H3	Effective accuracy of level measurement (mm)	> ± 30	± 20 - 30	± 10 - 20	± 6 - 10	≤ ± 6	0.8
	E-H4	Occurrence of unmeasured bypass flow	Severe / frequent		Infrequent / minor		Rare / none	1.2
	E-H5	Average annual number of gaugings at flows over 0.5 x QMED	≤ 0.1	0.1 - 0.5	0.5 - 1	1 - 2	> 2	0.8
	E-H6	Maximum gauged flow ÷ maximum archived flow	≤ 0.5	0.5 - 0.7	0.7 - 0.8	0.8 - 0.9	> 0.9	0.8
Low Flows	E-L1	Standard error of deviations for flows below Q95 (as a % Q95)	> 8	6 - 8	4 - 6	2 - 4	≤ 2	1.6
	E-L2	Significance of missing data	Significant		Some		Insignificant	0.5
	E-L3	Effective accuracy of level measurement (mm)	> ± 15	± 10 - 15	± 5 - 10	± 3 - 5	≤ ± 3	0.7
	E-L4	Average annual number of gaugings at flows below Q95	≤ 0.1	0.1 - 1.5	0.5 - 1	1 - 2	> 2	1
	E-L5	(Q95-minimum gauged flow) ÷ (Q95 - minimum archived flow)	≤ 0.5	0.5 - 0.7	0.7 - 0.8	0.8 - 0.9	> 0.9	1
General	E-G1	Standard error of deviations for full range of flows (as a % of daily mean flow)	> 8	6 - 8	4 - 6	2 - 4	≤ 2	1.2
	E-G2	Average annual number of missing daily flows	> 21	14 - 21	7 - 14	3 - 7	≤ 3	0.7
	E-G3	Effective accuracy of level measurement (mm)	> ± 15	± 10 - 15	± 5 - 10	± 3 - 5	≤ ± 3	0.7
	E-G4	Integrety of insulating membrane around coil	Poor condition		Unknown condition		Good condition	1.2
	E-G5	Average annual number of check gaugings	≤ 1	1 - 5	5 - 10	10 - 30	> 30	1

Table D.6. Level-only Scoring Scheme

	Attribute code	Attribute description	Look-up table					Weight
			1 (Worst)	2	3	4	5 (Best)	
General	L-G1	Type of instrument	Stage board	Chart/tape recorders	Ultrasonic types	Pressure transducers	Shaft encoder	2
	L-G2	Non-capture rate (percentage missing data)	> 10	5 - 10	1 - 5	0.5 - 1	≤ 0.5	0.8
	L-G3	Accuracy of level measurement (mm)	> ± 15	± 10 - 15	± 5 - 10	± 3 - 5	≤ ± 3	1.2
	L-G4	Average annual number of manual checks for level	≤ 6	6 - 12	12 - 20	20 - 40	>40	1.2
	L-G5	Truncation (of measured level) at high flows	Frequent		Occasional		Rare	0.4
	L-G6	Truncation (of measured level) at low flows	Frequent		Occasional		Rare	0.4
	L-G7	Siltation management (of intake pipe & stilling well)	Not managed		Partially managed		Well managed	0.8