Guidance for understanding and using hydrometric information received from SEPA

Contents

[Glossary 1](#_Toc58849466)

[1 Data use conditions and attribution statement 2](#_Toc58849467)

[2 Quality assurance 2](#_Toc58849468)

[3 Historic data quality – pre 2007 3](#_Toc58849469)

[4 Confidence in flow estimates 3](#_Toc58849470)

[4.1 Factors affecting confidence in flow estimation 3](#_Toc58849471)

[4.2 Sources of information on flow estimation confidence 4](#_Toc58849472)

[5 The hydrometric calendar 4](#_Toc58849473)

[6 Data file Formats and delivery method 4](#_Toc58849474)

[7 ZRXP format explained 6](#_Toc58849475)

# Glossary

|  |  |
| --- | --- |
| A2I | Access to Information – SEPA’s Access to Information Team |
| AMAX | Annual maximum flood series |
| EIR | Environmental Information Requests |
| FOI | Freedom of Information Requests |
| Hydrological day | The industry standard for daily figures is the hydrometric day, which starts at 9am GMT |
| Hydrological year | The industry standard for annual data is the hydrological year which is the period from 1st October to 30 September. For example, in the UK, the 2005 hydrological year is the period from the 1st October 2005, 0900 to 30th September 2006, 0845. |
| Hydrometric Information | A term that encompasses the data, meta data and any supporting comments and caveats on confidence etc. |
| Instantaneous 'event tips' | A tipping bucket records the time every 0.2mm of rainfall. As standard SEPA’s aggregates these into total rainfall in every 15minute period. At some stations an additional series exists that records the time of every tip (0.2mm). This higher resolution allows analysis of intense downpours. |
| Meta Data | Information that helps understand the data and put it into context |
| NetHelpDesk | The application for managing requests for information from SEPA |
| NRFA | National River Flow Archive |
| Parameter | Refers to what is being measured. For example, River Discharge/Flow (Q) and Event Rainfall (RE) are parameters |
| POT | Peaks over threshold series |
| SEPA | Scottish Environment Protection Agency |
| Stage | Another term for water level, often with the abbreviation SG |
| Times series | Each parameter has different time series (resolutions) available. For example, 15minute flow and daily mean flow are two different time series from the same parameter |
| WIN-FAP | [Commercial software that incorporates flood estimation methodology and Peak Flow Datasets, supplied by https://www.hydrosolutions.co.uk/](https://www.hydrosolutions.co.uk/) |

# Data use conditions and attribution statement

© SEPA, 20\*\*. SEPA hydrometric data is licensed under the [Open Government Licence 3.0](http://www.nationalarchives.gov.uk/doc/open-government-licence/version/3). SEPA will supply all available data requested but advises that it is the requesters responsibility to ensure the data is fit for their purpose. Whilst all reasonable effort has been made to ensure that these data are accurate for SEPA’s intended purpose, no warranty is given by SEPA.

# Quality assurance

SEPA hydrometric data is quality controlled and coded ( Table 1 ) on a monthly cycle. Recent data is tagged with the default ‘unchecked’ code 254 until the quality control is undertaken. The unchecked status will apply to any ‘recent’ data that is available from the self-help options via the [SEPA website](https://www.sepa.org.uk/) and this provisional data should be used with caution.

Table 1 – quality codes

|  |  |  |
| --- | --- | --- |
| Code number | Code name | Abbreviated code name |
| 50 | Good | G |
| 100 | Estimated | E |
| 140 | Provisional | Prov |
| 150 | Suspect | S |
| 200 | Unchecked (imported from legacy database) | V |
| 254 | Unchecked | U |

Data available via our partners at the [National River Flow Archive](https://nrfa.ceh.ac.uk/) (NRFA) has been through SEPA quality control and a secondary quality control process by the Centre for Ecology and Hydrology This data is typically suitable for use in flood studies and academic studies.

The quality flags assigned to flow data reflects both the quality of the level date and the confidence in the rating accuracy. Flows based on an acceptable extrapolation beyond the highest or lowest valid gauging are referred to as ‘extrapolated values’ and are coded as estimated (100). Flows beyond the acceptable extrapolation limit are coded as suspect (150). The coding reflects a “worst wins” basis between stage quality and the status of the flow rating. For example, poor stage and good rating (in validated range) = poor flow data.

# Historic data quality – pre 2007

The term *historic data* largely refers to data prior to when SEPA’s current national database became operational in 2007. Prior to this, any data in digital format was stored in legacy regional databases before being imported into the current database. Any data that had not been quality controlled in the legacy system is assigned with code 200 (unchecked) in the current database.

Nearly all historic data sets will have been quality controlled to some degree, but the methods used may not align with current national practice. Please also note that historically differences in quality control existed between regions.

If you request historic data from 5 or less stations, SEPA staff can typically provide bespoke supporting comments on data confidence if deemed pertinent by SEPA. If you request more than 5 sites, SEPA staff are not able to proactively comment on historic data confidence but may be able to investigate further for specific stations and time periods on request.

Please note that river level records before the 1970s, and as late as the1990s in some locations, were measured on clockwork instruments with analogue paper chart records. For a number of SEPA stations, only derived summary datasets such as mean daily flow are available for the early period of record before the installation of digital loggers.

# Confidence in flow estimates

## Factors affecting confidence in flow estimation

All river flow data are estimates of the true value and the accuracy is dependent on many factors that are unique to each station.

The station design and how it is operated, based on station purpose, are factors that account for the flow accuracy. For example, many stations built in the 1970s by the River Purification Boards were designed and operated for flow estimation in the low to mid-range only. It’s common for a station purpose and design to change through its life span.

Relatively lower confidence is assigned to extreme low and high flow estimations that are often beyond the capability of the gauging station infrastructure and are relatively more difficult to gauge. large floods can also dramatically change a river channel, resulting in periods of low confidence.

For flood estimation, the confidence in the derived flow is only good up to the highest gauged flow. This can be well below the highest recorded flood. Derived flows above the highest gauging are typically based on a simple extrapolation of the rating. Although a flow figure is calculated, it may be well beyond the calibration limit of the station and be significantly out from the true value. In these cases, you may wish to extend the rating, for example, by carrying out your own hydraulic modelling.

## Sources of information on flow estimation confidence

In addition to the codes assigned to the supplied values (see section 2), other sources of information on data confidence are provided below.

SEPA can usually provide the goodness of fit statistics and confidence bands for the rating curve that the flows are based on and may be able to provide additional information on request.

Metadata on the NRFA [website](https://nrfa.ceh.ac.uk/data/search) is useful for putting the flow data into context, understanding station performance and data confidence. For understanding confidence in flood estimations, useful metadata includes the highest recorded stage on record, the highest gauging, bankfull level, pictures, rating and gauging plots in addition to station notes. All such information is provided on the NRFA website for stations with Peak Flow status. The ‘indicative suitability’ status of the station assigned in Peak Flows also indicates whether there is some confidence in the very extreme flood events (pooling status) or if there is confidence to QMED levels only (no pooling status).

# The hydrometric calendar

All aggregate time series (daily/monthly/annual) are based on the industry standard UK hydrometric day, which begins at 9am GMT. Annual figures will be based on the UK hydrological year (1st October to 30 September) unless you specifically request calendar year

# Data file Formats and delivery method

SEPA hydrometric data will normally be sent by email and provided in CSV format which can be opened in Microsoft Excel.

**Large files**

The datasets may be condensed in a ‘zipped` folder to reduce the file size and this will have to be opened by right clicking on the file and choosing the extract or open option. If the files are over 5MB even when zipped, the information will be made available on the [SEPA disclosure log](https://www2.sepa.org.uk/disclosurelog/)

For very large data sets where 15min resolution is required from numerous sites, the data may have to be supplied in ZRXP format. This format is described in detail below and some basic coding will be required to extract data from the export file.

**What the data looks like**

The examples below illustrate the layout of a CSV format when uploaded into excel for various combinations of parameter and time series.

River Flow (Q) - 15 minute series

|  |  |  |
| --- | --- | --- |
| Team: | East Kilbride |  |
| StationName: | Daldowie |  |  |
| StationNumber: | 133074 |  |  |
| Easting: | 267184 |  |  |
| Northing: | 661676 |  |  |
| ParameterName: | Q |  |  |
| ParameterType: | Q |  |  |
| TimeseriesName: | 4/133074/Q/15m.Cmd |
| Date | Time | Value[m³/s] | State of value |
| 01/01/2020 | 00:00:00 | 29.46 | 50 (G) |
| 01/01/2020 | 00:15:00 | 29.306 | 50 (G) |
| 01/01/2020 | 00:30:00 | 29.229 | 50 (G) |
| 01/01/2020 | 00:45:00 | 29.152 | 50 (G) |

River Flow (Q) – mean daily flow

|  |  |  |
| --- | --- | --- |
| Team: | East Kilbride |  |
| StationName: | Daldowie |  |  |
| StationNumber: | 133074 |  |  |
| Easting: | 267184 |  |  |
| Northing: | 661676 |  |  |
| ParameterName: | Q |  |  |
| ParameterType: | Q |  |  |
| TimeseriesName: | 4/133074/Q/HDay.Mean |
| Date | Time | Value[m³/s] | State of value |
| 01/01/2020 | 09:00:00 | 27.695 | 50 (G) |
| 02/01/2020 | 09:00:00 | 37.933 | 50 (G) |
| 03/01/2020 | 09:00:00 | 51.456 | 50 (G) |
| 04/01/2020 | 09:00:00 | 54.008 | 50 (G) |

River Flow (Q) – AMAX series

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Team: | East Kilbride |  |  |  |
| StationName: | Daldowie |  |  |  |  |
| StationNumber: | 133074 |  |  |  |  |
| Easting: | 267184 |  |  |  |  |
| Northing: | 661676 |  |  |  |  |
| ParameterName: | Q |  |  |  |  |
| ParameterType: | Q |  |  |  |  |
| TimeseriesName: | 4/133074/Q/HYear.Max |  |  |
| Date | Time | Value[m³/s] | State of value | Occurrence time stamp | Occurrence count |
| 01/10/1976 | 09:00:00 | --- | 200 (255) |  | 0 |
| 01/10/1977 | 09:00:00 | --- | 200 (255) |  | 0 |
| 01/10/1978 | 09:00:00 | 351.064 | 200 (V) | 16/11/1978 13:00 | 1 |
| 01/10/1979 | 09:00:00 | 479.186 | 200 (V) | 26/11/1979 10:30 | 1 |
| 01/10/1980 | 09:00:00 | 710.215 | 50 (G) | 18/01/1981 00:30 | 1 |
| 01/10/1981 | 09:00:00 | 558.28 | 100 (E) | 04/01/1982 04:15 | 1 |

River level/stage (SG) – 15 minute resolution

|  |  |  |
| --- | --- | --- |
| Team: | East Kilbride |  |
| StationName: | Daldowie |  |  |
| StationNumber: | 133074 |  |  |
| Easting: | 267184 |  |  |
| Northing: | 661676 |  |  |
| ParameterName: | SG |  |  |
| ParameterType: | S |  |  |
| TimeseriesName: | 4/133074/SG/15m.Cmd |
| Date | Time | Value[m] | State of value |
| 01/01/2020 | 00:00:00 | 0.586 | 50 (G) |
| 01/01/2020 | 00:15:00 | 0.584 | 50 (G) |
| 01/01/2020 | 00:30:00 | 0.583 | 50 (G) |
| 01/01/2020 | 00:45:00 | 0.582 | 50 (G) |

Event Rainfall (RE) – 15minute resolution

|  |  |  |
| --- | --- | --- |
| Team: | East Kilbride |  |
| StationName: | Blackdyke Farm |  |
| StationNumber: | 115516 |  |  |
| Easting: | 253265 |  |  |
| Northing: | 628638 |  |  |
| ParameterName: | RE |  |  |
| ParameterType: | Precip |  |  |
| TimeseriesName: | 4/115516/RE/15m.Total |
| Date | Time | Value[mm] | State of value |
| 02/01/2020 | 11:45:00 | 0 | 50 (G) |
| 02/01/2020 | 12:00:00 | 0.2 | 50 (G) |
| 02/01/2020 | 12:15:00 | 0.6 | 50 (G) |
| 02/01/2020 | 12:30:00 | 0.4 | 50 (G) |

Event Rainfall (RE) – Daily total

|  |  |  |
| --- | --- | --- |
| Team: | East Kilbride |  |
| StationName: | Blackdyke Farm |  |
| StationNumber: | 115516 |  |  |
| Easting: | 253265 |  |  |
| Northing: | 628638 |  |  |
| ParameterName: | RE |  |  |
| ParameterType: | Precip |  |  |
| TimeseriesName: | 4/115516/RE/HDay.Total |
| Date | Time | Value[mm] | State of value |
| 01/01/2020 | 09:00:00 | 0 | 50 (G) |
| 02/01/2020 | 09:00:00 | 5.6 | 50 (G) |
| 03/01/2020 | 09:00:00 | 9 | 254 (U) |
| 04/01/2020 | 09:00:00 | 2.4 | 50 (G) |

# ZRXP format explained

ZRXP is a line orientated text format structured with a header block describing the data followed by lines of data. A file may contain data from many sources, data from separate sources being separated by a header block. A summary of the format with details relevant to transfer of most time series data follows.

**ZRXP Format**

Lines starting ## are comment lines.

Lines beginning with # are header information that describes the data that follows in the file.

The header is composed of key words and values separated by a field delimiter. The field delimiter is |\*|. For example the station number in the header would be described as #SANR3033373|\*|, (SANR is the key word for station number).

The key words relevant to transfer of data are;

|  |  |  |  |
| --- | --- | --- | --- |
| **Key Word** | **Meaning** | **Values** | **Description of Value** |
| CUNIT | unit | m mAOD m³/s m³/d  | metres (relative)metres above ordinance datumcubic metres per secondcubic metres per day |
| SNAME | station name | alphanumeric string |  |
| SANR | station number | Alphanumeric string |  |
| RTIMELVL | Time level of values | daily high-resolution monthly | daily valueshigh resolution values – eg 15 minute valuesMonthly values |
| CNAME | Parameter name | QS | FlowLevel |
| RTYPE | value type | amountsInstantaneous valuesmaxima | Amount – eg daily totalInstantaneous value – eg 15 minute stage readingMaximum values eg daily total |
| RINVAL | Value for invalid data | -777 | Flag for missing or invalid data |

The lines of data are plain text, space delimited.

The first group of numbers is the date and time in reverse order with no delimiters

The DateTime is a fixed width field in the form YYYYMMDDhhmmss. For example a reading from 09:00 on 15th June 2008 would be 20080615090000.

Characters 1 to 4 are the year.

Characters 5 and 6 are the month number.

Characters 7 and 8 are the day number.

Characters 9 and 10 are the hour.

Characters 11 and 12 are the minute.

Characters 13 and 14 are the second.

Position 15 is a space.

The positions between the first space and the second space hold the data values.

The last value on the line is the quality flag. For recent data this will normally be zero (unchecked).

An example of the first few lines of a file.

## Exported ZRXP Block for Boleside.SG.ir.O

##TSNAMEBoleside.SG.ir.O;\*;

#ZRXPVERSION2206.235;\*;ZRXPCREATORZEXP3.9.5;\*;

#ZRXPMODEextended;\*;

#CUNITm;\*;

#SNAMEBoleside;\*;SANR15008;\*;SWATERTweed;\*;CNR3033705;\*;

#CNAMESG;\*;CMW86400;\*;CTYPEn-min-ip;\*;

#RTYPEMomentanwerte;\*;RTIMELVLHochaufloesend;\*;RORPRoriginal;\*;CNTYPE1;\*;

##DAYSTART0000

#REXCHANGETS\_21006\_1\_10000\_5;\*;

#RSTATEW6;\*;

20110513000000 0.616 0

20110513001500 0.616 0

20110513003000 0.615 0

20110513004500 0.615 0