

# Package ‘winfapReader’

May 8, 2026

**Type** Package

**Title** Interact with Peak Flow Data in the United Kingdom

**Version** 0.1-7

**Maintainer** Ilaria Prosdocimi <prosdocimi.ilaria@gmail.com>

**URL** <https://ilapros.github.io/winfapReader/>

**BugReports** <https://github.com/ilapros/winfapReader/issues>

**Description** Obtain information on peak flow data from the National River Flow Archive (NRFA) in the United Kingdom, either from the Peak Flow Dataset files <<https://nrfa.ceh.ac.uk/data/peak-flow-dataset>> once these have been downloaded to the user's computer or using the NRFA's API. These files are in a format suitable for direct use in the 'WINFAP' software, hence the name of the package.

**License** GPL-3

**Imports** lubridate

**Depends** utils, R (>= 4.1)

**Suggests** testthat, utf8, rnrfa, httr2, curl (>= 6.4.0), knitr, rmarkdown, zoo

**LazyData** true

**RoxygenNote** 7.3.3

**VignetteBuilder** knitr

**Encoding** UTF-8

**Language** en-GB

**NeedsCompilation** no

**Author** Ilaria Prosdocimi [aut, cre] (ORCID: <<https://orcid.org/0000-0001-8565-094X>>),  
Luke Shaw [aut] (Luke developed the code to handle the missing and gap periods for Peaks over threshold records.)

**Repository** CRAN

**Date/Publication** 2026-02-20 11:30:02 UTC

## Contents

|                      |   |
|----------------------|---|
| get_amax . . . . .   | 2 |
| get_cd . . . . .     | 3 |
| get_pot . . . . .    | 4 |
| known_Oct1 . . . . . | 5 |
| read_amax . . . . .  | 6 |
| read_cd3 . . . . .   | 7 |
| read_pot . . . . .   | 7 |
| water_year . . . . . | 8 |

|              |           |
|--------------|-----------|
| <b>Index</b> | <b>10</b> |
|--------------|-----------|

---

|          |  |
|----------|--|
| get_amax | <i>A function to obtain annual maxima (AMAX) data using the NRFA API</i> |
|----------|--|

---

### Description

The function queries the NRFA API for the .AM file similar to the WINFAP file for a given stations. It then processes the file in a fashion similar to read\_amax.

### Usage

```
get_amax(station)
```

### Arguments

|         |  |
|---------|--|
| station | the NRFA station number for which the annual maxima records should be obtained. Can also be a vector of station numbers. |
|---------|--|

### Value

a data.frame with information on the annual maxima for the station with the following columns

**Station** NRFA station number (can be a vector of station numbers)

**WaterYear** the correct water year for the peak flow

**Date** date of maximum flow

**Flow** the maximum flow in m<sup>3</sup>/s

**Stage** the stage (height) reached by the river - this information is used to derive the flow via a rating curve

**Rejected** logical, if TRUE the water year has been flagged as rejected by the NRFA

### See Also

[read\\_amax](#). Information on river flow gauging in the UK and the annual maxima can be found at the National River Flow Archive website <https://nrfa.ceh.ac.uk>

**Examples**

```
a40003 <- get_amax(40003) # the Medway at Teston / East Farleigh
multipleStations <- get_amax(c(40003, 42003))
names(multipleStations)
summary(multipleStations$`42003`)
```

get\_cd

*A function to obtain information on the station and on the catchment upstream of the station using the NRFA API*

**Description**

The function queries the NRFA API for information of a given station. Unlike [get\\_amax](#) and [get\\_pot](#), the output of this function is not exactly the same from the output of the [read\\_cd3](#) function due to differences in the information made available by the NRFA API

**Usage**

```
get_cd(station, fields = "feh")
```

**Arguments**

|         |   |
|---------|---|
| station | the NRFA station(s) number for which the the information is required  |
| fields  | the type of information which is required. Can be "feh" (default), which outputs a subset of information typically used when applying the flood estimation handbook methods, or "all", which output all information made available in the NRFA API. |

**Value**

a data.frame of one row with different columns depending on whether fields = "all" or fields = "feh" was selected.

**See Also**

[read\\_cd3](#). Information on catchment descriptors river flow gauging in the UK can be found at the National River Flow Archive website <https://nrfa.ceh.ac.uk>

**Examples**

```
cdMult <- get_cd(c(40003,42003), fields = "all")
### lots of information on the catchment/station
### including information on rejected annual maxima
cdMult$`40003`$`peak-flow-rejected-amax-years` ## no rejections
cdMult$`42003`$`peak-flow-rejected-amax-years` ## several rejections
cd40003 <- get_cd(40003, fields = "feh")
# less information, mostly the FEH descriptors
dim(cd40003)
```

```
sapply(cdMult, ncol)
```

---

|         |  |
|---------|--|
| get_pot | <i>A function to obtain Peaks-Over-Threshold (POT) data using the NRFA API</i> |
|---------|--|

---

### Description

The function queries the NRFA API for the .PT file similar to the WINFAP file for a given stations. It then processes the file in a fashion similar to [read\\_pot](#).

### Usage

```
get_pot(station, getAmax = FALSE)
```

### Arguments

|         |   |
|---------|---|
| station | the NRFA station number for which peaks over threshold information should be obtained. It can also be a vector of station numbers |
| getAmax | logical. If TRUE information on the annual maxima values will be retrieved and attached to the WaterYearInfo table                |

### Value

Like [read\\_pot](#) a list of three objects tablePOT, WaterYearInfo and dateRange.

tablePOT contains a table with all the peaks above the threshold present in the record

WaterYearInfo a table containing the information on the percentage of missing values in any water year for which some data is available in the POT record. This is useful to assess whether the lack of exceedances is genuine or the result of missing data and to assess whether the threshold exceedances present in tablePOT can be deemed to be representative of the whole year

dateRange a vector with the first and last date of recording for the POT record as provided in the [POT Details] field. Note that this period might be different than the period for which annual maxima records are available

### See Also

[read\\_pot](#). Information on the peaks over threshold records and river flow gauging in the UK can be found at the National River Flow Archive website <https://nrfa.ceh.ac.uk>

## Examples

```
## Not run:
### the example take longer than 5 seconds to run
p40003 <- get_pot(40003) # the Medway at Teston / East Farleigh
p40003$tablePOT[p40003$tablePOT$WaterYear > 1969 &
  p40003$tablePOT$WaterYear < 1977,]
### no events in 1971 nor 1975
p40003$WaterYearInfo[p40003$WaterYearInfo$WaterYear > 1969 &
  p40003$WaterYearInfo$WaterYear < 1977,]
# in 1971 all records are valid,
# in 1975 no exceedances
# might be due to the fact that almost no valid record are available

p40003 <- get_pot(40003, getAmax = TRUE)
p40003$WaterYearInfo[p40003$WaterYearInfo$WaterYear > 1969 &
  p40003$WaterYearInfo$WaterYear < 1977,]
# the annual maximum in 1971 and 1975 was below the threshold
# no events exceeded the threshold

## End(Not run)
```

---

known\_Oct1

*Known events which happened on October 1st before 9am*

---

## Description

The Water Year in the UK runs from 9am of the 1st October of a given year to 8:59am of the 1st October of the next year. Since the WINFAP files contain information only on the date of the annual maximum (and not time) it is possible that an event is mis-classified when using the `water_year` function. This dataset lists the events which are known to have happened to October 1st before 9am. This is used to correct the `WaterYear` information in these known cases in the `read_amax` and `get_amax` functions. For some stations events on October 1st have been deemed as annual maxima only in some winfap releases. They are maintained in the dataset in the event that somebody read old winfap files.

## Usage

```
known_Oct1
```

## Format

A data frame with 36 rows and 3 variables:

**Station** NRFA station number

**Date** date of maximum flow (always the 1st October)

**WaterYear** the correct water year for the peak flow

**Source**

Derived manually by identifying events which happened on Oct. 1st and comparing it with information on <https://nrfa.ceh.ac.uk>

---

|           |                                     |
|-----------|-------------------------------------|
| read_amax | <i>A function to read .AM files</i> |
|-----------|-------------------------------------|

---

**Description**

The function reads .AM files once these are in a local folder: these files contain information on annual maxima (AMAX) records extracted from the instantaneous river flow measurements. The function checks for the presence of any [AM Rejected] information and includes it in the output.

**Usage**

```
read_amax(station, loc_WinFapFiles = getwd())
```

**Arguments**

|                 |  |
|-----------------|--|
| station         | NRFA station number(s) for which the .AM file (named station.AM) should be read.   |
| loc_WinFapFiles | the file.path of the WINFAP files, i.e. the location in which the station.AM file can be found. Default is the working directory |

**Value**

a data.frame with information on the annual maxima for the station with the following columns

**Station** NRFA station number (can be a vector of station numbers)

**WaterYear** the correct water year for the peak flow

**Date** date of maximum flow

**Flow** the maximum flow in m3/s

**Stage** the stage (height) reached by the river - this information is used to derive the flow via a rating curve

**Rejected** logical, if TRUE the water year has been flagged as rejected by the NRFA

**See Also**

Information on the .AM files and river flow gauging in the UK can be found at the National River Flow Archive website <https://nrfa.ceh.ac.uk>

---

read\_cd3                      *A function to read .CD3 files*

---

### Description

The function reads .CD3 files once these are in a local folder: these files contain information on the gauging station and on the catchment upstream the station.

### Usage

```
read_cd3(station, loc_WinFapFiles = getwd())
```

### Arguments

station                      the NRFA station number(s) for which the .CD3 file (names station.CD3) should be read

loc\_WinFapFiles              the file.path of the WINFAP files, i.e. the location in which the station.CD3 file can be found. Default is the working directory

### Value

a data.frame with information on the catchment descriptors for the station

### See Also

Information on the .CD3 files and river flow gauging in the UK can be found at the National River Flow Archive website <https://nrfa.ceh.ac.uk>. Specific information on the catchment descriptors can be found at <https://nrfa.ceh.ac.uk/data/about-data/catchment-information/feh-catchment-descriptors>

---

read\_pot                      *A function to read .PT files*

---

### Description

The function reads .PT files once these are in a local folder: these files contain information on Peaks-Over-Threshold (POT) records from the instantaneous river flow measurements. The function checks for the presence of any [POT GAPS] and [POT REJECTED] periods. If these are present, they are merged and information on the proportion of days with missing records in each water year is provided.

### Usage

```
read_pot(station, loc_WinFapFiles = getwd(), getAmax = FALSE)
```

**Arguments**

|                 |  |
|-----------------|--|
| station         | NRFA station number(s) for which the .PT file (names station.PT) should be read.   |
| loc_WinFapFiles | the file.path of the WINFAP files, i.e. the location in which the station.PT file can be found. Default is the working directory |
| getAmax         | logical. If TRUE the annual maxima values (extracted from a station.AM file) will be attached to the WaterYearInfo table         |

**Value**

a list of three objects tablePOT, WaterYearInfo and dateRange.

tablePOT contains a table with all the peaks above the threshold present in the .PT file

WaterYearInfo a table containing the information on the percentage of missing values in any water year for which some data is available in the POT record. This is useful to assess whether the lack of exceedances is genuine or the result of missing data and to assess whether the threshold exceedances present in tablePOT can be deemed to be representative of the whole year

dateRange a vector with the first and last date of recording for the POT record as provided in the [POT Details] field. Note that this period might be different than the period for which annual maxima records are available

**See Also**

Information on the .PT files and river flow gauging in the UK can be found at the National River Flow Archive website <https://nrfa.ceh.ac.uk>

---

|            |   |
|------------|---|
| water_year | <i>Derive water year value for a date</i> |
|------------|---|

---

**Description**

Derive water year value for a date

**Usage**

```
water_year(date, start_month = 10, start_hour = 9, tz = "UTC")
```

**Arguments**

|             |   |
|-------------|---|
| date        | the (vector of) dates for which the water year will be calculated |
| start_month | the month in which the water year starts, default is October      |
| start_hour  | the time of the day in which the water year starts, default is 9  |
| tz          | time zone for the record, defaults to "UTC"                       |

## **Value**

The water year value

## **Examples**

```
## Not run:
water_year(c(as.POSIXct("1958-10-01 09:00:00", tz = "UTC"),
             as.POSIXct("2021-10-01 12:00:00", tz = "UTC"),
             as.POSIXct("1918-10-01 08:55:00", tz = "UTC")))

water_year(c(as.Date("2021-01-01"), as.Date("2021-11-01"),
             as.Date("2019-02-01"), as.Date("1965-09-30")))

## End(Not run)
```

# Index

## \* datasets

known\_Oct1, 5

get\_amax, 2, 3

get\_cd, 3

get\_pot, 3, 4

known\_Oct1, 5

read\_amax, 2, 6

read\_cd3, 3, 7

read\_pot, 4, 7

water\_year, 8