Worked example 5.5: Threshold level method

Table of Contents

The threshold level method can be used to select drought events from time series of river flow as long as there are not too many missing values in the dataset and a meaningful threshold is chosen. Data from River Ngaruroro at Kuripapango (NZ) are used to demonstrate the procedure in the example below.

# Loading the Data

56 years of daily flow (20 September 1963 to 8 October 2019) are analysed. In this river the low flow period covers the turn of the calendar year. To avoid problems with allocating droughts to a specific calendar year because of drought events starting in one year and ending in another year, the start of the year is set to 1 September. An event is attributed to the year it starts.

library(tidyverse)  
library(hydroDrought)  
  
ngaruroro <- international %>%  
 filter(river == "Ngaruroro") %>%  
 select(data) %>%  
 unnest(data)

# Missing values

The time series, Ngaruroro, contains missing values. We do not know if a missing value (NA) represents a flow below the threshold or above the threshold, as the flow value itself is unknown. A single missing value will cause the function drought\_events() to terminate a dry spell (drought event) or similar, a wet spell. Accordingly, most characteristics derived for this event (e.g. drought duration, drought termination, drought volume, etc.) will not be correct.

A conservative approach would be to eliminate years with missing values completely. Instead, to avoid losing too many years of observations, we filled periods of missing data with linear interpolation if they are of short duration. Here short duration is defined as periods < 15 days, whereas years containing long periods of missing values (≥15 days) have been removed. This results in 49 years of daily flow (1 September 1964 to 31 August 2019). In total eight years are omitted from the series (1963/64, 1965/66, 1977/78, 1978/79, 1986/87, 1987/88, 2001/02 and 2019/20).

ngaruroro <- ngaruroro %>%  
 sanitize\_ts(approx.missing = 14) %>%  
 mutate(  
 year = water\_year(time, origin = "-09-01")  
 )  
  
coverage <- ngaruroro %>%  
 filter(!is.na(discharge)) %>%  
 pull(time) %>%  
 coverage\_yearly(origin = "-09-01")  
  
incomplete <- coverage %>%  
 filter(days.missing > 0)   
  
complete <- coverage %>%  
 filter(days.missing == 0)  
  
ngaruroro <- ngaruroro %>%  
 anti\_join(incomplete, by = "year")

The table below displays the year removed, the total number of days in the year (365 or 366 for leap years), the number of days with flow observations, the number of NA-values (days with missing data) and the remaining fraction of days.

print(incomplete)

## # A tibble: 8 x 5  
## year days.in.year days.with.data days.missing coverage  
## <dbl> <int> <int> <int> <dbl>  
## 1 1963 366 347 19 0.948  
## 2 1965 365 294 71 0.805  
## 3 1977 365 350 15 0.959  
## 4 1978 365 305 60 0.836  
## 5 1986 365 341 24 0.934  
## 6 1987 366 336 30 0.918  
## 7 2001 365 344 21 0.942  
## 8 2019 366 38 328 0.104

# Threshold selection and drought events

A sequence of drought events is obtained from the streamflow hydrograph by considering periods with flow below a certain threshold, . In this example m3s-1 is used as threshold. A table of drought characteristics is derived with the function drought\_events().

q90 <- lfquantile(ngaruroro$discharge, exc.freq = 0.9) %>%  
 print()

## Q90   
## 4.949

droughts <- ngaruroro %>%  
 drought\_events(threshold = q90, pooling = "none")

Table 5.8 Drought deficit characteristics, River Ngaruroro at Kuripapango, NZ.

## # A tibble: 210 x 7  
## event first.day last.day duration volume qmin tqmin   
## <int> <date> <date> <drtn> <dbl> <dbl> <date>   
## 1 1 1967-04-23 1967-04-23 1 days 6307. 4.88 1967-04-23  
## 2 2 1967-04-26 1967-04-26 1 days 4579. 4.90 1967-04-26  
## 3 3 1967-05-09 1967-05-10 2 days 17453. 4.80 1967-05-10  
## 4 4 1967-05-13 1967-05-14 2 days 18835. 4.76 1967-05-14  
## 5 5 1967-05-23 1967-05-23 1 days 3024. 4.91 1967-05-23  
## 6 6 1968-02-07 1968-02-08 2 days 34646. 4.69 1968-02-08  
## 7 7 1968-02-17 1968-03-08 21 days 1766621. 3.44 1968-03-05  
## 8 8 1968-03-11 1968-04-02 23 days 2349562. 3.23 1968-03-26  
## 9 9 1968-04-06 1968-04-09 4 days 293933. 3.76 1968-04-08  
## 10 10 1969-03-25 1969-03-30 6 days 103766. 4.68 1969-03-26  
## # … with 200 more rows

The table displayed above includes:

* first.day: the start date, defined as the first day below the threshold;
* last.day: the end date, defined as the last day below the threshold;
* duration: the drought duration (days), defined as last.day - first.day + 1
* volume: the deficit volume in m3, defined as the sum of the daily deficit flows times the duration in days;
* qmin: the minimum flow in m3s-1, defined as the minimum flow within a drought event;
* tqmin: the date of the minimum flow.

# Removing minor droughts (Filtering)

Several minor droughts, lasting for a few days only, can be observed. To reduce the problem of minor droughts two restrictions are imposed:

* a minimum drought duration, which removes droughts with duration less than a specified number of days;
* a minimum drought deficit volume (coefficient ), which removes droughts with a deficit volume less than a certain fraction of the maximum drought deficit volume observed in the complete series of drought events.

We will append a logical column called is.minor to the table of drought events. It is TRUE when drought duration is less than five days OR if the drought volume is less than 5% of the maximum drought deficit volume (i.e., 51 133.25 m3). In total 99 droughts are considered minor, and thus removed, based on these criteria.

droughts <- droughts %>%  
 mutate(is.minor = duration < 5 | volume < max(volume) \* 0.005)

print(droughts)

## # A tibble: 210 x 8  
## event first.day last.day duration volume qmin tqmin is.minor  
## <int> <date> <date> <drtn> <dbl> <dbl> <date> <lgl>   
## 1 1 1967-04-23 1967-04-23 1 days 6307. 4.88 1967-04-23 TRUE   
## 2 2 1967-04-26 1967-04-26 1 days 4579. 4.90 1967-04-26 TRUE   
## 3 3 1967-05-09 1967-05-10 2 days 17453. 4.80 1967-05-10 TRUE   
## 4 4 1967-05-13 1967-05-14 2 days 18835. 4.76 1967-05-14 TRUE   
## 5 5 1967-05-23 1967-05-23 1 days 3024. 4.91 1967-05-23 TRUE   
## 6 6 1968-02-07 1968-02-08 2 days 34646. 4.69 1968-02-08 TRUE   
## 7 7 1968-02-17 1968-03-08 21 days 1766621. 3.44 1968-03-05 FALSE   
## 8 8 1968-03-11 1968-04-02 23 days 2349562. 3.23 1968-03-26 FALSE   
## 9 9 1968-04-06 1968-04-09 4 days 293933. 3.76 1968-04-08 TRUE   
## 10 10 1969-03-25 1969-03-30 6 days 103766. 4.68 1969-03-26 FALSE   
## # … with 200 more rows

# Eliminating dependent droughts (Pooling)

The inter-event time criterion (IC) is used to pool dependent droughts, which are droughts separated by a short period of flow above the threshold. If the time between two droughts is less than a critical duration, , the two events are pooled.

In this example is set equal to two days.

pooled <- ngaruroro %>%  
 drought\_events(  
 threshold = q90, pooling = "inter-event",   
 pooling.pars = list(min.duration = 2, min.vol.ratio = Inf)  
 ) %>%  
 filter(duration >= 5, volume > max(volume) \* 0.005) %>%  
 arrange(desc(duration)) %>%  
 print()

## # A tibble: 100 x 9  
## event first.day last.day duration dbt volume qmin tqmin pooled  
## <int> <date> <date> <drtn> <drtn> <dbl> <dbl> <date> <dbl>  
## 1 166 2015-01-06 2015-03-15 69 days 69 days 10226650. 2.17 2015-03-04 0  
## 2 37 1974-01-20 1974-03-17 57 days 56 days 6506957. 2.88 1974-03-15 1  
## 3 138 2008-01-10 2008-03-01 52 days 51 days 6412608 2.64 2008-02-28 1  
## 4 30 1973-01-26 1973-03-12 46 days 46 days 6583939. 2.66 1973-03-03 0  
## 5 50 1983-02-16 1983-04-02 46 days 46 days 7381066. 2.46 1983-03-30 0  
## 6 148 2009-03-14 2009-04-26 44 days 43 days 6008774. 2.53 2009-04-19 1  
## 7 157 2013-02-07 2013-03-18 40 days 40 days 7089898. 2.38 2013-03-15 0  
## 8 124 2005-02-06 2005-03-16 39 days 38 days 4987094. 2.68 2005-03-13 1  
## 9 139 2008-03-09 2008-04-14 37 days 36 days 5215795. 2.65 2008-04-06 1  
## 10 59 1989-04-01 1989-04-29 29 days 29 days 2706480 3.3 1989-04-29 0  
## # … with 90 more rows

When drought events are pooled the table of drought events contain two more columns:

* dbt: the duration below the threshold, i.e. the drought duration minus short period(s) above the threshold (note: the ‘full’ duration can be derived from the start and end date of each event);
* pooled: the number of drought events.

The drought deficit characteristics of the ten longest (pooled) drought events are given in the table above. In total, there are 100 drought events, which equal an average of 2.04 events per year.

Key drought characteristics for all drought events occurring in the period (09.1963- 08.2020), can be summarized for different drought metrics. In the example below, for each year, the number of droughts in the year, the days below the threshold (summed over all events) in a year and the minimum flow in a year, are presented:

pooled %>%  
 mutate(  
 year = water\_year(first.day, origin = "-09-01")  
 ) %>%  
 group\_by(year) %>%  
 summarise(  
 n.droughts = n(),  
 real.duration = sum(dbt),   
 min.flow = min(qmin)  
 )

## # A tibble: 35 x 4  
## year n.droughts real.duration min.flow  
## \* <dbl> <int> <drtn> <dbl>  
## 1 1967 2 44 days 3.23  
## 2 1968 2 22 days 3.88  
## 3 1969 3 37 days 3.89  
## 4 1970 3 34 days 3.73  
## 5 1972 7 114 days 2.66  
## 6 1973 2 63 days 2.88  
## 7 1974 2 10 days 4.31  
## 8 1975 2 18 days 4.09  
## 9 1977 1 25 days 3.46  
## 10 1982 2 57 days 2.46  
## # … with 25 more rows

Time series of the drought duration are plotted in Figure 5.12. The longest drought durations (dbt) are found in 1972, 1973, 1982, 2007, 2008 and 2014.

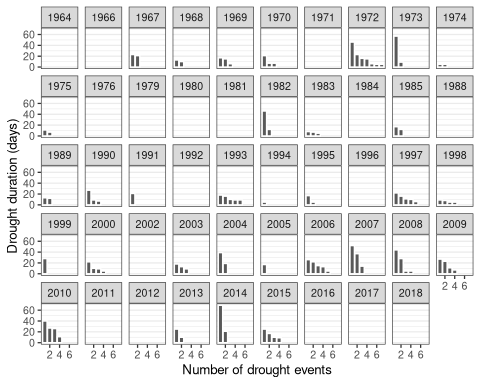


Figure 5.12 Time series of drought duration for River Ngaruroro at Kuripapango (NZ). Selection criteria: threshold level = , days, and days.

A histogram of the drought duration is seen in Figure 5.13, and a very skewed distribution is revealed. Short duration droughts are dominating with 43 events lasting less than 11 days. Only nine events lasted more than 30 days.

p %>%  
 # replace\_na(list(duration = 0)) %>%  
 ggplot(aes(duration)) +   
 geom\_histogram(binwidth = 5, boundary = 0, closed = "left",   
 size = 0.2, col = "black", fill = "grey90") +   
 scale\_x\_continuous(limits = c(0, NA)) +   
 scale\_y\_continuous(breaks = breaks\_integer()) +   
 labs(x = "Drought duration (days)", y = "Counts")

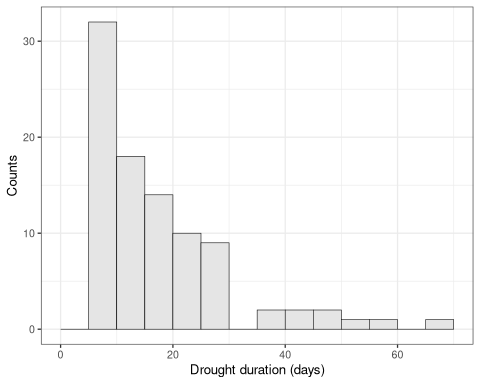


Figure 5.13 Histogram of drought duration for River Ngaruroro at Kuripapango (NZ). Selection criteria: threshold level = , days, and days.