The R-Package `smires`

Calculating Hydrological Metrics for Univariate Time Series

Tobias Gauster

Institute of Applied Statistics and Computing
BOKU, Vienna

Lyon-Villeurbane, June 2016

Outline

The project is hosted on github:
`https://github.com/mundl/smires`

Install the current development version (0.4.0) of smires from github. You have to install the package every time you want to update to the most recent version.

```
library(devtools)
install_github("mundl/smires")
```

Load the package to use it. You have to load it every time you restart R.

```
library(smires)
```

Currently there are only two data sets included:

- Balder at Balderhead Reservoir: `balder`
- Ampney Brook at Ampney St Peter: `ampneyBrook`

You can print a dataset by typing its name. An object of class `tibble` is used for data sets.

The first column is the index of the time series, it is named `time` and is of class `Date`. The second column named `discharge` contains the observed streamflow discharges. Throughout the analysis, further variables (columns) can be appended.

```
> ampneyBrook
# A tibble: 11,841 × 2
         time discharge
       <date>     <dbl>
1  1983-05-01      1.02
2  1983-05-02      1.18
3  1983-05-03      1.62
4  1983-05-04      1.99
5  1983-05-05      2.10
6  1983-05-06      2.02
7  1983-05-07      1.92
8  1983-05-08      1.79
9  1983-05-09      1.66
10 1983-05-10      1.55
# ... with 11,831 more rows
```

Determing Intermittency

Checking for intermittency according to the SMIRES definition:

```
> is.intermittent(x = ampneyBrook, ndays = 5,
                  consecutive = TRUE, threshold = 0.001)
[1] TRUE

> is.intermittent(ampneyBrook)
[1] TRUE
```

Preprocessing

```
> discharge <- check_ts(balder)
The time series contains 654 observations numerically
equal to zero (28.3 %).

The time series contains 245 missing observations (10.6 %)

Time series covers only 6.3 years.
A minimum length of 10 years is advised.
```

There are additional arguments to `check_ts`:

```
check_ts(x, minyear = 10, approx.missing = 5, accuracy = 0)
```

Metrics

Most hydrological metrics are constructed in a similar way. The general approach comprises four steps:

1. First the time series can be *preprocessed*, e.g. by interpolating missing values or by applying a moving average.
2. If necessary, an optional step involves the identification of *distinct events* such as low flow periods. For each event a set of new variables (e.g. event duration or event onset) is derived.
3. In a third step *summary statistics* are calculated for arbitrary periods (e.g. months, seasons, calendar years, hydrological years, decades).
4. Repeated step 3 until the original time series is aggregated to a single value.

Metrics: Mean Annual Maximum Duration of Dry Spells

1. Preprocessing
2. Event detection: find dry spells
3. only keep the max. duration of a period
   period = `'year'`
   aggregation function: `max()`
4. Aggregate to a single value
   aggregation function: `mean()`

```
> metric(balder, period = "year",
         agg1 = "max", agg2 = "mean")
```

Metrics: Mean Number of Annual Dry Days

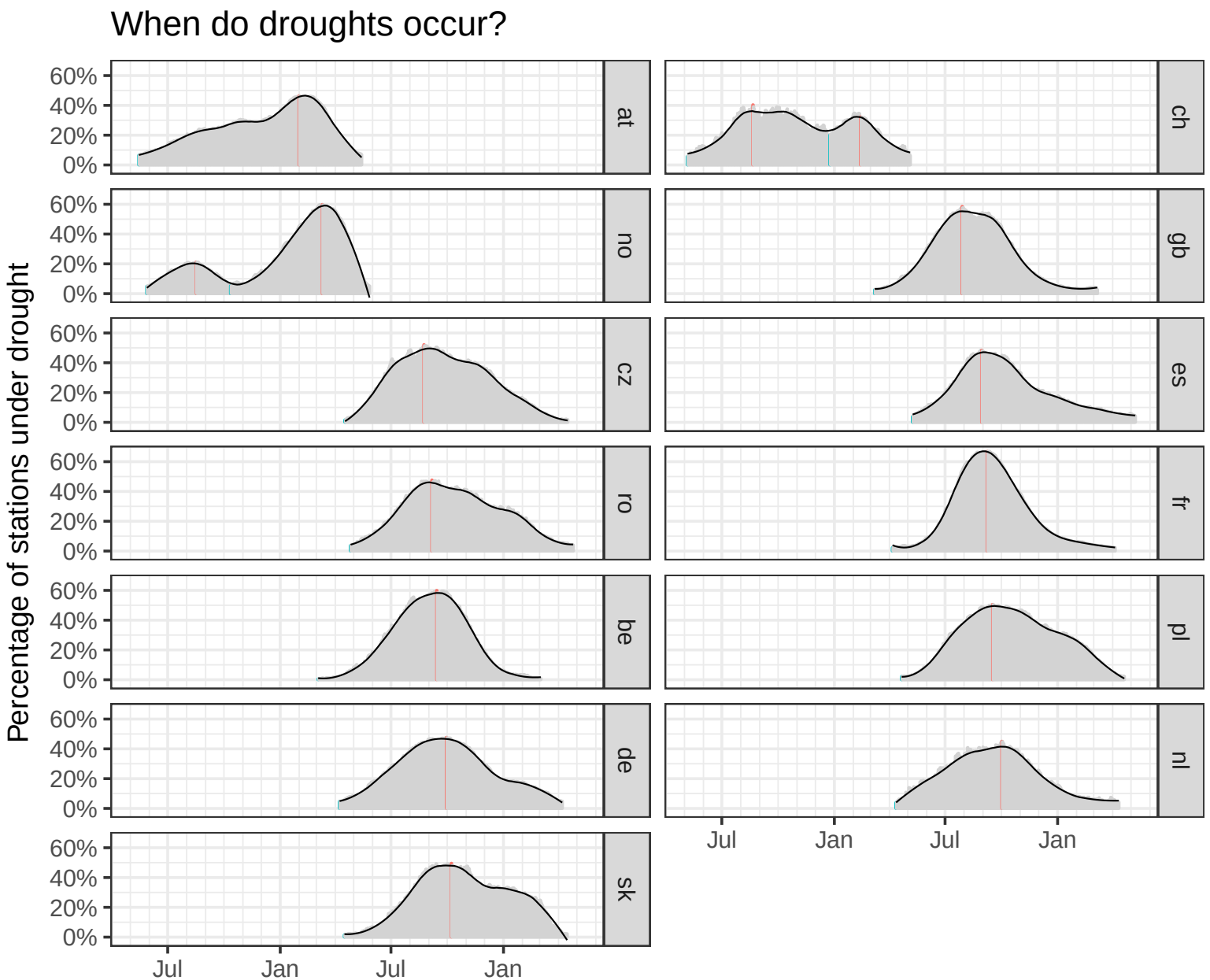
1. Preprocessing
2. Event detection: find dry spells
3. sum up all durations of a period
   period = `'year'`
   aggregation function: `sum()`
4. Aggregate to a single value
   aggregation function: `mean()`

```
> metric(balder, period = "year",
         agg1 = "sum", agg2 = "mean")
```

Convenience Functions

```
mean_annual_max_duration_dry <- function(x)
{
    y <- metric(x, period = "year",
                agg1 = "max", agg2 = "mean")
    y$duration[y$state == "no-flow"]
}

mean_annual_number_dry_days <- function(x)
{
    y <- metric(x, period = "year",
                agg1 = "sum", agg2 = "mean")
    y$duration[y$state == "no-flow"]
}
```


When do droughts occur?
Percentage of stations under drought
60%
40%
20%
0%
at
ch
no
gb
cz
es
ro
fr
be
pl
de
nl
sk
Jul
Jan
Jul
Jan