Short Introduction to RcmdrPlugin.lfstat:

An R Commander plug-in for package lfstat to calculate low flow statistics for daily stream flow data[1]

Gregor Laaha & Daniel Koffler

Institute of Applied Statistics and Computing (IASC),
University of Natural Resources and Life Sciences, BOKU Vienna, Austria
(daniel.koffler@boku.ac.at).

Summary:

The calculation of characteristic stream flow during dry conditions is a basic requirement for many problems in hydrology, ecohydrology and water resources management. As opposed to floods, a number of different indices are used to characterise low flows and streamflow droughts. Although these indices and methods of calculation have been well documented in the WMO Manual on Low-flow Estimation and Prediction [1], comprehensive software was missing which enables a fast and standardized calculation of low flow statistics. We give here a short introduction to our software packages *lfstat* and *RcmdrPlugin.lfstat* which were compiled to fill in this obvious gap. Our software package is based on the statistical open source software R, and expands it to analyse daily stream flow data records focusing on low flows. As command-line based programs are not everyone's preference, we offer a plug-in for the R-Commander, an easy to use graphical user interface (GUI) provided for R which is based on the tcl/tk package.

Reference:
[1] Gustard, A. & Demuth, S. (2009) (Eds) Manual on Low-flow Estimation and Prediction. Operational Hydrology Report No. 50, WMO-No. 1029, 136p.

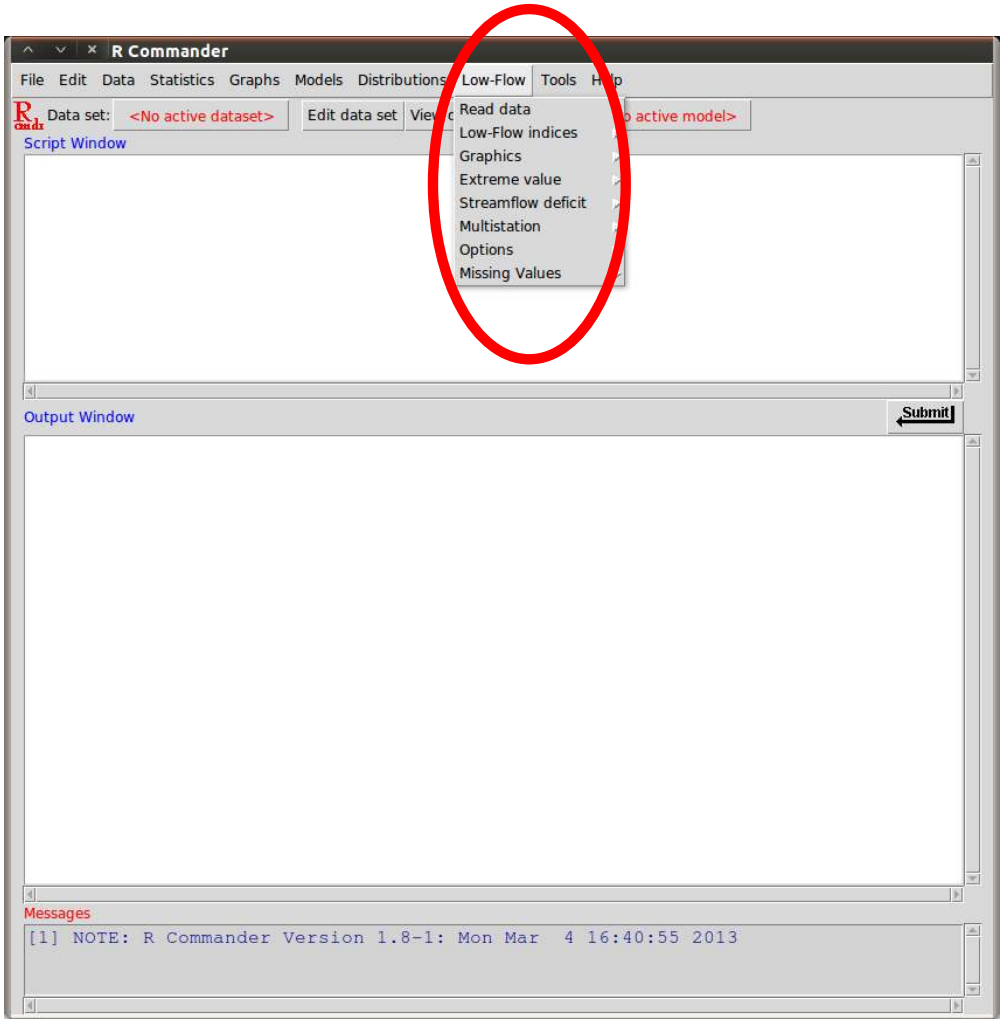
[1] The software packages are funded by the UNESCO and WMO funded project "Low flow and Drought software tool" and are a contribution to UNESCO's FRIEND-Water program. This short introduction is for version 0.2 (February/March 2013).

Introduction

The *RcmdrPlugin.lfstat* accesses the basic functionality of the low flow software package *lfstat* through a graphical user interface (GUI). The basic interface is John Fox's *R Commander*. The plug-in adds a "Low-flow" menu where the functions of *lfstat* are implemented, for easier application through menu control. After running a command the R syntax is displayed in the script window, what is a useful functionality for getting started with the powerful R programming language, and the specific syntax of package *lfstat*.

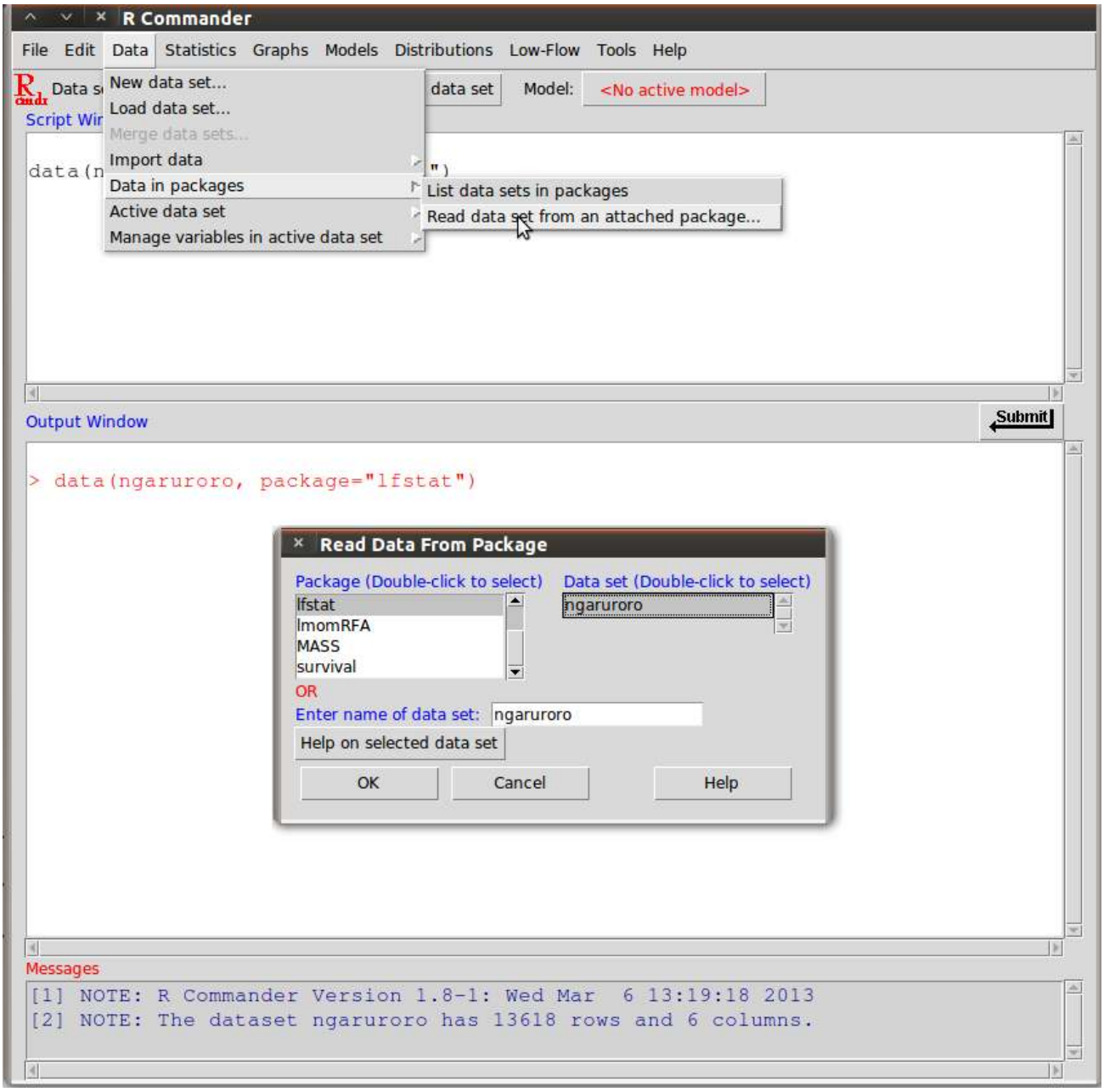
The package *RcmdrPlugin.lfstat* can be installed from the CRAN web repository and started by entering the following commands in the R command line:

```
install.packages("lfstat")
install.packages("RcmdrPlugin.lfstat")
require(lfstat)
require(RcmdrPlugin.lfstat)
```



Example data set

For exploring lfstat and the R Commander plug-in we have added an example dataset of station Ngaruroro to the software package. The data set consists of a daily discharge series from 1963-2000 and the information is stored in the variables (columns) day, month, year, flow. The figure below illustrates how the data set can be loaded. Use button "*View data set*" from R Commander GUI for viewing the data.



Data import

R and the R-Commander offer various methods of reading data in different formats like Excel, SPSS, csv, etc. For most countries, the national hydrological service provides streamflow data in standardised ASCII-files similar to the following Global Runoff Data Centre (GRDC) file format:

```
# Title: GRDC STATION DATA FILE
# Format: DOS-ASCII
# GRDC-No.: 9104020
# River: LABE
# Station: DECIN
# Country: CZ
YYYY-MM-DD;hh:mm; Original; Calculated; Flag
1887-11-01;--:--; 78.000; -999.000; -999
1887-11-02;--:--; 79.000; -999.000; -999
```

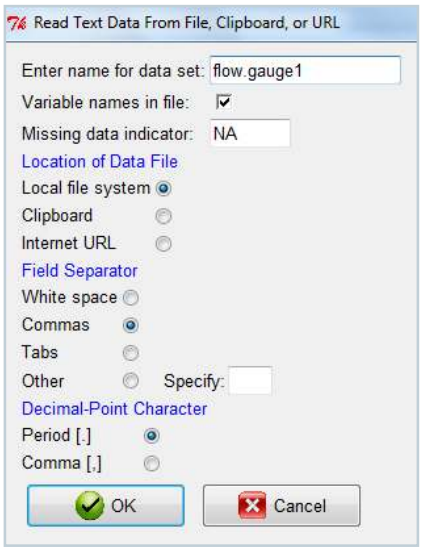
We are keen to provide a direct reading method for as much of these standards as possible. At the moment the following three national formats are supported:

GRDC; HZB (Austria); LfU-Bayern (Germany)

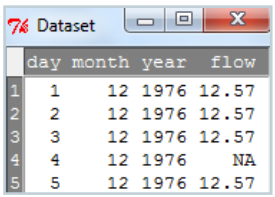
For all other cases, it is recommended to use flow data in csv format for import. The data need to have a matrix structure (common standard: row = day, column=variable) where the daily flow measurements are stored in one column. Further columns (e.g. date of measurement) are optional, but will not be used for data import (step 2). All days without measurement need to be contained in the dataset, properly coded by a missing value code (standard: no value = NA, but this can be customized during the data import). On this basis, the csv data import consists of two steps:

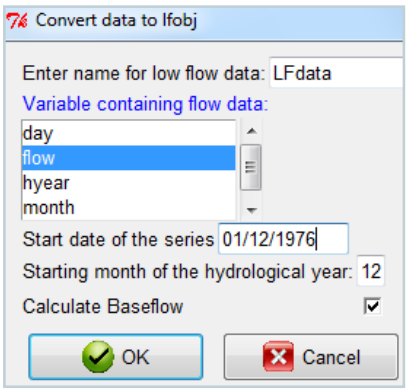
<u>Step 1:</u> Data -> Import data -> from text file

i. Chose a name for the new dataset (no spaces)
ii. Specify the import format (field separator, decimal-point)
    The figure (right) shows the settings for csv import.
iii. OK -> browse and select data file for import.
iv. Message window: are there any errors?
v. Do the number of rows and columns look as expected?
vi. View the data via *View data set* button.



<u>Step 2:</u> Low-Flow -> Read data -> Convert active data set to lfobj





Low flow indices

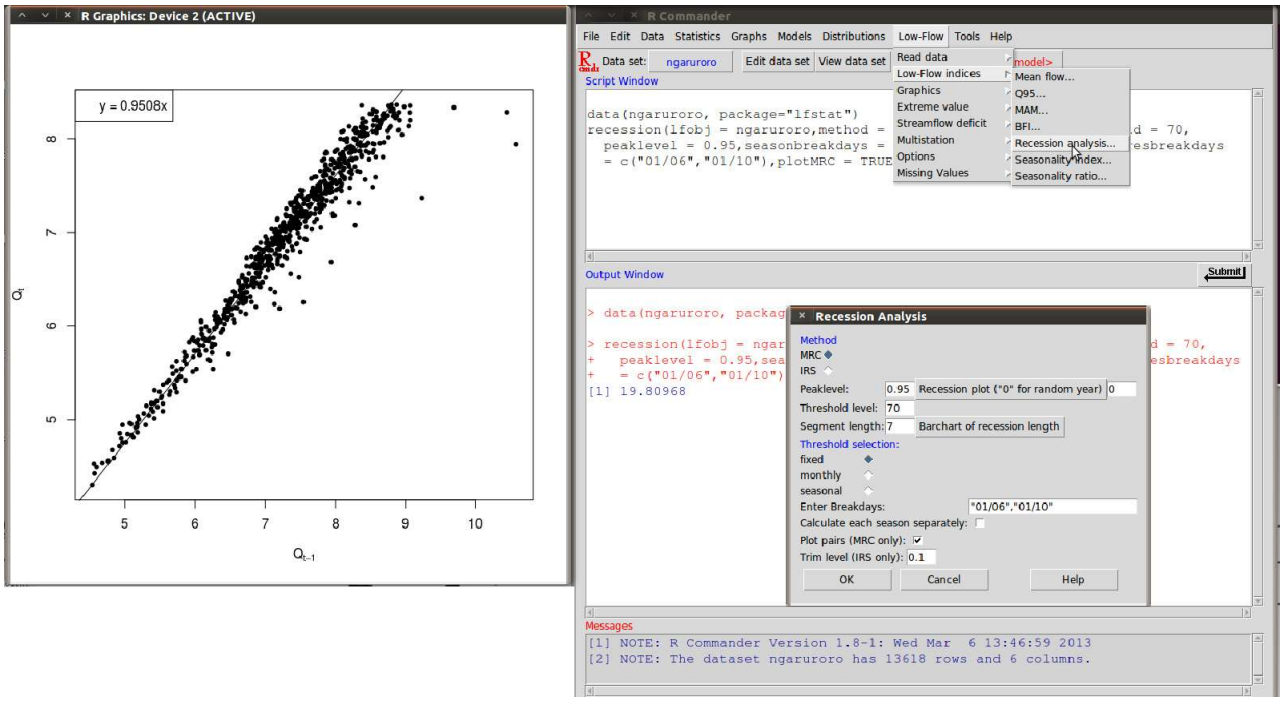
For all analyses the data need to be stored as class *lfobj*. How to crate a data object of class lfobj is explained in Section “Data import”. The lfobj can be explored for missing values using the utility function *Missing values –> Summary…* which provides summary statistics and plots of *run length of missing values*, and *number of missing values per year*. Missing values are standardly excluded from all calculations.

The functionality of lfstat includes estimation methods for low flow indices, extreme value statistics, deficit characteristics, and additional graphical methods to control the computation of complex indices and to illustrate the data.

The menu *Low flow indices* offers calculation routines of the following commonly used flow indices:

- Mean flow
- Q95
- MAM
- BFI
- Recession constants

In addition, the seasonality index and seasonality ratio can be computed[2]. The figure below illustrates the estimation of the base flow recession constant through the menu item *Recession analysis…*



[2] Seasonality methods of package lfstat are fully described in Laaha G. & Blöschl G. (2006) Seasonality indices for regionalizing low flows. Hydrol. Process. 20, 3851-3878.

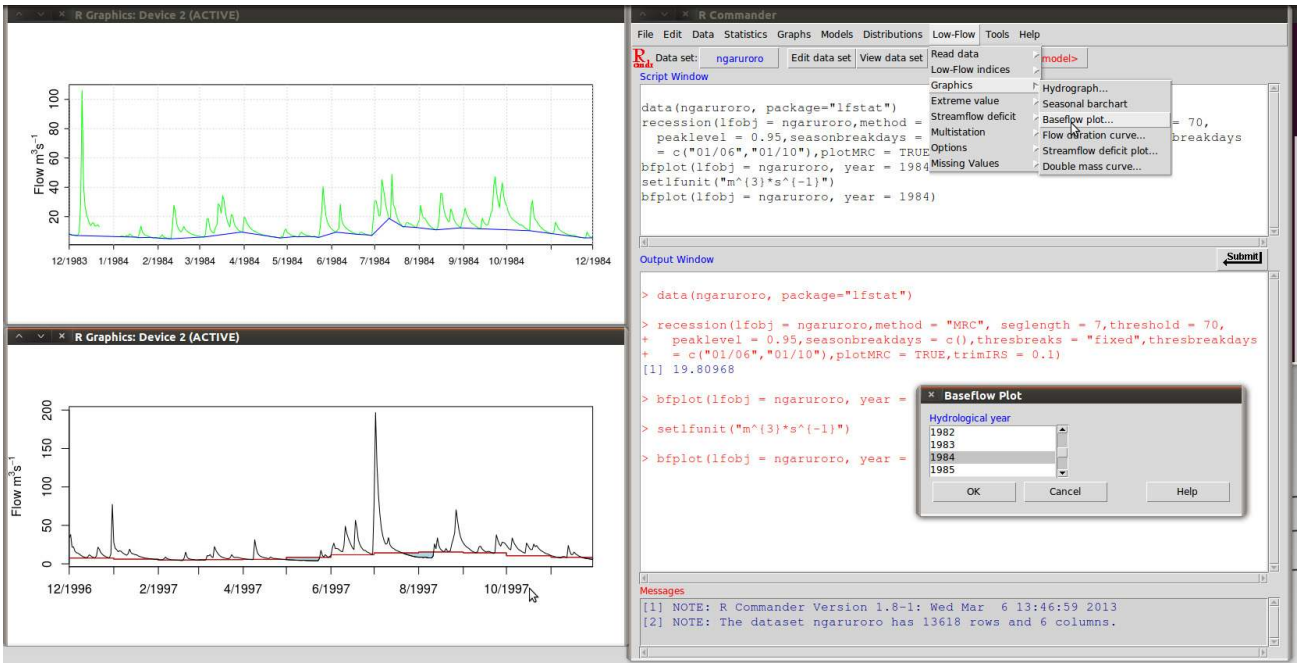
Graphics

The *Graphics* menu offers plotting functions for hydrographs for different periods, flexible streamflow deficit plots, baseflow visualisation, recession diagnostics, flow duration curves as well as double mass curves. These functions can be accessed through the following menu items:

- Hydrograph
- Seasonal barchart
- Baseflow plot
- Flow duration curve
- Streamflow deficit plot
- Double mass curve

All plots can be modified according to the user preferences by changing the function call in the script window. Units can be set using the utility function *Options -> Set unit in plots…*

The figure below illustrates the computation of the *Baseflow plot* for station Ngaruroro (GUI windows and upper graph). The lower graph shows the *Streamflow deficit plot* using monthly Q70 as threshold level. Areas marked in blue indicate the deficit volumes of dry spells, where the daily discharge fell below the threshold level.

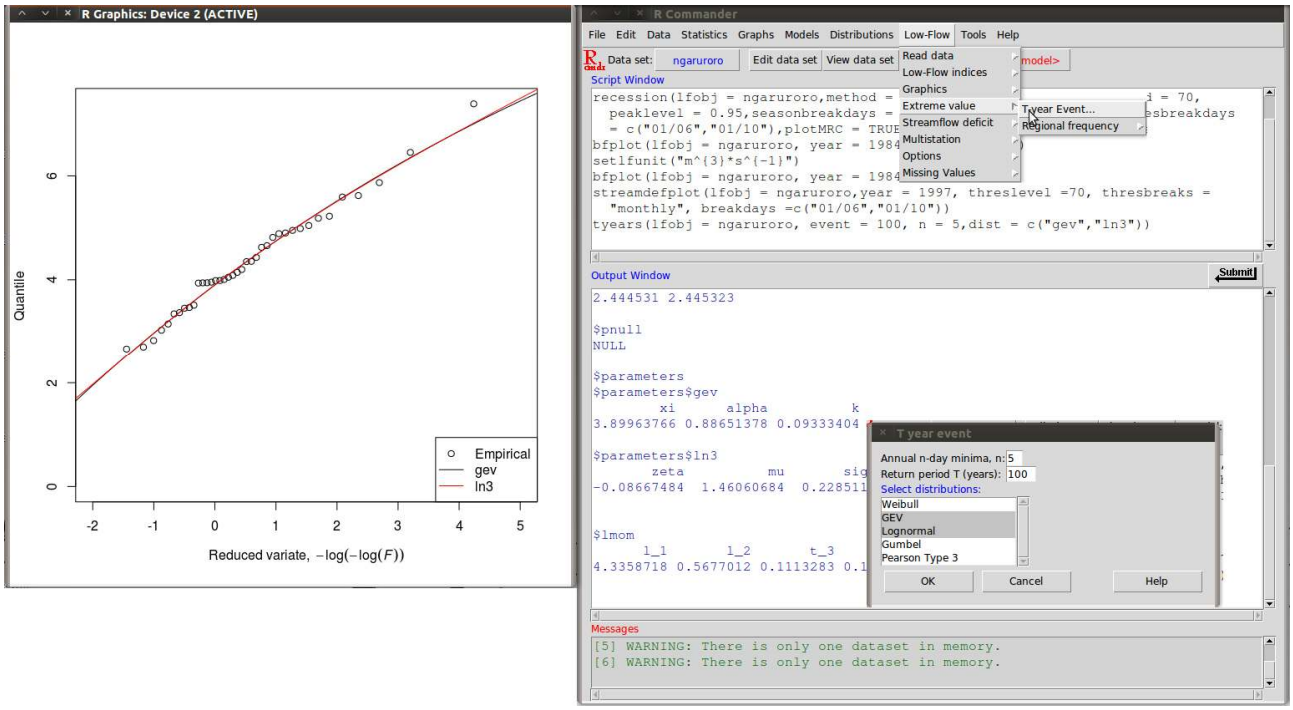


Extreme value statistics

For extreme value statistics, state-of-the-art methods for L-moment based local and regional frequency analysis (RFA) are available. Calculations are based on the annual minima approach of extreme value statistics. The function *Extreme value -> T-year Event…* computes a complete frequency analysis of low flow extreme values: it derives an annual n-day minimum (AMn) series from the daily discharge record, calculates its sample L-moments, fits a distribution model to the AMn series, which is finally applied to estimate an event with a return period of T-year s (i.e. flow quantile with an exceedance probability of p=1-1/T).

The function requires specification of the length of the smoothing window *n* (in *days*), the return period *T* (in *years*), and the distribution model. In cases when the AMn series contains zero flows a censored frequency analysis approach is automatically used for the calculation of the T-years events.

The figure below illustrates the calculation of the 100-year event for station Ngaruroro. The simultaneous selection of various distribution functions enables a straightforward sensitivity analysis of T-year events with respect to the probabilistic model. The function returns a Gumbel-scaled probability plot where empirical quantiles (traditionally called plotting-positions) are plotted as dots, and the quantiles of the distribution model are plotted as lines. The numerical output returned to the *Output window* contains the estimate for the T-year event, the probability of zero values (for censored AMn series), the parameters of the fitted distribution model(s), and the sample L-moments.

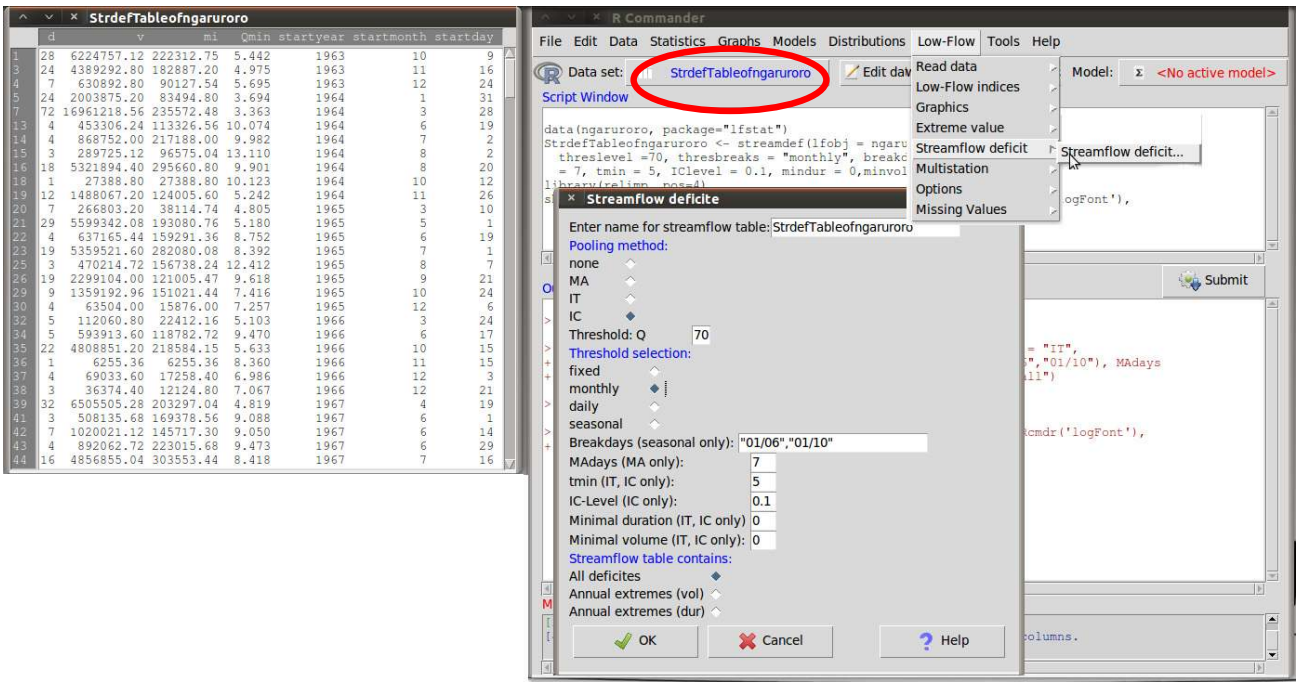


Streamflow deficit

The tools for deficit characteristics include various pooling and threshold selection methods to support the calculation of drought duration and deficit indices.

The figure below demonstrates how a streamflow deficit analysis for station Ngaruroro can be performed. For defining dry spells, the function allows fixed and temporally (monthly / daily / seasonally) varying flow quantiles to be used as threshold levels. Various pooling methods are available to remove minor interruptions of mutually dependent drought events. A set of parameters can be specified for customizing the pooling methods.

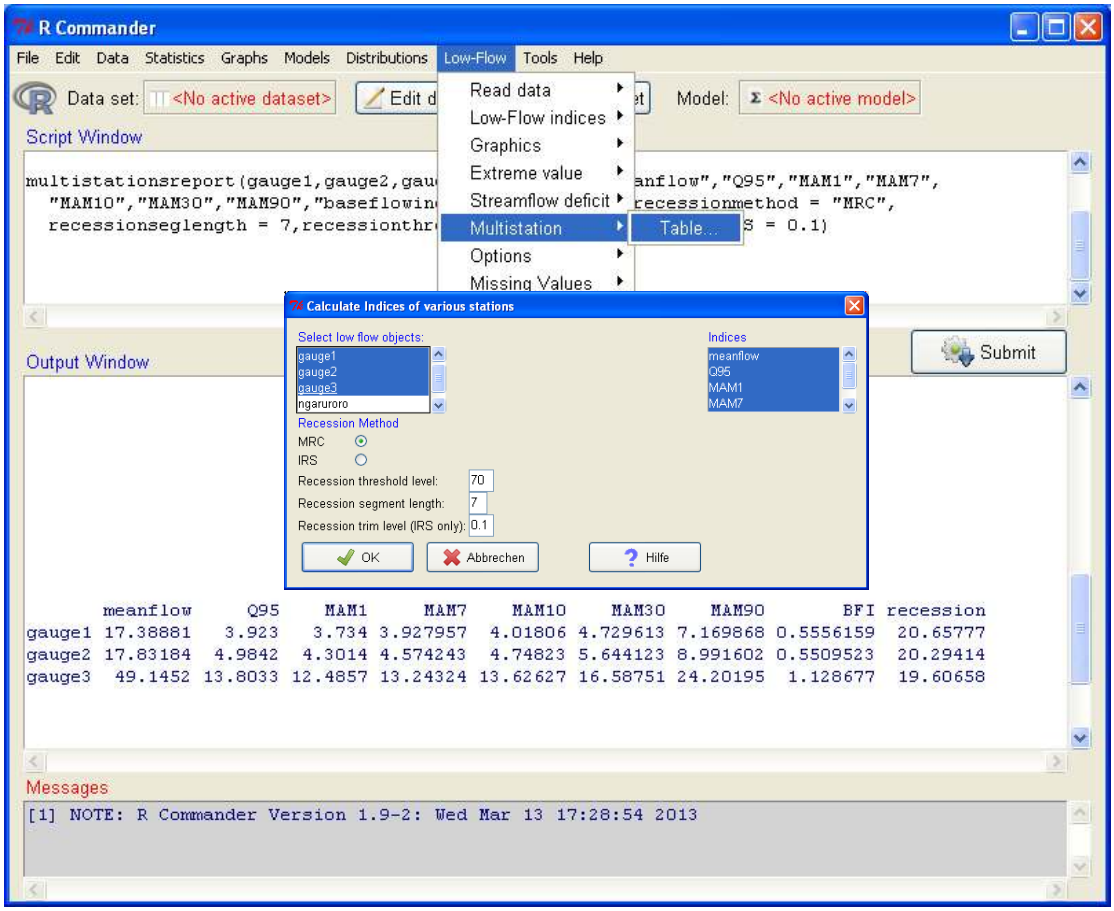
After running the function, a streamflow deficit table is returned to the R workspace (object of class "data frame"). It may be selected as active *Data set* and subsequently visualized using the *View data* button from the R Commander GUI (see table left). The table contains characteristics of each low flow period (dry spell), including duration, deficit volume, drought magnitude, minimum flow, and start date (year, month, day). The table can provide the necessary input for advanced statistics of dry spells (not fully supported by the package), including extreme value statistics of drought duration and deficit volume.



Utility functions

a) Multistation report

Multistation -> Table… enables a fast calculation of selected low flow indices for a set of stations. A table is returned to the Output window which can be copied into other applications.



b) Options

The GUI of RcmdrPlugin.lfstat has a built-in memory which stores the individual parameter settings of a (successful) function call. Hence, a repeated call of the same function will typically start with the same parameter setting as the previous call, what makes it easy for the user to refine an analysis, or to repeat the analysis for a different data set. From the *Options* menu, the parameter memory can be *restored* to standard settings. Furthermore, the parameter memory can be *saved* to disc and *loaded* into a running R session, to restart a new low flow analysis with same settings as the last session.