R: A statistical environment for hydrological analysis

Zambrano-Bigiarini, Mauricio and Bellin, Alberto

Dep. of Civil and Environmental Engineering, Faculty of Engineering, Università degli Studi di Trento, Trento, Italy

e-mail: mauricio.zambrano@ing.unitn.it

1) Introduction

The free software environment for statistical computing and graphics "R" has been developed and it is maintained by statistical programmers, with the support of an increasing community of users with many different backgrounds, which allows access to both well-established and experimental techniques.

Hydrological modelling practitioners spent large amount of time in pre- and post-processing data and results with traditional instruments. In this work "R" and some of its packages are presented as powerful tools to explore and extract patterns from raw information, to pre-process input data of hydrological models, and post-processing its results. In particular, examples are taken from analyzing 30-years of daily data for a basin of 85000 km2, saving a large amount of time that could be better spent in doing analysis.

## 2) Why an hydrologist should invest time in trying R ?

- Many ready-to-use algorithms.
- Existing functions, graphics and packages can be easily adapted to particular needs.
- Write once, use many times.
- Large and active user community.
- Documentation is available in several languages.
- Multi-platform (GNU/Linux, MacOS, Windows).
- Open Source.
- Free :)

3) hydroTSM and hydroGOF

hydroTSM is an R package with S3 functions devoted to management, analysis, interpolation and plot of hydrological time series, mainly oriented to provide support to hydrological modelling tasks.

hydroGOF is an R package with S3 functions providing both, numerical and graphical goodness-of-fit measures between observed and simulated time series.
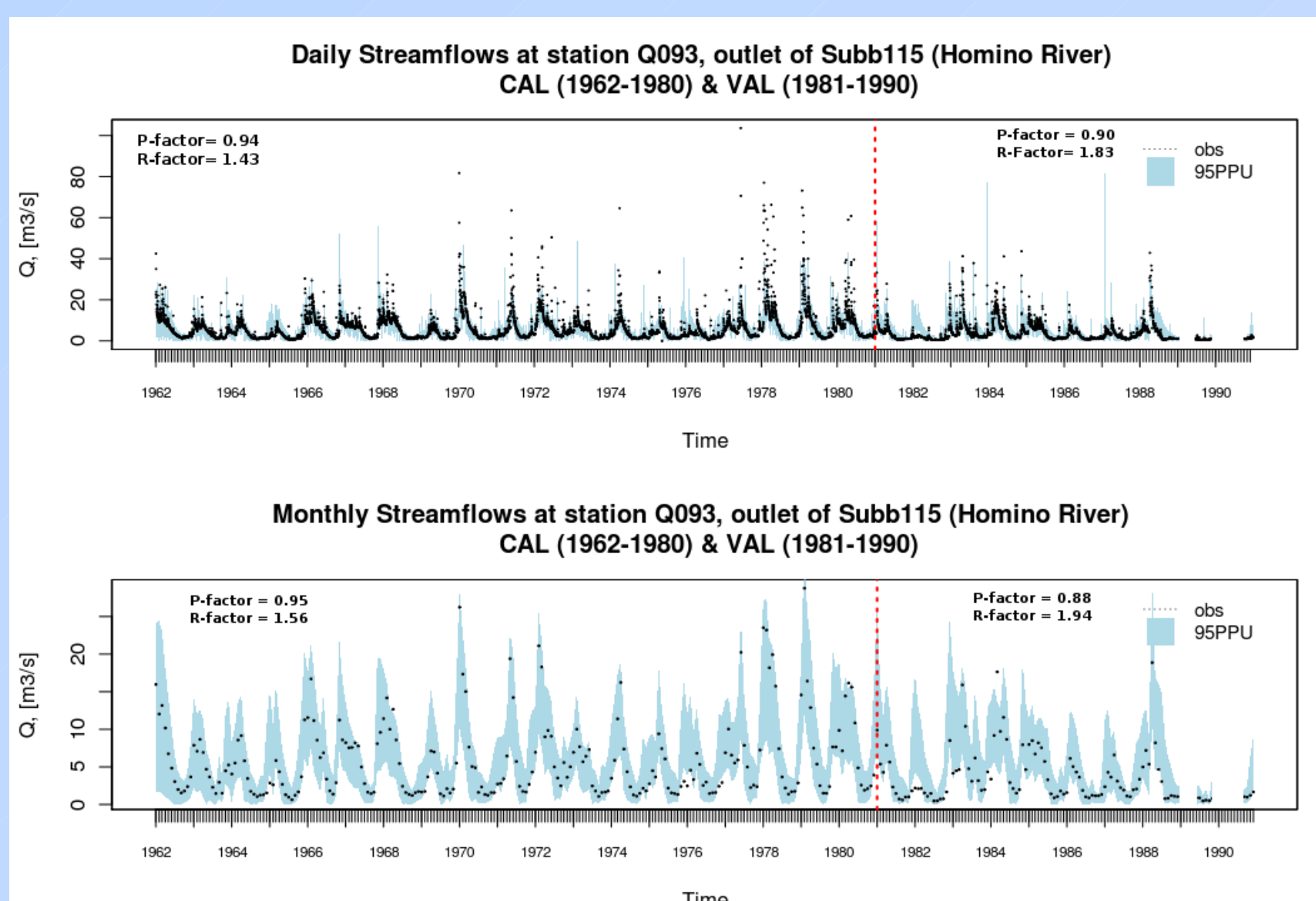
Both packages will be soon available on the R website. Contributions are particularly welcome.
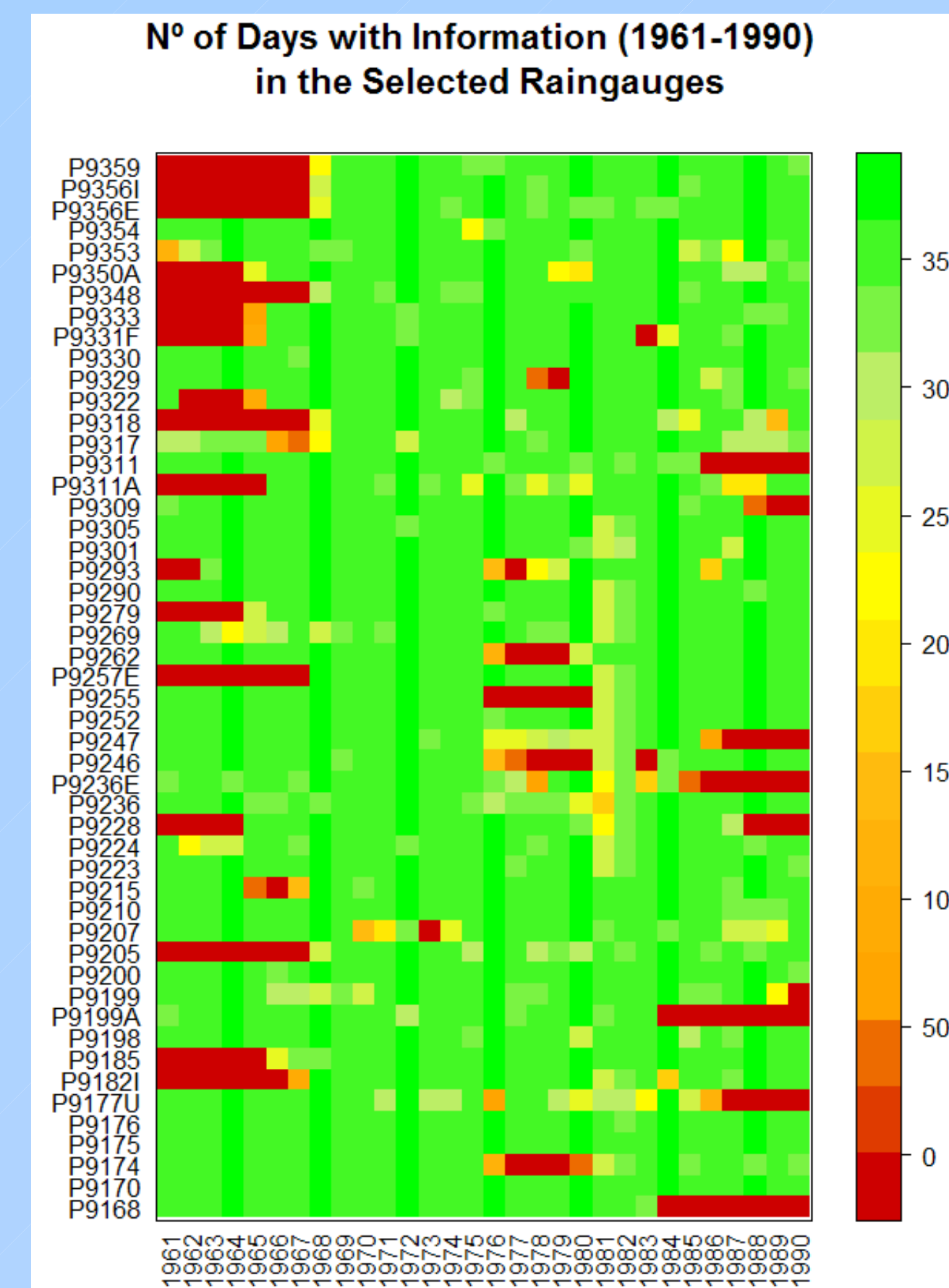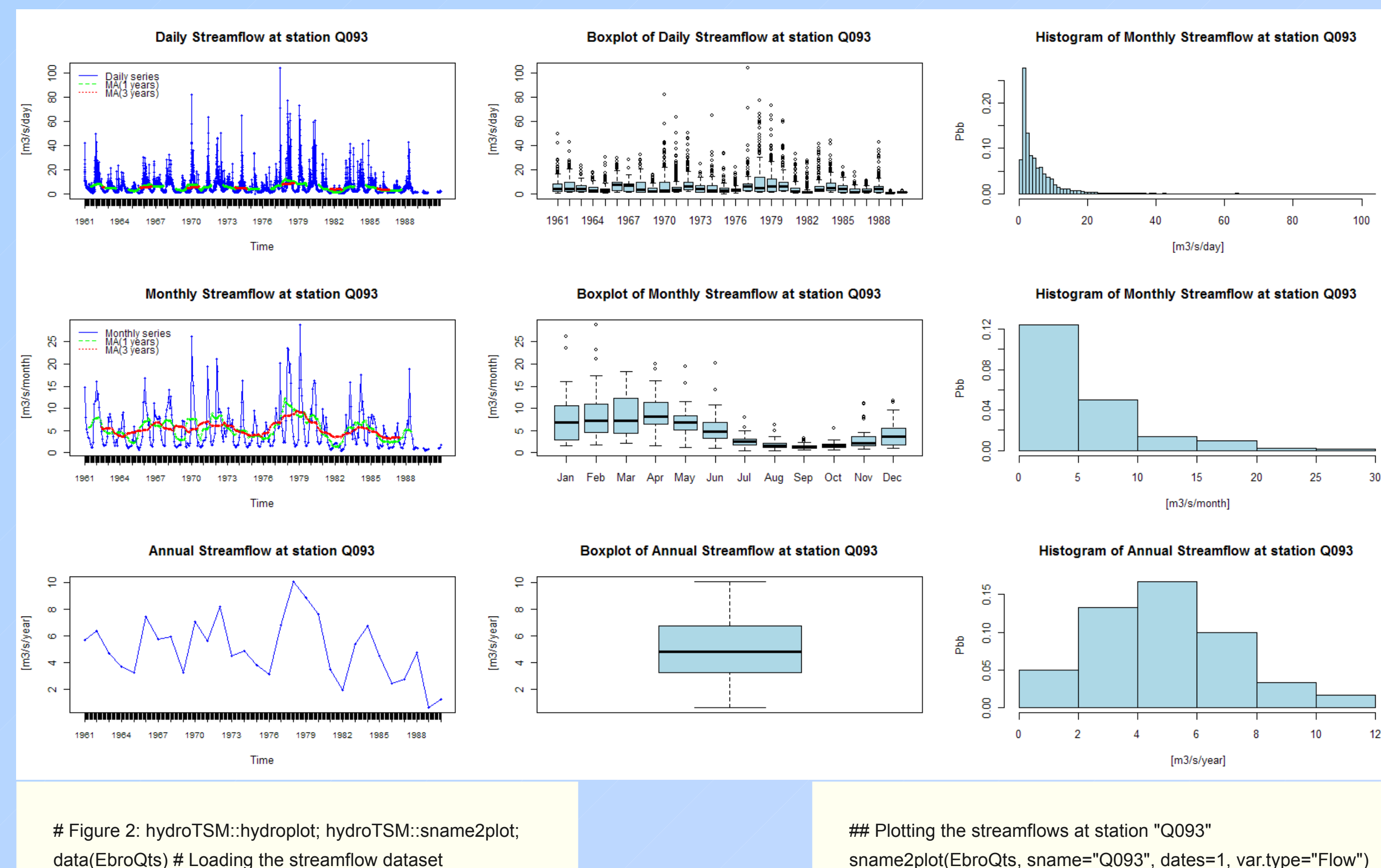

Figure 1: hydroTSM::matrixplot

# Matrix with the days with information per year in selected stations
info <- dwi(EbroQts[,61:110], out.unit="years", dates=EbroQts[,1])

# Plotting the previous matrix, with custom title
matrixplot(info, main="N° of Days with Information (1961-1990) \n in the Selected Streamgauges")
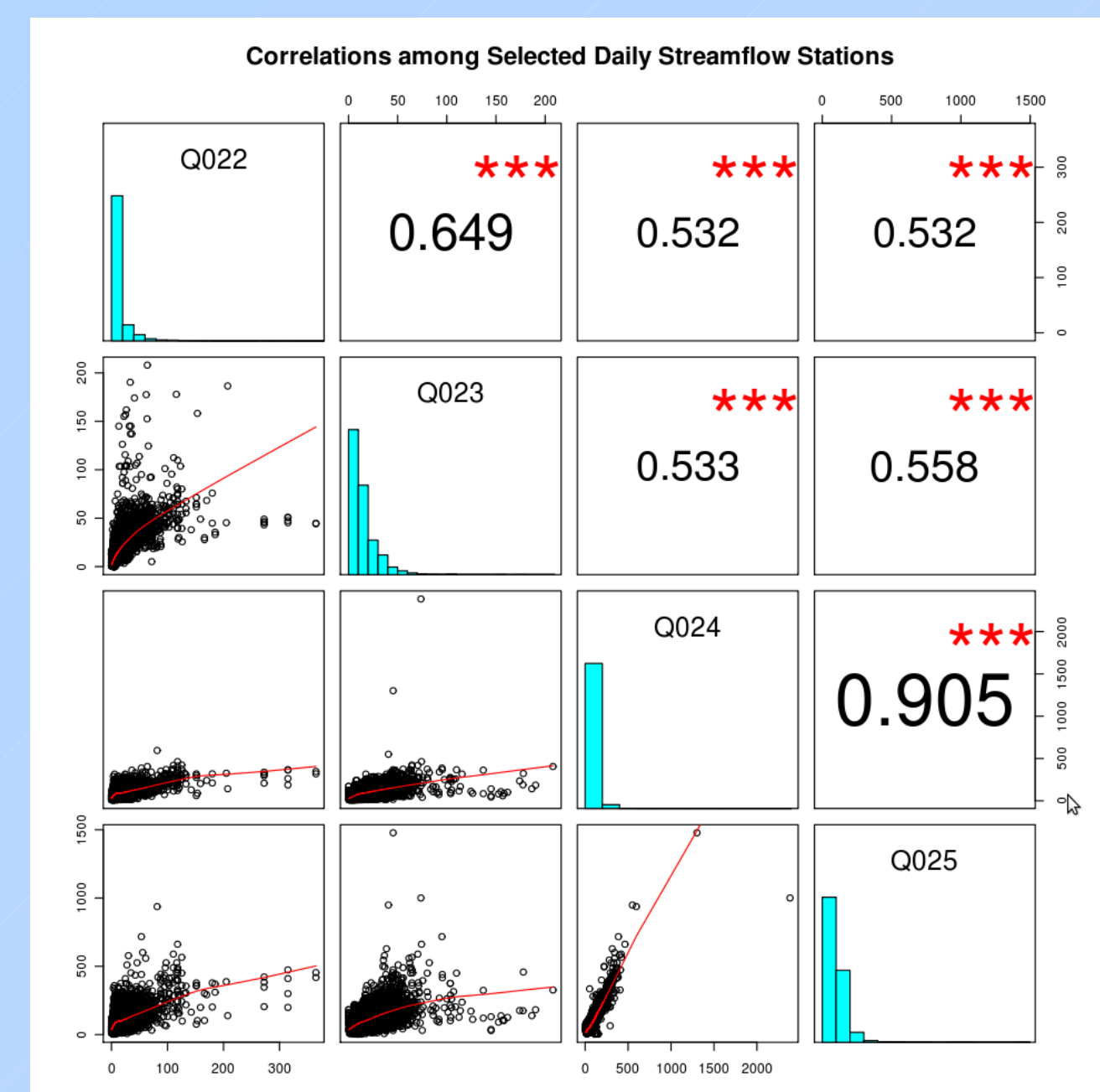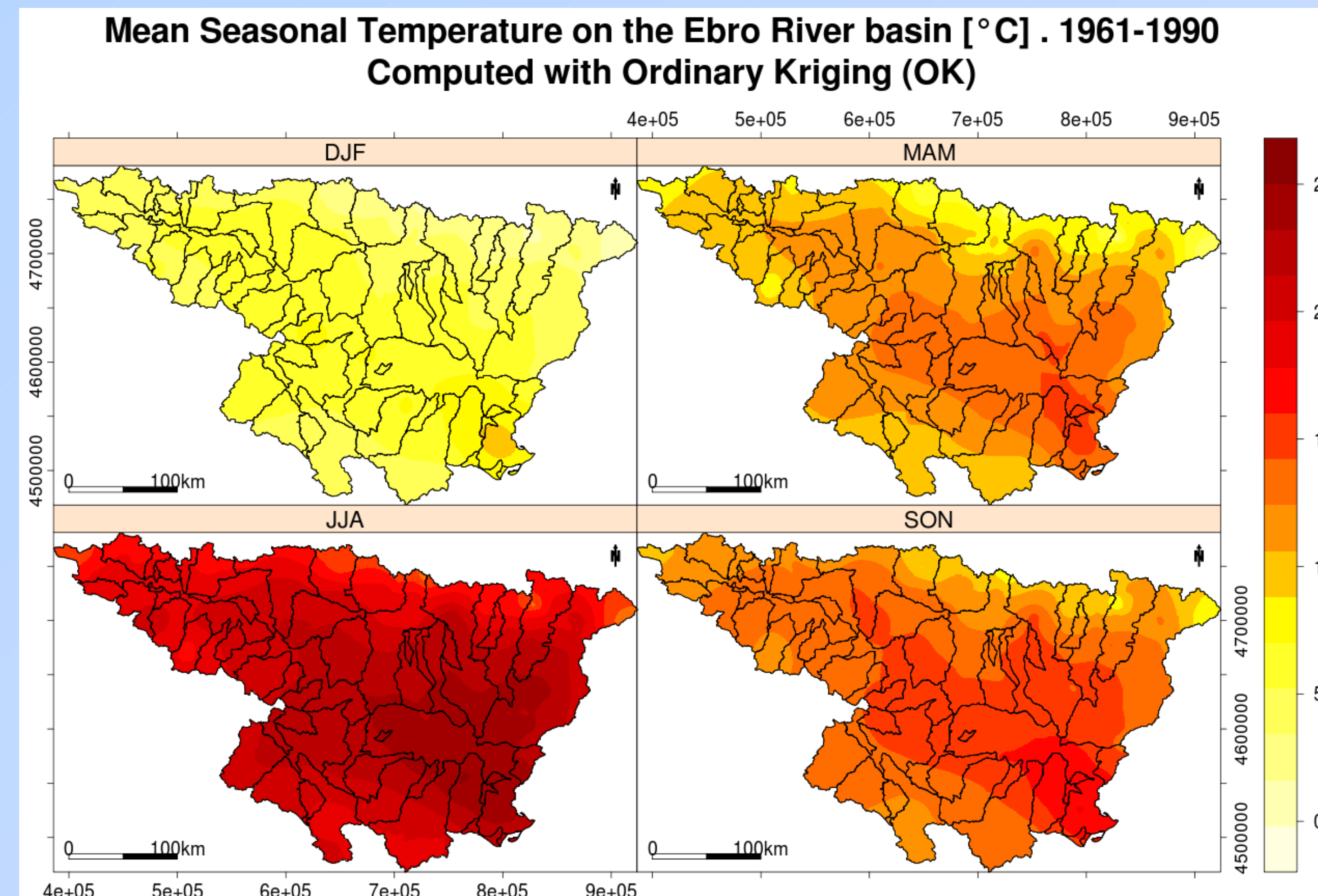

# Figure 3: hydroTSM::hydropairs
hydropairs(EbroQts[,21:24], main="Correlations among Selected Daily Streamflow Stations")
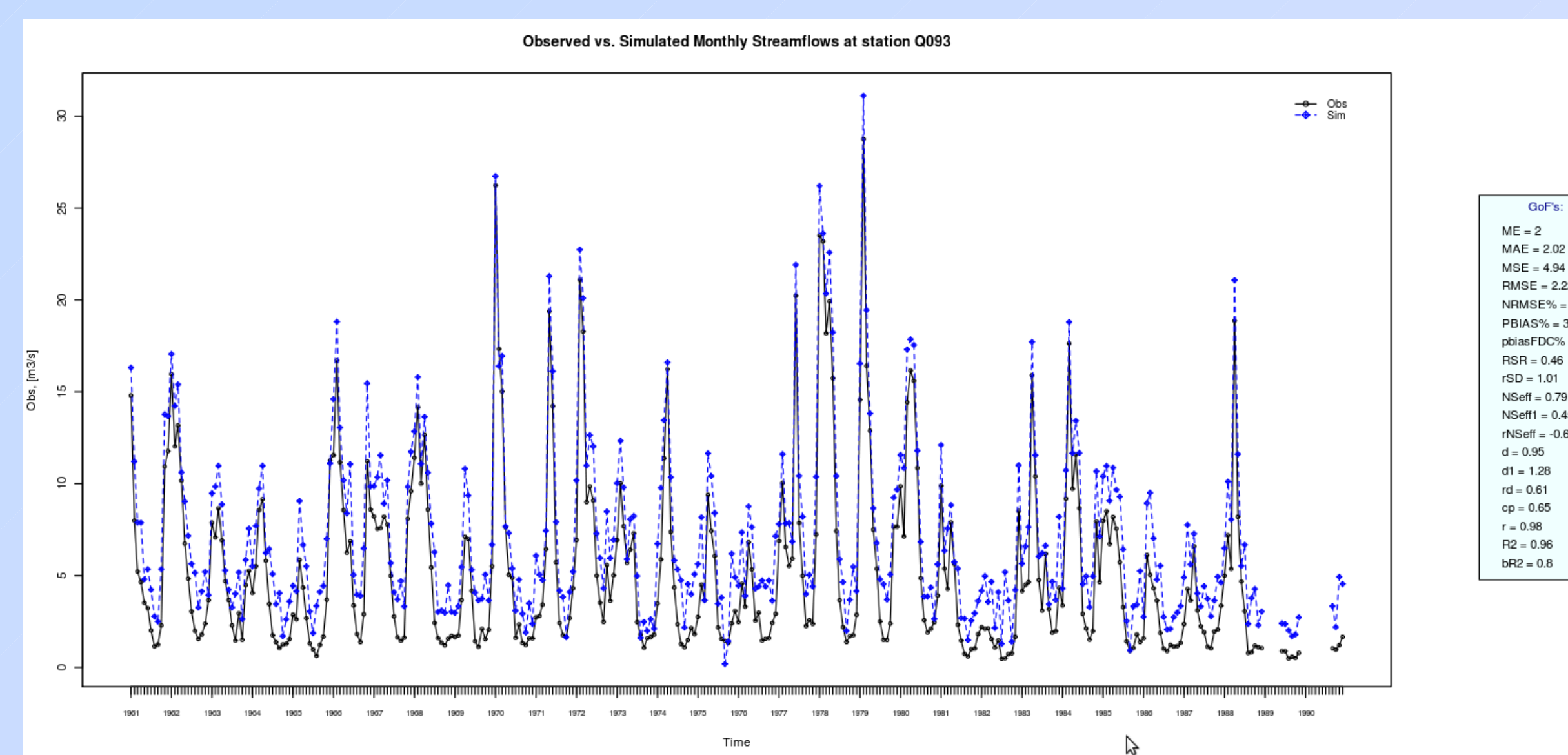

Figure 2: hydroTSM::hydroplot; hydroTSM::sname2plot; data(EbroQts) # Loading the streamflow dataset

Plotting the streamflows at station "Q093"
sname2plot(EbroQts, sname="Q093", dates=1, var.type="Flow")

Figure 4: hydroTSM::hydrokrige and hydroTSM::msplot

Fig 5: hydroGOF::plotbands

Fig 6: hydroGOF::ggof

4) Packages by thematic area:

- **Hydrology:** RHydro, HydroMe, hydrogeo, hydrosanity, topmodel, wasim, *hydroTSM, hydroGOF*
- **Geostatistics:** gstat, automap, geoR, fields, RandomFields
- **GIS**: spgrass6, RSAGA, rgdal, mapproj, sp, maptools RpyGeo, RGoogleMaps, RArcInfo
- **Flood frequency**: POT, evd, nsRFA, extremes, lmomco
- **Programming language interfaces**: C, Fortran, Python, Perl, Java (JRI, JGR,...)
- **Wavelets: wavelets**; wavethresh, wmtsa, Rwave
- **High Performance Computing**:snowfall, multicore, jit, nws, Rmpi, snow, taskPR
- **Optimization:** DEoptim, optim
- **Spreadsheets & DB**: RPostgreSQL, RMySQL, RSQLite, RNetCDF, RexcelInstaller, xlsReadWrite
- **Other statistical software** (e.g, S, SAS, SPSS, Stata, Systat, Minitab): foreign
- **Bayesian statistics**: BAS, BLR, ensembleBMA, evdbayes, LearnBayes, ramps, spBayes,...
- **Latex**: xtable,pgfSweave
- **Data Mining**: Rweka, rattle, party, RandomForest, ...

5) At the end...

"R" is a valuable environment to support undergraduate and graduate education in hydrology, since it is helpful to capture the main features of large amount of data; it is a flexible and fully functional programming language, able to be interfaced to existing Fortran and C code and well suited to the ever growing demands of hydrological analysis; and finally because it is a promising environment for tackling most of the practical problems that reality poses to the hydrological modeller.

6) Where to Start ?

- **Hydrology in R (wiki)**
- **R reference card**
- **Graphics**: http://addictedtor.free.fr/graphiques/
- **Quick R**: http://www.statmethods.net/index.html
- **R-Spatial**: http://r-spatial.sourceforge.net/gallery/
- **CSRL**: http://casoilresource.lawr.ucdavis.edu/drupal/node/100
- **Manuals**: http://cran.r-project.org/manuals.html
- **Packages**: http://cran.r-project.org/web/packages/
- **Search engine**: http://www.rseek.org/, R Site search

7) References

- Jones, O., R. Maillardet, A., Robinson. (2009). Introduction to Scientific Programming and Simulation Using R. 472pp. Chapman & Hall/CRC. Boca Raton, FL.
- Spector, P. 2008. Data Manipulation with R. 154pp. Springer-Verlag, Carey, NC. ISBN 978-0-387-74730-9.
- Applied Spatial Data Analysis with R. Series: Use R. Bivand, Roger S., Pebesma, Edzer J., Gomez-Rubio, Virgilio. 2008. ISBN: 978-0-387-78170-9