Detection and correction of the weekend effect in daily precipitation series

Jose A. Guijarro jaguijarro21@gmail.com

Retired from the State Meteorological Agency (AEMET, Spain) and associate member of the WMO Expert Team on Data Development and Stewardship

15^{th} EUMETNET Data Management Workshop (Oslo, Nov. 4–6, 2025)

Introduction

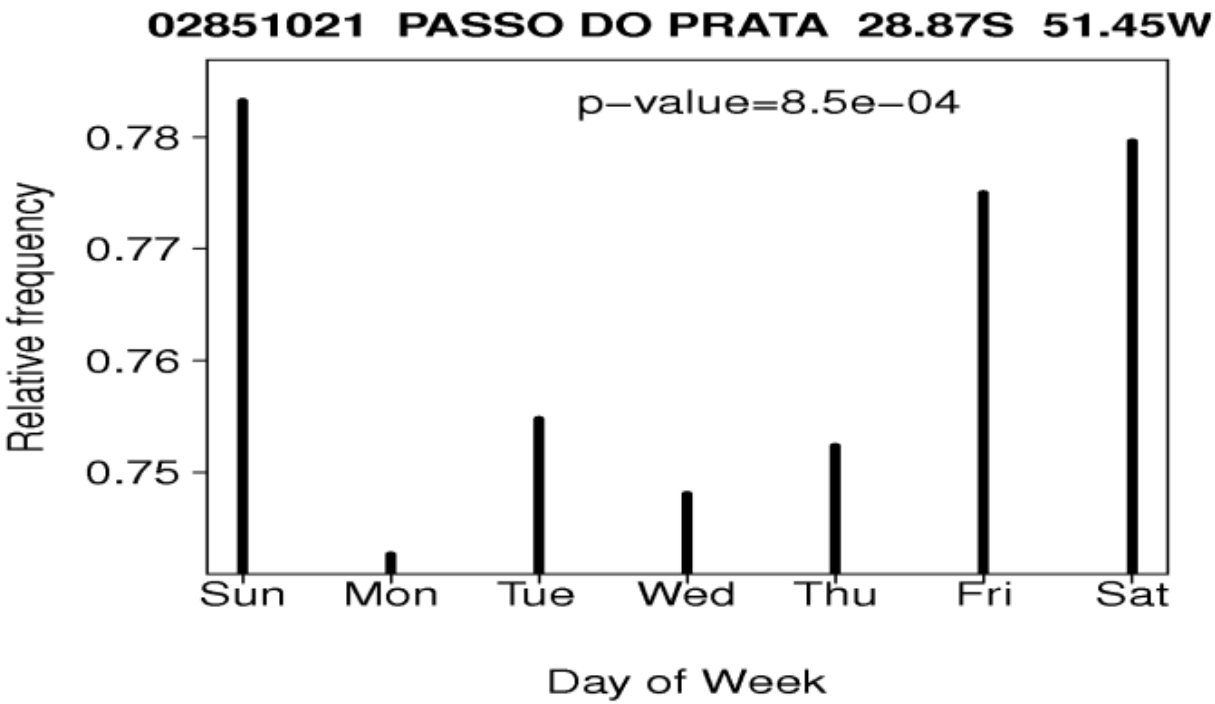
Manual precipitation observations have been traditionally made by measuring the water accumulated in Hellmann or similar rain gauges. This implies that, when the observer has not been able to take the measurement in one or more consecutive days, the following reading will include the total rainfall of those days. When the observer reports those incidents, the missing observations may be assigned an accumulation code in the precipitation series ('-4' in AEMET climate database) to account for these events in any further statistical treatment of the data.

The problem is that these accumulations preclude the correct evaluation of monthly statistical parameters such as the maximum amount of daily precipitation or the number of days with rainfall higher that preset thresholds. One way to overcome this difficulty is using the homogenization function `homogen` included in the *climatol* R package, which can be called setting the `cumc` parameter equal to the cumulation code used in the series. By doing so, `homogen` will redistribute the accumulated precipitation among their corresponding days proportionally to their estimated values calculated from nearby stations. While these adjusted values do not substitute the cumulation codes in the database, they facilitate the calculation of the aforementioned monthly statistical parameters.

Weekend accumulation example

However, sometimes observations in cooperative stations located in administrative offices may be skipped systematically on weekends without noting it in their monthly precipitation reports. This was illustrated in a real case by showing a barplot of the frequency of zero precipitation in every day of the week (Sugahara *et al.*, 2010):

Frequency of zero precipitation:



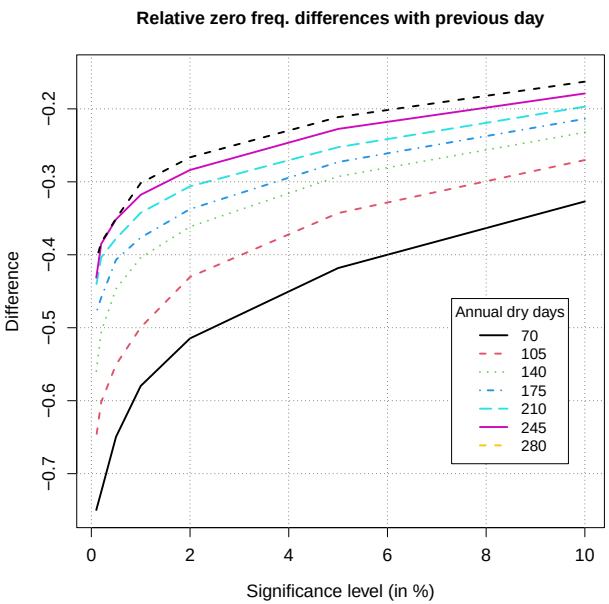
Simulation of null precipitation along the week

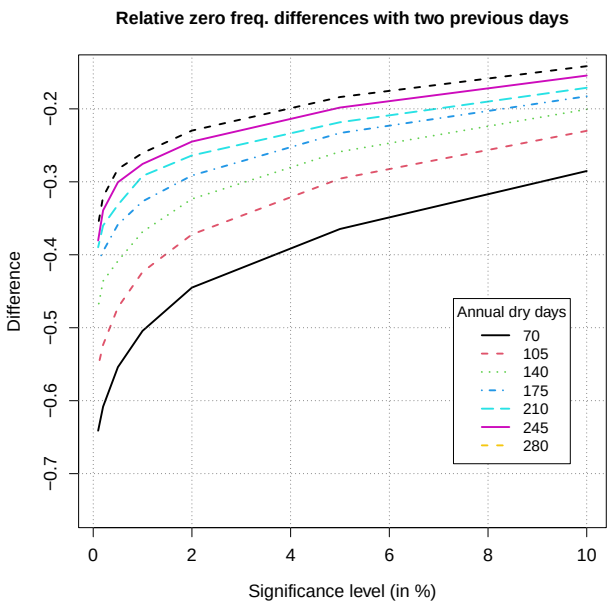
Following a suggestion of my colleague J.A. López, the uniform distribution was used to generate equiprobable frequencies of null precipitation on the seven days of the week and obtain confidence thresholds to detect those cases of missing observations that appear as zeros.

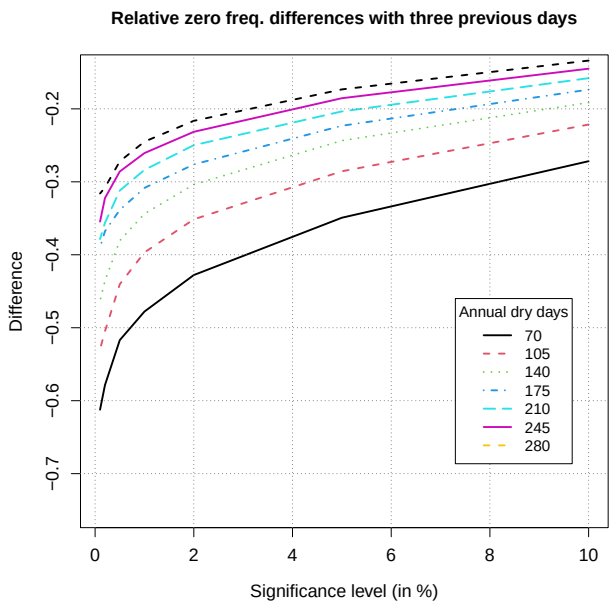
As the detection threshold could depend on the frequency of days without precipitation, means of 10, 15, 20, 25, 30, 35 and 40 random values of a uniform distribution between 0 and 2 were calculated with the help of the R function `runif` for every week day. This is equivalent to 70, 105, 140, 175, 210, 245 and 280 mean annual number of dry days, thus ranging from very humid to dry climates.

Simulation results

This was repeated 10,000 times, calculating differences between the simulated zero relative frequencies of zeros of every week day with the mean of 1, 2 and 3 previous days. Quantiles of the 10,000 differences in each case were calculated for probabilities 0.1, 0.05, 0.02, 0.01, 0.005, 0.002 and 0.001.







Adjusting a model

The implementation of these percentiles of differences of relative frequencies of dry days between any week day and the average of previous days in a function easily applicable to real data was done by adjusting linear models. Differences with the previous day and the average of two and three previous days were modeled separately in the following way.

Calling D the differences, X the probabilities 0.1, 0.05, 0.02, 0.01, 0.005, 0.002, 0.001 and Y the annual frequencies of dry days 70, 105, 140, 175, 210, 245, 280, multiple linear regression models were adjusted as:

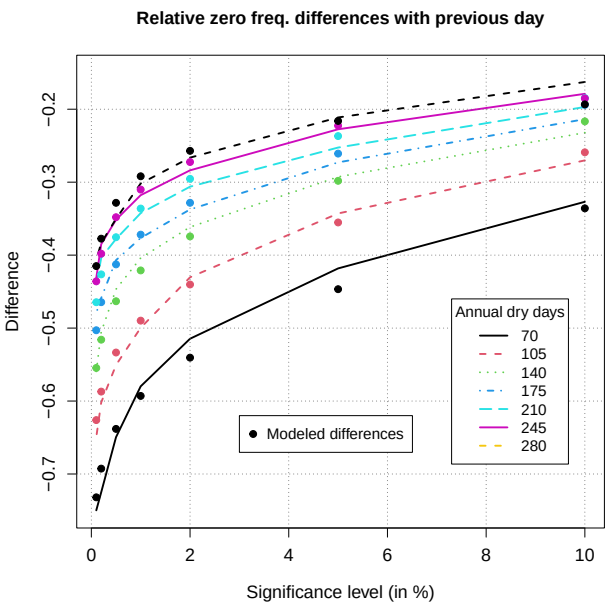
$$D = a + b_1 X + b_2 Y + b_3 log(X) + b_4 log(Y) + b_5 X^2 + b_6 Y^2 + b_7 XY$$

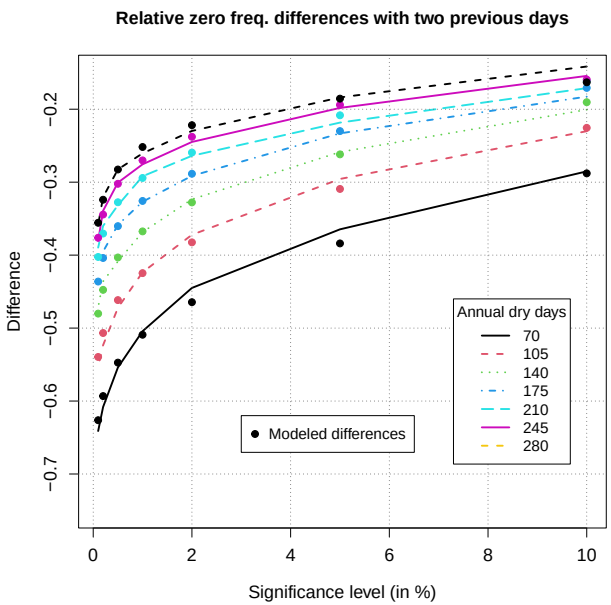
Then, a stepwise regression was applied to these models to remove the less significant variables, which were X^2 and Y. The adjusted coefficients of determination and residual standar errors of the final models were:

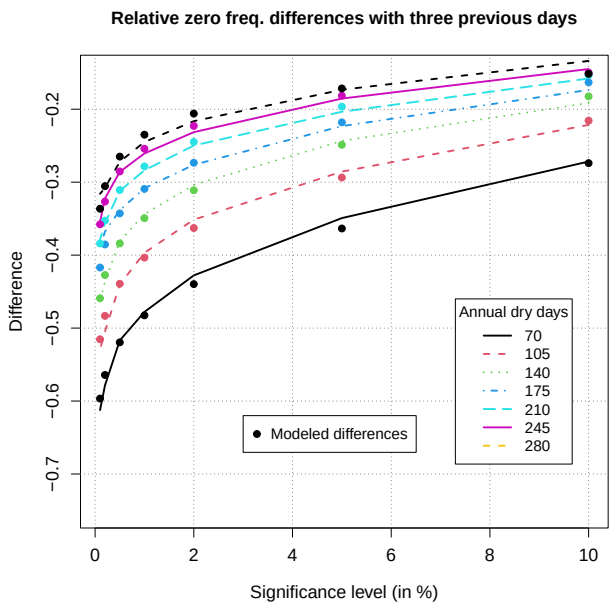
Differences with...	Adjusted R^2	Residual Squared Error
The previous day	0.9870	0.0159
The average of the two previous days	0.9922	0.0105
The average of the three previous days	0.9918	0.0104

Comparison with simulated frequencies

Although the models explain around 99% of the variance, the unexplained 1% is noticeable in some of the graphs as deviations of the modeled values (dots) from the Monte Carlo simulated frequencies (lines):







Application to climate model series

Modeled differences between zero frequencies in a week day and 1, 2 or 3 preceding days were implemented in a new function `weekendaccum`, which tests if these differences are lower than the present probability on a station and yearly basis.

This function was tested with 30 Slovenian daily precipitation series from the INDECIS project. Those series were produced by the historical run (1950-2005) of the Regional Atmospheric Climate Model version 2 (KNMI), so they must be free from spurious weekend accumulations.

However, the new function detected many accumulation cases, even when setting very low significant levels, pointing to bad results of the method used to derive the probabilities of anomalous weekend accumulations.

Therefore, new probabilities of zero frequency differences were derived from those 30 simulated Slovenian series of 20454 daily precipitation data.

Null daily precipitations in Slovenian modeled series

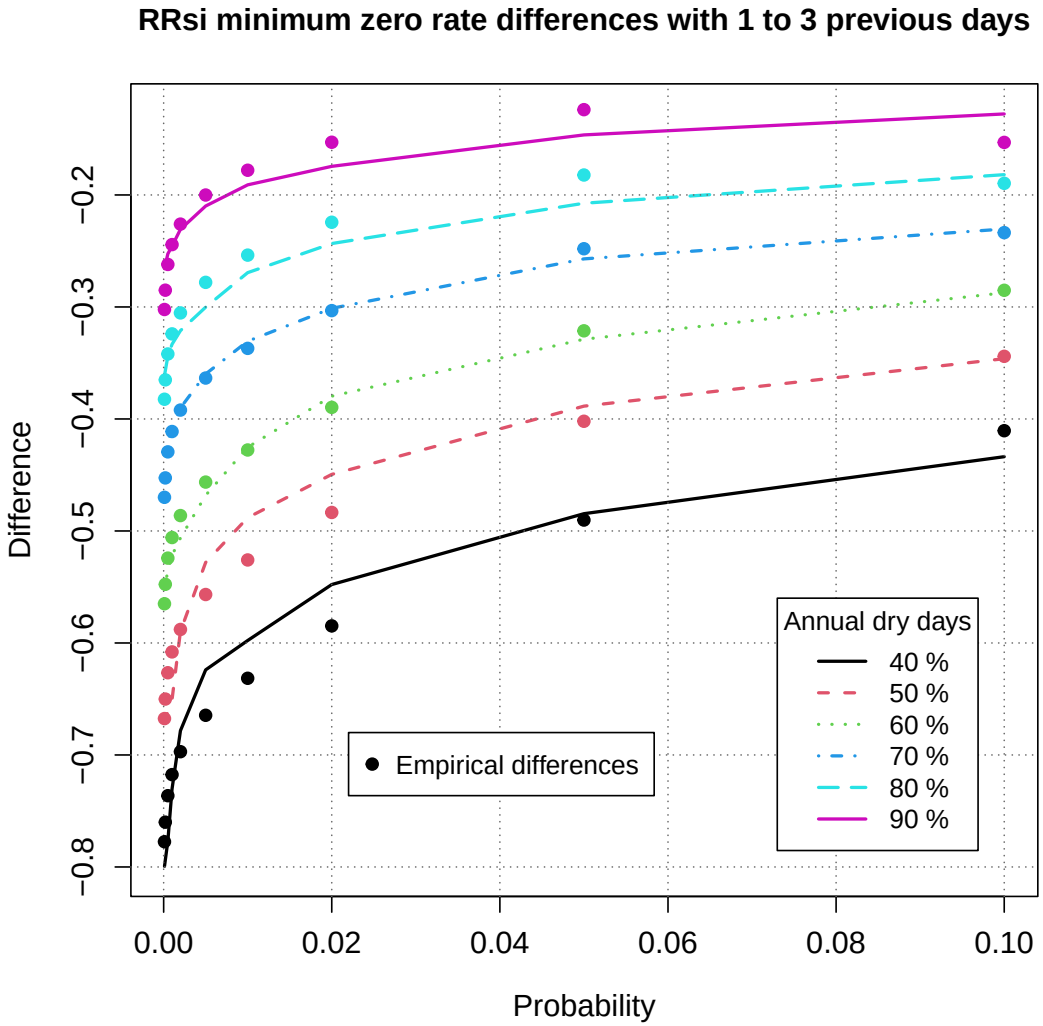
The modeled Slovenian precipitation series have a 39% of null precipitation days in average. As the percentage of zeros could influence de probability to detect anomalous differences of zeros between any week day and 1, 2 or 3 previous days, different versions of the data set were prepared with 50, 60, 70, 80 and 90% of null precipitation days.

The application of the `weekendaccum` function to these datasets yielded maximum absolute differences between a week day and the average of 1, 2 or 3 preceding days saved in `*.Rdat` binary files for every proportion of dry days. After loading those differences in memory, quantiles with probabilities 0.1, 0.05, 0.02, 0.01, 0.005, 0.002, 0.001, 0.0005, 0.0002 and 0.0001 were calculated and, in view of the negligible influence of the number of previous days, a new general function was adjusted (adj.R^2=0.9838, RSE=0.0222):

$$D = -1.065 + 5.139\,X + 1.436 \times 10^{-2}\,Y + 2.470 \times 10^{-2} log(X) - 14.28\,X^2 - 3.732 \times 10^{-5}\,Y^2 - 4.360 \times 10^{-2} XY$$

Empirical and adjusted differences of zero frequencies

This figure shows the differences of zero precipitation frequencies with 1 to 3 previous days for different annual proportions of dry days (lines) and the empirical differences derived from the modeled Slovenian series (points):



Note that with 70% or more dry days the adjusted equation yields lower thresholds for a probability of 10% than for a 5%

Tests of the `weekendaccum` function

Tests have been performed with three of the Slovenian modeled series for the period 1981-1995, setting to zero the 1986-1990 Saturday precipitations in one station and the 1989-1992 Friday and Saturday precipitations in another. These are the results for probabilities 0.01, 0.02, 0.05, 0.1, 0.2, 0.5, 1, 2, 5 and 10%:

Probabilities (%)	**Detection***
0.01 to 0.2	Correct
0.5 to 2	One year of false positives
5	Three years of false positives
10	Many false positives
(: Inserted problems were correctly detected in all cases)*	

Output of the weekendaccum function

```
Checking for weekend effect in Prec_1981-1995 data with wdsl = 1 :

--------- Station 1 (p064):
1986 : 1 day (Sat) difference -0.834 < -0.497
1987 : 1 day (Sat) difference -0.989 < -0.522
1988 : 1 day (Sat) difference -1.575 < -0.59
1989 : 1 day (Sat) difference -1.575 < -0.59
1990 : 1 day (Sat) difference -1.632 < -0.581
1994 : 2 days (Sat-Sun) difference -0.778 < -0.721
175 dates with suspect zero precipitation
--------- Station 2 (p084):
1989 : 2 days (Fri-Sat) difference -0.919 < -0.423
1990 : 2 days (Fri-Sat) difference -0.805 < -0.41
1991 : 2 days (Fri-Sat) difference -0.883 < -0.462
1992 : 2 days (Fri-Sat) difference -0.781 < -0.392
214 dates with suspect zero precipitation

No changes have been made to the original files in exploratory mode.
Rerun the process with expl=FALSE to change suspect zeros with the cumc code.

The output of this process has been saved to file Prec_1981-1995-wkn.txt
and most negative differences (maxdif) can be loaded with:
 load("Prec-wkn_1981-1995.Rdat")
```

Application to real data

The `weekendaccum` function was tested also on the 1883 complete Spanish real daily precipitation series of the period 1991-2000 with the same set of probabilities specified (in %) through the parameter `wdsl` (acronym for *Week Day Significant Level*). Example: after having prepared the input files for *climatol* with names `PREC_DIA_1991-2000.dat` (data) and `PREC_DIA_1991-2000.est` (stations), we can run:

```
library(climatol) #load version 4.3-1 or greater
weekendaccum('PREC_DIA',1991,2000,wdsl=1) #for 1% sig. level
```

Detections for the different tested probabilities (empirical significant levels):

wdsl	0.01	0.02	0.05	0.1	0.2	0.5	1	2	5	10
No. of years	26	26	32	37	43	66	140	357	1457	1188*
No. of series	8	8	9	11	15	33	94	280	970	814*

(: Note the lower detection with 10% probability due to bad adjustment in dry climates)*

Final remarks

The default value of the parameter `wdsl` has been set to the lowest 0.01 to avoid false positives, but results shown so far point to safe values up to `wdsl=0.2`. Values higher than 2 should be avoided.

Anyway, by default `weekendaccum` runs in exploratory mode, without modifying any value in the series. If you want to correct the detected accumulations, you must run the function again with the parameter `expl=FALSE`. In this way, the detected week days (the same each year) will be assigned the cumulative code `cumc` (-1 by default) in place of the anomalous zeros if there are positive accumulation in the following day. Then, running the `homogen` function with the same `cumc` code will redistribute the accumulated precipitation into their corresponding days proportionally to the values estimated from nearby stations.

Many thanks for your attention!!!

Climatol home page: https://climatol.eu

Reference: Sugahara S, Rocha RP, Silveira RPD (2010): Some comments on homogeneity of Brazilian instrumental rainfall data. In Eos Transactions AGU, v. 91, n. 26, Meeting of the Americas Supplement, abstract H41A-01, 2010.