# Homogenization of the ECA&D daily Essential Climatic Variables for the INDECIS project

J.A. Guijarro[1], E. Aguilar[2], G. van der Schrier[3], O. Skrynyk[2]

[1]Agencia Estatal de Meteorología (AEMET), Spain
[2]Universitat Rovira i Virgili, Center for Climate Change, Spain
[3]Royal Netherlands Meteorological Institute (KNMI), The Netherlands

10th Seminar for Homogenization and Quality Control in Climatological Databases and 5th Conference on Spatial Interpolation in Climatology and Meteorology (Budapest, 12-14 October 2020)

# Outline

# Introduction (i)

The Work Package 3 (WP3) of INDECIS is due to provide quality controlled and homogenized daily series of the Essential Climate Variables stored at the European Climate Assessment & Dataset (ECAD) to be used by the rest of the teams:

1. CC : Cloud Cover (oktas)
2. FG : Wind Speed (0.1 m s$^{-1}$)
3. HU : Humidity (1 %)
4. PP : Sea Level Pressure (0.1 hPa)
5. RR : Precipitation Amount (0.1 mm)
6. SD : Snow Depth (cm)
7. SS : Sunshine (0.1 hours)
8. TN : Minimum Temperature (0.1 °C)
9. TX : Maximum Temperature (0.1 °C)

# Introduction (ii)

- ▶ After a benchmarking exercise using daily series of a regional model as homogeneous database, only ACMANT and CLIMATOL provided complete results for all variables and benchmarks flavors.

- ▶ As none of them consistently provided better results in all benchmark and only CLIMATOL returned always complete results, with all missing data infilled, it was decided to use this package for the homogenization of the observed series stored at ECA&D.

- ▶ However, this task proved to be very difficult because of their:
    - ▶ Big number of series
    - ▶ Very different station density and geographical distribution
    - ▶ Highly variable time coverage in the series

# Map of RR stations

Example of broad geographical extension and high density variation of station distribution:



RR stations
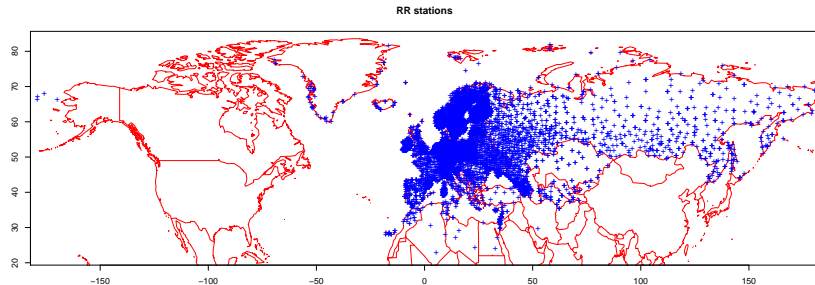
# First homogenization attempts

- ▶ To overcome these difficulties, an automatic homogenization by rectangular areas of 3x6° lat-lon was attempted, but in areas with few series the complete lack of data in some time steps halted the process.
- ▶ 20CR reanalysis series were then added to provide reference data in these cases, but many series are not updated and there were still frequent concurrent lack of data after 2014 (last year of 20CR data).
- ▶ In areas with high density of stations the homogenization took several weeks of processing, making it unpractical.

# First homogenization methodology

In view of the aforementioned difficulties, the final approach consisted in:

- ▶ Apply a thorough quality control of the series by means of the set of routines prepared by Enric Aguilar within this project.
- ▶ Delete data highly suspicious to be erroneous (from 0.02 to 0.29 %, depending on the variable).
- ▶ Homogenize the QC controled series one by one, using the closest 20CR series as reference for the period 1950-2014, and adding any existing raw data for the 2015-2018 period.

# First homogenization results

This table shows the number of series homogenized and the number of break-points and outliers detected and corrected. (Snow Depth was skipped because it was found terribly difficult to homogenize in the benchmarking tests.)

| Variable | Number of series | ≥10y in 1950-2014 | Repeated series | Homogen-ized series | No. of breaks | No. of outliers |
|----------|------------------|-------------------|-----------------|---------------------|---------------|-----------------|
| CC | 2147 | 1672 | 6 | 1672 | 1548 | 0 |
| FG | 1492 | 1232 | 18 | 1232 | 2697 | 1 |
| HU | 2395 | 1961 | 6 | 1961 | 2928 | 0 |
| PP | 1684 | 1293 | 42 | 1293 | 2482 | 176 |
| RR | 15962 | 13440 | 756 | 13439 | 1449 | 115 |
| SD | 8619 | – | 7 | – | – | – |
| SS | 1252 | 1006 | 12 | 1005 | 519 | 3 |
| TN | 6442 | 5163 | 744 | 5163 | 3439 | 0 |
| TX | 6287 | 5012 | 726 | 5012 | 1969 | 0 |

# Example of detection in a HU series

First break-point detected (at monthly scale):



**aux–m   1 (320)**
**BJELASNICA**
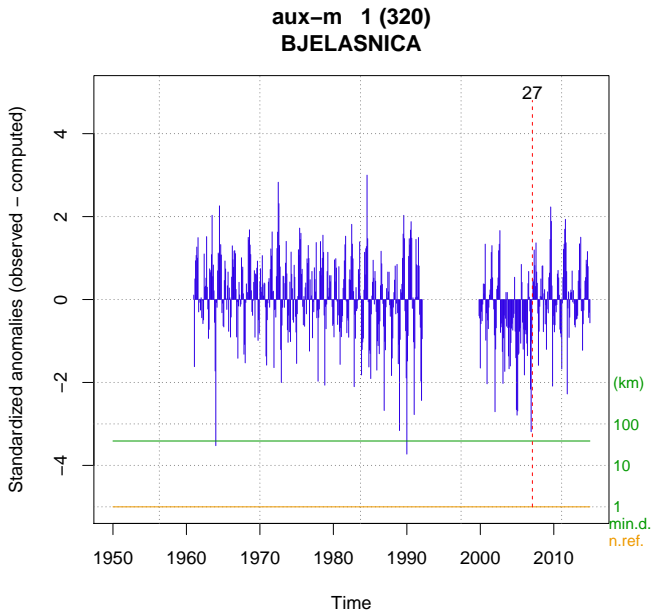
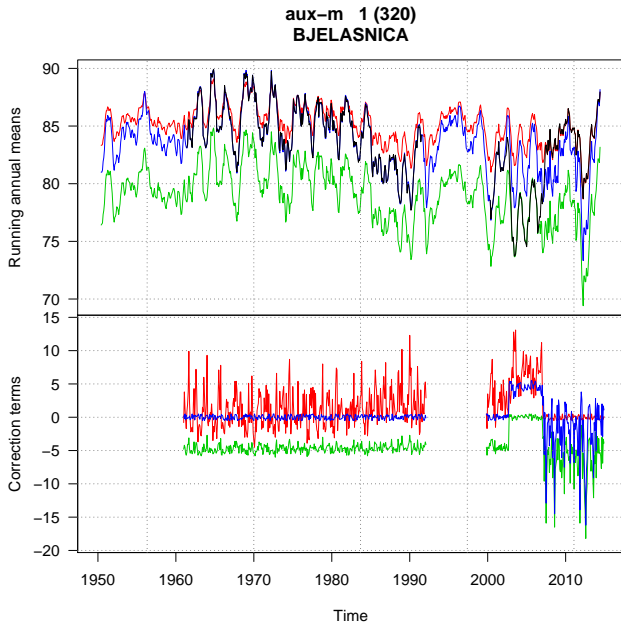# Example of detection in a HU series

Second break-point detected (at monthly scale):



**aux-m 3 (320-2)**
**BJELASNICA-2**
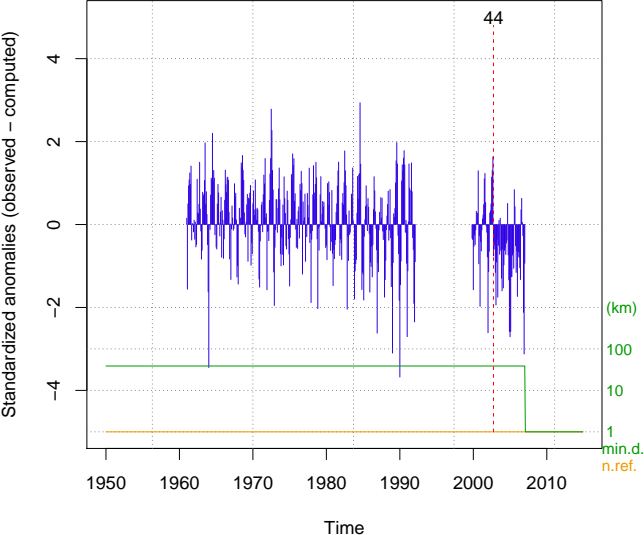
# Adjustment of the HU series

Original data in black. Adjustments in colors:



**aux–m   1 (320)**
**BJELASNICA**

# Homogenization update to 1950-2019

- ▶ At least yearly homogenizations (or two per year) are expected to keep the INDECIS products updated.

- ▶ An update of the 20CR reanalysis is under course, but we cannot expect it to keep pace with the growing ECA&D new data.

- ▶ Therefore, we are currently trying to use an upscaled version of ERA5 to help in the homogenization and missing data infilling after the last available year of 20CR, including 10 years of overlapping.

- ▶ This poses new problems due to the high spatial and temporal resolution of ERA5, and the need to compile derived fields (as relative humidity, from temperature and dew point grids).

- ▶ This current approach will need to be assessed, and a full automatization should be reached to keep the process operative in the future.

# Acknowledgments