Operational homogenization of daily climate series in Spain: experiences with different variables

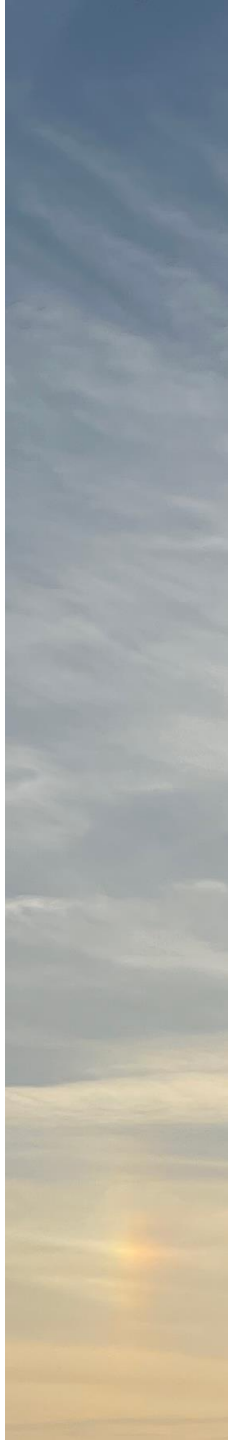
Belinda Lorenzo, José A. Guijarro, Andrés Chazarra, César Rodríguez-Ballesteros, José V. Moreno, Ramiro Romero-Fresneda, Maite Huarte, Ana Morata

Spanish State Meteorological Agency (AEMET)

blorenzom@aemet.es

11th Seminar for Homogenization and Quality Control in Climatological Databases and 6th Interpolation Conference jointly organized with the 14th EUMETNET Data Management Workshop

Hungarian Meteorological Service, Budapest, 9-11 May 2023



Overview

- Introduction
- Procedure
  - Data, regions and period selection
  - General steps using Climatol v4.0t
  - Different variables
  - Normal values
- Results
- Conclusions

Introduction

Calculation of the climatological normals for the new standard period 1991-2020

⇩

Need for homogenized series

Land observation network:

- Specialist observers
- Volunteer colaborators
- Automatic stations

⇨ Stored in National Climatological Database ⇨ widespread **daily** data available

Working with daily series:

Advantages: straightforward calculation of daily parameters and climatic indices, such as number of days over certain thresholds.

Disadvantages: their high variability and the huge calculation capacity that is required.

Introduction

Daily series of different variables

- Precipitation
- Maximum and min. temperature
- Sunshine hours
- Relative humidity
- Station level pressure
- Mean wind speed
- Maximum wind gust

Homogenization Climatol 4.0t

- Computed in AEMET's HPC system (Cirrus)
- 8 variables by region → 51 processes

Homogenized daily series

Derived variables

- Average temperature
- Sea level pressure
- Vapour pressure

Normal values with CLINO_tool v.1.5

Procedure

Data, regions and period selection

- Period extended to **1975-2020** ⟶ Allows CLINO calculations for 1981-2010 and 1991-2020.
- Minimum of **5 years** of daily data (10 years for precipitation).
- For precipitation and temperature, the **study area** is divided into 12 regions: mainland Spain (divided in turn into ten regions with similar climatic conditions and coinciding with the main hydrographic basins), the Balearic Islands and the Canary Islands, due to their spatial and climatic differentiation
- For the rest of the variables, the **study area** is divided into three regions: mainland Spain (including autonomous cities of Ceuta and Melilla), the Balearic Islands and the Canary Islands.

Procedure

General steps using Climatol v4.0t

1. **Preparing the input data (.dat and .est files).** csv2climatol().
2. **First exploratory analysis of the daily data for quality control.** homogen() with expl=TRUE.
3. **Only for precipitation: elimination of trace values and disaggregation of accumulated values.** homogen() with cumc=-4.
4. **Preparation of monthly values.** dd2m().
5. **Monthly homogenization.** homogen().
6. **Review of the monthly homogenization.**
7. **Daily homogenization.** homogen() with metad=TRUE.
8. **Calculation of the daily series of derived variables.**

	Monthly		daily		both			
	dz.max	inht	dz.max	nref	std	vmin	vmax	gp
Precipitation	15	20	25	1	2	-	-	4
Temperature	6	-	15	-	-	-	-	-
Sunshine hours	6	-	12	-	1	0	-	-
Relative humidity	6	-	15	-	-	0	100	-
Station level pressure	6	-	15	-	-	0	-	-
Mean wind speed and max. wind gust	6	-	15	-	2	-	-	-

Table 1. Parameters used in function homogen().

Procedure

Precipitation

- At least 10 years of daily data selected (6293 stations).

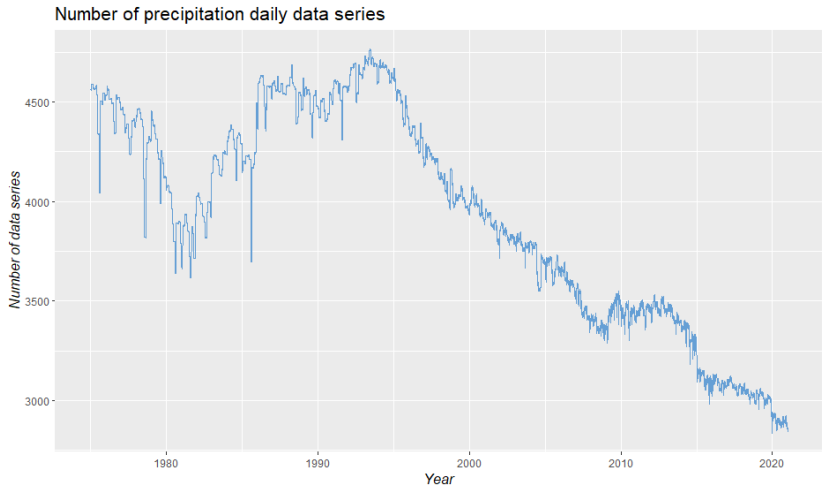


Fig. 1. Number of daily data series for precipitation.

- Territory is divided into 12 regions.

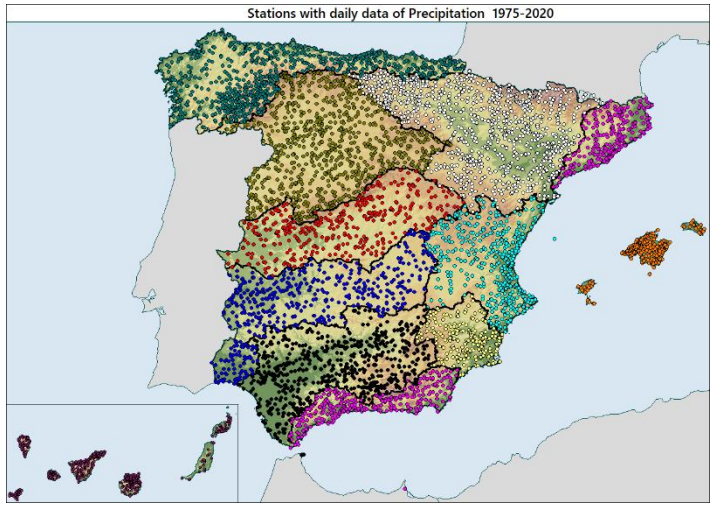


Fig. 2. Distribution of precipitation stations.

- Elimination of trace values and dissagregation of accumulated values.
- Daily data grouped into monthly values. dd2m() with valm=1 (totals).

Table 1. Parameters used in function homogen().

Procedure

Precipitation

- After monthly homogenization, 30 data series are rejected.

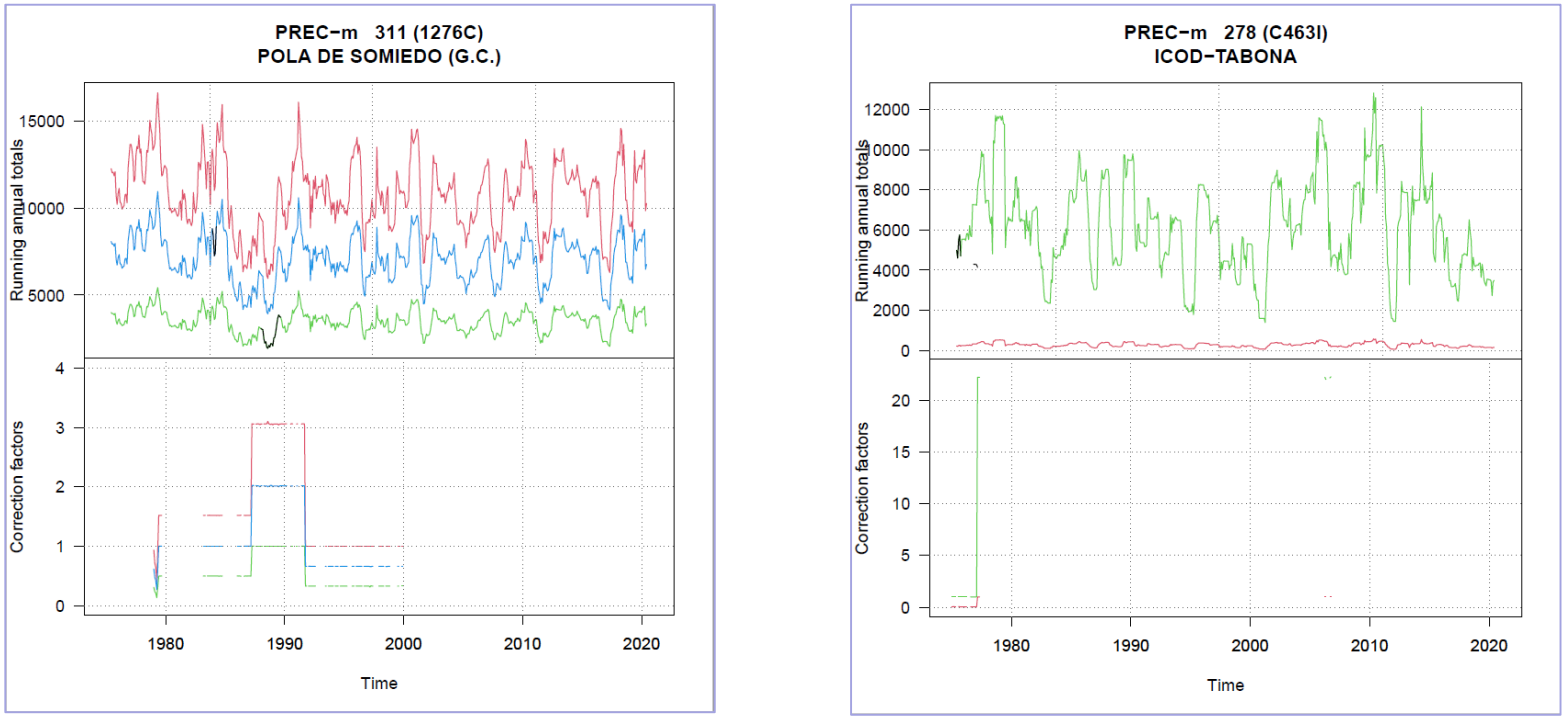


Fig. 3. Examples of series rejected for precipitation after monthly homogenization.

Procedure

Maximum and minimum Temperature

- At least 5 years of daily data selected (3704 stations).

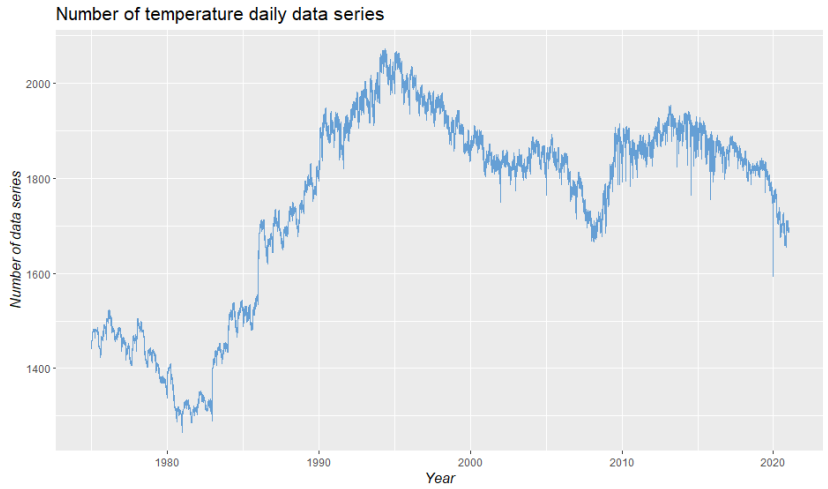


Fig. 4. Number of daily data series for temperature.

- Territory is divided into 12 regions.

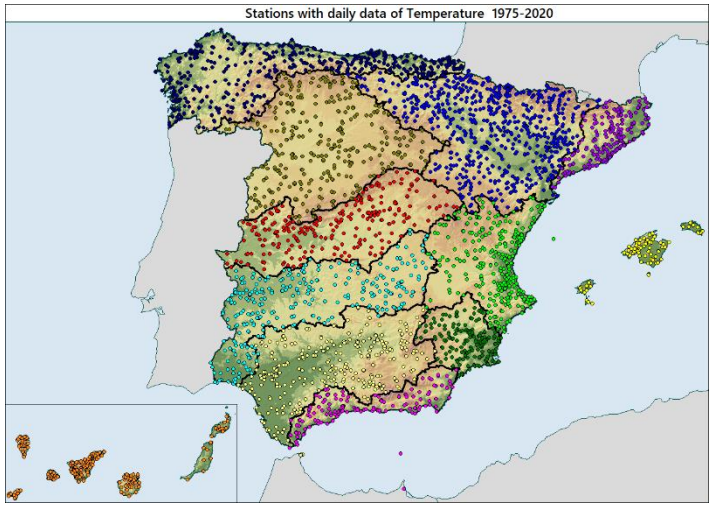


Fig. 5. Distribution of temperature stations.

- After homogenizing the grouped monthly data, some 34 data series are rejected.

Procedure

Sunshine hours

- 5 years of daily data selected (238 stations).

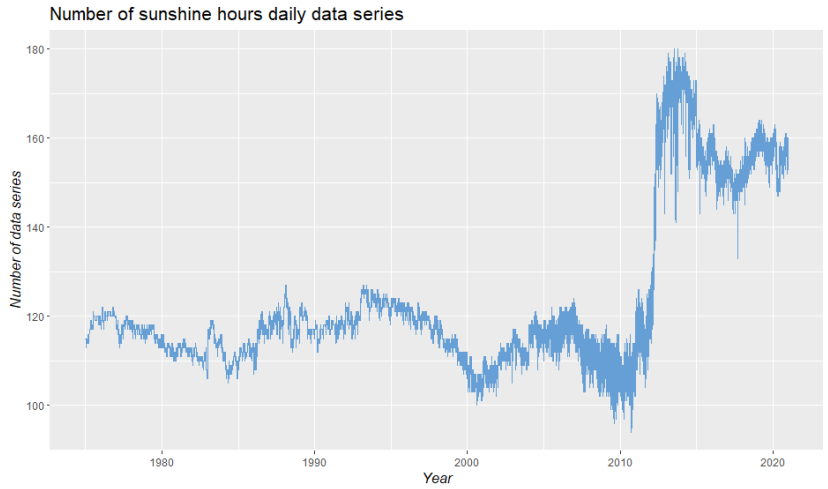


Fig. 6. Number of daily data series for sunshine hours.

- Territory is divided into 3 regions.

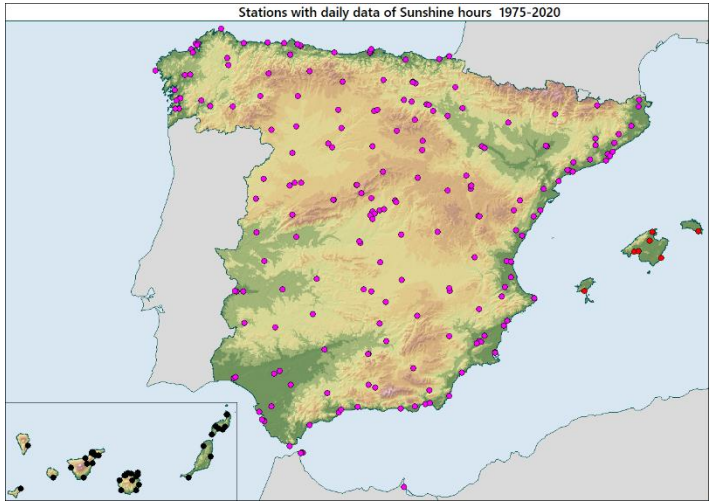


Fig. 7. Distribution of sunshine hours stations.

- After homogenizing the grouped monthly data, some 12 data series are rejected.
- After the daily homogenization, the total hours of sunshine can exceed the theoretical number of hours that a certain place should have. fix.sunshine().

Procedure

Relative humidity

- 5 years of daily data selected (914 stations).

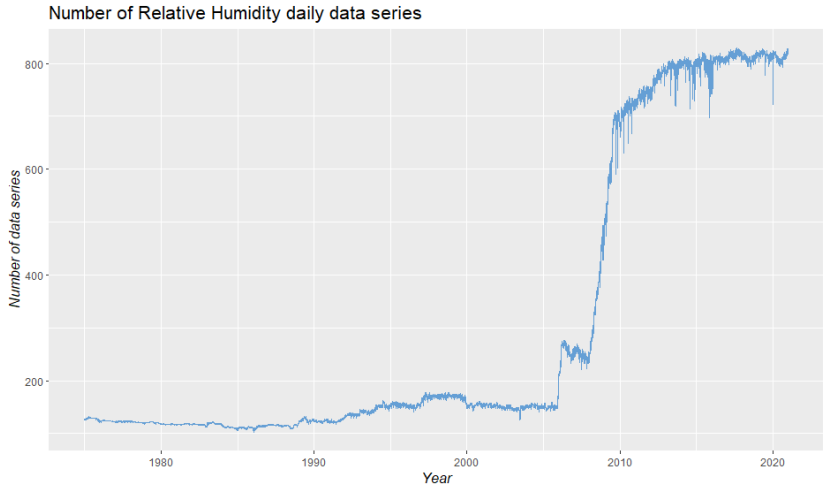


Fig. 8. Number of daily data series for relative humidity.

- Territory is divided into 3 regions.

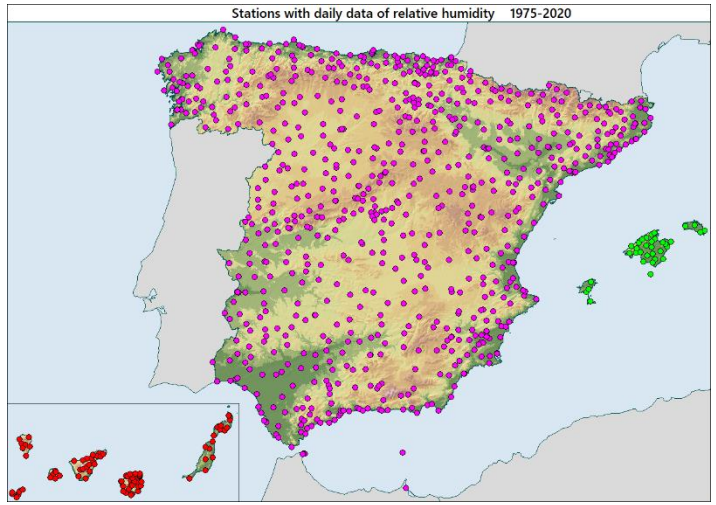


Fig. 9. Distribution of relative humidity stations.

- After homogenizing the grouped monthly data, about 15 data series are rejected.
- Vapour pressure is calculated as derived variable.

Procedure

Station level pressure

- 5 years of daily data selected (274 stations).

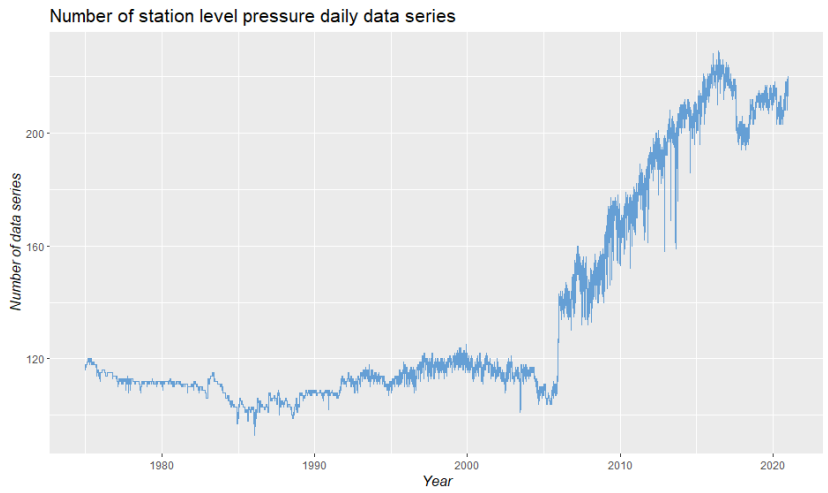


Fig. 9. Number of daily data series for station level pressure.

- Territory is divided into 3 regions.

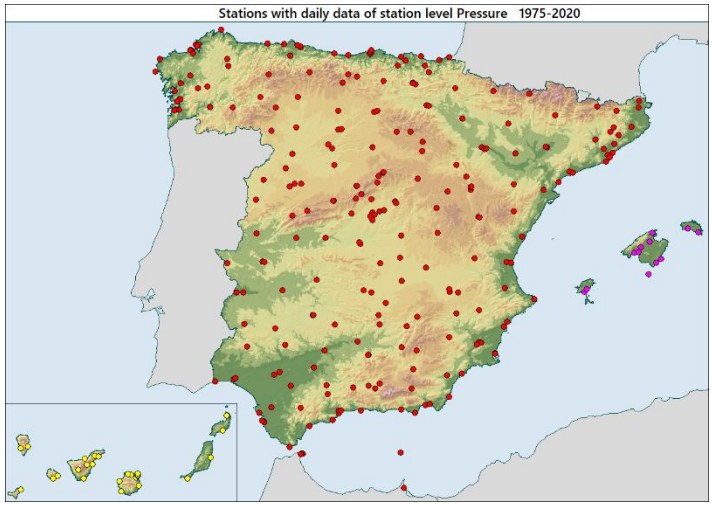


Fig. 10. Distribution of station level pressure stations.

- After homogenizing the grouped monthly data, about 13 data series are rejected.
- Pressure at sea level is calculated as derived variable.

Procedure

Mean wind speed and maximum wind gust

- 5 years of daily data selected (near 900 stations).

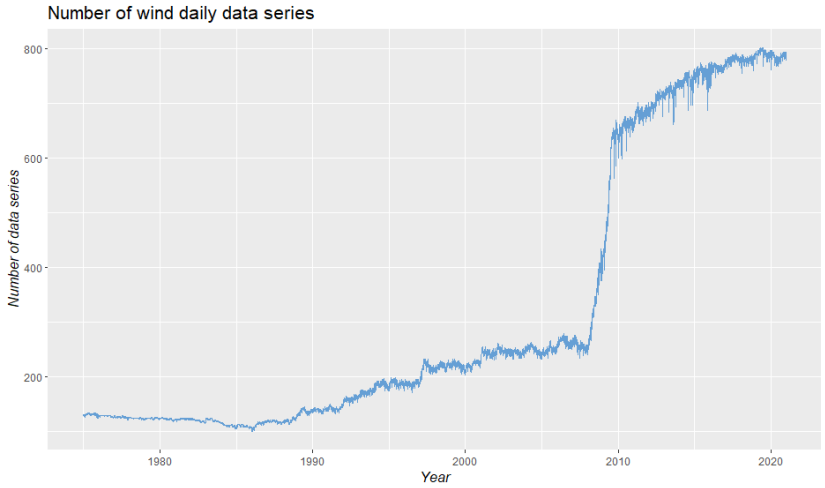


Fig. 11. Number of daily data series for wind.

- Territory is divided into 3 regions.

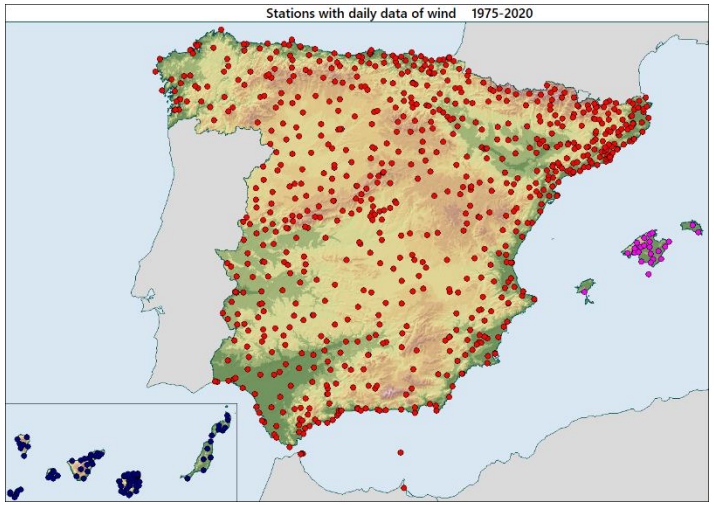


Fig. 12. Distribution of wind stations.

- After homogenizing the grouped monthly data, about 11 data series are rejected.

Procedure

Normal values

- Once all the relevant series have been obtained, the CLINO files can be calculated straightforward using the software **CLINO_tool v.1.5**.
- This CLINO files are requested with a specific structure.
- The CLINO_tool only calculates normals from series having the required minimum of 80% of data in the reference period 1991-2020.

```
World Meteorological Organization Climate Normals for 1991-2020,,,,,,,,,,,,,,,
Single Station Data Sheet For All Climatological Surface Parameters,,,,,,,,,,,,,,,,
,,,,,,,,,,,,,,,,
Station Header Record,,,,,,,,,,,,,,,,
,,,,,,,,,,,,,,,,
Country_Name,Spain,,,,,,,,,,,,,,,
Station_Name,A CORUÑA,,,,,,,,,,,,,,,
,,,,,,,,,,,,,,,,
WMO_Number,Latitude,Longitude,Station_Height,,,,,,,,,,,,,
8001,43|21|57|N,008|25|17|W,58.000000,,,,,,,,,,,,,
,,,,,,,,,,,,,,,,
WMO Integrated Global Observing System (WIGOS) Station Identifier (if available)
0-20000-0-08001,,,,,,,,,,,,,,,,
,,,,,,,,,,,,,,,,
,,,,,,,,,,,,,,,,
Principal Climatological Surface Parameters,,,,,,,,,,,,,,,,
,,,,,,,,,,,,,,,,
,,,,,,,,,,,,,,,,
Parameter_Code,Parameter_Name,Units,,,,,,,,,,,,,,
1,Precipitation_Total,mm,,,,,,,,,,,,,,
,,,,,,,,,,,,,,,,
WMO_Number,Parameter_Code,Calculation_Name,Calculation_Code,January,February,March,April
,May,June,July,August,September,October,November,December,Annual
8001,1,Sum,4,120.8,88.5,87.6,86.9,66.3,45.7,31.2,40.7,57.0,120.7,149.9,122.3,1017.6
```

Fig. 13. Example of CLINO file heading.

Results

Summary of breaks

	No. Series	No. Breaks	% Breaks/Series
Precipitation	6263	1759	28.1
Maximum Temperature	3704	5630	152.0
Maximum Temperature	3704	5577	150.6
Sunshine hours	238	126	52.4
Relative humidity	914	601	65.8
Station level pressure	274	353	128.8
Mean wind speed	900	586	65.1
Maximum wind gust	883	495	56.1

Table 2. Summary of the breaks by variable.

Results

Change in location

- The Fig. 14 shows a graph with the running annual means of **hours of sunshine** at the "Tenerife Sur" airport located in the Canary Islands.
- A break was detected for November 2014.
- According stored metadata, it has been verified that in November 2014, the location station at the airport was changed.
- In this case, this break produces a split in two series, where the last period undergoes an additional correction of more than 0.5 hours in the data prior to the break.

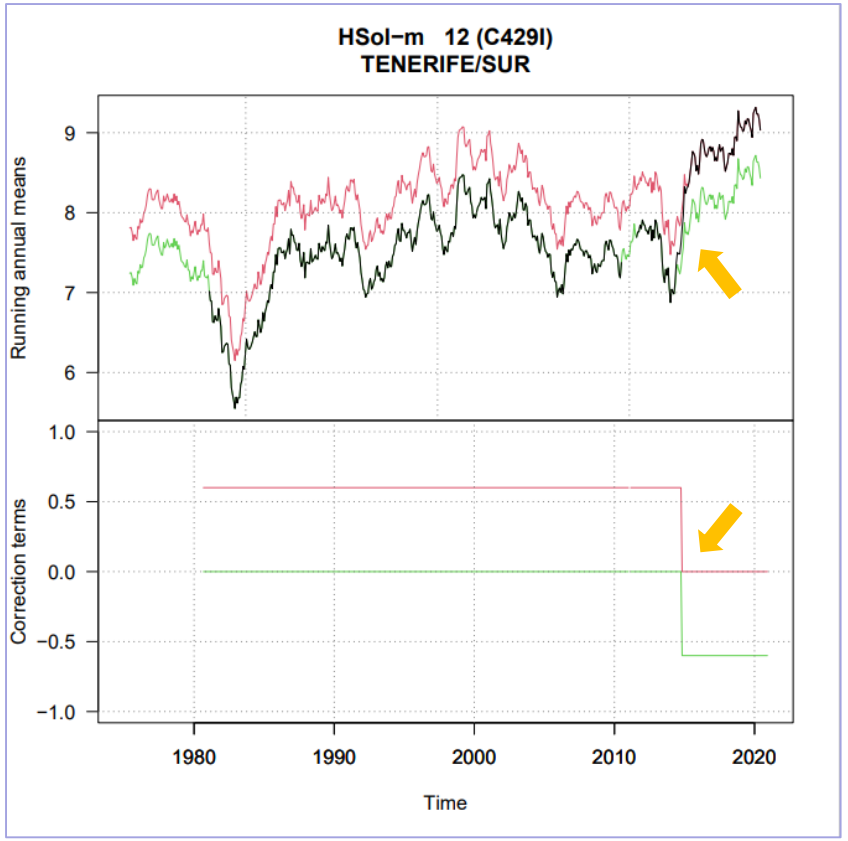


Fig. 14. Running annual means for sunshine in Tenerife Sur with break at November 2014.

Results

Change in location

- The Fig. 15 shows a graph with the running annual totals of **precipitation** at Llinars del Vallès (mainland).
- A break was detected in 1995.
- According to the registered metadata it is described that in the middle of the decade there was a change of location from a garden to an interior terrace 250 m apart.

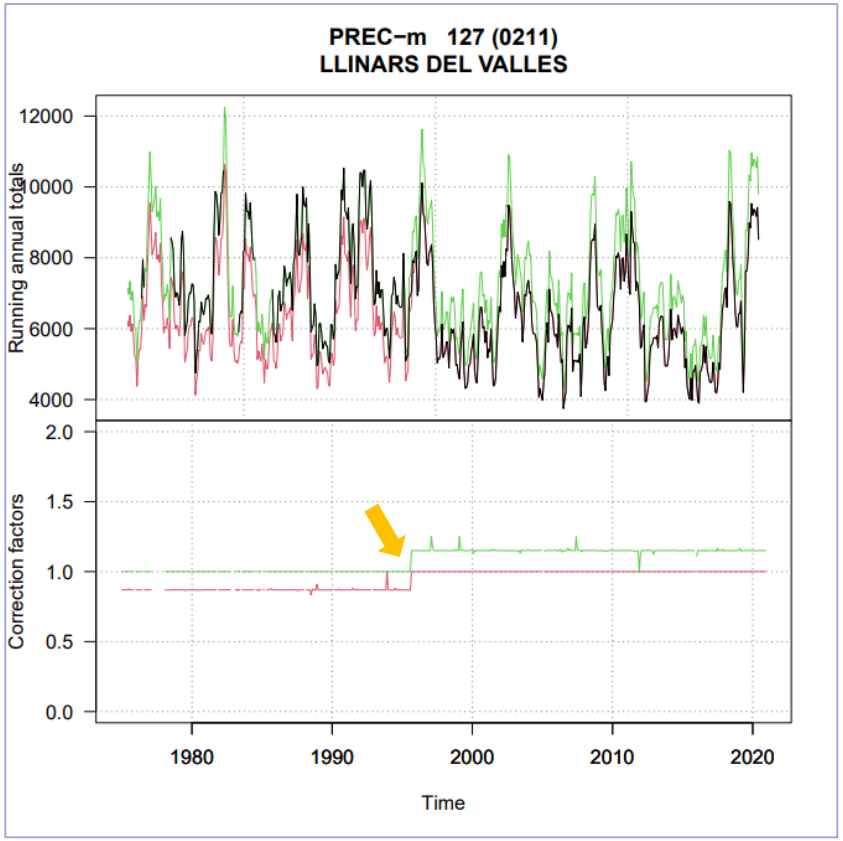


Fig. 15. Running annual totals for precipitation with break at 1995.

Results

Change in shelter

- The Fig. 16 shows a graph with the running annual means of **maximum temperature** at Hervás (mainland).
- It is a THIES automatic station that started operating in 2008.
- A break was detected in 2018.
- According to the records, in 2018, a change of shelter was carried out with a change of height.

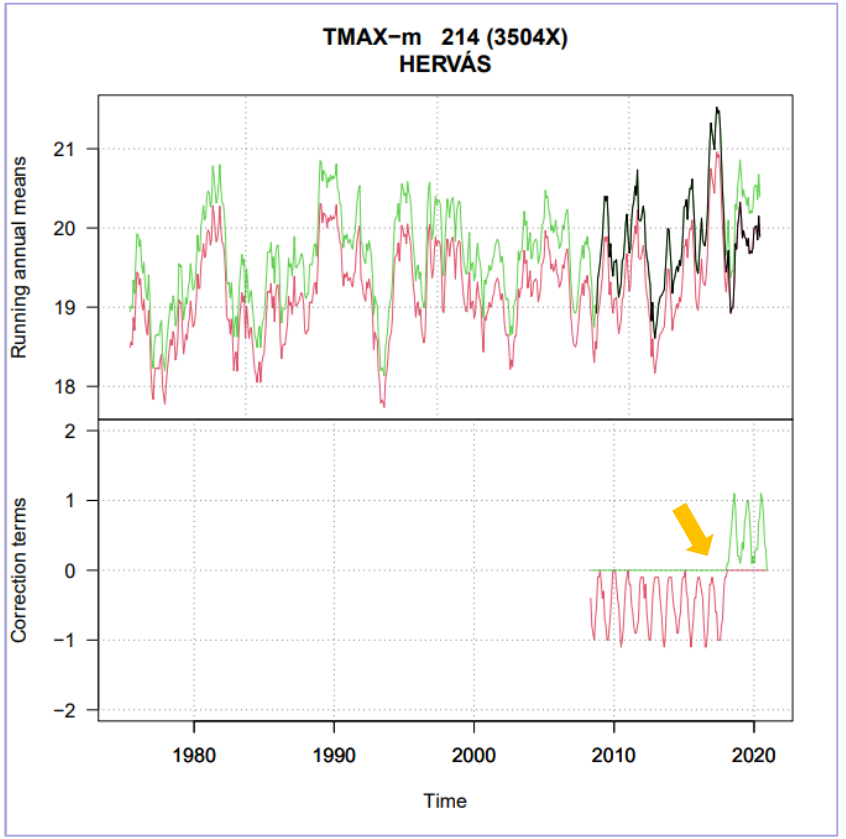


Fig. 16. Running annual means for maximum temperature with break at 2018.

Results

Change in instrumentation

- The Fig. 17 shows how a change of instrumentation could affect to measurements of **wind.**
- A break was detected in the late 2000s.
- It is a SEAC automatic station that at the end of the decade was replaced by a THIES-type station.

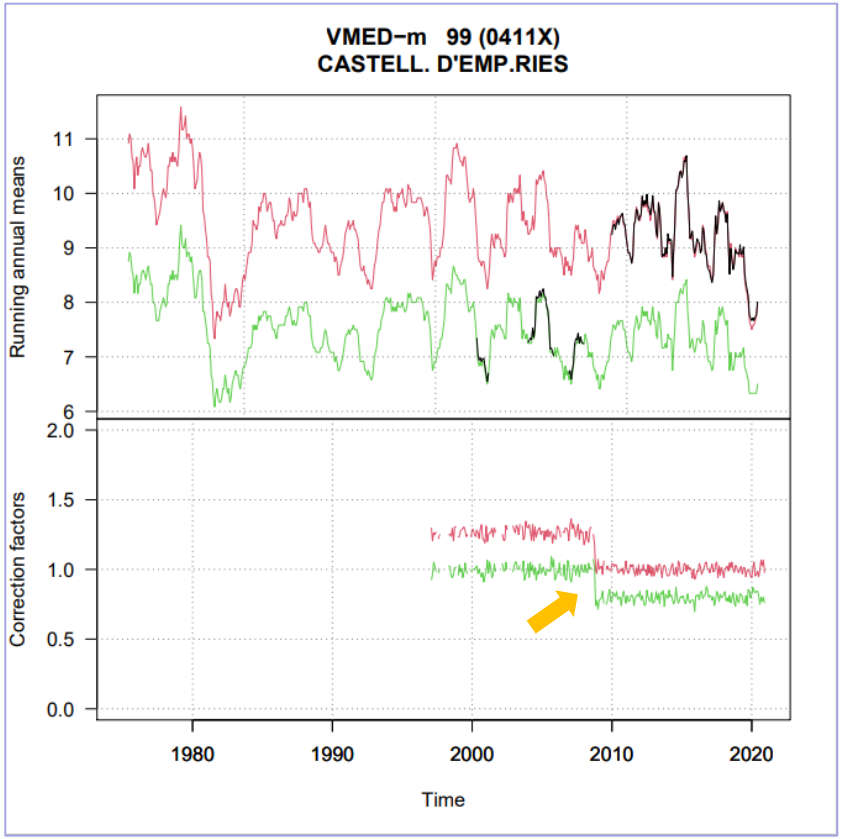


Fig. 17. Running annual means for mean wind speed.

Results

Change in calculation/measurement methods

- The Fig. 18 shows graph with the running annual means for the **station level pressure.**
- A break was detected in 1996.
- In this case, after 1996, the barometric reference was changed to reduce the pressure calculation.
- There is no further recorded metadata that could explain the other break detected in 2008.

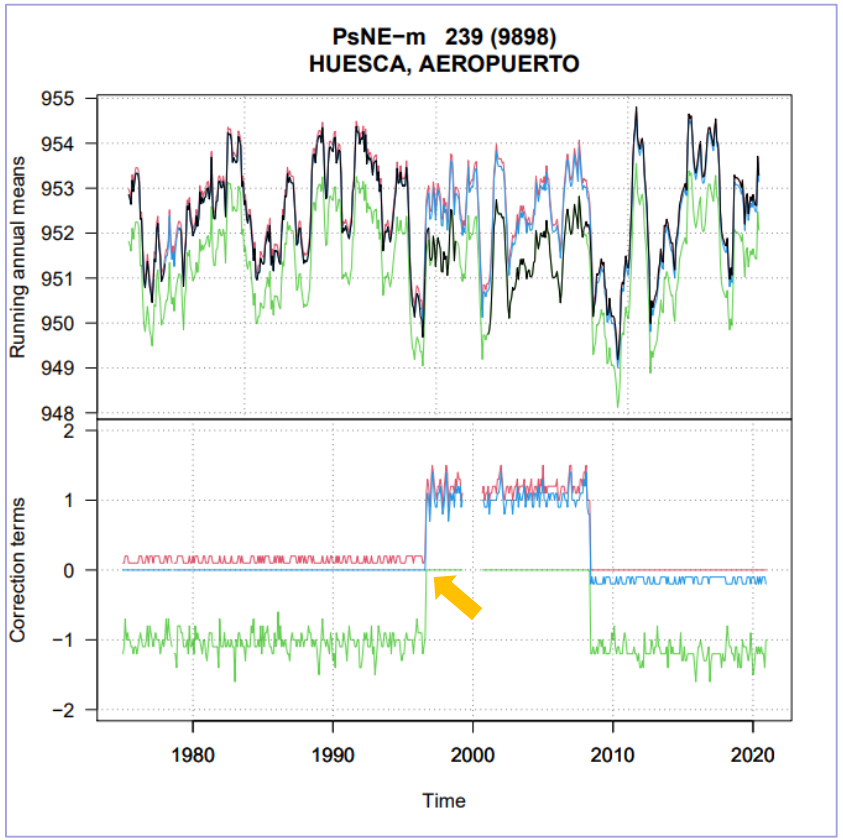


Fig. 18. Running annual means for mean for station pressure.

Conclusions

- Large amount of data and simultaneously running threads ⇨ heavy computational load ⇨ HPC ⇨ Reduced calculation time.
- Climatol has responded well to detecting jumps whose origin appears to be related to non-climatological changes recorded in the metadata.
- Obtaining the normal values from the homogenized series using CLINO_tool has been straightforward.
- As a continuation, the maps corresponding to the climatological normals for the period 1991-2020 will be obtained.
- An interesting perspective for the future could be obtaining daily data grids and maps of variables such as precipitation and temperature from the homogenized data. They could be compared with non-homogenized grids but validated by automatic methods.

Thank you
for your attention!!!