**DATA QUALITY PROGRAMMING TASKS**

completed | need more work

| | | | Function variables | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **Quality Criteria** | **Function name** | **Function description** | Dataframe | Station_ID | Latitude | Longuitude | altitude | station_StartDate | station_EndDate | Date | Variable | Theshold | **Function input** | **Function output** |
| **Dataframe preprocessing** | QA_preprocessing | Generates a dataframe with the required columns and column names for the analysis. It helps to process the next functions with less input variables | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☑ | ☐ | Raw Meteodata dataframe | Dataframe with standar column names and data types. |
| **Length of the time series** | QA_serieslenght_shortlist | Sortlist stations with time series shorter than the stablished thershold and generating a report. | ☑ | ☑ | ☐ | ☐ | ☐ | ☑ | ☑ | ☐ | ☐ | ☑ | Dataframe with meteodata | dataframe and report on txt, Excel, or Rmarkdown |
| | QA_serieslenght_plot | Plotting information from shortlisted stations to evaluate time series lenght | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | ☑ | ☐ | Shortlisted dataframe | visual Plots |
| | QA_serieslenght_clean | Select stations that are going to be removed. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | vector with the station_ID to remove and Meteo dataframe to clean | Clean meteo dataframe |
| **% NA values** | QA_NApc_shortlist | Sortlist stations with % of NA values higher than the stablished thershold and generating a report. | ☑ | ☑ | ☐ | ☐ | ☐ | ☑ | ☑ | ☐ | ☑ | ☑ | Dataframe with meteodata | dataframe and report on txt, Excel, or Rmarkdown |
| | QA_NApc_plot | Plotting information from shortlisted stations to evaluate NA% | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | Shortlisted dataframe | visual Plots |
| | QA_NApc_clean | Select stations that are going to be removed. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | vector with the station_ID to remove and Meteo dataframe to clean | Clean meteo dataframe |
| **Large data gaps** | QA_gaps_shortlist | Sortlist stations with data gaps larger than the stablished thershold and generating a report. | ☑ | ☑ | ☐ | ☐ | ☐ | ☑ | ☑ | ☑ | ☑ | ☑ | Dataframe with meteodata | dataframe and report on txt, Excel, or Rmarkdown |
| | QA_gaps_plot | Plotting information from shortlisted stations to evaluate data gaps | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | Shortlisted dataframe | visual Plots |
| | QA_gaps_clean | Select stations that are going to be removed. | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | vector with the station_ID to remove and Meteo dataframe to clean | Clean meteo dataframe |
| **Miquel/Cristina suggestion** | QA_yearly_filtering | Sortlist stations with data before 2000 with 5 or more years of data and after 2000 with 1 or more years of data and remove stations where all the values are NA (possible when many variables are included in the original df) | ☑ | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | ☑ | ☑ | Dataframe with meteodata | Filtered dataframe |
| **Removal of exteme values** | QA_outlier_precipitation | Sortlist stations with precipitation extreme data// Negative values and based on summary extreme precipitation (1500mm) | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | ☑ | Dataframe with meteodata | Filtered precipitation dataframe |
| | QA_outlier_temperature | Sortlist stations with extreme Temperature // Very low T for the are or very High or values where Tmax< Tmean<Tmin | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | ☑ | Dataframe with meteodata | Filtered temperature dataframe |
| **Outliers** | QA_outlier_variation | Function that helps to evaluate the monthly variability of data for each station: Calculates for each station the monthly mean of the difference with the monthly median value. | ☑ | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | ☑ | ☐ | Dataframe with meteodata | dataframe with median variability and graphs |
| | QA_outlier_shortlist | Sortlist potencial outliers: Shortlist measurement values that are outside a stablished variation range thershold and generating a report. | ☑ | ☑ | ☐ | ☑ | ☐ | ☐ | ☐ | ☑ | ☑ | ☐ | Dataframe with meteodata | Shortlisted dataframe |

| | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | QA_plot_outlier | Plotting information about the outlier in the specific month of the outler station | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | ☑ | ☐ | Shortlisted dataframe | Plots for comparison between close |
| | QA_buffer_stations | Generates a buffer around each outlier point and create a list with closeby stations (at least 2 stations) to the outlier one. | ☑ | ☑ | ☑ | ☑ | ☐ | ☐ | ☐ | ☑ | ☑ | ☑ | Dataframe with meteodata and shortlisted outlier stations | List with outlier stations and closeb |
| | QA_altitude_stations | Generates list with a station with higher and lower altitude for each outlier point and create a list with the outliers and the two stations with close altitude | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | Dataframe with meteodata and shortlisted outlier stations | List with outlier stations and closeb |
| | QA_outlier_clean | Select measurements that are going to be removed. | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | Dataframe with 2 columns: Station_ID and Date of the varable value that should be converted to NA. And Meteo dataframe to clean | Clean meteo dataframe |
| | | | | | | | | | | | | | | |
| | | | | | | | | | | | | | | |
| **Dont know if we need these functions** | QA_reliable_stations | Shortlist stations with reliable data in order to help us assing a threshold # Filter stations with observation counts greater than 240 (at least 20 years data), after 1990 (more reliable) | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | Dataframe with meteodata | Shortlisted dataframe |
| | QA_Stations_sampling | Samples in the reliable stations to select diverse stations in the study area, picking point with large distances between X, Y,Altitude | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | Shortlisted Reliable dataframe | 6 points |
| | QA_plot_yearly | Plotting for selected stations of a variable in months | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | ☑ | ☑ | Shortlisted Reliable dataframe | Plots |
| | QA_plot_yearly_meansd | Plotting for selected stations of a variable in months INCLUDES mean od the variable and 2sd | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | ☑ | ☑ | Dataframe | Plots |
| | QA_obs_plot | Plotting observation of a variable counts by number of stations through the years | ☑ | ☐ | ☐ | ☐ | ☐ | ☐ | ☐ | ☑ | ☑ | ☑ | Dataframe with meteodata | Plot with the Number of variable´s observation by number of Stations |