**WaPOR Precipitation Validation Exercise (Iraq)**

In this exercise, we will explore and analyse precipitation data from one gauging station in Iraq obtained from the [GHCN database](https://www.ncei.noaa.gov/products/land-based-station/global-historical-climatology-network-daily), namely IZ000040665 (Kut-Al-Hai), and compare it with corresponding data from WaPOR portal. The objective is to evaluate the performance of the WaPOR dataset in capturing precipitation patterns at this specific location.

**1. Prepare the daily Precipitation data:**

* Obtain the Precipitation data (daily) for station IZ000040665 from the OCW platform. Open file in excel and under “data” tab convert the table from text to column (delimited, comma separated). Create new excel file with the date and precipitation data for the period 2012-2023 (copy also quality flag column). Check correct unit of Precipitation (HINT check [read me file](https://www1.ncdc.noaa.gov/pub/data/ghcn/daily/readme.txt)), apply correction factor if relevant.
* Check the GHCN documentation for quality flags. What do the quality flags mean? Identify bad quality data points if any and remove them from the analyses (leave them blanc).
* Investigate the data and identify the location of the station (lat, long). HINT: check [GHCN documentation](https://www.ncei.noaa.gov/data/global-historical-climatology-network-daily/doc/GHCND_documentation.pdf) (Section A).
* Access the WaPOR portal and download the daily precipitation (point time series - Analysis Tool) for the mentioned station for the period of 2012 to 2023. Open file in excel and under “data” tab convert the table from text to column (delimited, comma separated).
* Organize the two precipitation datasets in a single excel sheet for comparative analysis for the period 2012-2023. Ensure that the dates in both datasets match accurately. Structure the columns in the following format: Column A: Date, Column B: Station data, Column C: WaPOR data. Add the metadata for the station (above the timeseries) to remember which station the analyses are for:



**2. Visualize daily comparison:**

* Create and compare the time series plots of station data and WaPOR data to visually assess their patterns and trends over time.
* Create scatter plot to visualize the relationship between the station data and WaPOR data. Include the equations for y and R-squared value.
* Format the graphs by adding axis labels, a chart title, and legend

**3. Calculate statistical metrics (see table in the annex with the formulas):**

1. Pearson’s correlation coefficient (CC): =ROUND(PEARSON(B6:B179;C6:C179);2)
2. Coefficient of determination (R2): =ROUND(PEARSON(B6:B179;C6:C179)^2;2)
3. Nash-Sutcliffe Efficiency (NSE): =ROUND(fit\_NSE(B6:B179;C6:C179);2)
4. Kling-Gupta Efficiency (KGE): =ROUND(fit\_KGE(B6:B179;C6:C179);2)
5. Bias: =ROUND(fit\_mbe(B6:B179;C6:C179);2)
6. Root mean square error (RMSE): =ROUND(fit\_rmse(B6:B179;C6:C179);2)
7. Mean Absolute error (MAE): =ROUND(fit\_mae(B6:B179;C6:C179);2)

**4. Prepare the monthly Precipitation data**

* Create Pivot table, select all data in the three columns, under “Insert” tab select pivot table (select OK for creating pivot table in new sheet). Select the following fields in the pivot table:



* Expand all the years (press on the “plus” sign)
* Create the monthly timeseries in a new table using following equation:

|  |  |  |  |
| --- | --- | --- | --- |
| A1 | B1 | C1 | D1 |
| Jan/2012 | =month(A1) | =year(A1) | =GETPIVOTDATA("Sum of Precipitation";$A$3;"Date";B1; "Years";C1) |
| Feb/2012 | =month (A2) | .. | .. |
| Mar/2012 | .. | .. | .. |
| \* “Precipitation” is the heading of the column |

* Repeat step 2 and 3 for the monthly data

**After doing the steps, reflect on these questions:**

1. Considering the duration of the reference dataset, do you believe it is extensive enough to enable robust statistical analysis?
2. From the obtained findings, what can be inferred about the accuracy of the WaPOR precipitation data? Would you have expected a 100% match? If not why not?
3. How do the daily and monthly evaluations compare, which one performs better and can you explain why?
4. Based on the Pearson's correlation coefficient (CC), how would you characterize the relationship between the station data and the WaPOR data? Is the correlation strong, weak, or negligible?

Annex

****