

Data Scientist recruitment test

Objectives of the test

This test has been defined for the candidates to demonstrate :

- their skills and knowledge about weather data handling,
- their skills to implement simple processing on weather data,
- their skills to present a scientific analysis of these data.

Usecase

The use case is the analysis of the variability of the onset date for the wet season in Angola.

The candidates are invited to present:

- A data analysis of the **spatio-temporal variability of the onset date** on the 1981-2010 period on a selected domain over Angola (see below for data set and onset computation method). This analysis is expected to give information about the spatial patterns of the onset date: what is the spatial organization of this onset date? is there a spatial gradient in any direction? Are there some locations in the domain that have significant behaviour from others? The candidates are invited to choose any method they think the most suitable.
- A comparison of the temporal variability of the onset date for the two cities Huambo and Malanje included the selected domain: what is the temporal distribution of the onset date for these 2 locations?
- the computing code they developed that should be shared on **github repository** (for example jupyter-notebook and dependencies).

Information about data and method to use:

- onset date computation: Liebmann's method ([Liebmann & Marengo, 2001](https://archive.eol.ucar.edu/projects/cppa/meetings/200608/posters/brant.pdf); <https://archive.eol.ucar.edu/projects/cppa/meetings/200608/posters/brant.pdf>)

Méthode de Liebmann (2006)

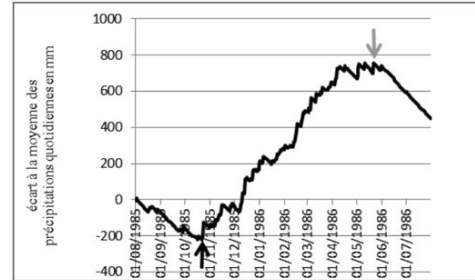
Liebmann propose une méthode basée sur la définition d'un indice appelé « **Anomalous Accumulation** » (AA) qui consiste à calculer la différence entre l'accumulation de précipitations journalières et une accumulation théorique moyenne pendant laquelle il précipiterait quotidiennement le taux journalier moyen, d'où :

$$AA(t) = \sum_{n=1}^t R(n) - \bar{R} \times t$$

- AA(t) : l'**Anomalous Accumulation** au jour t
- \bar{R} : moyenne de précipitations journalières = cumul annuel de précipitations / nb de jours dans l'année
- R(n) : précipitations au jour n

La valeur la plus basse (haute) rencontrée est le minimum (maximum) de pluie accumulée dans l'année qui correspond au début (à la fin) de la saison le jour suivant (inflexion positive/négative de la courbe).

Le début de la saison des pluies est alors déterminé par la date de minimum de la courbe d'« **anomalous accumulation** ».



Méthode de Liebmann pour la station de Rio Branco, Acre, Brésil. Les écarts à la moyenne mettent en évidence la progression et le rythme des saisons. La flèche noire identifie le début de la saison, à la valeur la plus basse de la courbe, tandis que la flèche grise indique la fin de saison avec la valeur la plus haute de la courbe. Auteur : V. Michot ; Sources : Hydroweb Agência Nacional de Água

- data:
 - ERA5-Land
<https://cds.climate.copernicus.eu/cdsapp#!/dataset/reanalysis-era5-land?tab=form>; data are available freely for manual download or via API documented on the Copernicus Climate Data Store website and download page.
 - **study domain**: rectangular domain [15°E; 17°E] and [-13°N; -9°N] (~30Mo per year of data for total precipitation) including Malanje (-9.55°N; 16.34°E) and Huambo (-12.77°N; 15.73°E).
 - **time period**: 1981-2010