

# **Prominent discords in climate data through matrix profile techniques: detecting emerging long term pattern changes and anomalous events**

**Hussein El Khansa**

**Prof. Carmen Gervet**

**Audrey Brouillet**

# Introduction

One of the main application of climate data analysis is outlier detection (anomalous behaviour detection)

Climate data outliers can be the source of important human and environment impacts

Examples of these impacts:

- Wildfires
- Droughts
- Floods

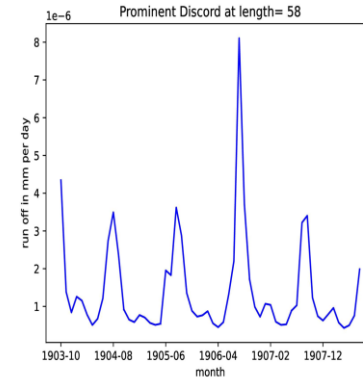
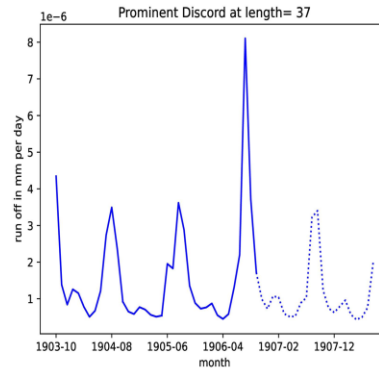
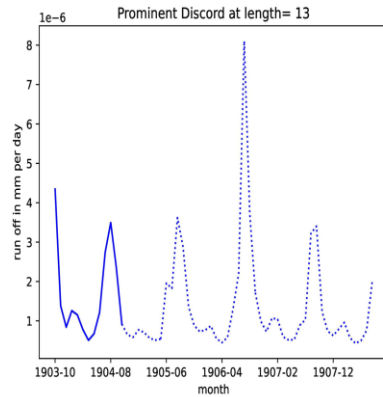
## **Goal**

Detecting insightful discords in  
climate data



# A new concept : Prominent discord

A prominent discord is a discord that subsumes multiple variable length discords



# Matrix Profile

- We are interested in detecting pattern anomaly that is able highlight an anomaly based on duration.
- We chose to use Matrix profile as our pattern anomaly detection algorithm because it is:
  - exact
  - parameter-free
  - Does not require the user to set similarity/distance threshold.

0	1	3	2	9	1	14	15	1	2	2	10	7
---	---	---	---	---	---	----	----	---	---	---	----	---

0	1	3	2	9	1	14	15	1	2	2	10	7
---	---	---	---	---	---	----	----	---	---	---	----	---

0	1	3	2	9	1	14	15	1	2	2	10	7
---	---	---	---	---	---	----	----	---	---	---	----	---

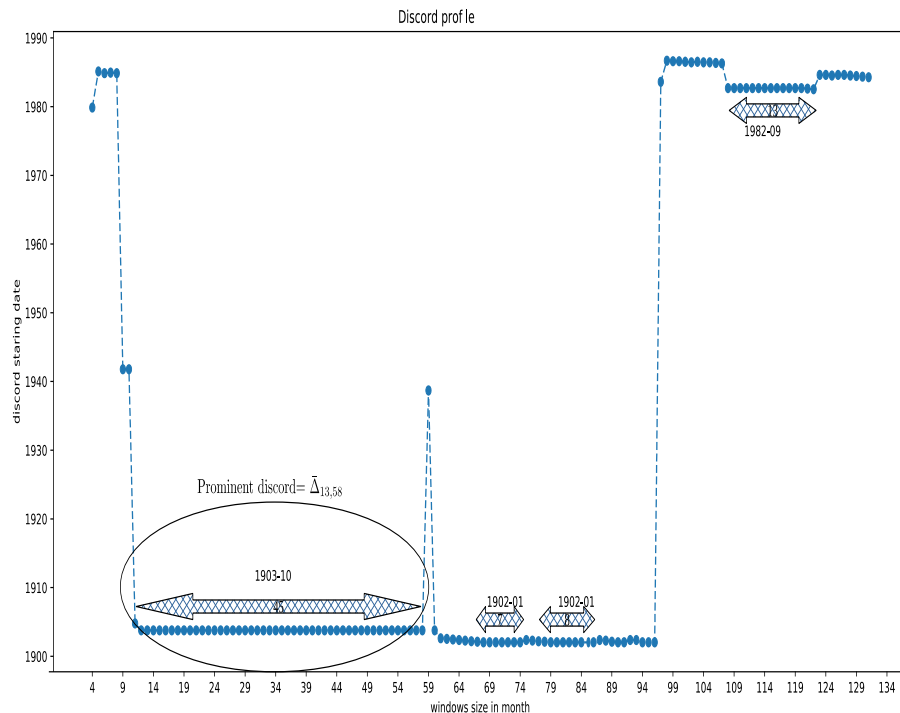
Distance Profile	0	7.4	6.9	14.7	19.3	17.7	19.9	15	8.2	8.9
------------------	---	-----	-----	------	------	------	------	----	-----	-----

0	7.4	6.9	14.7	19.3	17.7	19.9	15	8.2	8.9	6.9
7.4	0	10.9	7.9	15.7	18.8	19.1	158.8	1.4	8.4	14
6.9	10.9	0	16.8	16.1	13.6	18.8	14	11.6	6.2	6.2
14.7	7.9	16.8	0	16.8	19.8	18	19.4	8.2	13.4	7.9
19.3	15.7	16.1	16.8	0	20.7	23.6	18.7	15.3	14.4	11.4
17.7	18.8	13.6	19.8	20.7	0	19.2	23.1	19.8	1.4	13.6
19.9	19.1	18.8	18	23.6	19.2	0	14.1	20.1	20.5	14.1
15	15.8	14	19.4	18.7	23.1	14.1	0	16.2	16.1	14
8.2	1.4	11.6	8.2	15.3	19.8	20.1	16.2	0	8.6	1.4
8.9	8.4	6.2	13.4	11.4	14.4	20.5	16.1	8.6	0	6.2

Matrix Profile

# Our approach

- Discord profile:
  - Run matrix profile on different length.
  - Select the top discord from each matrix profile
  - Store them in vector in order.
- Group discord that share consecutive starting date
- Discord are ranked based on the number of subsumed discord(descending order)

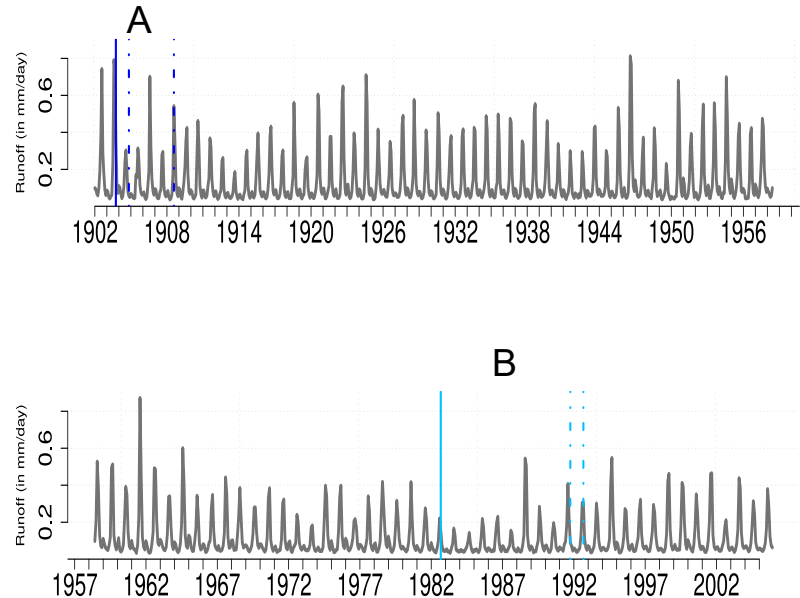


# Results

Two major outliers were detected :

**A** : matches the actual period from **1903-10** till **1908-08** which has the higher soil water amount and larger inter-annual variability, (occurring before a continuous long-term drying trend)

**B** : matches temporal pattern of the most severe drought from **1982-09** till **1992-09** ever recorded over the time series



Data: west Africa monthly runoff  
from 1902 till 2005

# Future Work

Exploring alternative ranking methods and concepts for discords

Idea: Integrating pointwise anomaly detection algorithm, with Matrix profile to create an ordering for discord based on the anomaly ratio

**Thank you very much  
for your attention**