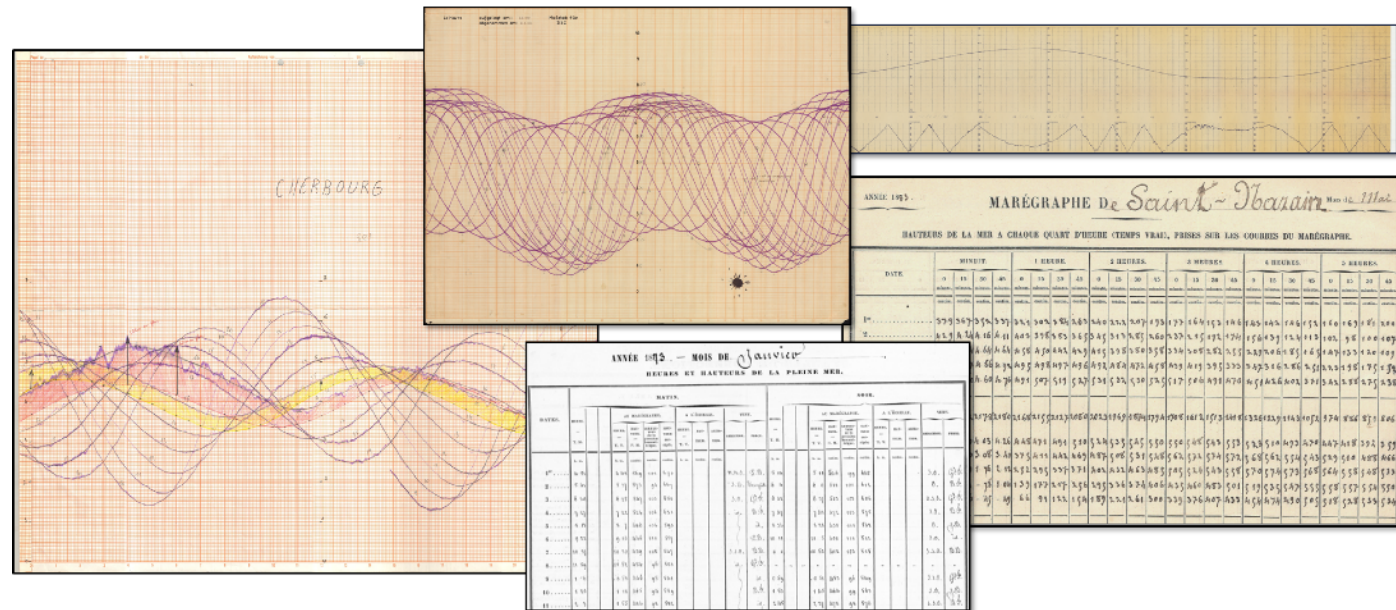


## Data Archeology

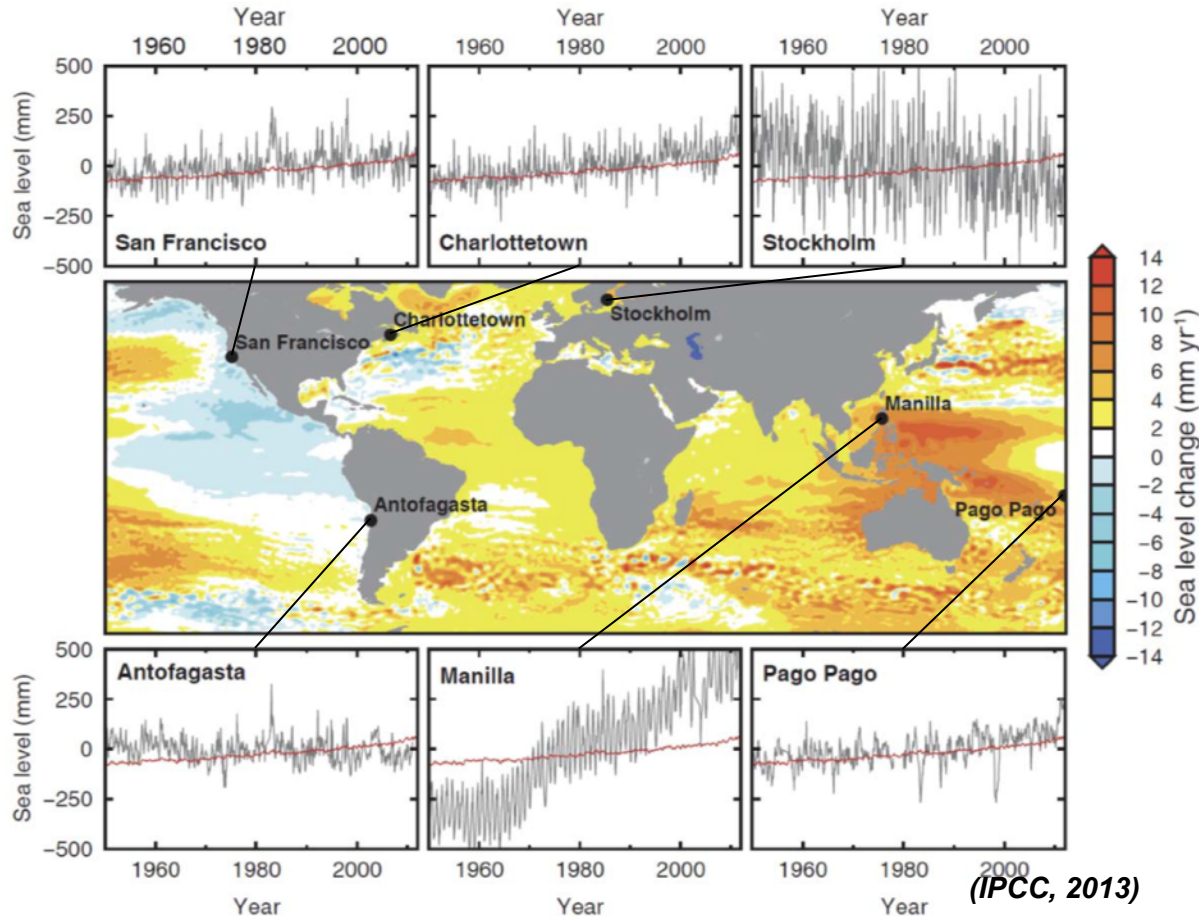


Y. Ferret (Shom)

V. Donato (Shom)

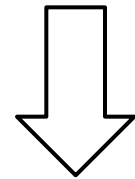
# BACKGROUND INTRODUCTION

Changes in temperatures (Ocean/atmosphere), climate patterns, ...  
**Affects Sea level evolution**



(IPCC, 2013)

Combined with modern  
satellite observations,  
historical measurements allow  
to assess secular changes in  
sea level



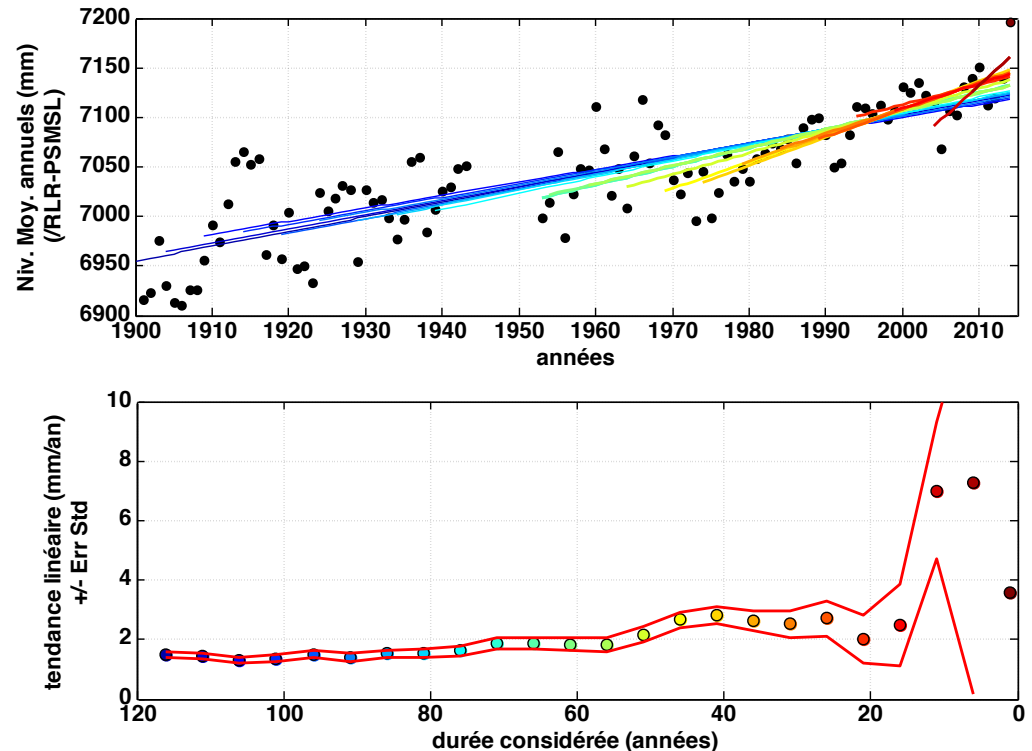
**Long and most continuous as  
possible sea level  
measurements are still needed**

# BACKGROUND INTRODUCTION

The estimation of the sea level rise varies according to the time series length, notably because of the influence of large scale atmospheric effects (ex: NINO, NAO, ...)

It has been estimated that 60 years of measurements are necessary to filter those effects (Douglas, 1991)

*Yearly mean sea levels, Brest since 1900:  
Trends depending on timeseries length*



## **OBJECTIVES OF DATA RESCUE**

---

- Rescuing legacy data
- Improving statistical modelling of extreme events
- Long term sea-level trends

# GENERAL STRATEGY

Time demanding work including different steps:

## FROM PAPER DOCUMENTS ...

## ... TO DIGITAL SEA LEVEL DATA

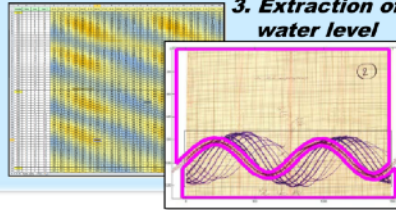
### 1. Inventory and recovery



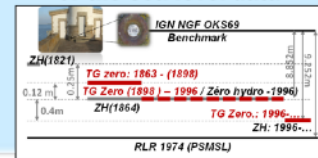
### 2. Scanning process



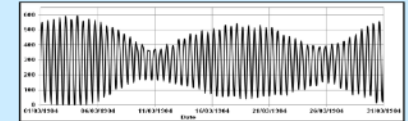
### 3. Extraction of water level



### 4. Vertical consistency of the sea level data



### 5. Quality check, data validation



→ **several type of documents to inventory**  
(tables, marigrams, metadata relative to measurements, ...)

Potentially complex depending on:

- **Size of the document**
- **Preservation state**

**Handwritten ledgers** manually digitised into tables  
**Marigrams** « automatically » digitised with the use of the NUNIEAU software (Pons, 2008) based on color recognition

→ **DIGITAL RAW DATA**

Based on the analysis of metadata linked to sea level observations:  
(levelling reports, technical notes, ...)

→ **Tide observatory history over times**

**Quality check**  
(spikes, consistency check, ...)

**Buddy checking**  
(comparison with sea level time series of nearby stations)

**Data flagging**  
depending on the data quality (Good/probably good/probably bad/bad)

# STATE OF THE INVENTORY OF SHOM ARCHIVES

## INVENTAIRE EN LIGNE



### Synthèse de l'inventaire des documents marégraphiques historiques : archives du Shom



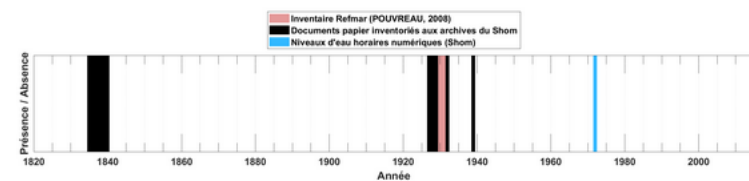
Le tableau ci-dessous reprend les sites pour lesquels le Shom dispose de documents marégraphiques papiers et dont la durée totale cumulée d'observations est supérieure à 1 an.

Pour chacun de ces ports, une synthèse a été réalisée, et il est possible d'accéder au détail des documents inventoriés en cliquant sur le nom du port souhaité.

(Les ports sont listés par ordre alphabétique : Ports français, d'*outre-mer*, puis *étrangers*.)

La totalité de l'inventaire des documents marégraphiques papiers conservés aux archives du Shom en cours de réalisation est téléchargeable [ici](#) (observations de courtes durées incluses).

Explication des frises temporelles synthétiques du tableau de synthèse.



[FIN DE PAGE](#)

PORT	PAYS	Année initiale	Année finale	Durée effective (ans)	Nombre de documents (Pourc. scanné)	Couverture temporelle synthétique
<a href="#">ABER WRACH</a>	FRANCE	1835	1939	12	86 (20 %)	
<a href="#">ADOUR</a>	FRANCE	1899	1904	6	298 (0 %)	
<a href="#">AJACCIO</a>	FRANCE	1978	1988	6	195 (100 %)	
<a href="#">ANTIFER</a>	FRANCE	1974	1986	10	301 (17 %)	

<http://refmar.shom.fr/dataRescue>



# STATE OF THE INVENTORY OF SHOM ARCHIVES

- More than **50.000 documents identified** and accurately inventoried
- **about 50 %** have already been **scanned**
- **But ... Still thousands of documents to carefully inventory/scan !**

## GEOGRAPHICAL EXTENT

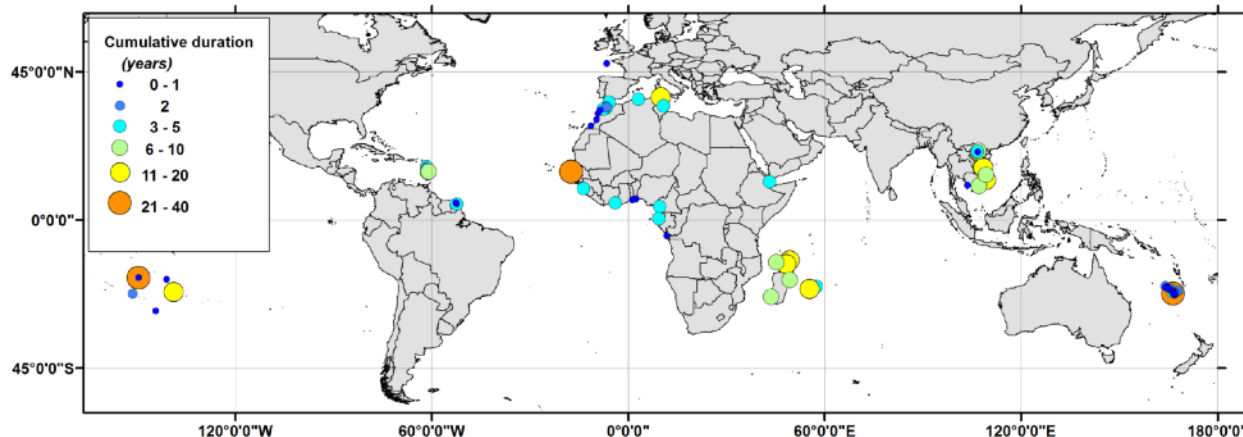
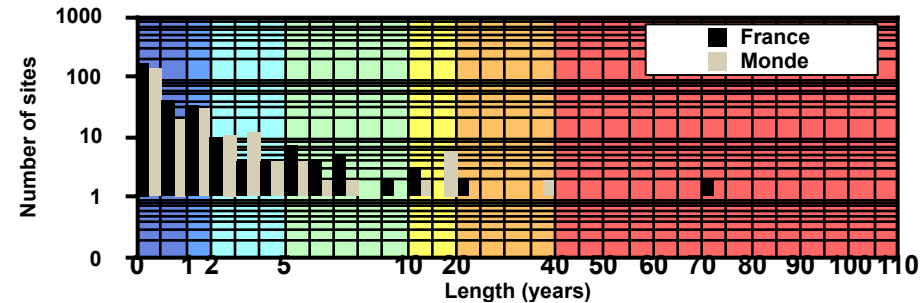
### **France (Fig. 3):**

- about 1.000 years of cumulated sea level measurements, ~ 300 sites
- Total duration per site ranging for few days/weeks/months (observations for sounding reduction purpose) to several decades
- Longest time series correspond to historical tide gauge network (Fig. 2)

### **Around the world (Fig. 3 & 4):**

- about 470 years of cumulated sea level measurements, ~ 240 sites
- Mainly short duration observations, but some « long » time series (fig. 4)

**Figure 3: Number of locations with sea level observations according to the total duration of the measurements from Shom archives**



**Figure 4: Partial view of the spatial distribution of the sea level data around the world, from Shom archives**

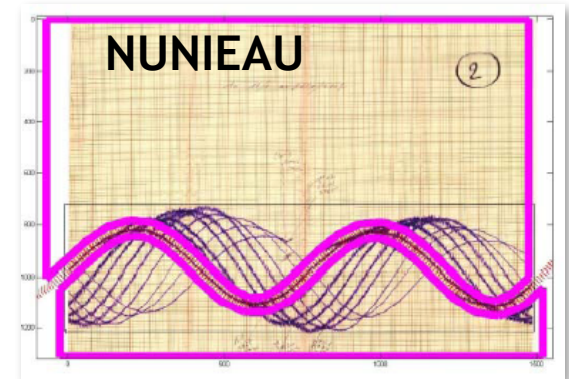
# DIGITALIZATION WORKS

## Scanning

## Vectorisation / retranscription (mostly manual)

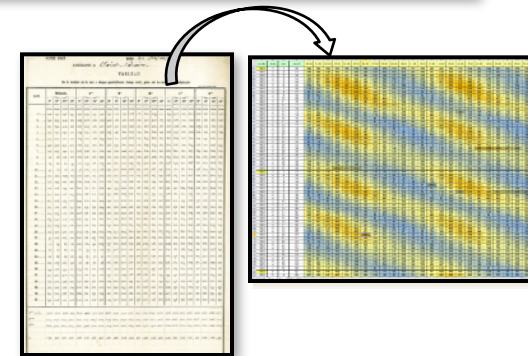
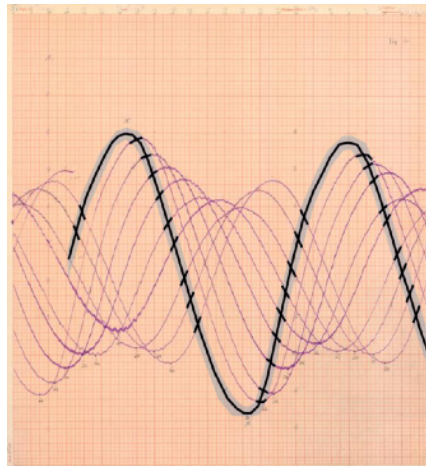
### Challenge :

- Volume of data
- Degradation of paper. Size of documents
- Important time of vectorisation
- ...



Way of improving vectorisation process

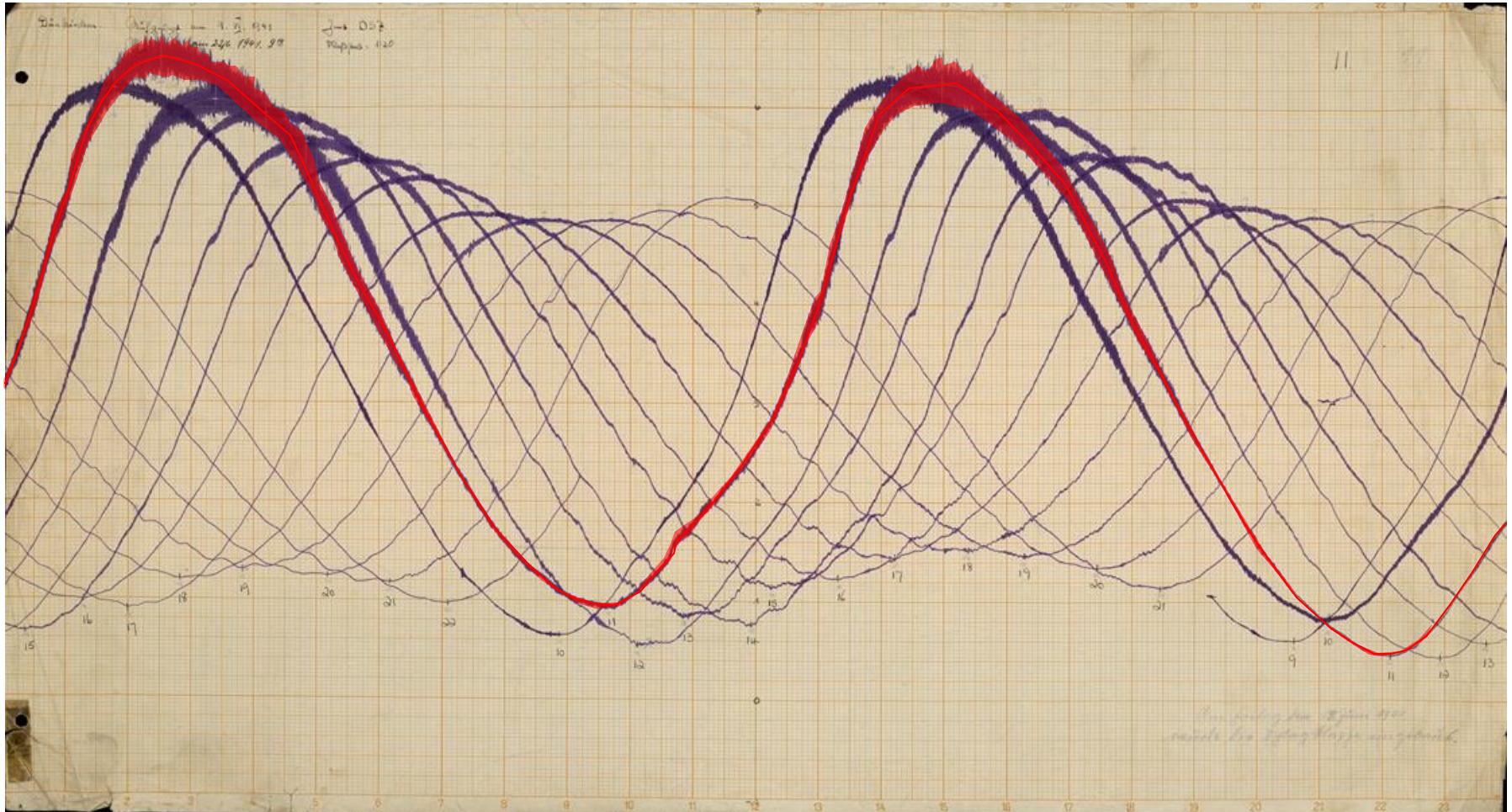
?





## Problem with high-frequency curves

→ What information do we want to retain ?  
(seiche, tsunami, ...)

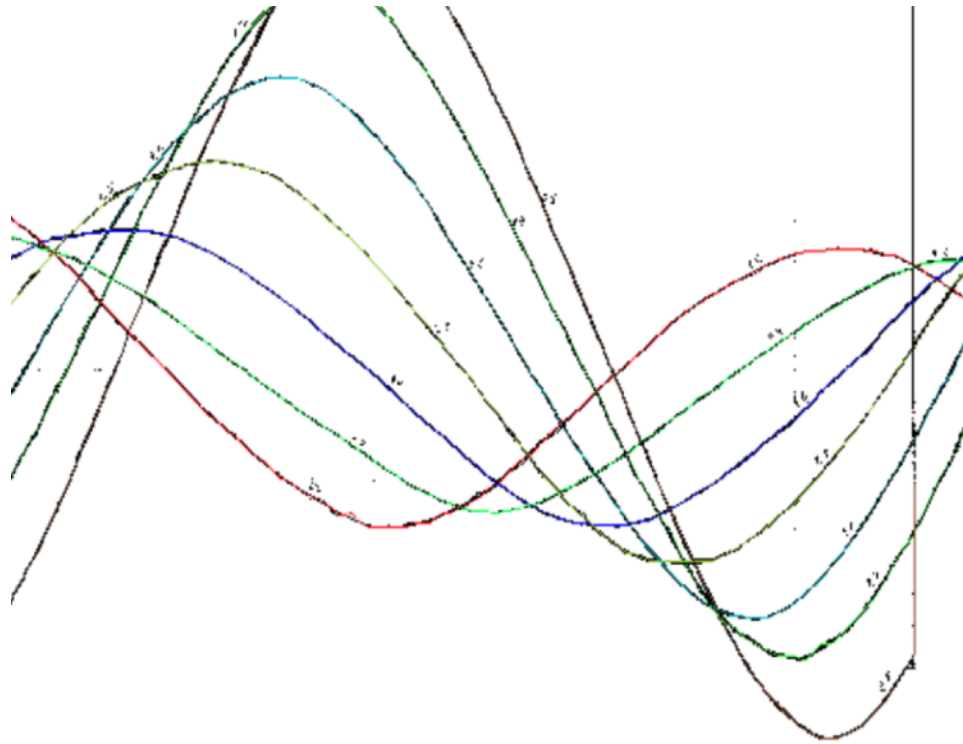
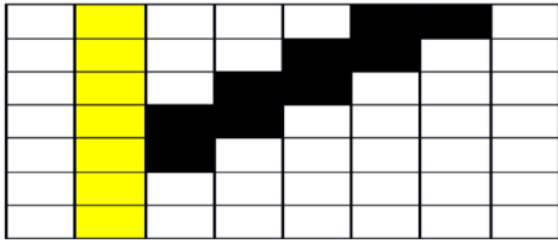


- M2 student project
- Image processing approach

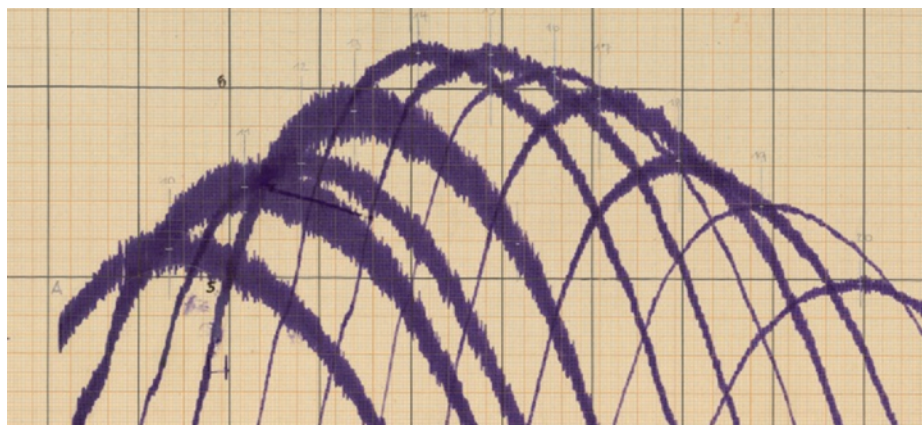


# IMPROVING VECTORISATION

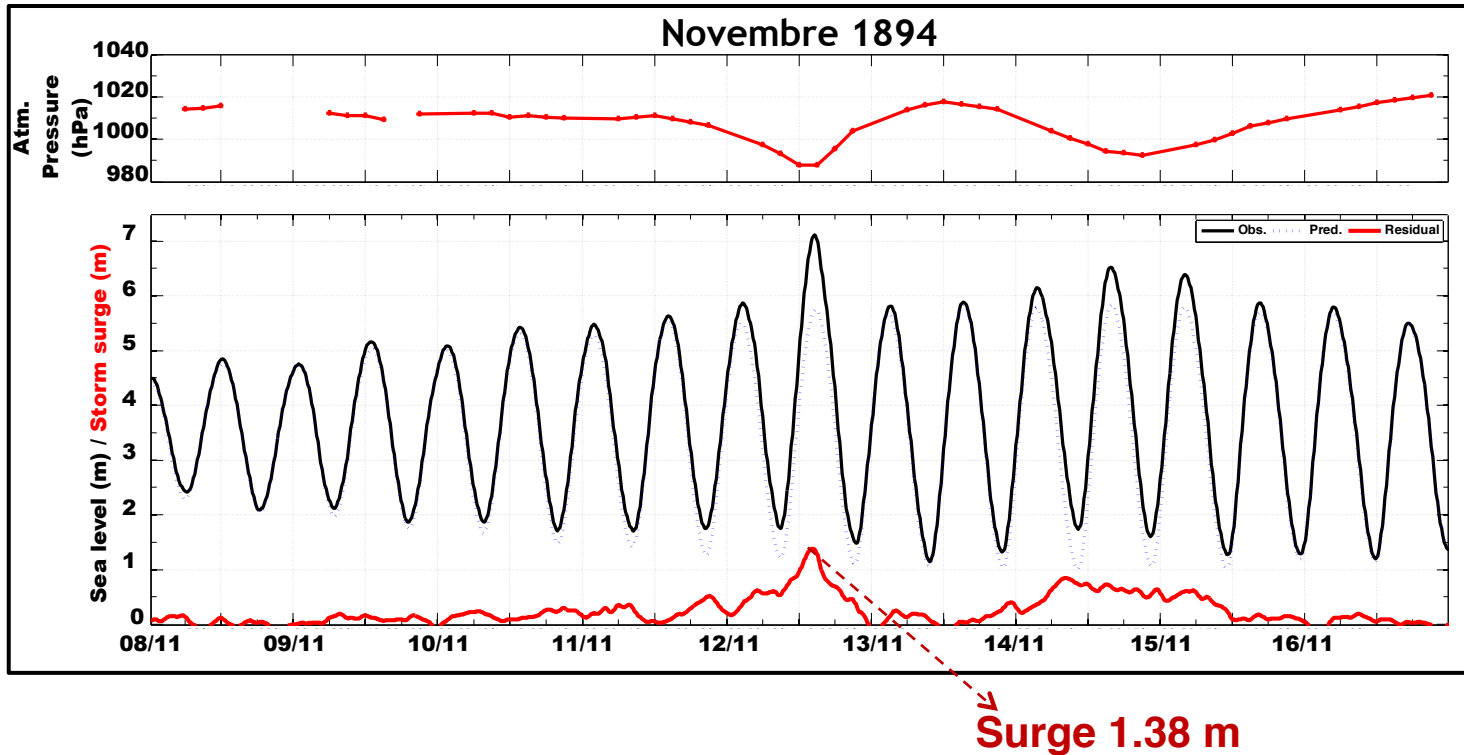
- Background pattern removing (stacking)
- Automatic curve detection and tracking



- To be continued
- Intelligent background removing (machine learning)
- Dealing with complex scenarios
  - High frequencies
  - Background and curves same colors



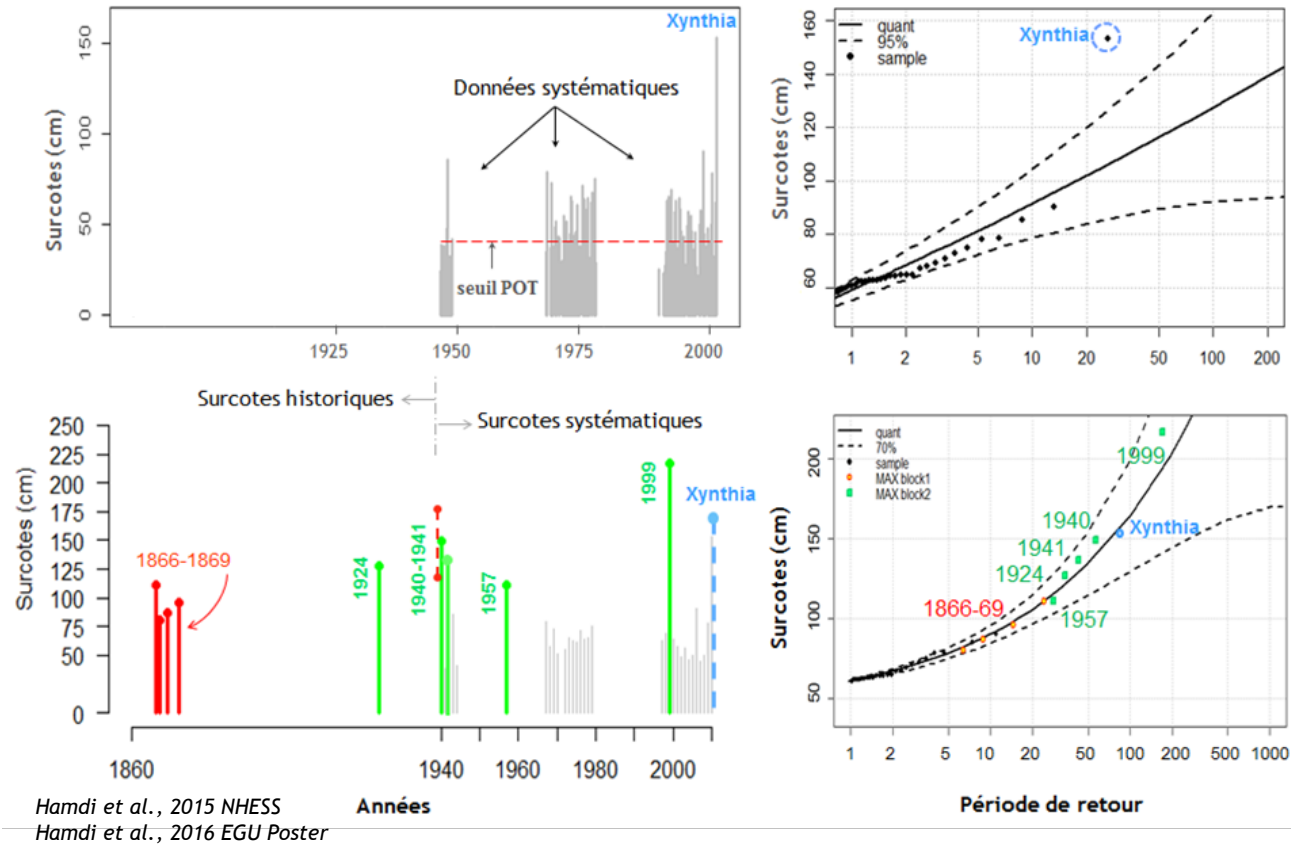
# EXTREME EVENTS STATISTICS



Short-time series may provide valuable information for extreme events knowledge



# EXTREME EVENTS STATISTICS



Outlier events may turn out to be not so exceptional, taking into account recovered events



# FUNDINGS

---

Most public fundings are related to extreme events recovering  
Risk management more appealing than (slow) sea-level rise

- Ministry of Civil protection
- Ministry of Ecology (Littoral risk management)

4 ongoing data rescue projects at Shom fuelled by extreme events statistics  
3 funded by national to regional bodies  
1 by EU project

1 permanent expert payed by ministry of Ecology



## DISCUSSION

---

- First important step is to inventory data
- Requires national coordination and ongoing effort
- International ?
- <http://refmar.shom.fr/dataRescue>