

# Etude Comparative de méthodes d'estimation fréquentielle de surcotes extrêmes

Yasser Hamdi, Claire-Marie Duluc, Lise Bardet, Vincent Rebour  
Institut de Radioprotection et de Sûreté Nucléaire



# Plan de la présentation

- Problématique
- La méthodologie générale - l'analyse fréquentielle
- Les approches proposées
- Extraction des observations extrêmes (+tests statistiques)
- Résultats & discussion
- Conclusion & perspectives

# La problématique

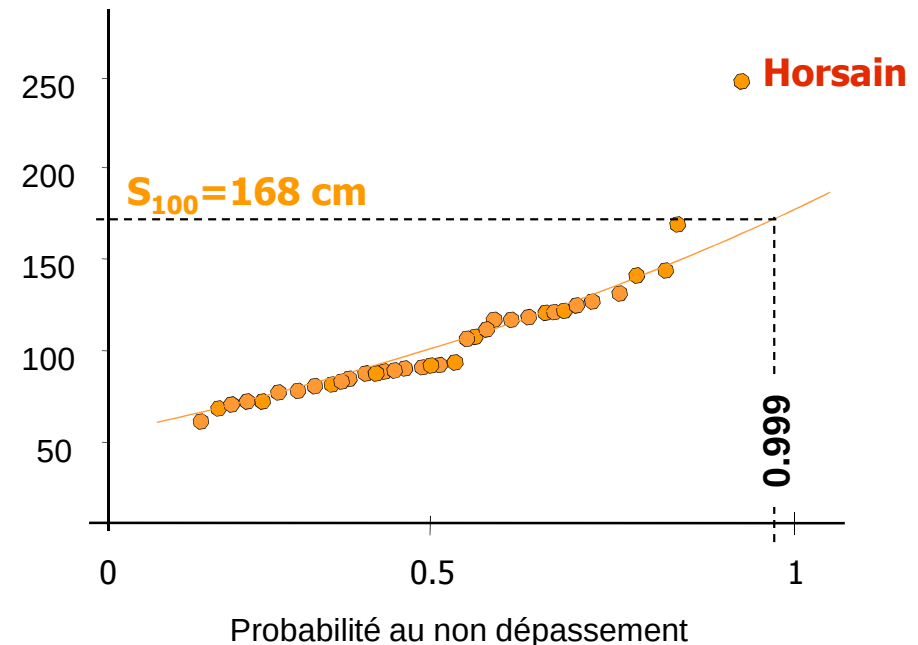
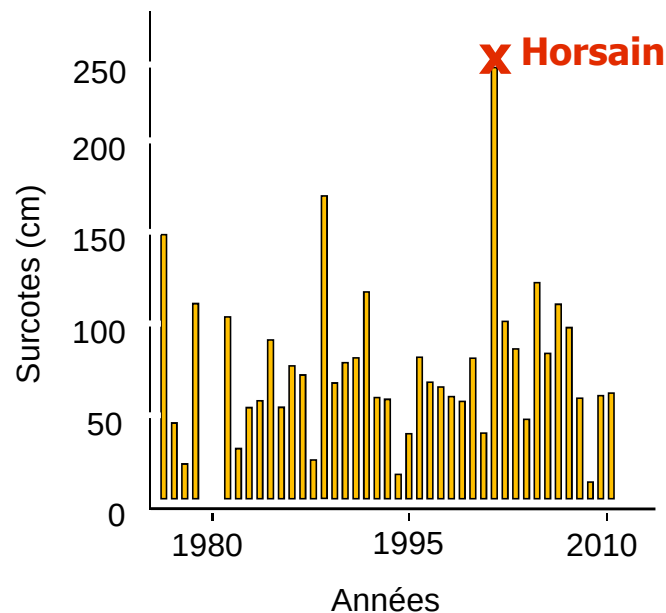
- Tempêtes de 1953, 1987, Martin (1999), Klaus (2009), Xynthia (2010) . . .
- Modélisation statistique des surcotes extrêmes (une analyse fréquentielle permet d'identifier le comportement statistique des événements;
- Plusieurs méthodes existent! quelle est la plus adéquate ?
- Modélisation très sensible à la qualité des données !!



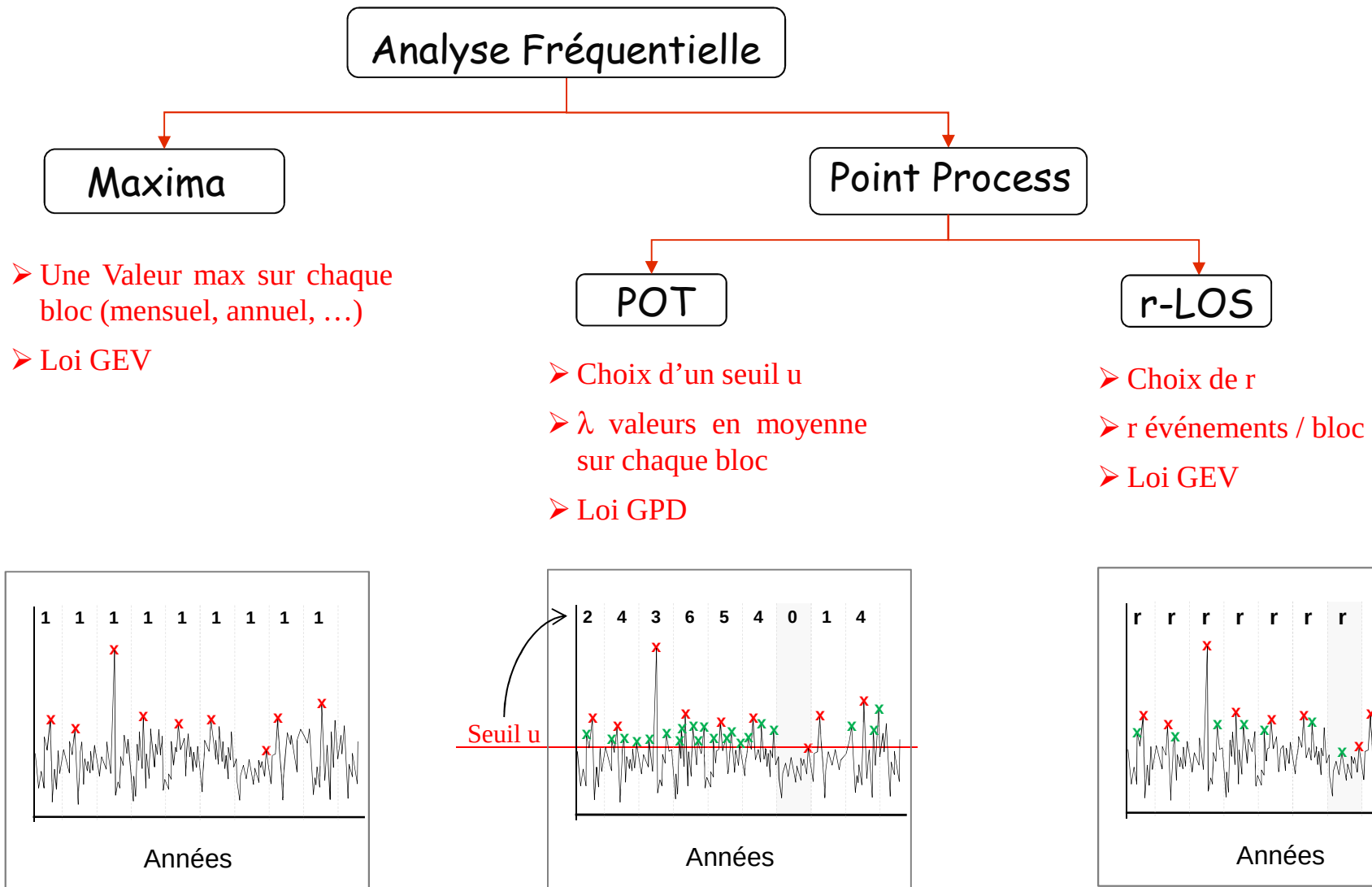
- Séries courtes (périodes de retour élevées!); données manquantes; présence d'horsains;
- Populations mixtes (experts!);
- Qu'est ce qu'une valeur extrême? Quelles sont les observations qui peuvent être considérées comme extrêmes ?

# Méthodologie générale

1. Méthode → extraction des données (stationnarité, homogénéité & indépendance)
2. Probabilités empiriques (plotting position)
3. Ajustement avec une fonction de distribution (EVT)
4. Comparaison des méthodes:
  - I. Incertitude sur les paramètres et les niveaux de retour (erreur standard et IC)
  - II. Critères et tests d'adéquation (Biais, RMSE, Chi-2, KS, AD)
  - III. Examen visuel

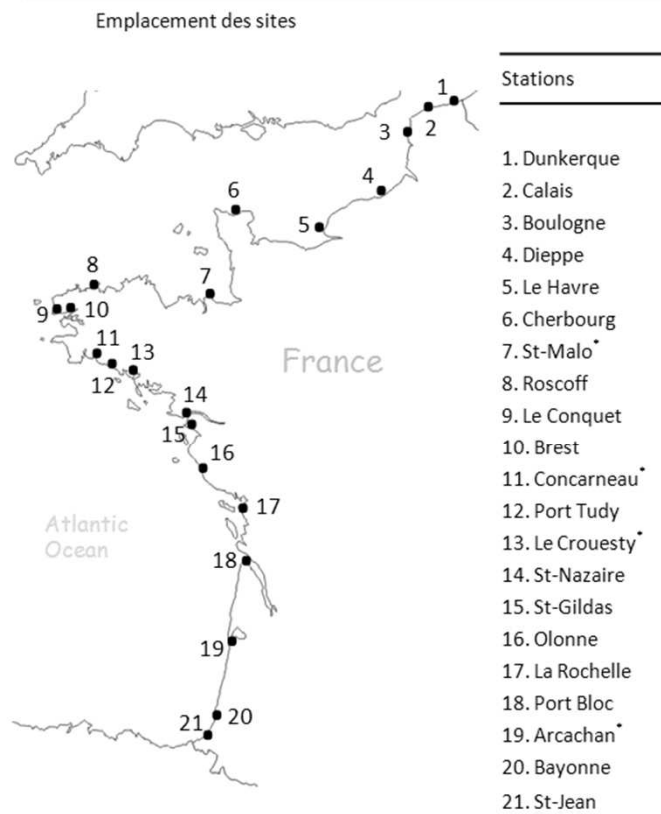


# Les approches



# Cas d'étude & extraction des données

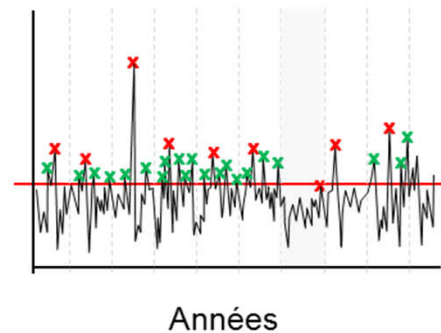
- Les sites d'intérêt – la France / Côte atlantique - Manche
- Un échantillon pour chaque méthode: **Max. Annuelle**; **POT**; **r-LOS**
- Choix du seuil (POT) & choix de  $r$  (r-LOS) ??



\*. Séries trop courtes

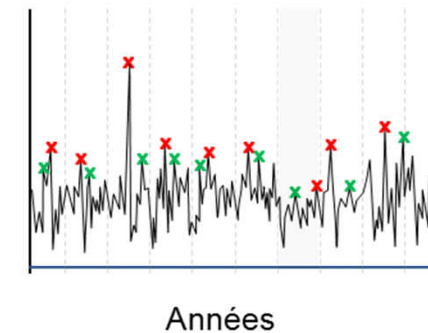
## POT

- Les observations supérieures à un seuil  $u \rightarrow$  **Choix du seuil  $u$**



## r-LOS

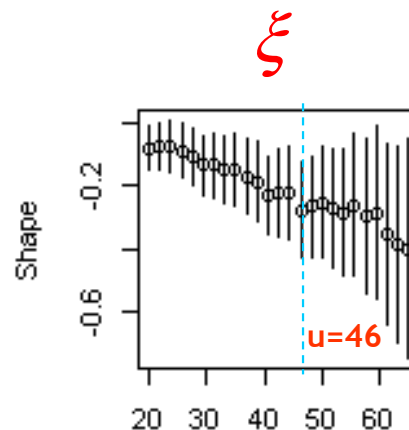
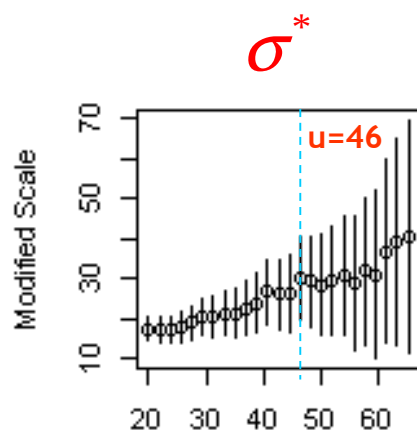
- $R$  observations par bloc  $\rightarrow$  **Choix de  $r$**



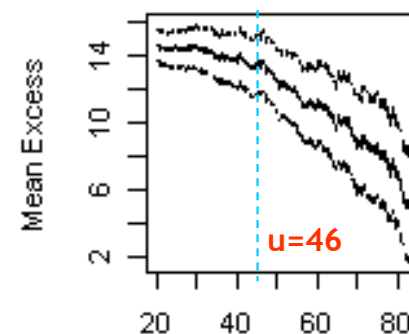
# Choix du seuil $u$ (POT) / courbes de stabilité

$$POT / GPD(\mu, \sigma, \xi) \xleftarrow{\mu = u} G(x) = \begin{cases} 1 - \left(1 - \xi x / \sigma\right)^{1/\xi} & \xi \neq 0 \\ 1 - e^{-x/\sigma} & \xi = 0 \end{cases}$$

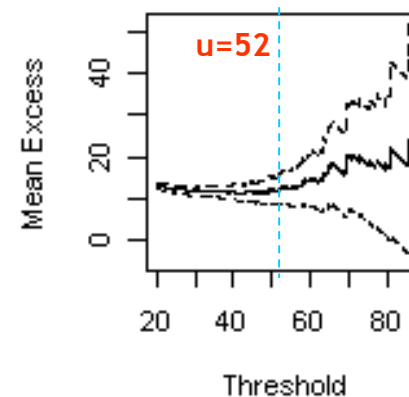
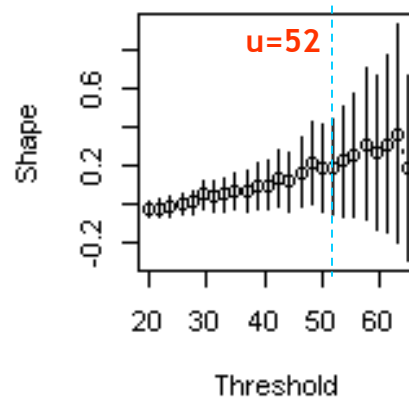
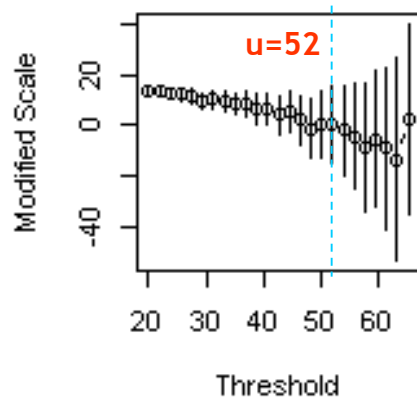
Calais



Mean Excess Life Plot



Dieppe



# Choix de r (r-LOS) / stabilité de l'erreur standard

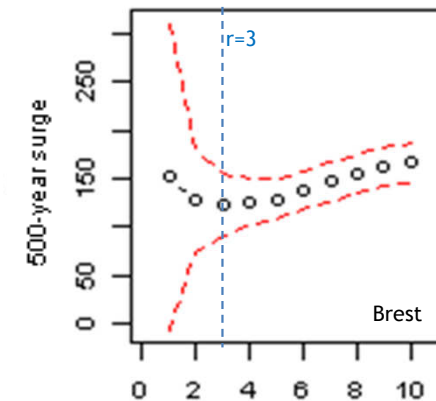
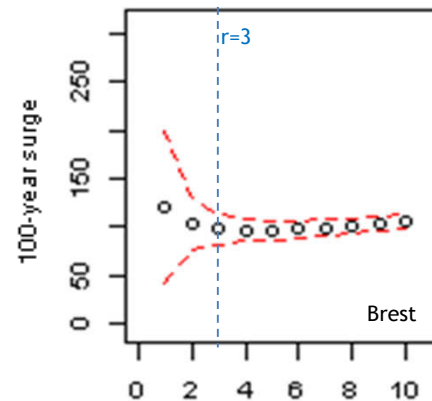
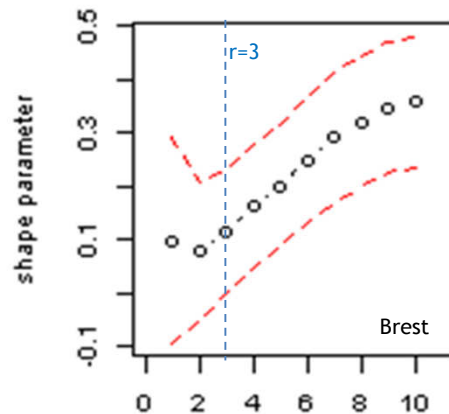
$$r-LOS / GEV(\mu, \sigma, \xi) \longleftarrow F(x) = \begin{cases} e^{-\left(1 + \xi \frac{x-\mu}{\sigma}\right)^{-1/\xi}} & \xi \neq 0 \\ e^{-e^{-(x-\mu)/\sigma}} & \xi = 0 \end{cases}$$

$\xi$

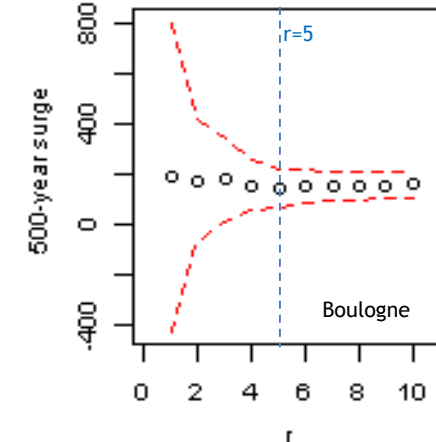
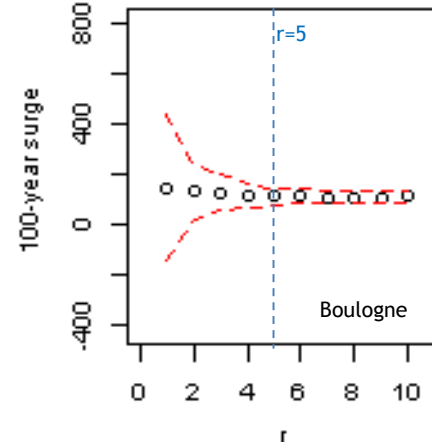
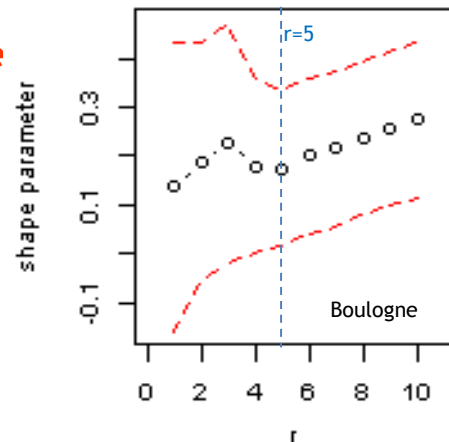
Surcotes (100-year)

Surcotes (500-year)

Brest



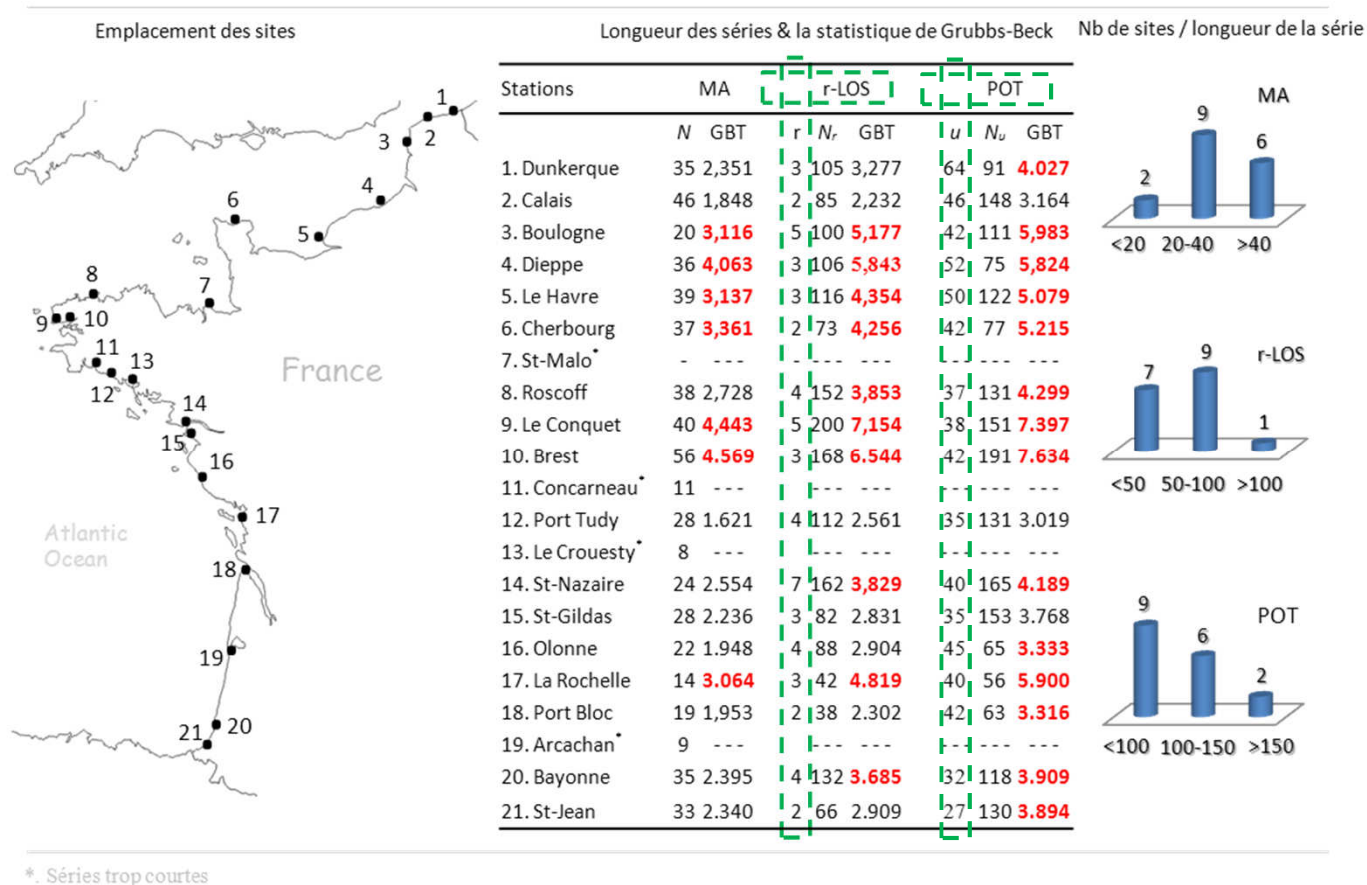
Boulogne





# Résultats des choix du seuil $u$ (POT) et de $r$ (r-LOS)

- 3-7 événements/année (concordant avec la littérature)
- Test de détection d'horsains (Grubbs-Beck) (concordant avec liste des tempêtes)



## Grubbs-Beck

Soient  $Y_1 < Y_2 < \dots < Y_n$  les observations classées en ordre croissant.

$$GBT = \frac{Y_n - \hat{Y}}{S}$$

$\hat{Y}$  est la valeur moyenne et  $S$  est l'écart-type.

➔ Horsain si GBT est supérieur à une valeur critique.

# Tests & hypothèses de base

Stationnarité / *Kendall* (validé avec *rho de Spearman*); Indépendance/*Wald Wolfowitz*; homogénéité / *Wilcoxon* (Niveau de confiance de 5%)

|                     | Sites            | Kendall |       |       | Wolfowitz |       |       | Wilcoxon |       |       |                     |
|---------------------|------------------|---------|-------|-------|-----------|-------|-------|----------|-------|-------|---------------------|
|                     |                  | MA      | r-LOS | POT   | MA        | r-LOS | POT   | MA       | r-LOS | POT   |                     |
| <b>Stat&lt;5%</b> → | 1. Dunkerque     | 0.541   | 0.029 | 0.279 | 0.176     | 0.195 | 0.372 | 0.443    | 0.030 | 0.551 | ← <b>Stat&lt;5%</b> |
|                     | 2. Calais        | 0.820   | 0.851 | 0.887 | 0.766     | 0.121 | 0.319 | 0.517    | 0.593 | 0.726 |                     |
| <b>Stat&lt;1%</b> → | 3. Boulogne      | 0.127   | 0.009 | 0.520 | 0.920     | 0.124 | 0.094 | 0.105    | 0.001 | 0.499 | ← <b>Stat&lt;1%</b> |
|                     | 4. Dieppe        | 0.151   | 0.154 | 0.552 | 0.790     | 0.078 | 0.818 | 0.369    | 0.679 | 0.929 |                     |
|                     | 5. Le Havre      | 0.913   | 0.609 | 0.612 | 0.847     | 0.062 | 0.284 | 0.380    | 0.217 | 0.580 |                     |
| <b>Stat&lt;5%</b> → | 6. Cherbourg     | 0.079   | 0.024 | 0.171 | 0.612     | 0.808 | 0.081 | 0.443    | 0.126 | 0.229 |                     |
|                     | 7. St-Malo*      | ---     | ---   | ---   | ---       | ---   | ---   | ---      | ---   | ---   |                     |
|                     | 8. Roscoff       | 0.791   | 0.666 | 0.488 | 0.984     | 0.709 | 0.529 | 0.954    | 0.772 | 0.422 |                     |
|                     | 9. Le Conquet    | 0.788   | 0.604 | 0.857 | 0.186     | 0.639 | 0.693 | 0.583    | 0.950 | 0.459 |                     |
|                     | 10. Brest        | 0.909   | 0.242 | 0.347 | 0.476     | 0.200 | 0.124 | 0.726    | 0.613 | 0.222 |                     |
|                     | 11. Concarneau*  | ---     | ---   | ---   | ---       | ---   | ---   | ---      | ---   | ---   |                     |
|                     | 12. Port Tudy    | 0.797   | 0.091 | 0.288 | 0.549     | 0.588 | 0.846 | 0.982    | 0.221 | 0.23  |                     |
|                     | 13. Le Crouesty* | ---     | ---   | ---   | ---       | ---   | ---   | ---      | ---   | ---   |                     |
|                     | 14. St-Nazaire   | 0.766   | 0.397 | 0.722 | 0.209     | 0.938 | 0.134 | 0.242    | 0.281 | 0.719 |                     |
|                     | 15. St-Gildas    | 0.950   | 0.984 | 0.868 | 0.274     | 0.249 | 0.424 | 0.302    | 0.132 | 0.908 |                     |
|                     | 16. Olonne       | 0.461   | 0.444 | 0.522 | 0.864     | 0.155 | 0.698 | 0.898    | 0.749 | 0.163 |                     |
|                     | 17. La Rochelle  | 0.826   | 0.182 | 0.932 | 0.174     | 0.834 | 0.835 | 0.902    | 0.132 | 0.801 |                     |
|                     | 18. Port Bloc    | 0.599   | 0.513 | 0.204 | 0.511     | 0.025 | 0.397 | 0.549    | 0.686 | 0.200 |                     |
|                     | 19. Arcachan*    | ---     | ---   | ---   | ---       | ---   | ---   | ---      | ---   | ---   |                     |
| <b>Stat&lt;1%</b> → | 20. Bayonne      | 0.029   | 0.001 | 0.005 | 0.702     | 0.079 | 0.823 | 0.079    | 0.035 | 0.002 | ← <b>Stat&lt;5%</b> |
| <b>Stat&lt;5%</b>   | 21. St-Jean      | 0.975   | 0.710 | 0.490 | 0.678     | 0.481 | 0.863 | 0.845    | 0.693 | 0.381 |                     |



Pourquoi r-LOS  
et non POT ??

# Comparaison des méthodes

## ➤ Comparaison basée sur les critères et tests d'adéquation

1. Les biais tendent tous vers 1 (difficile de discriminer)
2. **POT** (plus faibles RMSE)
3. Difficile de discriminer sur la base des tests de Chi-2, KS et AD

| Stations  |       | Bias                                                                            | RMSE   | $\chi^2$ | Kolmogorov<br>Smirnov (D) | Anderson<br>Darling (A <sup>2</sup> ) |
|-----------|-------|---------------------------------------------------------------------------------|--------|----------|---------------------------|---------------------------------------|
| Dunkerque | AM    | 0.998                                                                           | 22.588 | 0.238    | 0.896                     | 0.610                                 |
|           | r-LOS | Dunkerque station failed the <b>KT</b> and the <b>WT</b> at the level <b>5%</b> |        |          |                           |                                       |
|           | POT   | 0.996                                                                           | 17.39  | 0.239    | 0.837                     | 0.668                                 |
| Calais    | AM    | 1.004                                                                           | 21.609 | 0.238    | 0.785                     | 0.725                                 |
|           | r-LOS | 1.001                                                                           | 21.855 | 0.239    | 0.971                     | 0.596                                 |
|           | POT   | 0.999                                                                           | 15.229 | 0.239    | 0.990                     | 0.791                                 |
| Boulogne  | AM    | 0.980                                                                           | 23.349 | 0.240    | 0.955                     | 0.587                                 |
|           | r-LOS | Boulogne station failed the <b>KT</b> and the <b>WT</b> at the level <b>1%</b>  |        |          |                           |                                       |
|           | POT   | 0.993                                                                           | 18.349 | 0.239    | 0.912                     | 0.258                                 |
| Dieppe    | AM    | 0.983                                                                           | 24.789 | 0.238    | 0.996                     | 0.156                                 |
|           | r-LOS | 0.999                                                                           | 21.876 | 0.239    | 0.622                     | 0.962                                 |
|           | POT   | 0.989                                                                           | 16.956 | 0.239    | 0.976                     | 0.841                                 |
| Le Havre  | AM    | 0.999                                                                           | 30.228 | 0.238    | 0.366                     | 0.996                                 |
|           | r-LOS | 1.002                                                                           | 28.608 | 0.240    | 0.269                     | 1.000                                 |
|           | POT   | 0.994                                                                           | 22.067 | 0.239    | 0.774                     | 0.936                                 |
| Cherbourg | AM    | 1.002                                                                           | 14.848 | 0.238    | 0.819                     | 0.962                                 |
|           | r-LOS | Cherbourg station failed the <b>KT</b> at the level <b>5%</b>                   |        |          |                           |                                       |



Échantillons différents

# Comparaison des méthodes

Niveaux de retour

Les méthodes

Pour chaque paramètre

& r-LOS); GPD pour POT

| Stations  |       | $\mu$                                                                                                                             |      | $\sigma$           |      | $\xi$                 |       | Niveaux de retour |               |               |
|-----------|-------|-----------------------------------------------------------------------------------------------------------------------------------|------|--------------------|------|-----------------------|-------|-------------------|---------------|---------------|
|           |       | MLE(95% CI)                                                                                                                       | SE   | MLE (95% CI)       | SE   | MLE (95% CI)          | SE    | $S_{50}(SE)$      | $S_{100}(SE)$ | $S_{500}(SE)$ |
| Dunkerque | AM    | 75.40(69.49:81.32)                                                                                                                | 3.02 | 16.18(12.07:20.28) | 2.10 | -0.175(-0.380:0.030)  | 0.105 | 121.14(40.54)     | 126.51(52.18) | 136.67(80.06) |
|           | POT   | ---                                                                                                                               | ---  | 13.53(9.26:17.80)  | 2.18 | -0.033(-0.272:0.207)  | 0.122 | 113.68(4.52)      | 121.83(6.15)  | 140.07(10.75) |
|           | r-LOS | Dunkerque station failed the <b>KT</b> (p-value=0.029) and the <b>WT</b> (p-value=0.030) at the level 5% (for the r_LOS data set) |      |                    |      |                       |       |                   |               |               |
| Calais    | AM    | 61.85(56.38:67.33)                                                                                                                | 2.79 | 16.53(12.57:20.49) | 2.02 | -0.425(-0.623:-0.226) | 0.101 | 93.34(19.36)      | 95.25(22.82)  | 97.98(29.06)  |
|           | POT   | ---                                                                                                                               | ---  | 17.55(13.81:21.30) | 1.91 | -0.277(-0.424:-0.130) | 0.075 | 87.92(4.37)       | 91.67(5.40)   | 98.03(7.61)   |
|           | r-LOS | 55.94(52.22:59.66)                                                                                                                | 1.90 | 15.72(13.11:18.32) | 1.33 | -0.323(-0.453:-0.193) | 0.070 | 90.79(11.32)      | 93.57(13.79)  | 98.04(18.79)  |

## ➤ Comparaison basée sur l'incertitude (SE) sur les paramètres & niveaux de retour

1. L'erreur standard (SE) sur les paramètres  $\mu$  et  $\sigma$  est très petite par rapport aux estimations MLE
2. Réduction de l'incertitude avec les approches **r-LOS** et **POT**

## ➤ Comparaison basée sur les niveaux de retour

1.  $S_{100}$  et  $S_{500}$  (**AM**) sont systématiquement supérieures (pénalisant)
2. La méthode **PP (r-LOS & POT)** donne les plus faibles niveaux de retour

# Examen visuel

## Exemple de Brest

Ajustements à gauche et Q-Q plot à droite

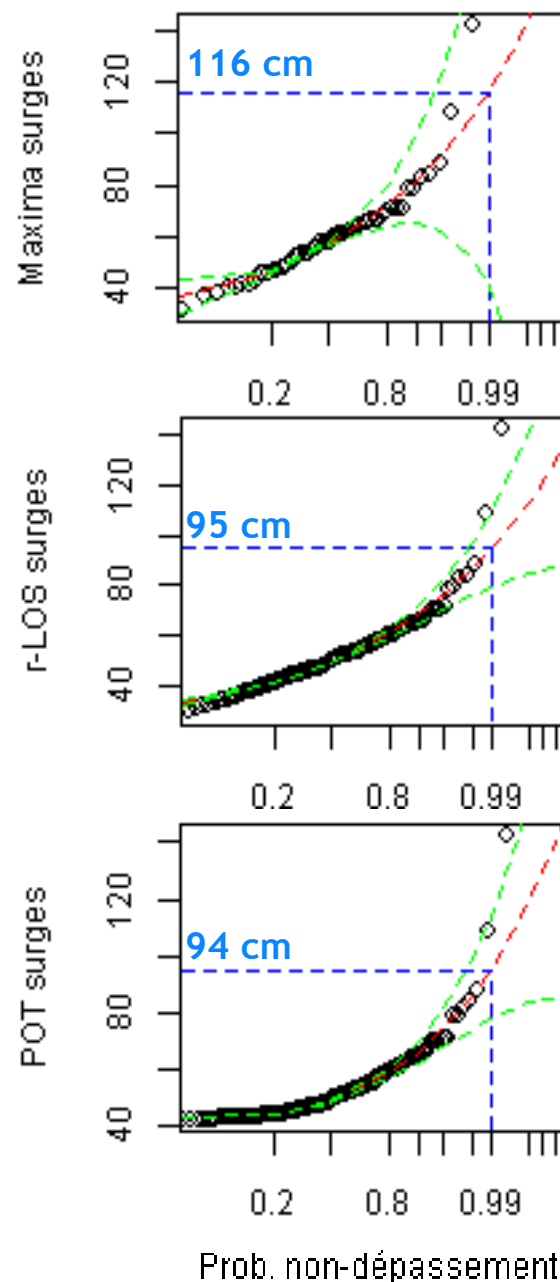
### 1. IC / SE / incertitude

→ POT & r-LOS (++)

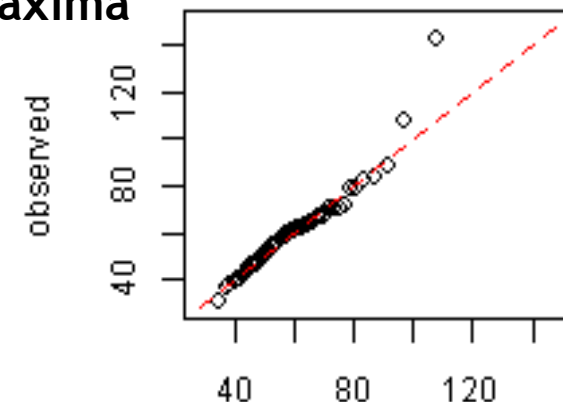
### 2. Niveaux de retour (queue droite, horsain)

→ Maxima (+)

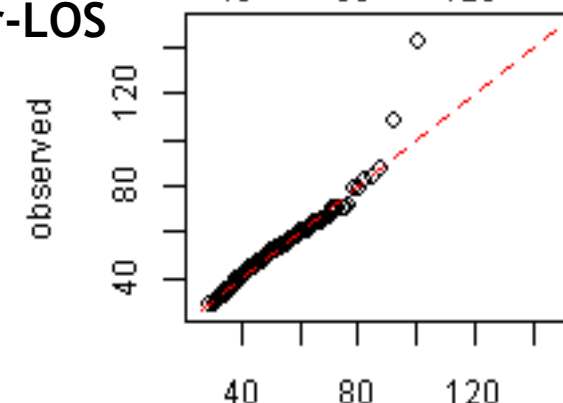
→ moins certaine!!



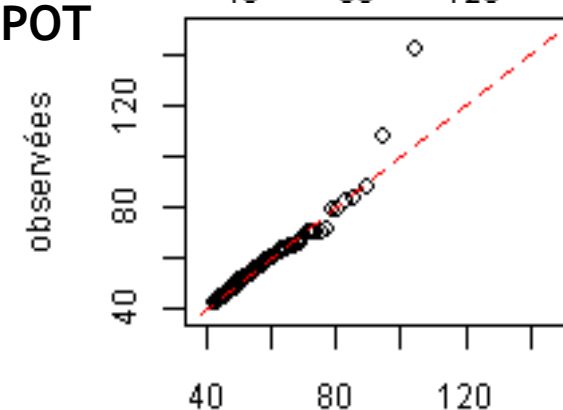
## Maxima



## r-LOS



## POT



# Conclusion

- Les approches PP (POT & r-LOS) estiment les événements extrêmes avec des variances plus faibles (++).
- En présence d'horsains, les probabilités empiriques au niveau de la queue droite de la fonction de distribution sont mal ajustées quelle que soit l'approche (méthode du Maxima +) ← La représentativité de l'horsain dans la série de données est faible.
- La méthode du Maxima donne des niveaux de retour plus élevés.

# Perspectives

- exploration des mises en œuvre propres aux méthodes (choix du seuil  $u$  avec POT...)
- Combiner les approches POT et r-LOS (chercher les  $r$  plus grandes observations qui sont supérieurs à un seuil )
- L'information historique est une piste à explorer même en absence d'horsains.

# Perspectives

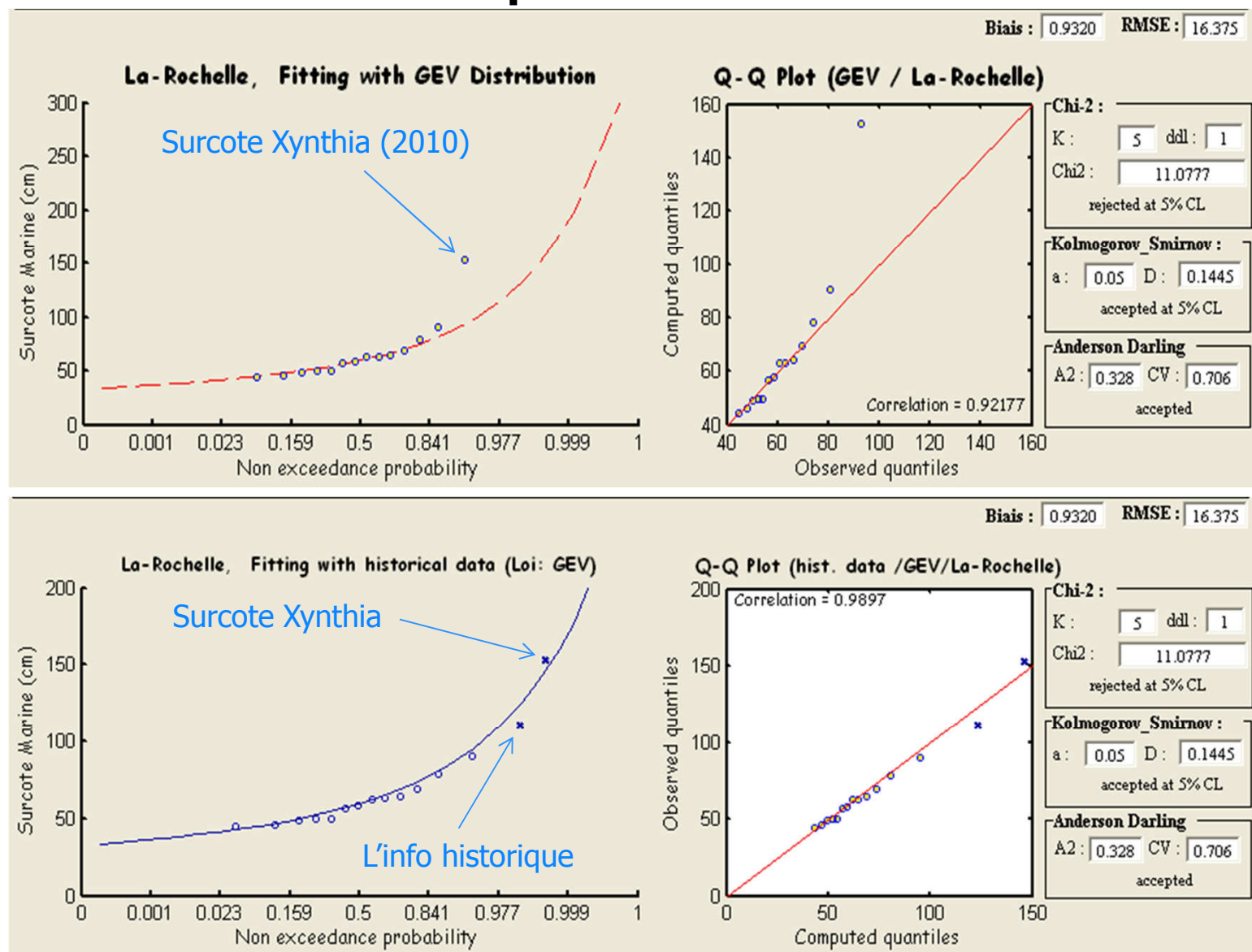
## Utilisation de l'information historique (IH)

(Travail présenté à EGU 2013, Vienne)

- Le fait d'ajouter des observations (tempêtes importantes) qui ont eu lieu bien avant la période de mesures pourra améliorer la représentativité d'un **horsain** dans la série de données;
- Sources de l'**IH**: Archives, vieux journaux, personnes âgées, anciens rapports non publiés, paléo, sédimentologie . . .
- (i) Différents types d'**IH** (Valeur exacte, borne inférieure, intervalle, seuil de perception) (ii) Les observations systématiques > au seuil de perception sont traitées de la même manière que l'**IH**;
- des probabilités empiriques et une vraisemblance différentes de l'approche sans IH.



# Exemple d'application de l'information historique : La Rochelle





**IRSN**

INSTITUT  
DE RADIOPROTECTION  
ET DE SÛRETÉ NUCLÉAIRE



# Discussion



Système de management  
de la qualité IRSN certifié

Yasser Hamdi, Claire-Marie Duluc, Lise Bardet, Vincent Rebour  
Institut de Radioprotection et de Sûreté Nucléaire