DOI 10.1051/LHB/2010047

Théorie statistique du renouvellement pour la détermination des houles extrêmes. Partie 1 : le point sur les méthodes disponibles

The Peaks-over-Threshold statistical theory for the estimation of extreme sea-states. Part 1 : A review of available methods

FRANCK MAZAS
Ecole des Ponts ParisTech
présentement Sogreah Consultants
6 rue de Lorraine, 38130 Echirolles, France
Tél : +33 (0)4 76 33 40 00, e-mail : franck.mazas@ponts.org

LUC HAMM
Sogreah Consultants
6 rue de Lorraine, 38130 Echirolles, France
Tél : +33 (0)4 76 33 41 88, e-mail : luc.hamm@sogreah.fr

La théorie statistique du renouvellement est internationalement recommandée depuis maintenant 15 ans pour procéder à la détermination de la hauteur significative des états de mer extrêmes utilisée en ingénierie côtière pour la conception des digues et autres ouvrages en mer. Une revue des avancées théoriques de cette approche est présentée ici conduisant à une amélioration de sa mise en application pratique par rapport aux recommandations publiées jusqu'à présent dans la profession. Celles-ci portent sur une amélioration de la définition des tempêtes, l'utilisation de la loi GPD avec une méthode de maximum de vraisemblance pour procéder à l'ajustement théorique, l'optimisation du seuil de censure. L'élargissement de l'analyse à un grand nombre de lois statistiques est une piste d'amélioration suggérée à condition de pouvoir y associer des critères de choix objectifs.

The Peaks-over-Threshold (POT) statistical theory is internationally recommended for the last fifteen years to get the significant wave heights of extremes sea-states used in coastal engineering to design rubble-mound breakwaters and other marine structures. A review of theoretical advances of this method is presented here leading to significant improvements in practical applications compared to previously published recommendations. A new method to define and select storm peaks using a double threshold is introduced and settled. The GPD distribution coupled with the likelihood maximum estimator (LME) fitting method is recommended together with an optimisation of the censored threshold. An enlargement of the analysis to several statistical distributions is proposed in association with unbiased best-fitting indicators.

I ■ INTRODUCTION

Sur un site maritime, prévoir la hauteur significative des états de mer extrêmes pour des périodes de retour élevées (de l'ordre de quelques dizaines d'années) est primordial pour le dimensionnement des ouvrages portuaires, mais relève de la gageure. Une sous-estimation conduira à une fragilisation, voire à une destruction des ouvrages et pourra provoquer des dégâts importants, alors qu'une surestimation du danger engendrera des surcoûts importants et superflus, tant financiers qu'environnementaux.

Les méthodes statistiques développées depuis quelques décennies ambitionnent de mettre à la disposition de l'analyste des outils objectifs. La méthode la plus fiable serait évidemment l'interpolation sur une trentaine de données : pour déterminer une houle centennale, cela nécessiterait de disposer de mesures précises sur 3 000 ans ! Une très large extrapolation est donc inévitable, à tel point que la recherche d'une prédétermination fiable peut s'apparenter à une cause perdue. Et pourtant, l'ingénieur chef de projet doit prendre une décision.

À la fin des années 1980, l'adoption par la communauté maritime de la méthode proposée par le Professeur Goda, de l'Université de Tokyo, a représenté un réel progrès (Goda, 1988b [1], Goda et Kobune, 1990 [2], Goda, 2000 [3]). Le document de synthèse de 1994 du Groupe de Travail sur les

Statistiques des Houles Extrêmes [4] et le Manuel enrochement (CIRIA *et al.*, 2007 [5]) ont entériné son utilisation générale. Toutefois, de nouveaux outils, notamment informatiques, sont apparus depuis et permettent, sans remettre en cause le fond de la méthode, de l'améliorer significativement sur plusieurs points. Un rapide tour d'horizon international ([6], [7], [8]) montrant que Goda est encore très suivi, le but de cet article est de sensibiliser la communauté des analystes à ces améliorations théoriques. En nous fondant sur la théorie des statistiques des valeurs extrêmes, nous présentons donc ici une méthode globale en mettant l'accent sur les améliorations apportées.

II ■ UN PEU DE STATISTIQUES DES VALEURS EXTRÊMES

II.1 PÉRIODE DE RETOUR ET VALEUR DE RETOUR

Considérons un processus aléatoire X. Par définition, la *période de retour T* (ou *temps/durée de retour*) de la valeur x est l'espérance mathématique de la variable aléatoire définie comme la durée séparant deux dépassements successifs (dépassements vers les valeurs supérieures) de cette valeur.

En d'autres termes, la période de retour correspond à la durée moyenne séparant deux occurrences successives de l'événement $X \geq x$. Cette définition rigoureuse permet pour un processus donné d'attribuer à toute valeur x sa période de retour $T(x)$ et, inversement, d'attribuer une valeur $x(T)$ à toute période de retour T.

Ces notions sont bien sûr primordiales dans la détermination des houles extrêmes : si X désigne une hauteur significative de houle, $x(T = 100)$ désignera la valeur de la houle centennale, à condition que la période de retour soit exprimée en années.

Cette remarque nous amène d'ailleurs à introduire un paramètre d'importance dans l'analyse des échantillons de valeurs extrêmes : le taux moyen des évènements de tempêtes, ou *mean rate*. Soit N_T le nombre de tels évènements sur une période de K années. Alors, le taux moyen λ_T est défini comme suit :

$$\lambda_T = \frac{N_T}{K} \tag{1}$$

La valeur de retour est reliée à la fonction de répartition de la loi $F(x)$ par la formule :

$$x(T) = F^{-1}\left(1 - \frac{1}{\lambda_T T}\right) \tag{2}$$

où T est exprimé en années. On en déduit immédiatement la relation entre la période de retour et la fonction de répartition :

$$T(x) = \frac{1}{\lambda_T\left[1 - F(x)\right]} \tag{3}$$

II.2 LES DISTRIBUTIONS DES VALEURS EXTRÊMES

Considérons $X_1, X_2, \ldots, X_n$ un échantillon de n variables aléatoires réelles, indépendantes et identiquement distribuées selon une loi de fonction de répartition F. On s'intéresse aux valeurs extrêmes, ici aux maxima, d'un tel échantillon.

Soit $M_n = \max\{X_1, X_2, \ldots, X_n\}$ le maximum d'un tel échantillon. Fréchet (1927) [9] et Fisher et Tippett (1928) [10] ont montré qu'en introduisant deux constantes de normalisation $a_n > 0$ et b_n, la loi de $a_n(M_n - b_n)$ tend vers l'une de ces trois lois : les lois de Gumbel, Fréchet et Weibull.

Jenkinson (1955) [11] a montré que ces trois familles de lois des valeurs extrêmes pouvaient être écrites sous une unique forme : la *loi généralisée des valeurs extrêmes* (GEV, *Generalized Extreme Value distribution*), qui a pour fonction de répartition :

$$F_{x_0,\psi,k}(x) = \exp\left[-\left(1 + k\frac{x - x_0}{\psi}\right)^{-\frac{1}{k}}\right] \tag{4}$$

Le cas $k > 0$ correspond à la loi de Fréchet ; le cas $k < 0$ correspond à la loi de Weibull ; enfin, en faisant tendre k vers 0, on obtient la distribution de Gumbel par passage à la limite. x_o est appelé paramètre de localisation (*location parameter*) car il fixe la localisation de l'axe des x. Le paramètre ψ, appelé paramètre d'échelle (*scale parameter*), régit l'échelle linéaire de x. Enfin, k (sans dimension, strictement positif) est appelé paramètre de forme (*shape parameter*) car il détermine la forme fonctionnelle de la distribution.

II.3 DÉPASSEMENTS DE SEUIL ET MODÈLE POISSON-GPD

On s'intéresse à présent à la loi régissant le dépassement d'un seuil au sein d'un échantillon. Soit X une variable aléatoire réelle de fonction de répartition F. Soit u le seuil fixé et posons $Y = X - u$ sous condition que $X > u$ (Y est donc la variable aléatoire représentant le dépassement du seuil u par la variable X). La loi des dépassements de u est donnée par :

$$\Pr\left\{Y \leq y\right\} = \Pr\left\{X \leq u + y \middle| X > u\right\} = F_u(y) = \frac{F(u+y) - F(u)}{1 - F(u)} \tag{5}$$

Lorsque u approche le point terminal (fini ou infini), l'expression *(5)* peut être approchée par la *distribution généralisée de Pareto* (GPD, *Generalized Pareto Distribution*), dont la fonction de répartition est donnée par :

$$F_{\psi,k}(y) = 1 - \left(1 + k\frac{y}{\psi}\right)^{-\frac{1}{k}} \tag{6}$$

Le théorème de Pickands (1975) [12] justifie cette approximation pour une taille d'échantillon assez grande, et pour un seuil u assez élevé. De même que pour la loi GEV, le cas $k < 0$ peut se rattacher au modèle de Weibull de la théorie classique des valeurs extrêmes. Quant au cas $k = 0$ (par passage à la limite), il correspond en fait à la distribution exponentielle d'espérance ψ.

Dans le cas où le nombre N_1 de dépassements du seuil u au cours d'une année obéit à une loi de Poisson d'espérance λ, on suggère le modèle suivant, appelé *modèle Poisson-GPD* et défini ainsi : les dépassements de seuil obéissent à une loi GPD et sont i.i.d., et N_1 suit une loi de Poisson :

$$\Pr\left\{N_1 = n\right\} = e^{-\lambda \frac{\lambda^n}{n!}} \qquad (7)$$

Cependant, l'analyste travaille souvent sur des durées de l'ordre de la douzaine d'années, et un processus poissonien peut être délicat à mettre en évidence. Aussi pourra-t-on travailler avec le taux moyen d'évènements par année, λ_T. Notons que cette moyenne empirique est souvent l'estimateur de λ, ce qui justifie ce choix.

III ■ TRAITEMENT DE L'ÉCHANTILLON

● III.1 LES TROIS TYPES DE JEUX DE DONNÉES

L'ingénieur analyste travaille à partir d'échantillons de données environnementales, réelles ou simulées, comme ici la hauteur significative des vagues. Il existe alors trois approches de ces jeux de données.

La première approche est celle de l'échantillon complet, ou *total sample method*. Dans cette approche, toutes les données observées ou enregistrées sont utilisées. On essaye alors de les analyser sous la forme d'une fonction de répartition à ajuster à une distribution particulière. La distribution qui s'ajustera le mieux aux données (*best-fitting distribution function*) permettra alors de déterminer les quantiles désirés.

La deuxième approche, appelée *block maxima method*, ne garde de l'échantillon que les valeurs des maxima sur une période de temps donnée, le plus souvent un an. On parle alors de méthode des maxima annuels. Cela réduit bien entendu considérablement la taille de l'échantillon (typiquement, pour un enregistrement sur une douzaine d'années, on peut passer de 20 000 données environ à 12 !).

La troisième approche est celle de la méthode du renouvellement, plus souvent nommée d'après son acronyme anglais POT (*peaks-over-threshold method*). Cette méthode ne retient que les valeurs maximales des épisodes de tempêtes, grâce à la fixation d'un seuil (*threshold*). Autrement dit, si le seuil est fixé à trois mètres, la méthode POT ne gardera de l'échantillon de données que les valeurs maximales de chaque épisode de tempête ayant généré des hauteurs significatives de houle supérieures à trois mètres. Cette méthode de sélection demande néanmoins réflexion, et peut être améliorée par l'emploi combiné de deux valeurs de seuil, comme nous le verrons plus loin.

Chaque méthode a ses partisans et ses contempteurs. On ne peut faire un choix totalement objectif entre ces trois méthodes. Cependant, deux conditions doivent être réunies par l'échantillon pour se livrer à une analyse statistique rigoureuse : l'*indépendance* et l'*homogénéité*. Autrement dit, les valeurs de l'échantillon étudié doivent être indépendantes et identiquement distribuées (i.i.d.). Or, les hauteurs de vagues successives sont fortement corrélées à court et moyen terme (jusqu'à quelques jours). De ce fait, la plupart des auteurs rejettent la première méthode.

Un inconvénient majeur de la deuxième méthode est d'écarter des valeurs qui apportent une information valorisante. Le choix de ne garder que les maxima annuels, ou même mensuels, apparaît donc trop tranché et trop restrictif pour des phénomènes tels que les tempêtes, qui surviennent plusieurs fois par an (le choix du maximum annuel se justifierait plus aisément, par exemple, pour l'étude des crues de grands fleuves qui peuvent durer plusieurs semaines, voire plusieurs mois). C'est pourquoi il existe aujourd'hui un consensus assez large pour privilégier la méthode POT. Le choix de ne garder que la valeur maximale d'un épisode de tempête et de fixer une durée minimale entre deux épisodes (de l'ordre de 2 à 3 jours) garantit une quasi-indépendance des données.

Enfin, il faut veiller à l'homogénéité de l'échantillon, i.e. au fait que toutes les valeurs proviennent d'une même loi (inconnue) de distribution. L'analyse des causes physiques des houles extrêmes est alors précieuse : par exemple, dans les régions concernées, il faut prendre soin de séparer les valeurs appartenant à la saison de la mousson des autres, car il est plus que probable qu'elles ne proviennent pas de la même loi de distribution. Un tri géographique par secteur de provenance est également souvent effectué, de façon à prendre en compte les différents régimes de vent.

La méthode POT sera ainsi retenue dans toute la suite de cette étude.

● III.2 CENSURE DES DONNÉES ET DOUBLE SEUIL

Les N_T tempêtes retenues de cette façon sont d'intensités très diverses, si le seuil est assez bas : les tempêtes de grande ampleur côtoient les tempêtes d'importance faible ou moyenne. Cette constatation n'est pas anodine : ces dernières peuvent en effet distordre l'ajustement à une distribution en apportant trop de poids aux faibles valeurs de pics, donc en introduisant un biais négatif. On préfère donc pratiquer un ajustement rigoureux sur les plus fortes valeurs des pics extrêmes. Cependant, les tempêtes de faible ou moyenne intensité apportent une information valorisante sur la fréquence de leur occurrence qu'il est bon de prendre en compte si cela est possible. Pour cela, on applique alors un *processus de censure* : seules N données correspondant aux N plus importantes tempêtes seront prises en compte dans l'analyse ($N < N_T$). En revanche, le paramètre λ_T reste défini comme spécifié en *(1)*, avec le paramètre N_T. On définit alors le *paramètre de censure* de la sorte :

$$\nu = \frac{N}{N_T} \qquad (8)$$

Pour ce faire, on fixe alors un premier seuil u_1 assez bas permettant de distinguer chaque tempête, même mineure : on obtiendra alors une estimation de N_T, qui n'a pas besoin d'être très précise. Un second seuil u_2 plus élevé fixera alors la valeur de N, qui correspondra à la taille de l'échantillon auquel l'analyse statistique rigoureuse sera appliquée. Ce type de censure est appelé *censure de type I à gauche*. Toute la difficulté du travail de l'analyste résidera alors dans le choix de la valeur de u_2, car celui-ci déterminera l'échantillon analysé rigoureusement. Des outils objectifs seront évoqués dans la seconde partie de l'article illustrant l'application pratique de cette méthode. Nous appellerons ce doublet (u_1, u_2) *double seuil*.

On peut s'interroger sur l'importance du phénomène de censure. En effet, l'équation *(3)* montre que l'influence de la censure est asymptotiquement nulle et que l'on peut la négliger pour des périodes de retour suffisamment grandes. Cependant, beaucoup d'analystes (dont le Groupe de Travail) recommandent de travailler sur des données censurées, de façon à ne pas perdre toute l'information donnée par les tempêtes de faible intensité, c'est-à-dire leur fréquence d'apparition, sans pour autant perturber l'ajustement pour les valeurs extrêmes.

III.3 PROPOSITION DE CRITÈRES POUR LA DÉFINITION DE TEMPÊTES INDÉPENDANTES

Le seul choix d'un seuil, simple ou double, ne suffit pas pour distinguer une tempête. Prenons l'exemple de la tempête des 20 et 21 décembre 1999 avec des mesures de vagues toutes les heures au large de Cannes et étudions le résultat pour des valeurs de seuil à 2 et 3,5 mètres (cf. *fig. 1*).

Un simple examen visuel de la courbe permet d'affirmer qu'il s'agit là d'un seul et unique évènement extrême, avec une valeur maximale à 4,4 mètres. Dans le premier cas, la valeur du seuil permet bien de sélectionner cette seule et unique valeur. Mais si l'on décide de relever le seuil, on risque d'augmenter de façon artificielle et erronée le nombre de tempêtes du fait des fluctuations de la hauteur significative des vagues. Dans le cas de la tempête du 20 décembre 1999, on obtient ainsi quatre valeurs. La question de la définition précise d'une tempête est donc plus complexe qu'il n'y paraît, et un simple critère de seuil se révèle insuffisant.

Il serait commode de disposer d'une définition et de critères de sélection précis. Or, force nous est de constater que la littérature élude le plus souvent le sujet. Sur le modèle des Italiens [5], nous proposons donc les critères suivants :

1. la hauteur significative des vagues se maintient au-dessus du seuil u_1 pendant plus de N_H heures consécutives
2. la hauteur des vagues peut néanmoins descendre sous le seuil u_1 pendant moins de n_H heures consécutives ($n_H < N_H$)
3. un intervalle minimum de l'ordre de 1 à 3 jours doit être observé entre deux tempêtes, de façon à assurer l'indépendance des données

Ces trois critères, en adaptant bien sûr les valeurs quantitatives aux conditions météorologiques et maritimes locales (par exemple, en Méditerranée, u_1 = 1 m, N_H = 12 h et n_H = 6 h), permettent une sélection plus rigoureuse des pics de tempêtes.

IV ■ ANALYSE RIGOUREUSE DE L'ÉCHANTILLON

IV.1 CHOIX DES DISTRIBUTIONS CANDIDATES

La théorie nous a dit plus haut que la loi correspondant à des échantillons POT est la loi GPD. Dans une analyse simple, c'est donc bien cette loi qu'il s'agit d'utiliser, et non la loi de Gumbel ou de Weibull comme on le voit souvent. Il est alors pertinent de mettre en place un modèle Poisson-GPD.

Mais on peut également approfondir l'analyse. En effet, la théorie des valeurs extrêmes est certes très séduisante, mais il est primordial de garder à l'esprit qu'elle a un caractère *asymptotique*. Pour que son utilisation soit vraiment pertinente, il faut des échantillons de taille beaucoup plus grande que ce dont l'on dispose habituellement, c'est-à-dire quelques dizaines de valeurs. En outre, nous n'avons aucune information sur la vitesse de convergence de la loi de l'échantillon vers ces lois asymptotiques : or rien ne garantit qu'elle ne soit pas très faible ! Ce caractère asymptotique est confirmé par un autre point soulevé par Bouleau (1991) [13] : les trois types de lois classiques (Gumbel, Weibull, Fréchet) sont incompatibles avec des transformations non linéaires. Or, les variables environnementales (hauteurs de houle, débits de crue…) dépendent de processus fortement non linéaires. Nous ne pouvons donc avoir aucune information sur la vitesse de convergence de leurs lois vers les lois extrêmes.

En conséquence, les lois des valeurs extrêmes (GEV, GPD) sont bien des candidates privilégiées pour modéliser les valeurs maximales et/ou les dépassements de seuil d'un échantillon. Mais la taille de ces échantillons comme la gamme des probabilités considérées dans les applications hydrologiques font que d'autres distributions peuvent *a priori* fournir une meilleure modélisation. Rien ne permet donc de rejeter les lois log-normale, log-Pearson de type III, Gamma, χ^2…

Une analyse plus approfondie utilisera donc avec bonheur un grand nombre de distributions candidates : bien que beaucoup plus lourde, c'est l'approche la plus justifiable à condition de pouvoir mettre en place des critères de choix objectifs.

V ■ AJUSTEMENT

La méthode d'ajustement par les moindres carrés a longtemps été utilisée car elle est légère et assez simple à coder. L'amélioration des outils informatiques permet cependant d'opter pour des ajustements plus sophistiqués et conduit donc à s'interroger sur le plus adapté.

Il faut disposer d'un estimateur *robuste* et *efficace*. Un estimateur est dit robuste s'il est très peu perturbé par une valeur rare et extrême (*outlier*) dont la probabilité qu'on lui affecte est très incertaine. La moyenne empirique, par exemple, n'est pas un estimateur robuste. Un estimateur est d'autant plus efficace que sa variance est faible. Enfin, on cherche à ce que cet estimateur ait un biais le plus fai-

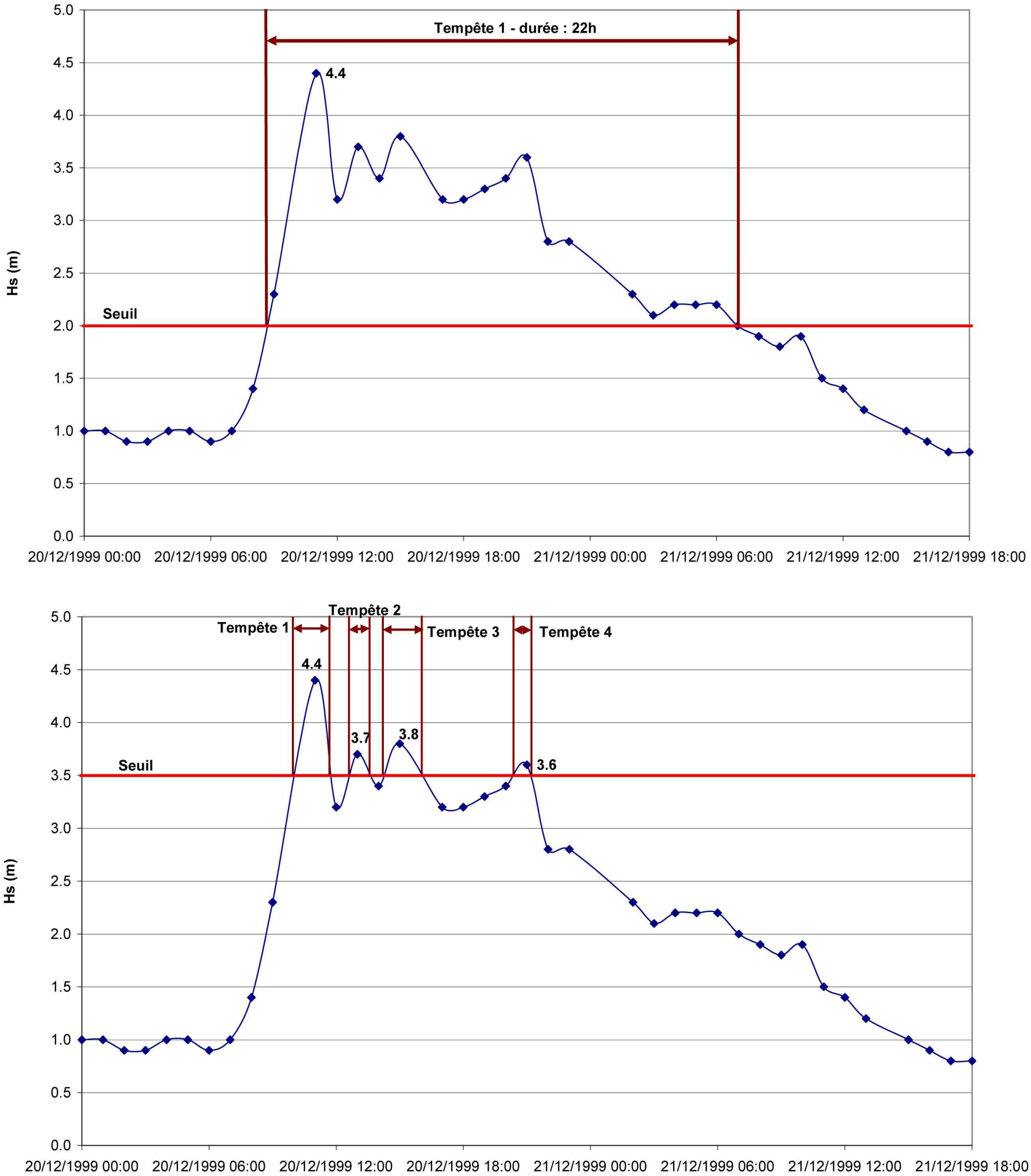

Figure 1 : L'épisode de tempête du 20 décembre 1999 pour des seuils à 2 et 3,5 m

ble possible, et notamment qu'il soit asymptotiquement non biaisé, i.e. que le biais tende vers 0 lorsque la taille de l'échantillon tend vers l'infini.

La méthode des moindres carrés présente le grave défaut de donner beaucoup trop de poids aux évènements rares, ce qui conduit à des ajustements biaisés. Pour des processus non linéaires, elle est aujourd'hui fortement déconseillée par les statisticiens.

La méthode des moments est la plus intuitive. Elle consiste à utiliser les relations entre les moments de l'échantillon et les paramètres de la loi que l'on cherche à ajuster. La méthode des moments construit certes des estimateurs convergents, mais ceux-ci sont souvent entachés de biais négatifs importants pour les petits échantillons.

La méthode des moments pondérés (Hosking et Wallis, 1987) [14] tente d'y remédier en pondérant les moments par leur probabilité. Dans le cas d'échantillons de taille réduite (inférieure à 500), pour l'ajustement à une loi GPD, Hosking et Wallis ont montré que cet estimateur était plus efficace que le maximum de vraisemblance pour $k < 1/2$. Des généralisations (Diebolt *et al.*, 2005b [15]) permettent d'étendre cette méthode à $k < 3/2$. Dans la pratique, cette condition est souvent vérifiée… mais ce résultat ne concerne que la loi GPD.

De manière générale, l'estimateur du maximum de vraisemblance (EMV) est considéré comme le plus rigoureux par les statisticiens. Il repose sur la fonction de vraisemblance de l'échantillon, qui est une mesure de la crédibilité de l'affirmation : « cette distribution est celle qui a donné naissance à cet échantillon ». L'EMV consiste à maximiser la fonction de vraisemblance en fonction des paramètres de la famille de lois choisie pour l'ajustement. La méthode du maximum de vraisemblance repose sur des bases théoriques plus solides que celles de la méthode des moments. En particulier, on montre que, sous des conditions très générales, un estimateur MV est convergent, asymptotiquement normal et efficace. La méthode du maximum de vraisemblance est aujourd'hui la principale méthode d'estimation. En particulier, elle semble s'adapter beaucoup plus facilement à l'utilisation de données censurées, ce qui nous intéresse particulièrement.

L'estimateur du maximum de vraisemblance est donc recommandé, même si une pondération judicieuse des moments peut donner de meilleurs résultats dans les domaines de validité *ad hoc*. Enfin, la détermination des intervalles de confiance sera une étape cruciale de l'analyse.

VI ■ CONCLUSIONS

De telles analyses sur des échantillons de données environnementales recueillies sur une grande période de temps sont très délicates. De nombreux tests sont nécessaires pour appréhender les difficultés de toute sorte qui interviennent. Celles-ci sont généralement de deux types : des difficultés intrinsèques à l'échantillon et des difficultés purement statistiques. Les premières sont dues au caractère non indépendant mais surtout non homogène et non stationnaire de l'échantillon. C'est la plus grande source d'imprécision et, partant, la plus grande source potentielle d'améliorations. Des pistes existent (homogénéisation des données par diverses méthodes – découpage géographique, séparation des différents systèmes de vagues par analyse spectrale des états de mer –, analyses régionale, historique…) mais la tâche est très difficile.

Parallèlement, les outils numériques nous permettent aujourd'hui d'utiliser des méthodes statistiques plus performantes et plus justifiables théoriquement. Cette étude a dégagé une méthode rigoureuse : identification et sélection des pics de tempête, censure et double seuil, utilisation d'une loi théoriquement justifiée (loi GPD) ou d'une approche multi-distributions plus lourde mais plus judicieuse, ajustement par l'estimateur du maximum de vraisemblance. À l'avenir, les méthodes bayésiennes de type MCMC offrent également une piste prometteuse d'amélioration de l'ajustement en prenant en compte l'information *a priori* dont on dispose avant l'analyse.

Les avancées présentées ici permettent d'améliorer la méthode de Goda sur plusieurs points cruciaux : meilleure sélection des pics de tempête, optimisation des seuils, utilisation d'un estimateur plus efficace et de lois plus justifiées… ce qui nous apporte une moindre sensibilité aux *outliers* et au biais, la possibilité d'intégrer des données censurées, et en définitive des résultats plus fiables.

Enfin, l'accent doit être mis sur l'importance des intervalles de confiance, dont la largeur peut très fortement varier. Un bon analyste sait décomposer judicieusement son échantillon en fonction des spécificités météorologiques et maritimes du site étudié, l'analyser rigoureusement et surtout prendre le recul nécessaire face aux résultats obtenus.

Dans la seconde partie de l'article, cette méthode sera testée sur différents sites de manière à la mettre à l'épreuve de la réalité. Nous verrons ainsi apparaître les avantages et les inconvénients des deux approches proposées ainsi qu'une illustration parlante des difficultés restant à résoudre.

VII ■ REMERCIEMENTS

Les auteurs souhaitent remercier tous ceux qui les ont aidés durant leur étude, et en particulier M. Éric Gaume pour sa relecture et ses précieux conseils.

VIII ■ BIBLIOGRAPHIE

[1] Goda Y. (1988) — On the methodology of selecting design wave height. *Proc. 21st Int. Conf. Coastal Engrg., Malaga*. 899-913

[2] Goda Y., Kobune K. (1990) — Distribution function fitting for storm wave data. *Proc. 22nd Int. Conf. Coastal Engrg., Delft*. 18-31

[3] Goda Y. (2000) — Random Seas and Design of Maritime Structures. *Advanced Series on Ocean Engineering –World Scientific*. **15**

[4] Mathiesen M., Goda Y., Hawkes P.J., Mansard E., Martin M. J., Peltier E., Thompson E.F., Van Vledder G. (1994) — Recommended practice for extreme wave analysis. *J. Hydraul. Re*. **32(6)** 803-814

[5] Ciria, Cur, & Cetmef (2009) — *Guide Enrochement. L'utilisation des enrochements pour les ouvrages hydrauliques. Version française du Rock Manual version 2007, P09-01, CETMEF, Compiègne, France*.

[6] Franco L., Piscopia R. (2004) — *Atlante delle onde nei mari italiani – Italian wave atlas. Agenzia per la Protezione dell'Ambiente e per i Servizi Tecnici*.

[7] Thompson E.F. (2002) — Hydrodynamic Analysis and Design Conditions. *Coastal Engineering Manual, J. Pope (ed.), US Army Corps of Engineers*. **II(8)**

[8] Rom 0. (1992) — *OLEAJE – Anejo I. Clima Marítimo en el Litoral Español. Obras Marítimas Tecnologia*.

[9] Frechet M. (1927) — Sur la loi de probabilité de l'écart maximum. *Annales de la Société polonaise de Mathématique, Cracovie*. **6**

[10] Fisher R.A., Tippett L.H.C. (1928) — Limiting forms of the frequency distributions of the largest or smallest member of a sample. *Proceedings of the Cambridge Philosophical Society*. **24** 180-190

[11] Jenkinson A.F. (1955) — The frequency distribution of the annual maximum (or minimum) values of meteorological

events. *Quaterly journal of the Royal meteorological society*. **81** 158-171

[12] Pickands J., Iii (1975) — Statistical inference using extreme order statistics. *Annals of Statistics*. **3** 119-131

[13] Bouleau N. (1991) — Splendeurs et misères des lois des valeurs extrêmes. *Risques*. **3** 85-92

[14] Hosking J.R.M., Wallis J.R. (1987) — Parameter and quantile estimation for the generalized Pareto distribution. *Technometrics*. **29** 339-349

[15] Diebolt J., Guillou A., Rached I. (2005) — Approximation of the distribution of excesses using a generalized probability weighted moment method. *Comptes rendus. Mathématique*. **340(5)** 383-388