



Étude des corrélations statistiques entre la houle, la marée et les surcotes. Application au site de Dieppe.

Vidal Cohen¹, Bernard Lapeyre, Christian Monnet,
Alain Toubol et Zhang Ruotao²

*École Nationale des Ponts et Chaussées,
Centre d'Enseignement et de Recherche
en Mathématiques, Informatique et Calcul Scientifique*

Résumé : Le but de cet article est de présenter les résultats obtenus lors du contrat entre le STCPMVN et le CERMICS sur le thème de l'étude de corrélation entre houle, marée et surcote. Dans ce travail l'accent a été mis sur l'obtention d'une procédure de dimensionnement. Cette procédure est basée sur la modélisation du couple (hauteur d'eau (D), hauteur significative de houle H). A partir de cette modélisation nous avons cherché à déterminer, dans le plan (D, H) , des régions englobant le couple $(D(t), H(t))$ tout au long de la durée de vie escomptée de l'ouvrage. Un des points essentiels que nous avons souhaité particulièrement éclairer est l'incidence de la corrélation entre D et H sur la détermination des conditions de projet à période de retour donnée.

1. Introduction

Le maître d'ouvrage de ce travail était le STCPMVN. (D.P.N.M., Ministère de l'Équipement). Ce contrat s'est déroulé parallèlement à un groupe de travail dans lequel sont intervenus M. Kovarik et M. Raout du STCPMVN, M. L'Her et M. Goasguen du STNMTE, M. Simon du SHOM, M. Teisson, M. Peltier, M. Galland de l'EDF-LNH. Nous les remercions ici d'avoir largement inspiré et orienté les travaux. Nous tenons, tout particulièrement, à remercier le STCPMVN pour avoir financé et suivi cette étude dans ses détails.

Lors de cette étude nous avons testé plusieurs modélisations différentes (analyses en composantes principales, divers modèles linéaires). Nous ne présenterons, dans ce qui suit, que l'une d'entre elles qui nous paraît la plus pertinente. Avant de la décrire en détail, nous commençons par résumer les points importants résultant de l'étude statistique générale de la loi du triplet (hauteur caractéristique de houle H , marée théorique M , surcote mesurée S).

Il existe une corrélation notable entre la hauteur caractéristique de houle H et la surcote S (ce point avait déjà été noté dans [2]). La corrélation entre H et la hauteur d'eau $D = M + S$ est nettement moins sensible, bien qu'appréciable statistiquement. Notons cependant que la corrélation entre H et S est difficile à utiliser

1. Ecole Nationale des Ponts et Chaussées, Département de Mathématiques

2. Ecole Nationale des Ponts et Chaussées, Centre d'Enseignement et de Recherche en Mathématiques, Informatique et Calcul Scientifique

concrètement en l'absence de critère de défaillance des ouvrages à dimensionner ne faisant intervenir que les variables H et S .

On peut exploiter la corrélation entre H et D pour le dimensionnement des ouvrages extérieurs des ports maritimes (digues), en se basant sur la formule de Goda, par exemple. L'impact de la corrélation est cependant peu sensible : on obtiendra des résultats proches en supposant H et D indépendants. L'idée d'un dimensionnement probabiliste à partir de la loi du couple (H, D) reste cependant intéressante (et très accessible techniquement) même lorsque les variables aléatoires H et D sont indépendantes.

La corrélation temporelle entre les données est très importante. Cette corrélation *doit* être utilisée pour identifier les modèles. Si la corrélation temporelle n'est pas essentielle lorsque l'on se pose des problèmes de dimensionnement, on peut, sans doute, exploiter cette caractéristique pour mettre en œuvre des systèmes d'alerte.

Pendant la durée de l'étude, nous nous sommes trouvés confrontés au problème des données disponibles : nous ne disposons de données que sur une période cumulée d'environ 1 an et nous souhaitons faire des prévisions sur des périodes de plusieurs dizaines d'années. Ce type de situation conduit à faire des prévisions ayant un degré de confiance très faible : il est souhaitable de posséder des données sur une période dont l'ordre de grandeur est comparable à la période de prévision. En particulier, la validation du modèle et de son exploitation pour le dimensionnement nécessite beaucoup plus de données que celles auxquelles nous avons pu utiliser.

2. Les données utilisées

Les données utilisées dans ce travail provenaient du site de Dieppe. Elles regroupaient les paramètres suivants : la hauteur significative de houle H représentant la moyenne du 1/3 supérieur des vagues; la hauteur maximum de houle; la marée calculée; la surcote instantanée. Nous disposons de 902 données comprises entre fin 85 et la mi-88. Elles sont largement lacunaires puisque l'on s'attend à avoir de l'ordre de $2,5 \times 365 \times 24 \approx 22000$ données toutes les heures sur cette période ! Notons que nous ne disposons pas de données sur certaines périodes, en particulier sur certaines saisons (notamment l'été). De ce fichier nous avons extrait des données toutes les 3 heures, sensiblement moins lacunaires. C'est sur ces dernières que nos modèles ont été calés.

Les limites que nous venons de souligner font que les résultats de notre travail doivent être pris avec prudence. Nous pensons cependant que la démarche décrite (ou des variantes) peut mener à une méthode de dimensionnement raisonnable si on l'utilise avec un jeu de données plus abondant.

3. Des points essentiels pour bâtir un modèle réaliste

A l'aide de ces données nous avons cherché à bâtir un modèle réaliste décrivant le triplet (H, M, S) . Avant de préciser le choix de notre modèle nous commençons par insister sur quelques points essentiels qui ont guidé ce choix.

Nécessité d'un modèle linéaire avec auto-corrélation Nous avons naturellement cherché à exploiter la corrélation entre les valeurs (D_t, H_t) (valeurs à l'instant t du couple (D, H)) en des instants successifs. Cette corrélation est claire si l'on songe à l'origine physique du phénomène. Dans un souci de simplicité nous nous sommes orientés vers un modèle linéaire avec autocorrélation pour le couple $(\log(H), D)$ (nous décrivons le modèle de façon plus précise dans le paragraphe 4). L'utilisation de $\log(H)$ provient de l'examen attentif des valeurs de H sur lesquelles nous avons pu ajuster de manière plus satisfaisante une loi lognormale que normale (voir, à ce propos, la figure 1.).

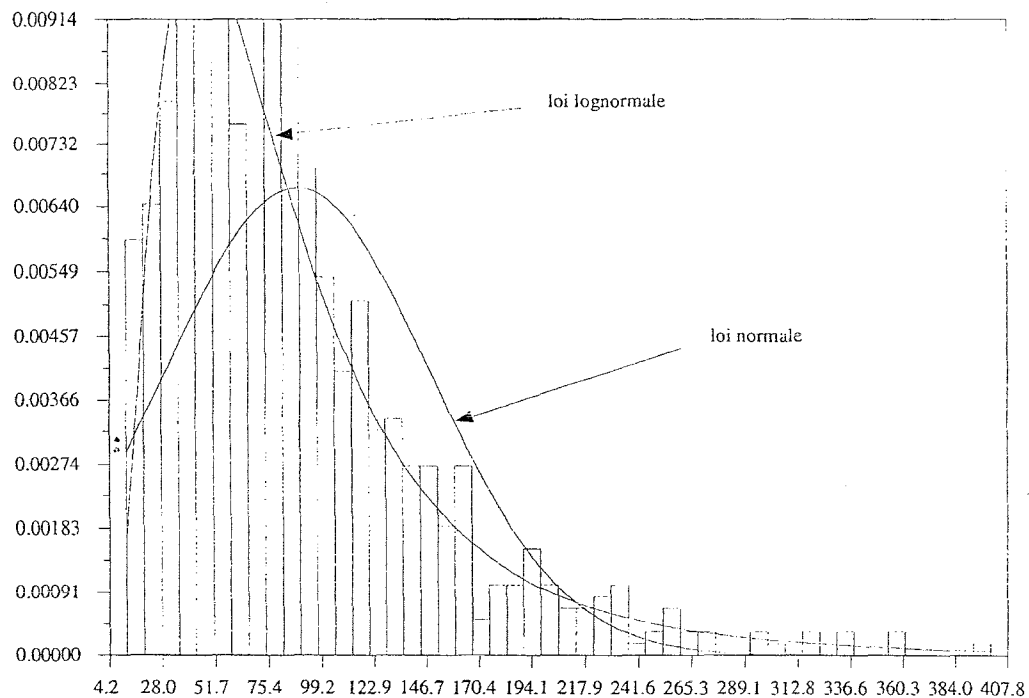


FIG. 1 – Loi de la H : approximation de normale et lognormale

Les formules de dimensionnement font intervenir le couple (H, D) La plupart des formules de dimensionnement classiques utilisent la formule de Goda (voir [3]). Cette formule fait intervenir la hauteur totale d'eau $D = M + S$ (et non de la surcote S). Ce fait nous a conduit à privilégier une approche cherchant

à modéliser le couple (H, D) et non le couple (H, S) (même si la corrélation est nettement plus importante sur le couple (H, S) que sur le couple (H, D)).

Le problème de la dépendance temporelle Il importe de soulever un point qui rend délicate l'exploitation des résultats. Le but de l'étude est de déterminer dans le plan (H, D) un domaine dont le point représentatif (H_t, D_t) s'échappe peu, c'est-à-dire avec une probabilité que l'on estime acceptable, pendant la durée de vie escomptée de l'ouvrage. Il faut souligner que, conformément à l'intuition, la probabilité de séjourner pendant une période longue dans un domaine déterminé peut être beaucoup plus faible que la probabilité de s'y trouver à un instant donné.

L'indépendance complète peut être réfutée. L'hypothèse d'indépendance entre deux instants successifs distants de 3 heures est clairement fausse. Elle est intuitivement très pénalisante dans l'estimation de la probabilité de séjour car elle ne prend pas en compte les corrélations qui ont tendance à éviter une évolution trop erratique des valeurs. Cette hypothèse permet toutefois d'indifier une zone de sécurité (qui sera sans doute plus grande que le strict nécessaire). Notons N le nombre de pas de temps t correspondant à la période que l'on veut étudier, c'est-à-dire la durée de vie escomptée de l'ouvrage. Par exemple, pour une période de 20 ans, le nombre de pas de 3 heures vaut $N = 8 \times 365 \times 20 = 58400$. L'hypothèse d'indépendance conduit à choisir une zone associée telle que la probabilité de ne pas s'y trouver est extrêmement faible. Un calcul simple montre que cette probabilité est d'environ d'une chance sur un million ! On appellera, dans la suite, cette probabilité la *probabilité d'évasion* de la zone.

Quelle que soit l'option dimensionnante retenue, le faible nombre de données rend très difficile l'estimation de probabilités inférieures à 10^{-3} . Aussi proposons-nous une autre approche, par simulation, qui dans le cadre du modèle retenu, ne nécessite que l'estimation de quantités plus en rapport avec le nombre de données dont nous disposons. Pour une utilisation opérationnelle de ce travail, il serait souhaitable de disposer d'un échantillon permettant d'estimer ce niveau de probabilité (un échantillon de l'ordre de quelques milliers d'observations espacées de trois heures devrait convenir).

Nécessité d'une approche par simulation L'extrême faiblesse de la probabilité d'évasion de l'exemple précédent est essentiellement due à l'hypothèse d'indépendance entre les pas de temps. Pour contourner ce problème, nous avons proposé une approche par simulation, qui consiste à estimer directement dans le cadre d'un modèle simple la loi du maximum de D_t sur un an notée D_{\max} , puis en notant τ l'instant où ce maximum est atteint, la loi de H_τ conditionnelle en $D_\tau = D_{\max}$. L'indépendance entre D_t et le résidu du modèle linéaire permet d'évaluer cette loi comme si τ était un instant déterministe. Pour mener à bien la simulation, nous calculons pour chaque instant t la valeur théorique de la marée M_t , à laquelle nous ajoutons la valeur de S_t simulée suivant le modèle choisi. Pour répondre à la question du dimensionnement, nous avons supposé que les années successives étaient indépendantes, ce qui paraît réaliste. Avec cette méthode, pour obtenir un temps de retour moyen de 20 ans, nous sommes conduits à choisir un domaine de probabilité d'évasion de $1/20 = 5 \cdot 10^{-2}$. Cette valeur est nettement plus raisonnable que dans

le cas “indépendant”.

4. Un modèle et une procédure pour le dimensionnement

L'ensemble des considérations précédentes nous a conduit à choisir le modèle suivant, qui s'il reste critiquable, fait apparaître les caractéristiques principales du phénomène.

Le modèle d'évolution du triplet (H_t, M_t, S_t) Le modèle est le suivant :

$$\begin{cases} \log(H_t) &= \beta_0 + \beta_1 D_t + u_t \\ u_t &= \rho u_{t-1} + \varepsilon_t \end{cases} \quad (1)$$

où ε_t est supposé être un bruit blanc gaussien dont on a estimé la variance. On sait alors estimer les coefficients ρ , β_0 et β_1 par une procédure itérative classique dite de Cochran Orcut (voir [5] ou [4] pour des précisions sur cet algorithme). Les divers coefficients ainsi que la variance du bruit intervenant dans le modèle ont été estimés à partir des données. Nous avons, de plus, constaté l'indépendance statistique entre D_t et les résidus u_t . Nous avons ainsi pu identifier la loi de H_t conditionnellement en D_t comme une loi gaussienne de moyenne $\alpha + \beta D_t$ et de variance σ^2 , les paramètres α , β et σ pouvant être estimés sur les données.

Afin de pouvoir mener à bien notre procédure de simulation de D_{\max} nous avons fixé un modèle d'évolution pour le paramètre S_t . Le modèle le plus simple permettant de rendre compte des fortes corrélations temporelles constatées est un modèle auto-régressif d'ordre 1, c'est à dire du type :

$$S_t = \gamma_0 + \gamma_1 S_{t-1} + \eta_t, \quad (2)$$

en supposant que η_t est un bruit blanc gaussien dont on a estimé la variance sur les données.

A l'aide du modèle précédent, on peut effectuer les calculs suivants qui visent à identifier la loi du couple (D, H) .

1. *Estimation de la distribution statistique de D_{\max}* : la surcote S étant considérée comme aléatoire et M étant assimilée à un phénomène déterministe, on peut déterminer la distribution statistique de la hauteur d'eau maximale sur une année D_{\max} par simulation (en utilisant le modèle de surcote (2) et la hauteur d'eau théorique) sous la forme d'une densité $f_1(d)$ que nous avons supposée gaussienne. Il s'agit, bien sûr, d'une approximation mais elle nous paraît tout à fait réaliste dans le cadre de nos hypothèses et elle conduit à des calculs très simples.
2. *Estimation de la loi conditionnelle de H en fonction de la hauteur d'eau D* : Le modèle linéaire expliquant $\log(H)$ par D choisi est (1). En s'en tenant aux modèles linéaires avec perturbation centrée, gaussienne d'écart-type inconnu σ , nous pouvons estimer la loi de probabilité de cette variable H conditionnellement à $D = d$, soit $f_2^d(h)$.

La densité de la loi du couple (D_{\max}, H) , $f(d, h)$ s'exprime alors sous la forme :

$$f(d, h) = f_1(d)f_2^d(h).$$

La procédure de dimensionnement. Une fois la loi du couple (D, H) déterminée, la méthode de dimensionnement que nous proposons est la suivante :

1. *Détermination des courbes de niveau de la loi de probabilité du couple (D_{\max}, H) .* Dans le plan (D_{\max}, H) , nous avons déterminé les points tels que :

$$f(d, h) = z \text{ où } z \text{ est une constante positive.}$$

Ceci nous donne une famille de courbe de niveau dans le plan (D_{\max}, H) , $A_z = \{(d, h), f(d, h) = z\}$ et une famille de zones C_z , enserrées par les courbes de niveau :

$$C_z = \cup_{y \leq z} A_y = \{(d, h), f(d, h) \leq z\}$$

2. *Détermination d'un domaine de probabilité d'évasion p .* Toujours dans le plan (D_{\max}, H) on a identifié un domaine C_z tel que $\mathbf{P}((D, H) \in C_z) = 1 - p$.

On obtient, ainsi, pour une probabilité d'évasion fixée p (ou si l'on préfère pour un temps de retour $1/p$ vu l'hypothèse d'indépendance), une zone de confiance caractérisée par p . Nous donnons dans ce qui suit quelques résultats obtenus par cette méthode.

5. Un exemple de résultats pour le couple $(\log(H), D_{\max})$

Nous avons donc identifié un modèle linéaire de type (1) expliquant $\log(H)$ par D_{\max} . L'hypothèse de nullité des coefficients β_0, β_1 a pu être rejetée sans hésitation. La valeur du test de Durbin-Watson (voir [5] pu [4]) pour ϵ_t est ici proche de 2, on a donc pu considérer la décorrélation comme très satisfaisante. La loi conditionnelle de $\log(H)$ sachant D_{\max} est alors gaussienne et s'écrit très simplement en fonction de β_0 et β_1 . On utilise, de plus, la méthode de simulation précédemment décrite pour estimer la loi de D_{\max} comme une variable gaussienne. Ceci conduit à des zones de confiance de temps de retour fixés données dans la figure 2.

Sur cette figure nous avons tracé des courbes telles que la houle de Goda est constante (iso-Goda). Pour cela nous avons utilisé la formule de Goda tirée de [3]. Cette formule est utilisée pour calculer une hauteur de houle de dimensionnement des digues marines. Elle dépend non seulement de D la hauteur totale d'eau, de H la hauteur caractéristique de houle, mais aussi de T la période de houle et de i la pente du talus. Nous avons choisi de fixer les valeurs de i et T de la façon suivante : la pente vaut forfaitairement $i = 3\%$; la période T est calculée en fonction de la hauteur H à l'aide d'une fonction déterministe $T = f(H)$. Cette fonction a été estimée à l'aide de la loi conditionnelle empirique de T sachant H par maximum

de vraisemblance et en supposant f linéaire. Ceci nous a conduit à choisir : $T = f(H) = 1.49H + 2.70$, si H est exprimé en mètres et T en secondes.

Sur la figure 2, nous avons superposé les courbes de Goda et les zones de probabilité pour des temps de retour de 5, 10, 20, 50, 100 définies à partir de la loi estimée du couple (H, D_{\max}) . Nous avons, de plus, fait figurer les valeurs de H associées aux périodes de retour des coordonnées, sous forme de droites horizontales. Nous avons utilisé comme droites horizontales les valeurs de H correspondant à des périodes de retour de 1 et 5 ans, calculées sous l'hypothèse lognormale. Les valeurs de retour pour H à 10 ans (7.68 m), 20 ans (8.71 m), 50 ans (10.20 m) et 100 ans (11.43 m) sont trop grandes pour figurer sur le graphique. Les droites verticales sont calculées sur la loi de D_{\max} sous les hypothèses que D_{\max} gaussien et que les années sont indépendantes. On a, enfin, superposé les nuages de points mesurés (D, H) (et non (D_{\max}, H) qui sont des points simulés et non mesurés).

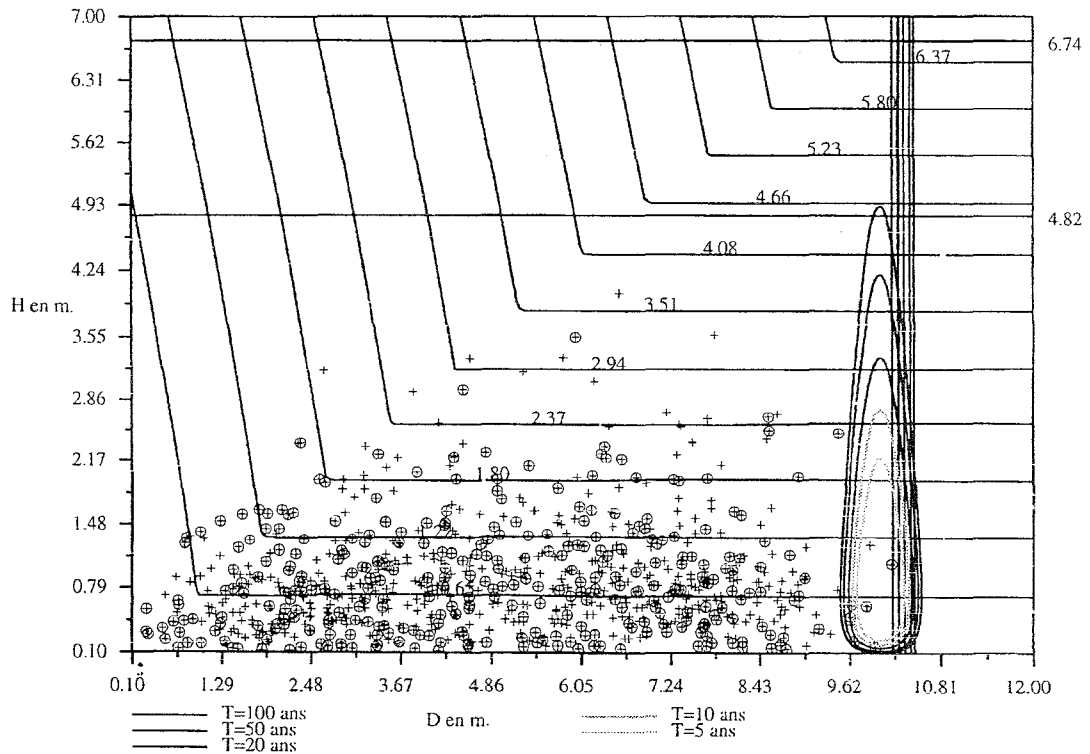


FIG. 2 – Courbes de Goda dans le plan (D_{\max}, H) , période de retour de 5, 10, 20 et 50 ans

6. Limites de l'approche et perspectives

Il paraît important de rappeler que ce travail doit être envisagé plutôt comme l'ébauche d'une méthode que présentant des résultats définitifs, la faiblesse de

l'échantillon des données ne permettant pas de valider ces derniers de façon satisfaisante. Notre procédure d'identification de D_{\max} , en particulier, si elle nous paraît la seule accessible techniquement sur nos données, est sans doute critiquable et devrait être conduite sur des données réelles et non par simulation.

On peut envisager de prolonger et de perfectionner le travail décrit de plusieurs façons. Il est certain que les données météorologiques jouent un rôle important dans le phénomène physique. Leur intégration dans les modèles est donc souhaitable.

Nous n'avons pas cherché à identifier la loi de l'évolution du triplet (H, M, S) . Ceci est techniquement possible en fixant une classe de modèles susceptible de décrire le phénomène. On peut, ainsi, envisager d'utiliser et perfectionner notre modèle pour une utilisation en alerte.

Enfin, si l'on doit mettre l'accent sur la loi du couple (H, D) , il est important d'étudier et d'identifier directement une loi de valeur extrême pour le couple s'inspirant de la méthode de renouvellement utilisée dans le cas unidimensionnel.

Références

- [1] Études des corrélations entre houles et surcotes, Le cas de Dieppe et de Bayonne, STNMTE, Division Océanographie, février 1995.
- [2] B. Simon, Statistique de niveaux marins extrêmes le long des côtes de France, Notice STC ER PM 94/01 (STCPMVN).
- [3] Goda, Random seas and design of maritime structures, University of Tokyo Press, 1985.
- [4] V. Cohen "Analyse des données et Econométrie", Notes de cours ENPC, Paris 1995-1996
- [5] J. Johnston "Econometric Methods" McGraw Hill, Tokyo 1972
- [6] Etude des corrélations statistiques entre la houle, la marée et les surcotes, Application au site de Dieppe, Rapport du Contrat CERMICS/STPMVN, 1997.