

Méthodologie de combinaison d'informations statistiques et d'avis d'experts dans le cadre de la théorie de Dempster et Shafer : Application au dimensionnement des ouvrages côtiers

Nadia Ben Abdallah¹ Nassima Mouhous Voyneau² et Thierry Denoeux³

Résumé L'estimation des niveaux marins et des houles extrêmes associés à des périodes de retour de l'ordre du siècle revêt une importance majeure dans la conception des ouvrages de protection côtière et les analyses de risque d'inondation. L'approche commune pour l'estimation de ces niveaux consiste en une inférence statistique à partir des mesures disponibles, en supposant que la distribution des extrêmes est stationnaire. Cependant, durant les dernières décennies, une attention croissante a été accordée à l'intégration de l'effet du changement climatique dans les analyses de risques environnementaux. L'adaptation des ouvrages de protection au réchauffement climatique implique la combinaison d'une information additionnelle aux observations historiques dans l'estimation du forçage hydraulique lors du dimensionnement. L'information sur les prévisions des impacts climatiques futurs sur le niveau marin et la houle, fournie par des experts, est généralement très incertaine du fait de la compréhension très limitée des processus physiques qui régissent le climat.

Dans cet article, nous présentons une méthodologie basée sur la théorie des fonctions de croyance pour représenter et combiner des informations statistiques et des opinions d'experts, afin d'estimer un niveau de dimensionnement de période de retour de 100 ans. On représente l'information statistique par des fonctions de croyance construites à partir de la vraisemblance [7], et les projections des experts de l'élévation du niveau marin futur (SLR) par des distributions de possibilité trapézoïdales. Une simulation de Monte Carlo permet ensuite de combiner ces

¹ Nadia Ben Abdallah
Université de Technologie de Compiègne, CNRS, UMR 7253 Heudiasyc
nadia.ben-abdallah@hds.utc.fr

² Nassima Mouhous-Voyneau
Université de Technologie de Compiègne, Avenues-GSU
nassima.voyneau@utc.fr

³ Denoeux Thierry
Université de Technologie de Compiègne, CNRS, UMR 7253 Heudiasyc
thierry.denoeux@hds.utc.fr

deux éléments d'information pour fournir une estimation du niveau futur d'adaptation ainsi que le niveau d'incertitude associé.

1. Introduction

Une analyse des incertitudes consistant à identifier, représenter et propager les incertitudes devrait être une partie intégrante dans les procédures d'évaluation des risques et du dimensionnement afin de fournir des éléments fiables d'aide à la décision. En hydrologie, communiquer l'incertitude autour du risque d'inondation futur aux décideurs devient une pratique courante [1, 12]. S'il existe un consensus général sur la nature des sources d'incertitude les plus pertinentes dans une analyse de risque d'inondation, le cadre mathématique le plus adéquat pour quantifier ces incertitudes fait aujourd'hui débat parmi les analystes. Le cadre probabiliste communément utilisé a été sujet à des critiques concernant notamment sa façon de traiter indifféremment les incertitudes aléatoires, qui caractérisent la variabilité naturelle, et les incertitudes épistémiques qui elles caractérisent une connaissance limitée [1]. Sachant que, dans les analyses de risque environnementaux, les incertitudes émanent de sources différentes (avis d'experts, observations, modèles...), le besoin de cadres alternatifs qui reconnaissent la différence de nature entre les types d'incertitudes et les abordant différemment a émergé. Des recherches importantes sont menées depuis quelques décennies pour étudier la pertinence de nouvelles approches telles que la théorie floue, les probabilités imprécises ou encore la théorie de Dempster–Shafer dans les analyses de risques et de fiabilité.

Dans cet article, nous nous intéressons à la représentation de l'incertitude relative à l'estimation d'un paramètre de dimensionnement des ouvrages de protection côtière dans un contexte de changement climatique, dont un des principaux impacts est l'élévation du niveau marin moyen. Les éléments d'évidence disponibles proviennent de deux sources différentes. La première est un ensemble d'observations statistiques traditionnellement traitées par des outils fréquentistes ou bayésiens dont la pertinence est sujet à des critiques diverses [5, 8]. La seconde consiste en des projections des impacts du changement climatiques, notamment l'élévation du marin moyen, fournies par des experts du climat. La compréhension limitée du système climatique et de son évolution ainsi que l'ignorance du développement humain futur mènent, comme sera nous le verrons, à un haut niveau d'incertitude attachée aux projections fournies dans la littérature. On se propose de représenter et combiner les deux sources d'évidence dans le cadre de la théorie des fonctions de croyances qui constitue un cadre mathématique adapté à la fois à la représentation d'avis d'experts et d'informations statistiques.

L'article est organisé comme suit. Dans la première section, nous présentons les principaux éléments pour le dimensionnement des ouvrages côtiers et rappellerons les fondements des statistiques des valeurs extrêmes utilisées en hydrologie. Nous abordons ensuite brièvement la question des impacts du changement climatique et présentons quelques projections sur la hausse de niveau de la mer future existant dans la littérature. Dans la deuxième partie, nous justifions et expliquons l'inférence basée sur la vraisemblance pour représenter l'évidence statistique et abordons brièvement son extension au cadre des fonctions de croyance. Finalement, la dernière partie décrit l'application de la méthodologie à l'estimation du niveau de retour centennal pour le site du Havre (France) et récapitule les résultats principaux.

2. Éléments pour l'adaptation au changement climatique

Les structures de protection côtière doivent résister aux événements exceptionnels et leur conception nécessite l'estimation des niveaux extrême de mer et de houles qui ont rarement, ou jamais été atteints. L'outil principal pour modéliser les événements extrêmes est la théorie des valeurs extrêmes (TVE). Cette théorie est basée sur l'hypothèse que les valeurs maximales d'un échantillon indépendant et identiquement distribué (i.i.d) suivent asymptotiquement une loi particulière dans la famille dite des valeurs extrêmes généralisées (GEV) [11]. La fonction de répartition d'une loi GEV est de la forme :

$$F(z, \mu, \sigma, \xi) = \begin{cases} \exp\left(-\left[1 - \xi \frac{z-\mu}{\sigma}\right]^{\frac{1}{\xi}}\right) & \text{si } \xi \neq 0 \\ \exp\left(-\exp\left[-\frac{z-\mu}{\sigma}\right]\right) & \text{si } \xi = 0, \end{cases} \quad (1)$$

où $\mu, \sigma > 0, \xi$ sont, respectivement, les paramètres de location, d'échelle et de forme. Selon le signe du paramètre ξ , la distribution est dite de Fréchet ($\xi > 0$), Weibull ($\xi < 0$) ou Gumbel ($\xi = 0$).

Dans la théorie des valeurs extrêmes, la probabilité qu'une variable Z dépasse une certaine valeur z est communément exprimée en termes de période de retour notée T et définie comme étant le nombre d'années moyen séparant deux dépassements successifs de la valeur z . Si la méthode d'échantillonnage des extrêmes est la méthode des maxima annuels (consistant à ne garder qu'une seule valeur par année), alors la période de retour d'un niveau z est directement reliée à sa probabilité annuelle de non dépassement p par la relation : $T = 1/(1-p)$. Ainsi,

découle de (1) l'expression suivante du niveau de retour z_T de période de retour T :

$$z_T = \begin{cases} \mu - \frac{\sigma}{\xi} \left[1 - \left(-\text{Log} \left(1 - \frac{1}{T} \right) \right)^{-\xi} \right] & \text{si } \xi \neq 0 \\ \mu - \sigma \text{Log} \left(-\text{Log} \left(1 - \frac{1}{T} \right) \right) & \text{si } \xi = 0. \end{cases} \quad (2)$$

Les quantiles z_T sont déterminés à partir des mesures historiques de la variable Z qui est, dans notre cas d'étude, le niveau marin.

Les structures de défense dans les régions côtières sont généralement conçues pour résister à des événements extrêmes avec une période de retour de 100 ans. Cependant, à cause du changement climatique, les structures seront soumises durant leur cycle de vie à des forçages hydrauliques plus importants. Le principal impact climatique est lié à l'élévation du niveau moyen des mers qui affecte la fréquence et l'intensité des niveaux extrêmes. Pour des besoins d'adaptation, les quantiles extrêmes du niveau marin z_T devront intégrer les projections de l'élévation du niveau marin moyen sur la même période ; le niveau de dimensionnement, considérant le réchauffement, sera alors : $z'_T = z_T + SLR$, où SLR (Sea level Rise) représente l'augmentation du niveau moyen de la mer sur la période T .

Dans son dernier rapport d'évaluation [10], le GIEC (Groupe Intergouvernemental des Experts du Climat) estime que l'élévation du niveau marin moyen sur la période 1990-2095 sera comprise entre 0.18 m et 0.79 m; il est indiqué cependant que des valeurs supérieures à 0.79 ne sont pas exclues. Cet intervalle prend en compte les incertitudes sur les émissions futures des gaz à effet de serre associées à différents scénarios d'émission définie par le GIEC, (scénarios qui couvrent une large gamme d'états du monde futurs caractérisés par des niveaux divers de développement technologique, économique et énergétique), sur les modèles climatiques dits de circulation globale qui estiment les températures futures ainsi que sur les modèles d'impact (fonte de l' Antarctique et du Groenland, l' expansion des océans, etc.).

Depuis la parution du dernier rapport du GIEC, d'autres projections de l'élévation du niveau des mers basées sur des modèles semi-empiriques ont été proposées, fournissant souvent des prévisions plus pessimistes pour la fin du siècle. Ainsi, Rahmstorf [15], se basant sur un modèle statistique simple, obtient l'intervalle [0.5, 1.4]. D'autres part, des études plus récentes montrent qu'il existe une limite physique à l'élévation induite par la fonte massive des glaciers : le seuil de 2 m ne pourrait pas être dépassé avant la fin de ce siècle [13].

Les méthodes traditionnelles utilisées pour intégrer l'élévation du niveau marin dans les analyses de risque d'inondation et la conception des structures ont négligé la grande incertitude sur ce paramètre et ont été basées sur une valeur fixée (la

valeur moyenne sur les projections du GIEC, la valeur maximale, etc.). Cependant, comme il a été démontré par Purvis [14], qui a mené une analyse de risque d'inondation en intégrant le changement climatique, il s'avère que l'approche déterministe dans la description de l'élévation du niveau moyen pourrait significativement sous-estimer les conséquences en termes d'inondation et de coûts et mènerait ainsi à des prises de décisions inconsistantes.

Pour estimer le niveau de dimensionnement z'_T , on procède en deux étapes: en premier lieu, on représente la connaissance du niveau z_T obtenue à partir des mesures disponibles dans le formalisme des fonctions de croyance. On représente également dans le même cadre les avis d'experts sur la composante SLR . La seconde étape consiste à combiner les deux mesures pour évaluer le niveau d'adaptation z'_T ($z'_T = z_T + SLR$).

3. Représentation de l'information statistique

Pour une période de retour donnée, le niveau z_T est obtenu à partir de la relation (2) en remplaçant les paramètres de la loi d'ajustement par leurs meilleures estimations. Ces paramètres sont estimés par des méthodes fréquentistes basées sur des propriétés asymptotiques et leurs performances sont limitées dans le cas de séries de mesures courtes. Comme pour les estimations, les intervalles de confiance supposés donner une information sur l'incertitude des estimations des paramètres sont parfois peu fiables à cause notamment d'approximations dans le calcul des bornes inférieure et supérieure [18]. En effet, le principe du calcul des intervalles est basé sur l'hypothèse de l'échantillonnage répété qui consiste à répéter hypothétiquement l'expérience et calculer indépendamment les bornes de l'intervalle. Dans des cas où la cette hypothèse n'est pas justifiable, cette approche peut être remise en question et l'utilisation d'approches alternatives dans la représentation de l'incertitude des paramètres estimés statistiquement justifié.

Fisher [8], Cox [5], Barnard *et al.* [3] et Edwards [7] ont critiqué l'approche fréquentiste basée sur les tests de significations et les intervalles de confiance. Ils ont proposé une approche alternative basée uniquement sur la vraisemblance comme unique et totale description de l'évidence disponible et utilisable pour l'inférence. Ce principe, appelé "principe de vraisemblance", stipule que, étant donné une observation X représentée par un modèle statistique paramétré $f(X; \theta)$, l'information relative sur le paramètre inconnu θ ($\theta \in \Theta$) est entièrement contenue dans la fonction de vraisemblance $L(\theta; X) = f(X; \theta)$.

L'extension de l'inférence statistique du cadre bayésien au cadre de la théorie des fonctions de croyance est motivée par le fait que les fonctions de croyance constituent un ensemble plus riche que les mesures de probabilités.

Shafer [17] a été le premier à proposer de représenter le paramètre à inférer par une fonction de croyance consonante. L'idée de Shafer a été justifiée axiomatiquement par Wasserman [19] ; des arguments supplémentaires en sa faveur ont été fournis par Aickin [2]. Fisher [8] interprète la fonction de vraisemblance comme une expression de la plausibilité relative du paramètre quand aucune autre information, à part les observations, n'est disponible. Il paraît alors raisonnable de définir la fonction contour de plausibilité, qui peut être interprétée comme une mesure de crédibilité, par :

$$pl(\theta; X) = \frac{L(\theta; X)}{L(\hat{\theta}; X)}, \quad (3)$$

où $\hat{\theta}$ est l'estimateur de maximum de vraisemblance (MLE) de θ .

La fonction contour (3) a une interprétation simple : $pl(\theta; x)$ représente la probabilité relative d'observer x si la valeur du paramètre est θ (la probabilité maximale étant obtenue pour $\theta = \hat{\theta}$). Une valeur du paramètre de faible plausibilité, par exemple 0.001, indique qu'il existe d'autres valeurs de θ pour lesquels la probabilité d'observer x est 1000 fois plus grande.

Rappelons ici la définition dans le cadre de la théorie des croyances de la mesure de plausibilité. Si l'ensemble des valeurs possible d'un paramètre θ est Θ , alors on attribue à chaque sous-ensemble $A \subset \Theta$, via une fonction de masse $m: 2^\Theta \rightarrow [0,1]$, une mesure du degré de confiance en l'affirmation 'La vraie valeur du paramètre θ est exactement dans A que l'on note $m(A)$. Les ensembles pour lesquels la masse est non nulle sont appelés éléments focaux.

Une fonction de masse dont les éléments focaux sont linéairement ordonnés par inclusion est dite consonante.

On associe à la fonction de masse m deux mesures : une fonction de croyance (notée Bel , comme *Belief*) et une fonction de plausibilité (notée Pl), définies sur 2^Ω dans $[0, 1]$ comme suit :

$$\begin{aligned} Bel(A) &= \sum_{B \subseteq A} m(B) \\ Pl(A) &= \sum_{A \cap B \neq \emptyset} m(B). \end{aligned} \quad (4)$$

$Bel(A)$ s'interprète comme un degré de support en l'hypothèse A , tandis que $Pl(A)$ indique dans quelle mesure les éléments d'évidence ne discréditent pas cette hypothèse.

Dans le cas où la fonction de masse est consonante, la plausibilité est reliée à la plausibilité contour $pl(pl(\theta) = Pl(\theta))$ par la relation :

$$Pl(A) = \sup\{pl(\theta; X); \theta \in A\}. \quad (5)$$

4. Application et Résultats

On se propose d'appliquer l'inférence étendue au cadre de Dempster –Shafer pour l'estimation du paramètre z_{100} à partir d'un échantillon d'observations correspondant à 15 années de mesures horaires du niveau marin (mesures recueillies par la bouée du port du Havre). Ce paramètre est estimé en ajustant les maxima annuels de l'échantillon à la loi de Gumbel (2), μ étant considéré comme le paramètre structurel et σ comme le paramètre de nuisance. La fonction contour de z_{100} est représentée dans la Figure 1. La valeur la plus plausible de ce paramètre est caractérisée par un niveau de plausibilité égal à 1 et correspond à l'estimateur du maximum de vraisemblance.

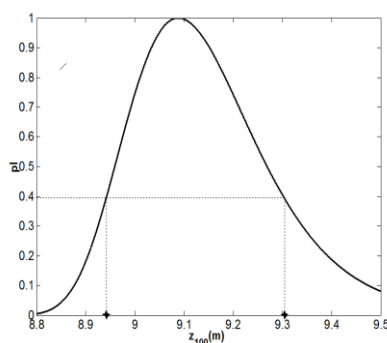


Fig.1 Mesure de plausibilité du paramètre Z_{100} .

En second lieu, on s'intéresse à intégrer l'effet du réchauffement en terme d'élévation du niveau marin SLR et à évaluer le niveau d'adaptation z'_{100} . Pour cela, on représente l'évidence disponible fournie par les experts sur le paramètre SLR. Comme sources d'information dans la littérature, les projections du GIEC et celles de Rahmstorf [15], ainsi que les limites physiques préconisées par Pfeffer [13] ont été retenues comme les plus pertinentes. Elles sont d'ailleurs les plus utilisées dans les analyses de risques environnementaux incorporant la composante due au changement climatique. On formalise chacun des trois éléments d'évidence par une fonction de plausibilité trapézoïdale (Figure 2) : une telle mesure, par exemple pour les estimations du GIEC ([0.18, 0.79]), attribuera une valeur de 1 à chaque valeur dans l'intervalle [0.18,0.79], une valeur de 0 en dehors de l'intervalle physiquement admissible (i.e. [0,2]) et le résultat d'une interpolation linéaire pour le reste des valeurs dans [0,2]. Puisque les deux sources sont fiables, une règle de combinaison conjonctive est applicable. Parmi celles-ci, le minimum est le plus approprié quand les avis des experts sont basés sur une

connaissance commune : ainsi, le niveau de plausibilité d'un niveau donné du SLR sera le minimum des deux mesures de plausibilités attribuées par chaque expert.

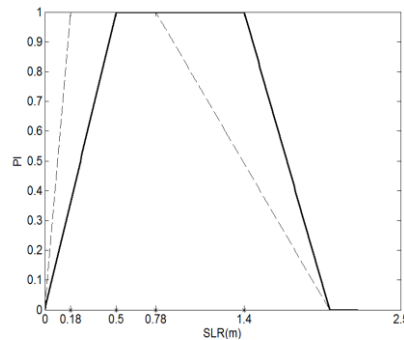


Fig.2 Mesures de plausibilité sur le paramètre SLR (en ligne continue épaisse : inférence basée sur les projections de Rahmstorf ; en ligne discontinue : inférence basée sur les projections du GIEC).

Finalement, on évalue les fonctions de croyance du niveau d'adaptation $z'_{100} = z_{100} + SLR$ via une simulation type Monte Carlo. Cette procédure consiste en un tirage aléatoire de niveaux de plausibilité α et ω dans $[0,1]$. Soient $[\underline{z}_{100}^{\alpha}, \overline{z}_{100}^{\alpha}]$ et $[\underline{SLR}^{\omega}, \overline{SLR}^{\omega}]$ les coupes correspondantes. Utilisant le fait que les deux paramètres z_T et SLR sont indépendants, le niveau z'_{100} associé sera dans $[\underline{z}_{100}^{\alpha} + \underline{SLR}^{\omega}, \overline{z}_{100}^{\alpha} + \overline{SLR}^{\omega}]$. Cette procédure est répétée un millier de fois pour obtenir mille intervalles sur z'_{100} . Il est alors possible d'évaluer les fonctions de plausibilité et de croyance cumulées (d'après les définitions dans (4)) comme suit : la plausibilité cumulée d'un niveau fixé correspond à la fréquence relative, calculée sur les 1000 intervalles tirés aléatoirement, de l'évènement : "la borne inférieure de l'intervalle est en dessous du niveau fixé", alors que la croyance cumulée correspond à la fréquence relative de l'évènement : "la borne supérieure de l'intervalle est en dessous du niveau fixé". La Figure 3 visualise les fonctions de plausibilité et de croyance cumulées du niveau de dimensionnement z_{100} (sans considérer l'effet du réchauffement) et celui intégrant le SLR, z'_{100} . La courbe supérieure correspond à la plausibilité et la courbe inférieure à la croyance.

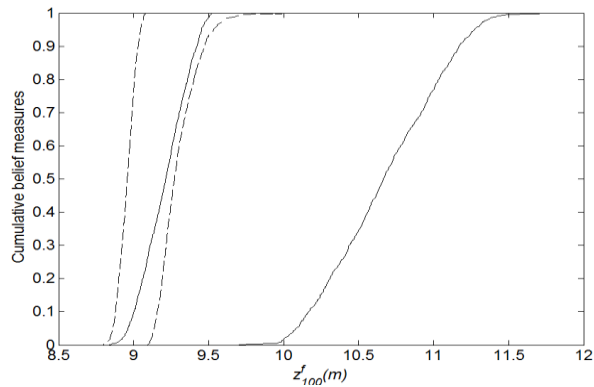


Fig. 3 Fonctions cumulées pour les niveaux z_{100} (trait discontinu) et z'_{100} (trait continu) : la distribution inférieure est la croyance, la supérieure est la plausibilité.

La surface entre les deux distributions de croyance et de plausibilité peut être interprétée comme une mesure du niveau d'incertitude sur le paramètre en question. Quand le changement climatique est considéré dans l'estimation du niveau de dimensionnement sur une période de 100 ans, cette incertitude devient très importante du fait de la très grande incertitude sur le paramètre SLR.

5. Conclusion

La généralité et la flexibilité de la théorie des fonctions de croyance la rendent adaptée pour représenter et combiner des éléments d'évidence provenant de sources diverses. Dans cet article, nous avons présenté une méthodologie qui permet de combiner des informations statistiques et des avis d'experts et nous avons illustré son application dans l'estimation du niveau de retour marin en tenant compte du changement climatique. Les travaux en cours visent à prendre en compte d'autres phénomènes, comme la houle, et à propager l'incertitude dans des modèles hydrauliques et de dimensionnement des ouvrages.

References

1. Apel, H., Thikeken, A.H.: Flood risk assessment and associated uncertainty. *Natural Hazards and Earth System Sciences*. **4**, 295-308 (2004)
2. Aickin, M.: Connecting Dempster-Shafer belief functions with likelihood based inference. *Synthese*. **123**, 347-364 (2000)
3. Barnard, G. A., Jenkins, G. M. and Winsten, C. B. Likelihood inference and time series. *Journal of the Royal Statistical Society*, 125(3) : 321-372, 1962.

4. Coles, S. G., Dixon, M. J.: Likelihood based inference for extreme value models. *Extremes*. **2**, 5-23 (1999)
5. Cox, D.R. : Some problems connected with statistical inference. *Ann. Math. Statistics*. **29**, 357-372 (1958)
6. Denoeux, T.: Maximum likelihood estimation from uncertain data in the belief function framework. *IEEE Transactions on Knowledge and Data Engineering* (à paraître), doi: 10.1109/TKDE.2011.201.
7. Edwards, A.W.F.: Likelihood. University Press, Baltimore, USA (1972)
8. Fisher, R. A : Inverse Probability and the use of likelihood. *Proceedings of the Cambridge Philosophical Society*. **28**, 257-261. CP3 (1932)
9. Gumbel, E. J.: The statistics of extremes. New York, Colombia University Press (1958)
10. IPCC Forth Assessment Report (2007), http://www.ipcc.ch/publications_and_data/publications_and_data_reports.shtml
11. Jenkinson, A.F.: The frequency distribution of the annual maximum (or minimum) values of meteorological elements, *Quart. J. Roy. Meteo. Soc.* **81**, 158-171 (1955)
12. Merz, B. and al.: Flood risk curves and uncertainty bounds. *Natural hazards*. **51**, 437-458 (2009)
13. Pfeffer, W., Harper, J. and O'Neel, S.: Kinematic constraints on glacier contribution to 21st century sea level rise. *Science*. **321**, 1340-143, doi: 10.1126/science.1159099 (2008)
14. Purvis, M.: Probabilistic methodology to estimate future coastal flood risk due to sea level rise. *Coastal Engineering*. **55**, 1062-1073 (2008)
15. Rahmstorf, S.: A semi empirical approach to projecting future sea level rise. *Science*. **315**, 368-370 (2007)
16. Shafer, G. : A mathematical Theory of Evidence, Princeton University Press (1976)
17. Shafer, G.: Belief Functions and Parametric Models, *Journal of the Royal Statistical Society.Series B*. **44**, 322-352 (1982)
18. Xu, P. and al.: Uncertainty analysis in statistical modeling of extreme hydrological events. *Stochastic environmental research and risk assessment*. **24**, 567-578 (2010)
19. Wasserman, L.: Belief functions and statistical inference. *The Canadian journal of Statistics*. **18**, 183-196 (1990)