



HAL
open science

La distribution de Weibull, un outil utile ou dépassé? Application au confort au vent

Mateusz Bogdan, Edouard Walther, Antoine Hubert

► To cite this version:

Mateusz Bogdan, Edouard Walther, Antoine Hubert. La distribution de Weibull, un outil utile ou dépassé? Application au confort au vent. IBPSA France, La Rochelle, France, mai 2024, May 2024, La Rochelle, France. hal-04697093

HAL Id: hal-04697093

<https://hal.science/hal-04697093>

Submitted on 13 Sep 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

La distribution de Weibull, un outil utile ou dépassé ? Application au confort au vent

Mateusz Bogdan¹, Edouard Walther^{2,3,*}, Antoine Hubert¹

¹AREP L'hypercube, 16 avenue d'Ivry, Paris, France

²INSA Strasbourg, Département Génie Thermique Énergétique et Environnement
24 Boulevard de la Victoire - 67000 Strasbourg

³Université de Strasbourg, INSA Strasbourg, CNRS, ICube Laboratory UMR 7357, Strasbourg, France

*edouard.walther@insa-strasbourg.fr

RÉSUMÉ. Ce document présente une analyse critique de l'utilisation de la traditionnelle distribution de Weibull pour le calcul des probabilités de dépassement d'un seuil de vitesse de vent, en moyenne horaire. On présente deux densités de probabilités alternatives pour représenter les distributions de vitesses de vent, qui sont comparées à la distribution de Weibull sur des fichiers météorologiques réels. Ces distributions sont plus performantes que la loi de Weibull mais ne concordent pas pour autant avec les données réelles. En ouverture, l'utilisation des données réelles pour calculer un seuil de dépassement dans le cadre du calcul de confort au vent montre que la loi de Weibull donne des résultats qui sous-estiment largement l'inconfort.

MOTS-CLÉS. Weibull, distribution de fréquence, Confort au vent

ABSTRACT. The present work deals with the use of the Weibull distribution for the determination of the probability of exceeding of a given wind speed threshold. Two alternative probability distribution function are studied and compared with historical wind weather data. Those distributions perform better than Weibull's, however they do not represent conform with the actual wind distributions. The application to pedestrian wind comfort of the explicit determination of the probability of exceeding the wind annoyance threshold shows that Weibull's distribution exhibits large under-estimation of wind discomfort.

KEYWORDS. Weibull, frequency distribution, pedestrian wind comfort

1. CONFORT AU VENT

La quantification du confort au vent est fondée sur les vitesses de vents locales en présence d'un environnement bâti. Cette caractérisation se déroule en trois grandes étapes :

1. L'analyse des données météorologiques d'un site, avec l'estimation d'une densité de probabilité représentative des vents réelles, par direction de vent incident dans le but de pouvoir prédire la probabilité d'un dépassement. *Lorsque le vent souffle du Nord, il y a 12% de chances que ce dernier soit supérieur à 4 m/s.*
2. Un jeu de simulation CFD (*Computational Fluid Dynamics*), qui permet d'obtenir pour chaque incidence de vent la répartition des vitesses d'air locales dans la zone d'intérêt. *Lorsque le vent souffle du Nord à 4m/s, en ce point la vitesse d'air atteint 2,1 m/s.*
3. Enfin, de façon probabiliste, on reconstruit un indicateur annuel qui donne en tout point d'une zone d'intérêt la probabilité de dépasser un seuil de gêne fixé. *À cet endroit précis, il y a une probabilité annuelle de 14% que le vent dépasse 4 m/s.*

De nombreux points de la méthodologie d'étude du confort au vent ont déjà été étudiés : le choix de indicateurs (Delpech et al., 2005) et seuils de gênes associés (Janssen et al., 2013),

l'intégration des vents rafales explicites (Vita et al., 2020), le types de simulations CFD (Jacob and Sagaut, 2018), qualité et précision des simulations CFD (Tominaga et al., 2008), (Franke et al., 2007), nombre de directions de vents considérées (Hågbo and Giljarhus, 2024), etc., mais l'objet de cet article est de se concentrer sur la première étape de la méthodologie, fondement de la méthode : l'identification des paramètres climatiques liés au vent. Plus précisément la représentation statistique des occurrences de vents à l'aide d'une distribution probabiliste.

Ainsi, en utilisant la méthode la plus répandue en France, développée par le CSTB (présentée dans (Delpech et al., 2005)) on ne s'intéresse qu'à des données météorologiques en moyenne horaire, ce qui de fait exclut les vents dits de *rafale*. À date, ces derniers ne sont considérés que de façon indirecte par une modification des seuils de vitesses liée à l'intensité de la turbulence.

$$P_D(v > v^*) = \sum_{\theta} f_{\theta} P_{D,\theta}(v > v^*) \quad (1)$$

On s'intéresse ensuite à cartographier la probabilité de dépasser un seuil de gêne, fixé dans la méthode à 3 m/s. Pour obtenir cette cartographie sur une année, on réalise un ensemble de simulations CFD (*Computational Fluid Dynamics*), pour des directions de vents équiréparties, desquelles on tire des cartographies de facteur d'amplification $\Psi = v_{\text{local}}/v_{\text{meteo}}$. L'indicateur de confort au vent est finalement reconstruit en pondérant ces cartographies par les fréquences d'occurrences directionnelles des vents, puis localement en regardant la probabilité de dépasser une vitesse seuil : $v_{\text{local}} < \text{ou} > v_{\text{seuil}}/\Psi$, à l'aide de l'Équation (1). Le dépassement de ce seuil est ensuite analysé au regard de l'occupation de l'espace urbain.

1.1. PRINCIPE DU CALCUL

Le confort au vent se définit comme la probabilité de dépassement d'un seuil de gêne sur une période de l'ordre d'un an. On l'estime à l'aide de l'Équation (1), qui somme de façon pondérée la contribution de chaque direction de vent au dépassement d'un seuil de gêne v^* .

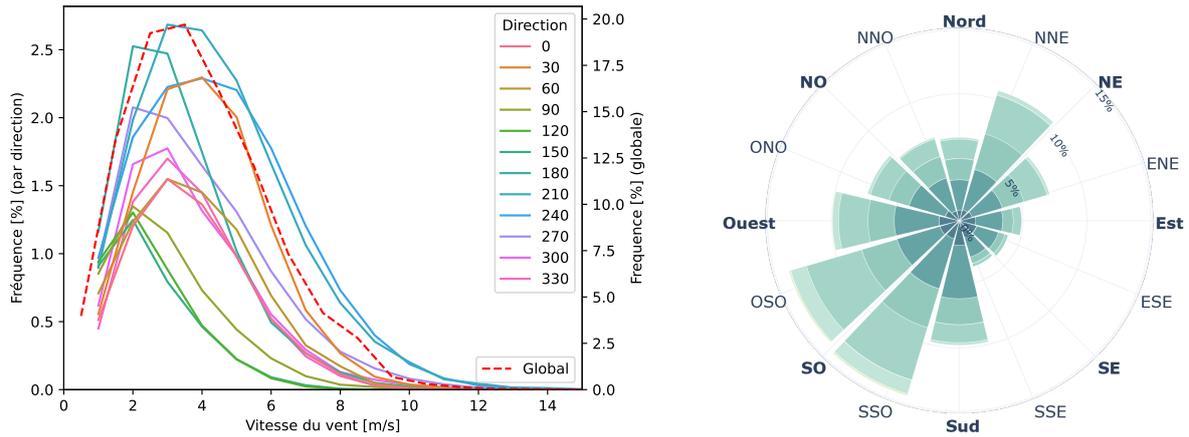
Pour une direction de vent incident donnée, θ , f_{θ} représente sa fréquence d'occurrence et $P_{D,\theta}$ la fonction qui définit la probabilité de dépassement pour une vitesse v^* .

Traditionnellement, la fonction de distribution utilisée pour décrire façon continue l'occurrence des vitesses de vents est une fonction de Weibull. Cette distribution a été choisie pour deux raisons : d'une part sa capacité à décrire de nombreuses « formes » de distributions à l'aide de deux paramètres (k et λ dans l'Eq. 2) et d'autre part la forme analytique que prend sa fonction de distribution cumulée P_c (Eq. 3) ainsi que la « facilité » d'accès à la probabilité de dépassement d'une vitesse v^* nommée P_d (Eq. 4).

$$f(v, k, \lambda) = \frac{k}{\lambda} \left(\frac{v}{\lambda}\right)^{k-1} e^{-\left(\frac{v}{\lambda}\right)^k} \quad (2)$$

$$P_c(v^*) = 1 - \exp \left[- \left(\frac{v^*}{\lambda} \right)^k \right] \quad (3)$$

$$P_d(v > v^*) = 1 - P_c(v^*) = \exp \left[- \left(\frac{v^*}{\lambda} \right)^k \right] \quad (4)$$



(a) Distribution des fréquences d'occurrences des vitesses de vent

(b) Rose des vents

Figure 1 : Paris Orly, discrétisation des vents en 12 directions équiréparties - données Météo-France horaires - 2000 à 2020

Les paramètres k et λ peuvent être identifiés sur les données météorologiques sans distinction d'angle d'incidence des vents (utilisé dans l'estimation du potentiel éolien (Wais, 2017), ou dans le cas du confort au vent « par direction » de vent (indice θ dans l'Eq. 1).

1.2. MOTIVATION

Certaines distributions réelles de vitesses de vent par classe se prêtent mal à une identification par un modèle de Weibull, notamment pour des valeurs qui, ramenées à hauteur d'occupation, peuvent être pénalisantes pour le calcul du confort (vitesses élevées), ou pour des valeurs faibles avec une mauvaise estimation des périodes calmes. Ces éléments sont longuement discutés dans (Wais, 2017), notamment les *méthodes* d'identification, cependant on se focalise plutôt ici sur des méthodes standard d'identification, accessibles en ingénierie avec des outils de traitement de données *simples*, en allant chercher d'autres distributions probabilistes, ou en testant directement ce qu'on appelle un calcul « explicite » de fréquence, rendu accessible par la disponibilité des données. En effet, les outils et méthodes disponibles aujourd'hui permettent aisément de traiter 20 ou 30 années de données horaires pour déterminer la probabilité de dépasser X m/s dans la direction de vent Y° .

On se propose donc d'étudier trois types de fonctions de distributions puis de les comparer des calculs « explicites » basés sur 20 années de données horaires MétéoFrance, en différents lieux : la distribution de *Weibull* détaillée en préambule, la distribution *log-normale tronquée mixte* et la distribution *sigmoïde*, décrites dans la section qui suit.

1.3. DENSITÉS DE PROBABILITÉS ALTERNATIVES

1.3.1. Fonction sigmoïde

Les auteurs (Reiminger et al., 2020) proposent une variante de sigmoïde originale qui permet de mieux représenter certaines distributions singulières, notamment utiles sur des données

historiques avec « peu » de classes de vitesses de vents. Sa densité de probabilité prend la forme de l'équation 5. Elle contient cinq paramètres positifs à identifier $\eta, \xi_1, \xi_2, \gamma_1, \gamma_2$:

$$f(v) = \eta \left(-1 + \frac{1}{1 + \xi_1 e^{-\gamma_1 v}} + \frac{1}{1 + \xi_2 e^{\gamma_2 v}} \right) \quad (5)$$

Sa distribution cumulée prend la forme de l'intégrale (6), dont le calcul analytique s'avère délicat, car on ne dispose pas d'une primitive réelle pour ses fractions. On la calculera numériquement par la suite.

$$p = \int_0^{v^*} \eta \left(-1 + \frac{1}{1 + \xi_1 e^{-\gamma_1 v}} + \frac{1}{1 + \xi_2 e^{\gamma_2 v}} \right) \cdot dv \quad (6)$$

1.3.2. Fonction log-normale tronquée mixte

L'identification de lois log-normales permettrait de s'épargner une intégration numérique, du fait de leur formulation normée et analytique, ainsi que présenté dans (Carta and Ramírez, 2007). Les travaux de (Carta et al., 2009) montrent que ces fonctions auraient, au global, les meilleurs résultats d'identifications parmi une vingtaines de variantes, et sont les seules à adaptées à des distributions de vent multi-modales. On étudie ici la fonction suivante, paramétrée en $\alpha_1, \beta_1, \alpha_2, \beta_2, \omega$:

$$f(v, \alpha_1, \beta_1, \alpha_2, \beta_2, \omega) = \omega \frac{\alpha_1}{\beta_1 (v/\beta_1)^{\alpha_1-1}} \exp \left[\left(\frac{-v}{\beta_1} \right)^{\alpha_1} \right] + (1 - \omega) \frac{1}{I_0(\alpha_2, \beta_2) \sqrt{2\pi}} \exp \left[\left(\frac{-(v - \alpha_2)^2}{2\beta_2^2} \right) \right] \quad (7)$$

$$I_0(\alpha, \beta) = \frac{1}{\beta \sqrt{2\pi}} \int_0^{+\infty} \exp \left[\left(\frac{-(v - \alpha)^2}{2\beta^2} \right) \right] dv \quad (8)$$

2. COMPARAISON DES APPROCHES

Dans cette section, on présente une comparaison qualitative puis quantitative des écarts constatés entre modèles. À cette fin, quatre jeux de données météorologiques ont été utilisés : Dunkerque, Le Havre, Nice, Paris et Lyon, dont les caractéristiques principales sont présentées sur la Figure 2. Il s'agit de données MétéoFrance horaires, prises entre 2000 et 2020.

2.1. IDENTIFICATIONS DES LOIS DE PROBABILITÉS

Afin de comparer semi-qualitativement les fonctions utilisées dans l'estimation de la probabilité de dépassement, on présente Figure 3 un cas détaillé pour une localisation, Dunkerque.

- Sous forme de distribution, les allures des courbes sont les bonnes, mais seule la loi « log-normale mixte » tend à reproduire le second pic d'occurrence de vent, autour de 7 m/s. Ce site météorologique présente en effet une distribution bi-modale des vitesses de vent.
- Sur les trois identifications présentées, on constate en considérant les RMSE que le meilleur modèle semble être la loi sigmoïde, suivi de la loi log-normale tronquée, et enfin la distribution de Weibull.
- La loi de Weibull tend à sous-estimer systématiquement les vents forts, et seule la fonction sigmoïde semble les reproduire fidèlement.

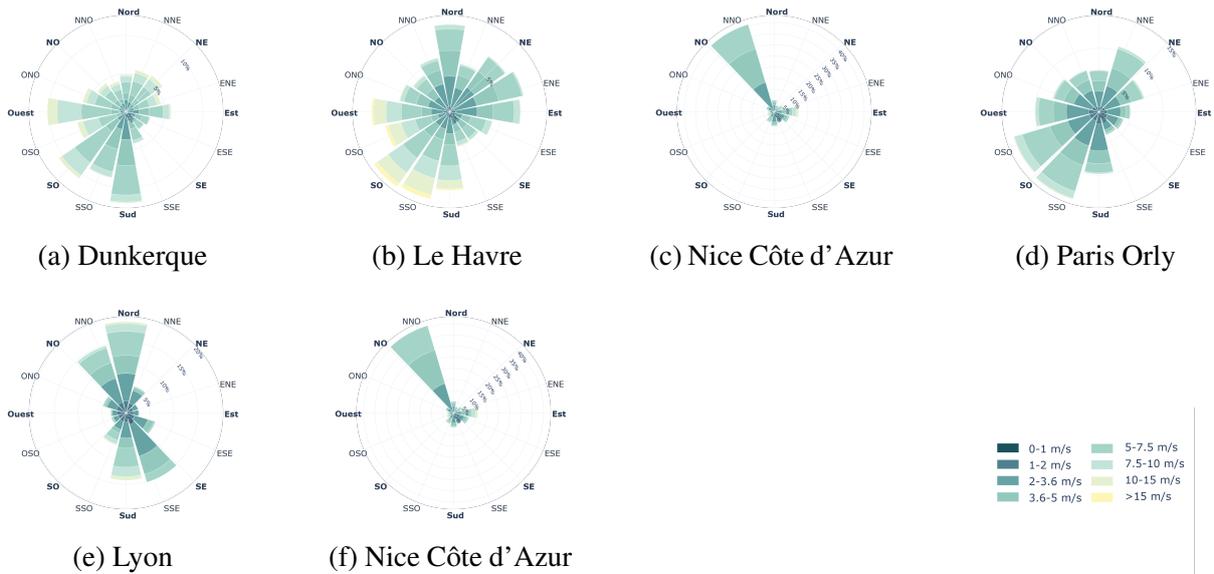


Figure 2 : Roses des vents - données MétéoFrance horaires - 2000 à 2020

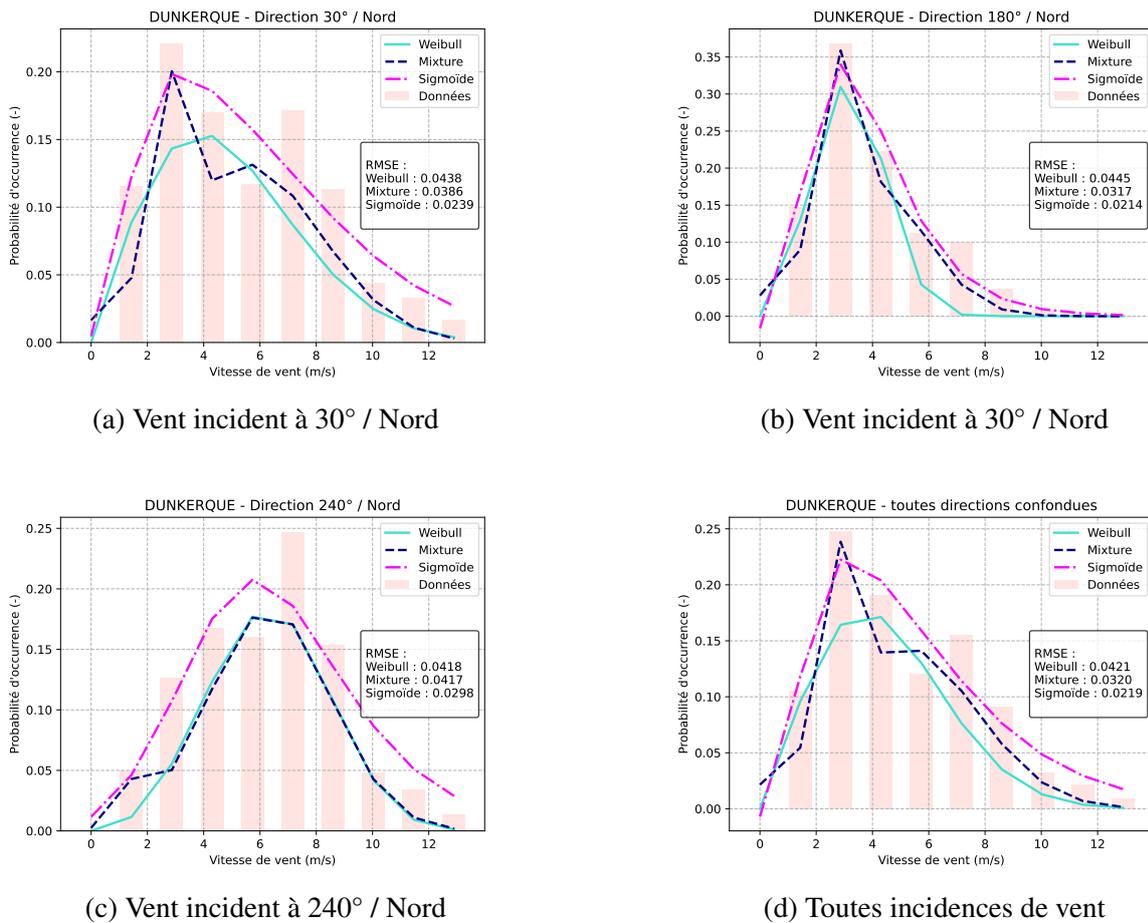


Figure 3 : Dunkerque : Distribution des fréquences d'occurrence de vent par classe de vitesses pour quelques directions de vent, et au global, avec comparaison aux différents modèles.

Villes	Fonctions	RMSE moyenne (12 directions)	RMSE Globale
Dunkerque	mixture	0.0029	0.0320
	weibull	0.0038	0.0421
	sigmoïde	0.0022	0.0219
Le Havre	mixture	0.0020	0.0437
	weibull	0.0055	0.0496
	sigmoïde	0.0014	0.0247
Nice	mixture	0.0015	0.0162
	weibull	0.0037	0.0401
	sigmoïde	0.0043	0.0353
Paris-Orly	mixture	0.0010	0.0123
	weibull	0.0011	0.0049
	sigmoïde	0.0005	0.0057
Lyon	mixture	0.0017	0.0071
	weibull	0.0025	0.0155
	sigmoïde	0.0013	0.0124

Tableau 1 : Erreurs d'identification (RMSE) moyennes sur les 12 directions de vents considérées, et globale (toutes incidences confondues)

On présente ensuite l'ensemble des résultats d'identification dans le Tableau 1, où l'erreur est calculée par rapport aux données réelles. On constate que pour un site donné, la RMSE moyenne de l'ensemble des directions est du même ordre de grandeur que la RMSE ($= \sqrt{\frac{1}{n} \sum_{i=1}^n (v_i - f_i)^2}$) obtenue sur l'ensemble des incidences de vents confondues (somme des fréquences de vent pour toutes les directions). La fonction sigmoïde semble se détacher comme la meilleure candidate à la représentation des vitesses réelles, suivie de la fonction log normale tronquée mixte.

2.2. DÉPASSEMENT DE DIFFÉRENTS SEUILS DE VITESSE

Dans le cas des études de confort au vent, les seuils de vitesses utilisés sont relativement faibles. Traditionnellement, en France, avec la méthode CSTB, on prend un seuil à 3 m/s, alors que d'autres critères européens ou américain (*Lawson, Davenport, NEN 8100*) utilisent plutôt 5 m/s. On se fixe comme référence le nombre d'heures de dépassement de plusieurs seuils de vitesses $v^* = [1..10]$ m/s sur les vingt dernières années des fichiers météorologiques, calculé comme suit :

$$P_{D\text{explicite}}(v > v^*) = \frac{\sum_{i=1}^{\text{nb de données}} 1 \text{ si } v_i > v^*}{\text{nb de données}} \quad (9)$$

La comparaison se fait en reconstruisant la probabilité à partir des lois de distributions (Equation (1)), et en calculant l'erreur commise par rapport à l'Equation (9) pour les seuils de 3 et 5 m/s. Les résultats sont présentés dans le Tableau 2. On constate que les meilleures estimations sont données en majorité par les fonctions log-normale tronquée mixtes ou sigmoïde. Cependant aucune fonction analytique ne présente de concordance nette avec le pourcentage réel du temps de dépassement calculé à partir des données météorologiques, avec des erreurs de l'ordre de la dizaine de pourcents au minimum.

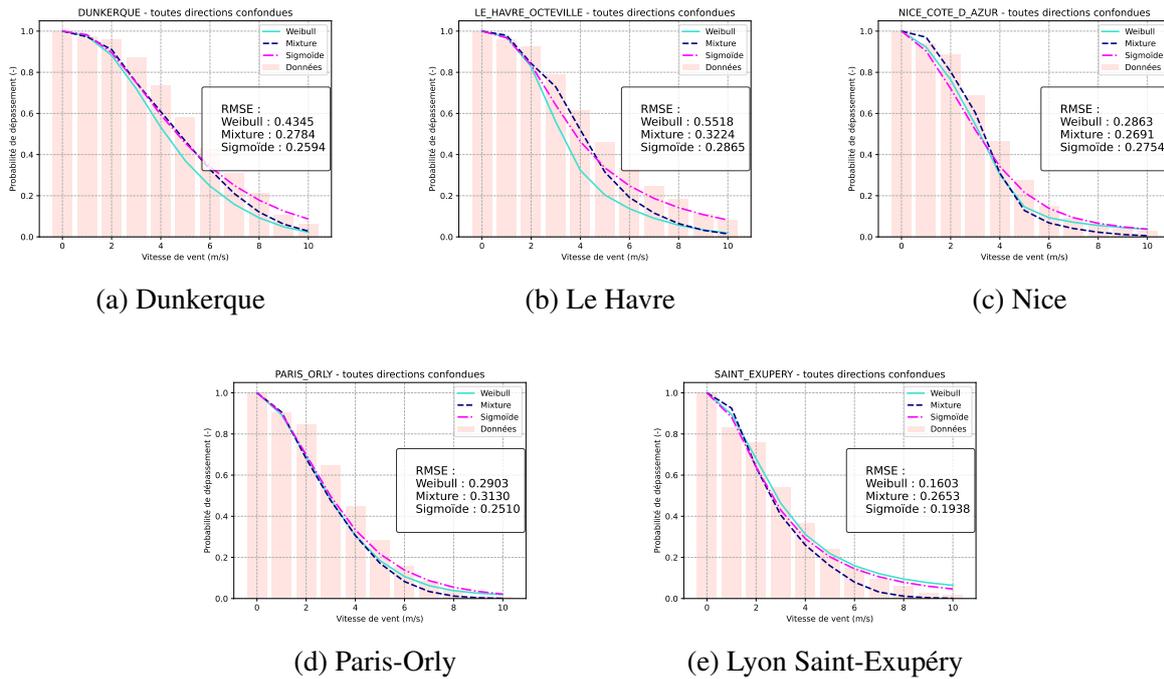


Figure 4 : Probabilités de dépassements de différentes vitesses seuils avec RMSE associées.

Villes	Fonctions	3 m/s	5m/s
Dunkerque	mixture	0.1367	0.1300
	weibull	0.1714	0.2406
	sigmoïde	0.1382	0.1420
Le Havre	mixture	0.0762	0.1843
	weibull	0.2912	0.3254
	sigmoïde	0.1883	0.1581
Nice	mixture	0.1228	0.2107
	weibull	0.2069	0.1854
	sigmoïde	0.2444	0.0834
Paris	mixture	0.2653	0.1697
	weibull	0.2583	0.1505
	sigmoïde	0.2338	0.1006
Lyon	mixture	0.2485	0.1455
	weibull	0.1418	0.3626
	sigmoïde	0.2037	0.0662

Tableau 2 : Erreurs sur les probabilités de dépassements de deux seuils de vitesse communs : 3 m/s et 5 m/s (toutes incidences confondues)

3. COMPARAISON DES MÉTHODES – APPLICATION AU CONFORT AU VENT

On propose ici de se passer des distribution de probabilité analytiques en calculant explicitement un seuil de dépassement sur N années de fichiers météorologiques, aujourd’hui largement disponibles en open-data. À partir des simulations par direction de vent incident réalisées en CFD, il est aisé de reconstruire les champ de confort au vent résultants, en analysant successivement toutes les heures d’un fichiers météo, et en sommant point par point les occurrences ou $v_{\text{local}} > v_{\text{seuil}}$. Ceci a été réalisé sur un projet en région parisienne, avec les données météorologiques de Paris-Orly. La Figure 5 présente les résultats obtenus avec la méthode explicite (à gauche), à savoir le pourcentage du temps où l’on dépasse un seuil de 3 m/s. On constate que les valeurs obtenues par la méthode de Weibull présentent jusqu’à 11% d’écart (Figure 5 droite). Ainsi l’utilisation de la loi de Weibull conduit à sous-estimer largement l’inconfort, avec des maxima autour de 5-6%, tandis que la méthode « explicite » atteint localement 16% d’inconfort sur une année type. Les seuils cibles de confort varient entre 2% pour une occupation immobile prolongée, 5% pour une situation immobile de courte durée et 10% pour une situation de promenade, ainsi la prédiction est entachée d’une erreur significative.

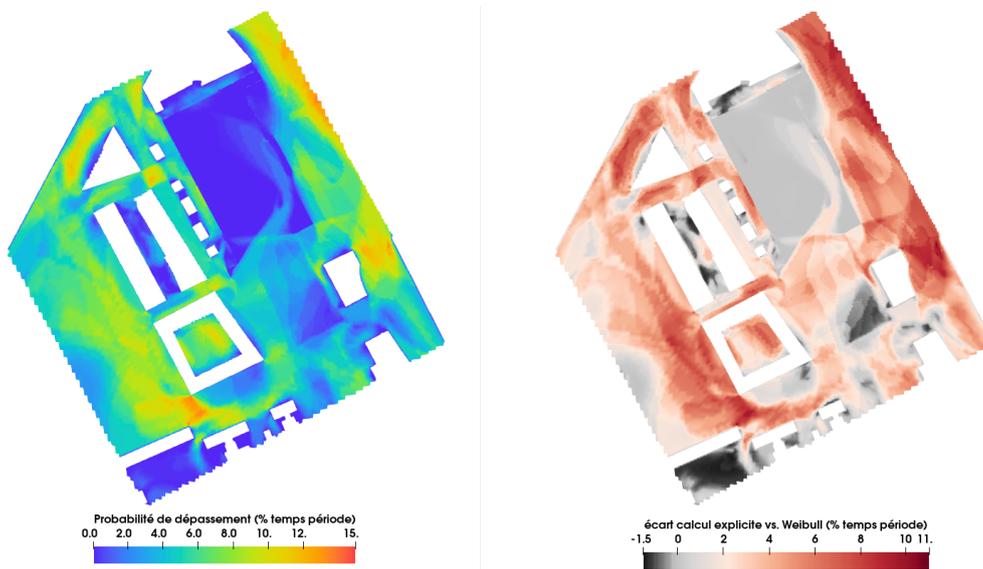


Figure 5 : Exemple de cartographie de confort au vent « explicite » (gauche), et carte des écarts par rapport à celle obtenue avec la « méthode de Weibull » (droite)

Note : Pour 10 ans de données, soit 87.600 données météorologiques horaires (direction et vitesse de vent), avec un une coupe à hauteur d’occupation d’environ 70.000 cellules, le temps de calcul explicite semble dérisoire par rapport aux temps d’étude en ingénierie, avec seulement 13 secondes de post-traitement (hors calcul CFD).

4. CONCLUSION

Dans ce travail, la loi de Weibull et deux autres distributions statistiques ont été étudiée sous l’angle de leur capacité à reproduire les mesures de vitesse de vent. L’étude a montré que les erreurs d’identification par direction de vent sont faibles pour chaque fonction considérée (de l’ordre de quelques pourcents), avec cependant de meilleures performances pour les lois log-normale mixte tronquée et sigmoïde par rapport à la loi de Weibull. Cependant, dans le cas du calcul de la probabilité cumulée, afin de déterminer un pourcentage de dépassement d’un

seuil de vitesse de vent, l'erreur commise par rapport aux données météorologiques réelles est considérable, de l'ordre de ~ 10 à 20% en fonction des stations météorologiques. Une application sur un cas de confort au vent montre qu'il n'est pas judicieux d'utiliser la loi de Weibull pour le calcul du seuil de dépassement d'une vitesse limite.

La conséquence pratique de cette étude est la suivante : eu égard à la disponibilité croissante de la donnée météorologique et à la rapidité de calcul du seuil explicite de dépassement, y compris sur de grands modèles géométriques, il est préférable de délaissier l'approche de Weibull (et plus généralement les approches d'identifications statistiques) pour le confort au vent. Des études complémentaires pourraient servir à déterminer les profondeurs historiques minimales nécessaires. Il en résulte également que pour une estimation du potentiel éolien d'un site, l'utilisation des probabilités analytiques de distribution de vent relève de l'application pédagogique plus que du calcul du productible.

BIBLIOGRAPHIE

- Carta, J. A. and Ramírez, P. (2007). Use of finite mixture distribution models in the analysis of wind energy in the canarian archipelago. *Energy Conversion and Management*, 48(1) :281–291.
- Carta, J. A., Ramirez, P., and Velazquez, S. (2009). A review of wind speed probability distributions used in wind energy analysis : Case studies in the canary islands. *Renewable and sustainable energy reviews*, 13(5) :933–955.
- Delpech, P., Baker, C., Blackmore, P., Koss, H., Sanz-Andres, A., Stathopoulos, T., and Willemsen, E. (2005). Pedestrian wind comfort assessment criteria : A comparative case study. *Proceedings of 4EACWE, Prague*, pages 82–83.
- Franke, J., Hellsten, A., Schlünzen, H., and Carissimo, B. (2007). Best practice guideline for the cfd simulation of flows in the urban environment.
- Hågbo, T.-O. and Giljarhus, K. E. T. (2024). Sensitivity of urban morphology and the number of cfd simulated wind directions on pedestrian wind comfort and safety assessments. *Building and Environment*, 253 :111310.
- Jacob, J. and Sagaut, P. (2018). Wind comfort assessment by means of large eddy simulation with lattice boltzmann method in full scale city area. *Building and Environment*, 139 :110–124.
- Janssen, W., Blocken, B., and van Hooff, T. (2013). Pedestrian wind comfort around buildings : Comparison of wind comfort criteria based on whole-flow field data for a complex case study. *Building and Environment*, 59 :547–562.
- Reiminger, N., Jurado, X., Vazquez, J., Wemmert, C., Blond, N., Wertel, J., and Dufresne, M. (2020). Methodologies to assess mean annual air pollution concentration combining numerical results and wind roses. *Sustainable Cities and Society*, 59 :102221.
- Tominaga, Y., Mochida, A., Yoshie, R., Kataoka, H., Nozu, T., Yoshikawa, M., and Shirasawa, T. (2008). AIJ guidelines for practical applications of CFD to pedestrian wind environment around buildings. *Journal of wind engineering and industrial aerodynamics*, 96(10-11) :1749–1761.
- Vita, G., Shu, Z., Jesson, M., Quinn, A., Hemida, H., Sterling, M., and Baker, C. (2020). On the assessment of pedestrian distress in urban winds. *Journal of Wind Engineering and Industrial Aerodynamics*, 203 :104200.
- Wais, P. (2017). A review of weibull functions in wind sector. *Renewable and Sustainable Energy Reviews*, 70 :1099–1107.