



HAL
open science

Développement de méthodes statistiques et probabilistes en corrosion par piqûres pour l'estimation de la profondeur maximale: application à l'aluminium A5

Adil Jarrah

► **To cite this version:**

Adil Jarrah. Développement de méthodes statistiques et probabilistes en corrosion par piqûres pour l'estimation de la profondeur maximale: application à l'aluminium A5. Sciences de l'ingénieur [physics]. Arts et Métiers ParisTech, 2009. Français. NNT : 2009ENAM0024 . pastel-00005700

HAL Id: pastel-00005700

<https://pastel.hal.science/pastel-00005700>

Submitted on 28 Jan 2010

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.



Ecole doctorale n° 432 : Sciences des Métiers de l'Ingénieur

THÈSE

pour obtenir le grade de

Docteur

de

l'École Nationale Supérieure d'Arts et Métiers

Spécialité "Mécanique-Matériaux"

*présentée et soutenue publiquement
par*

Adil JARRAH

le 8 décembre 2009

**DEVELOPPEMENT DE METHODES STATISTIQUES ET
PROBABILISTES EN CORROSION PAR PIQUES POUR
L'ESTIMATION DE LA PROFONDEUR MAXIMALE : APPLICATION
A L'ALUMINIUM A5**

Directeur de thèse : Alain IOST

Co-encadrement de la thèse : Jean-Marie NIANGA

Jury :

Michel TRAISNEL , Professeur, USTL, Lille	Examineur
Maxence BIGERELLE , Professeur, UTC, Compiègne.....	Rapporteur
Jean-Bernard VOGT , Professeur, ENSCL, Lille	Rapporteur
Benjamin FOURNIER , Docteur, CEA, Saclay	Examineur
Alain IOST , Professeur, LMPGM, Arts et Métiers ParisTech, Lille	Examineur
Jean-Marie NIANGA , Enseignant Chercheur, HEI, Lille	Examineur
Gildas GUILLEMOT , MCF, Arts et Métiers ParisTech, Lille.....	Invité
Denis NAJJAR , MCF HDR, Arts et Métiers ParisTech, Lille	Invité

Laboratoire de Métallurgie Physique et Génie des Matériaux
Arts et Métiers ParisTech, centre de Lille

Développement de méthodes statistiques pour la prévision des valeurs extrêmes : Application à la corrosion par piqûres de l'aluminium A5

RESUME: La corrosion par piqûres est l'une des formes de corrosion les plus répandues. Elle touche tous les matériaux et se place dans un contexte économique très important. Elle peut se manifester à des endroits spécifiques de la structure et mener à sa détérioration en particulier en présence de sollicitations mécaniques. L'aspect stochastique du phénomène a conduit au développement de méthodes statistiques pour le caractériser. Cette caractérisation est souvent faite via l'estimation de la profondeur maximale des piqûres afin d'évaluer le risque de perforation de la structure. Pour cela, la méthode de Gumbel est l'approche la plus utilisée. L'objectif de ce travail est de revenir sur la vérification des conditions d'application de cette méthode notamment l'indépendance et de la comparer avec les autres approches basées sur la loi des valeurs extrêmes généralisée et la loi de dépassement de seuil. La condition d'indépendance est vérifiée à l'aide des processus spatiaux. Une adaptation de l'analyse spectrale en corrosion par piqûres est aussi proposée. La comparaison entre les approches est basée sur des simulations numériques dont les paramètres sont issus de l'expérimentation.

Mots-clés : corrosion par piqûres, théorie des valeurs extrêmes, lois de probabilités, tests statistiques, analyse et statistique spatiale des données, analyse spectrale.

Development of statistical method to expect extreme values: Application to pitting corrosion of aluminum A5

ABSTRACT: pitting corrosion is one of the most prevalent forms of corrosion. It affects all materials and takes place in a very important economic context. Pits can manifest locally over the structure and leads to its deterioration particularly in the presence of mechanical solicitations. The stochastic aspect of the phenomenon led to the development of statistical methods in order to characterize it. This characterization is often done through the estimation of maximum pit depth in the aim to assess the risk of perforation. For this aim, the method of Gumbel is often used. The objective of this work is to check the conditions of application of this method notably the independence and compare it with the approaches based on law of generalized extreme values and the peak over threshold. The condition of independence is verified using spatial process. An adaptation of the spectral analysis in the context of pitting corrosion is also proposed. The comparison between the approaches is based on numerical simulations which the parameters come from the experimentation.

Keywords: pitting corrosion, extreme value theory, law of probabilities, statistical test, statistic of spatial data, spectral analysis.

Sommaire

Table des matières	5
Liste des figures	9
Liste des tableaux	14
Introduction	15
1 Corrosion par piqûres	18
2 Théorie des Valeurs Extrêmes	31
3 Analyse et Statistique Spatiales des Données	55
4 Analyse des données expérimentales et Méthode de Gumbel	77
5 Approches GEV et POT	108
Conclusion générale	121
Annexes	126
A	126
B	131
Glossaire	137
Bibliographie	141

Remerciement

A mon père,

Je tiens à remercier en premier lieu Alain IOST qui a dirigé cette thèse. Ses conseils, son expérience et ses qualités humaines m'ont beaucoup apporté tout le long de ces trois années.

Je tiens à remercier l'école des Hautes études d'Ingénieur pour le financement de cette thèse et en particulier Jean-Marie NIANGA de l'avoir codirigé.

Ce travail n'aurait pas été possible sans la grande implication et les conseils de Denis NAJJAR et Gildas GUILLEMOT que je tiens à remercier tout particulièrement. J'étais très ravi de travailler avec eux.

Je remercie également les membres du jury de cette thèse. Michel TRAISNEL le président du jury, Maxence BIGERELLE et Jean-Bernard VOGT d'avoir rapporté ce travail, et Benjamin FOURNIER de l'avoir examiné.

Je remercie mon épouse Souad de son soutien permanent surtout dans les moments difficiles. J'ai toujours pu compter sur elle. Je remercie également mes deux frères Mohamed et Hicham, mes deux soeurs Fatima et Hicham, pour leurs encouragements, sans oublier ma belle famille et mes amis Morad, Zakaria, Réda, Ghassane et Abdellah.

Je remercie plus particulièrement ma mère qui n'a cessé de me soutenir. L'amour que tu portes pour tes enfants m'a toujours impressionné et surtout motivé.

Cette thèse je la dédie à mon père qui aurait tellement aimé être à mes côtés le jour de la soutenance, mais que le destin en a décidé autrement. Tu resteras à jamais dans mon cœur.

Table des matières

Table des matières	5
Liste des figures	9
Liste des tableaux	14
Introduction	15
1 Corrosion par piqûres	18
1.1 Généralités	18
1.2 Caractérisation de la sensibilité à la corrosion par piqûres	20
1.2.1 Potentiel de piqûres	20
1.2.2 Temps d'amorçage	21
1.2.3 Approche statistique : Amorçage + Propagation	22
1.3 Facteurs favorisant la corrosion par piqûres	22
1.3.1 Matériau	22
1.3.2 Milieu	23
1.4 Mécanisme de la corrosion par piqûres	23
1.4.1 Rupture de la passivité	23
1.4.2 Croissance des piqûres	24
1.4.3 Repassivation	25
1.5 Lois de probabilités en corrosion par piqûres	25
1.5.1 Loi de Poisson	25
1.5.2 Loi Exponentielle	25
1.5.3 Loi Normale	26
1.5.4 Loi log-normale	26
1.5.5 Loi de lambda distribution généralisée	26
1.6 Tests classiques d'ajustement et d'analyse de données	26
1.6.1 Tests classiques d'ajustement	26
1.6.2 Tests d'analyse de données	28
2 Théorie des Valeurs Extrêmes	31
2.1 Vers la théorie des valeurs extrêmes	31
2.2 L'approche des maxima par blocs	32
2.2.1 Théorème fondamental des valeurs extrêmes	32
2.2.2 Loi des valeurs extrêmes généralisée GEV	34
2.2.3 Domaine d'attraction du maximum	34
2.2.4 Méthodes d'estimation des paramètres et des quantiles extrêmes	38

2.3	L'approche des dépassements de seuil	45
2.3.1	Modèle de renouvellement (ou modèle POT : Peaks Over Threshold)	46
2.4	Méthodes pour le choix du seuil	46
2.4.1	Méthodes graphiques	46
2.4.2	Méthodes numériques	47
2.5	Méthodes d'estimation des paramètres et des quantiles extrêmes	48
2.6	Autres méthodes d'estimation du paramètre ξ	51
2.6.1	Estimateur de Hill	51
2.6.2	Estimateur des moments de Dekkers	51
2.6.3	Estimateur UH	52
3	Analyse et Statistique Spatiales des Données	55
3.1	Introduction	55
3.2	Analyse spatiale des données : Cas stationnaire	56
3.2.1	Structure aléatoire	57
3.2.2	Structure Agrégée	58
3.2.3	Structure Régulière	59
3.2.4	Méthodes d'investigation des structures spatiales	60
3.2.5	$R - Spectre$ et $\theta - Spectre$	67
3.3	Analyse spatiale des données : Cas non-stationnaire	71
3.3.1	Méthodes d'ajustement d'une structure spatiale	71
3.3.2	Test de la validité de l'ajustement	73
3.3.3	Simulation d'un processus non-stationnaire	73
3.4	Statistique spatiale des données	74
3.4.1	Fonction de corrélation des marques K_{mm}	74
3.4.2	Estimation de la Fonction de corrélation des marques	75
4	Analyse des données expérimentales et Méthode de Gumbel	77
4.1	Analyse expérimentale des données	77
4.1.1	Préparation des échantillons de la première tôle	77
4.1.2	Courbes à potentiel libre pour la première et la deuxième tôle	80
4.1.3	Analyse spatiale pour la troisième tôle	82
4.1.4	Mesure des profondeurs	84
4.1.5	Etude complète sur une plaque de 75 cm ²	87
4.2	Méthode de Gumbel	90
4.2.1	Paramètres des simulations numériques	91
4.2.2	Méthode de Gumbel et régression linéaire	92
4.2.3	Choix de l'approximation empirique	94
4.2.4	Choix de la surface d'analyse	100
4.2.5	Comparaison des méthodes d'estimation	103
4.2.6	Application pour la plaque de 75 cm ²	103
5	Approches GEV et POT	108
5.1	Loi des valeurs extrêmes généralisée GEV	108
5.1.1	Comparaison des méthodes d'estimation PWM et MV pour la loi GEV	109
5.2	Loi des dépassements de seuil POT	113
5.3	Comparaison entre les approches GEV et POT	114
5.4	Application à la plaque de 75 cm ²	117

Conclusion générale	121
Annexes	126
A	126
B	131
Glossaire	137
Bibliographie	141

Liste des figures

1.1	Bilan des différentes formes de corrosion [X-PER-X]	18
1.2	Pile galvanique de corrosion : + Cathode, - Anode ([Lan03] modifiée)	19
1.3	Attaque par corrosion : a- Localisée, b- Généralisée ([Kow94] modifiée)	19
1.4	Exemples de morphologie des piqûres de corrosion ([Bar04])	20
1.5	Schéma d'une courbe Intensité-Potentiel	21
1.6	Schéma de détermination du potentiel de piqûration par la méthode galvanostatique	21
1.7	Schéma du mécanisme d'amorçage par pénétration ([DBB94])	24
1.8	Schéma du mécanisme d'amorçage par rupture du film ([DBB94])	24
2.1	Exemple de fonctions de densité de probabilité avec différentes valeurs de ξ	34
2.2	(u, \hat{e}_n) pour 1000 valeurs issues des lois : Weibull(1,1) \in MDA(Gumbel), Burr(1,2) \in MDA(Fréchet) et Uniforme(0,1) \in MDA(Weibull)	47
2.3	$UH_{k,n}(u)$ pour 1000 valeurs issues des lois : Weibull(1,1) \in MDA(Gumbel), Burr(1,2) \in MDA(Fréchet) et Uniforme(0,1) \in MDA(Weibull)	53
3.1	Exemple de réalisation d'un processus HPP	58
3.2	Exemples de réalisations d'un processus MTCP avec $\lambda_p = 40$, $\mu_c = 4$ et : (a) $\sigma_c = 0,01$, (b) $\sigma_c = 0,1$	59
3.3	Exemple de réalisations d'un processus d'inhibition séquentiel avec : (a) $\delta = 0,02$, (b) $\delta = 0,08$	60
3.4	Exemple de fonction Ripley pour les trois types de structures	61
3.5	Illustration du problème des effets de bord	62
3.6	Comparaison des estimations de $K(r)$ avec et sans correction de bord pour cent points générés avec un process HPP sur le carré unité	62
3.7	Illustration de la méthode des quadrats : découpage 4×4 pour un processus HPP à cent points dans le carré unité	64
3.8	Périodogramme d'un processus HPP à 500 points dans un carré unité	66
3.9	Périodogramme d'un processus MTCP à 497 points dans un carré unité avec : $\lambda_p = 50$, $\mu_p = 10$ et $\sigma_c = 0,05$	66
3.10	Périodogramme d'un processus <i>PSI</i> à 500 points dans un carré unité avec $\delta = 0,025$	67
3.11	Spectres d'un processus HPP à 500 points dans un carré unité : (a) <i>R - Spectre</i> , (b) $\theta - Spectre$. La valeur attendue est égale à 1.	69
3.12	Spectres d'un processus MTCP à 487 points dans un carré unité avec $\lambda_p = 50$, $\mu_p = 10$ et $\sigma_c = 0,05$: (a) <i>R - Spectre</i> , (b) $\theta - Spectre$	69
3.13	Spectres d'un processus <i>PSI</i> à 500 points dans un carré unité avec $\delta = 0,025$: (a) <i>R - Spectre</i> , (b) $\theta - Spectre$	70
3.14	Exemple de réalisation d'un processus <i>NHPP</i> d'intensité $1 + 100(x + y)$ dans le carré unité	74

4.1	Illustration du découpage de la tôle en échantillons	78
4.2	Relevé d'un échantillon d'environ 10 cm ² sur une plaque montrant l'état de surface et le phénomène de coalescence des piqûres	78
4.3	Schéma du masque à cinq carrés de 1 cm ²	79
4.4	Boîtes à moustaches et résultats du test de Kruskal-Wallis pour la comparaison du nombre de piqûres par carrés sur chacune des bandes A,B et C	79
4.5	Boîtes à moustaches et résultats du test de Kruskal-Wallis pour la comparaison du nombre de piqûres par carrés sur chacune des positions 1 à 6	80
4.6	Schéma du dispositif à trois électrodes utilisé pour la réalisation des courbes à potentiel libre	81
4.7	Variations du potentiel en fonction du temps pour les deux tôles : a-plaque non-piquée pour la deuxième tôle, b-plaque piquée pour la première tôle	81
4.8	Variations du nombre de piqûres en fonction de la position de la plaque pour les trois bandes	82
4.9	Image binaire de la plaque de position 4 issue de la bande 3	83
4.10	Spectres pour la plaque de position 3 issue de la bande 1 : a- <i>R-Spectre</i> , b- <i>θ-Spectre</i>	83
4.11	Spectres pour la plaque de position 4 issue de la bande 1 : a- <i>R-Spectre</i> , b- <i>θ-Spectre</i>	83
4.12	Profilomètre optique VEECO WYKO NT9300	84
4.13	Exemples de morphologie de piqûres (issues de la tôle 1) caractérisée à l'aide du profilomètre optique VEECO avec leurs mesures de profondeurs : a-mesure par profilomètre, b-mesure par microscope	85
4.14	Histogramme des profondeurs mesurées par les deux appareils	85
4.15	Quantile-quantile plot pour les résidus des mesures issues des deux appareils	86
4.16	Boîtes à moustaches et résultat du test ANOVA pour les deux mesures	87
4.17	Plaque de 75 cm ²	88
4.18	Répartition spatiale des centres de masse des piqûres	88
4.19	Histogramme des 603 mesures de profondeur de piqûre	89
4.20	Carte des profondeurs (en µm) en fonction des positions des piqûres	89
4.21	Fonction de corrélation des marques pour la plaque de 75 cm ²	90
4.22	Illustration des étapes de la méthode de Gumbel avec comme méthode d'estimation des paramètres la régression linéaire : a- Surface étudiée, b-Découpage en n blocs, c-Choix aléatoire de m blocs parmi n ($m=5$), d- Estimation pour T souhaitée avec intervalle de confiance (IC)	91
4.23	Moyennes des 1000 quantiles estimés à partir des profondeurs générées selon une loi normale avec $m = 10$ blocs : a-Moyenne des quantiles, b-Ecarts-types des quantiles	95
4.24	Moyenne des quantiles et coefficients de variations (CV) pour les approximations de Hazen et Gringorten : a-moyenne des quantiles, b- Coefficient de variation	96
4.25	Moyennes des 1000 espérances du maximum (notées EM) estimées à partir des profondeurs générées selon une loi normale : a-axes inversés, b-axes classiques	97
4.26	Moyenne des quantiles dans le cas de « bloc-égal » : a- axes inversés, b-axes classiques	98
4.27	Test ANOVA pour comparer les estimations calculées à partir de « bloc égal » et « bloc différent » : a- $m = 10$ blocs, b- $m = 50$ blocs	98
4.28	Moyennes des 1000 quantiles estimés à partir des profondeurs générées selon une loi log-normale : a-axes inversés, b-axes classiques	99
4.29	Méthode des quadrats pour les positions 3 et 4 de la bande 1	101
4.30	Comparaison des estimations en fonction de la taille des blocs (méthode de la régression) : a-Moyennes des quantiles , b-Moyennes des écarts-types des quantiles	102

4.31	Comparaison des méthodes d'estimations : a-moyennes des paramètres de localisation, b-écarts-types des paramètres d'échelle, reg : régression, mv : maximum de vraisemblance, mom : moment	103
4.32	Méthode des quadrats pour la plaque de 75 cm ²	104
4.33	Comparaison entre les quatre approximations empiriques : en abscisse, pourcentage de la surface totale analysée qui est égale à la taille du bloc \times leur nombre	104
4.34	Impact de la taille du bloc d'analyse sur les estimations des quantiles pour la méthode de la régression linéaire	105
4.35	Comparaison des méthodes d'estimations des paramètres pour la plaque de 75 cm ²	105
5.1	a- Moyenne des quantiles, b- Ecart-types des quantiles	109
5.2	Variation de la fonction quantile de la loi GEV normalisée en fonction de l'indice extrême sur une échelle Logarithmique (en oy)	110
5.3	Moyennes des estimations pour les échantillons simulés avec $\xi = 0,4$ en fonction de leurs tailles : (a)- Moyennes des indices extrêmes, (b)- Moyennes des paramètres de localisation μ , (c)- Moyennes des paramètres d'échelle σ , (d)- Moyennes des quantiles	111
5.4	Echantillons pour $\xi = 0,4$: a- Biais des quantiles, Ecart-types des quantiles	111
5.5	Moyennes des estimations pour les échantillons simulés avec $\xi = -0,4$ en fonction de leurs tailles : (a)- Moyennes des indices extrêmes, (b)- Moyennes des paramètres de localisation μ , (c)- Moyennes des paramètres d'échelle σ , (d)- Moyennes des quantiles	112
5.6	Echantillons pour $\xi = -0,4$: a- Biais des quantiles, Ecart-types des quantiles	112
5.7	Echantillons pour $\xi = 0,4$: a- Moyennes des quantiles, b-Ecart-types des quantiles	113
5.8	Echantillons pour $\xi = -0,4$: a- Moyennes des quantiles, b- Ecart-types des quantiles	113
5.9	Intervalle de confiance à 10% des estimations des quantiles pour la loi normale. De gauche à droite les quantiles de : Gumbel estimés par mv, Gumbel estimés par pwm, GEV estimés par mv, GEV estimés par pwm, POT estimés par mv, POT estimés par pwm, Dekkers, UH.	115
5.10	Intervalle de confiance à 10% des estimations des quantiles pour la loi lognormale. De gauche à droite les quantiles de : Gumbel estimés par mv, Gumbel estimés par pwm, GEV estimés par mv, GEV estimés par pwm, POT estimés par mv, POT estimés par pwm, Dekkers, UH.	115
5.11	Intervalle de confiance à 10% des estimations des quantiles pour la loi lognormale. De gauche à droite les quantiles de : Gumbel estimés par mv, Gumbel estimés par pwm, GEV estimés par mv, GEV estimés par pwm, POT estimés par mv, POT estimés par pwm, Dekkers, UH.	116
5.12	Intervalle de confiance à 10% des estimation de l'indice extrême. IC inf : borne inférieure de l'intervalle de confiance, IC sup : borne supérieure	117
5.13	Intervalle de confiance à 10% des estimations des quantiles pour la plaque pour le choix de 10 blocs. De gauche à droite les quantiles de : Gumbel estimés par mv, Gumbel estimés par pwm, GEV estimés par mv, GEV estimés par pwm, POT estimés par mv, POT estimés par pwm, Dekkers, UH.	118
5.14	Intervalle de confiance à 10% des estimations des quantiles pour la plaque pour le choix de 20 blocs. De gauche à droite les quantiles de : Gumbel estimés par mv, Gumbel estimés par pwm, GEV estimés par mv, GEV estimés par pwm, POT estimés par mv, POT estimés par pwm, Dekkers, UH.	118

A.1	Schéma descriptif de la boîte à moustache : Q_1 Q_3 représentent respectivement le premier et le troisième quartile. Les outliers désignent des valeurs considérées comme extrêmes	127
A.2	Spectres pour la plaque de position 3 issue de la bande 2 : a- <i>R-Spectre</i> , b- θ - <i>Spectre</i>	127
A.3	Spectres pour la plaque de position 4 issue de la bande 2 : a- <i>R-Spectre</i> , b- θ - <i>Spectre</i>	128
A.4	Spectres pour la plaque de position 3 issue de la bande 3 : a- <i>R-Spectre</i> , b- θ - <i>Spectre</i>	128
A.5	Spectres pour la plaque de position 4 issue de la bande 3 : a- <i>R-Spectre</i> , b- θ - <i>Spectre</i>	129
A.6	Moyennes des 1000 quantiles estimés à partir des profondeurs générées selon une loi normale avec $m = 30$ blocs : a-axes inversés, b-axes classiques	129
A.7	Moyennes des 1000 quantiles estimés à partir des profondeurs générées selon une loi normale avec $m = 50$ blocs : a-axes inversés, b-axes classiques	130
B.1	Moyennes des estimations pour les échantillons simulés avec $\xi = 0, 2$ en fonction de leurs tailles : (a)- Moyennes des indices extrêmes, (b)- Moyennes des paramètres de localisation μ , (c)- Moyennes des paramètres d'échelle σ , (d)- Moyennes des quantiles	131
B.2	Echantillons pour $\xi = 0, 2$: a- Biais des quantiles, Ecarts-types des quantiles	132
B.3	Moyennes des estimations pour les échantillons simulés avec $\xi = -0, 2$ en fonction de leurs tailles : (a)- Moyennes des indices extrêmes, (b)- Moyennes des paramètres de localisation μ , (c)- Moyennes des paramètres d'échelle σ , (d)- Moyennes des quantiles	132
B.4	Echantillons pour $\xi = -0, 2$: a- Biais des quantiles, Ecarts-types des quantiles . . .	133
B.5	Moyennes des estimations pour les échantillons simulés avec $\xi = 0$ en fonction de leurs tailles : (a)- Moyennes des indices extrêmes, (b)- Moyennes des paramètres de localisation μ , (c)- Moyennes des paramètres d'échelle σ , (d)- Moyennes des quantiles	133
B.6	Echantillons pour $\xi = 0$: a- Biais des quantiles, Ecarts-types des quantiles	134
B.7	Echantillons pour $\xi = 0, 2$: a- Moyennes des quantiles, b- Ecarts-types des quantiles	134
B.8	Echantillons pour $\xi = -0, 2$: a- Moyennes des quantiles, b- Ecarts-types des quantiles	135
B.9	Echantillons pour $\xi = 0$: a- Moyennes des quantiles, b- Ecarts-types des quantiles .	135

Liste des tableaux

1.1	Exemple de tableau de données	28
2.1	Exemples de distributions de type Fréchet	36
2.2	Exemples de distributions de type Weibull	36
2.3	Exemples de distributions de type Gumbel	37
4.1	Estimation et indice de confiance à 95% des paramètres de localisation et d'échelle pour les lois normale et log-normale	92
4.2	Expression des quatre distributions empiriques utilisées	94

Introduction

Le phénomène de corrosion touche la majorité des matériaux métalliques. Il peut être défini comme étant une dégradation de la structure du matériau ou de ses propriétés par réactions électrochimiques avec l'environnement ([Lan03]). Parce que les métaux et alliages métalliques se distinguent des autres matériaux par un ensemble de propriétés attractives (rigidité, ductilité, ténacité, conductibilité électrique, résistance à la traction,...), cela fait d'eux des éléments de structures essentiels dans l'industrie et mène à leur vaste utilisation dans tous les secteurs (transport, alimentation, construction,...). Par conséquent, le phénomène de corrosion qui touche ces matériaux de structure se place dans un contexte économique très important. Son coût annuel global en France¹ est estimé à environ 50 milliard d'euros. En général, le coût englobe les pertes directes liées aux remplacements des matériaux corrodés et les équipements dégradés et les pertes indirectes liées à la réduction de la production et à l'accroissement du coût de fonctionnement. Il en résulte un enjeu scientifique de taille à la fois pour la conception (nouveaux matériaux, techniques de protection,...) et pour la prévention en terme de prédiction et évaluation de risques.

Les techniques de protection contre la corrosion généralisée sont multiples et de plus en plus élaborées. Néanmoins, elles ne permettent pas de se prémunir contre une autre forme particulière de corrosion plus pernicieuse qui est susceptible de se manifester : la corrosion localisée. Cette dernière qui apparaît à des endroits spécifiques de la structure est souvent dûe à une hétérogénéité locale du matériau (inclusions, précipités,...) ou de l'environnement (variation de pH, température, concentration en oxygène, ...). Elle peut prendre plusieurs formes parmi lesquelles on distingue la corrosion galvanique, sélective, caverneuse, par piqûres, ...

Dans ce travail nous nous sommes intéressés à la corrosion par piqûres qui est l'une des formes les plus répandues de la corrosion localisée. Son caractère insidieux, fait qu'elle peut être très dangereuse et difficilement prévisible. En effet, même si celle-ci ne se manifeste que sous forme d'un petit point à la surface, une cavité peut se propager en volume jusqu'à la perforation du matériau et la ruine de la structure. Ces piqûres sont d'autant plus néfastes que les structures sont sollicitées mécaniquement en traction ou en fatigue et qu'elles peuvent constituer de véritables sites d'amorçage de fissures.

Devant cette problématique, plusieurs travaux ont été menés dans l'objectif de caractériser le phénomène et de quantifier le risque de perforation. Cette caractérisation peut être faite à l'aide de différents critères. Il s'agit notamment de potentiels de piqûration E_p et de repassivation E_r des piqûres, de la température critique de piqûration, du temps d'amorçage, du nombre, de la taille et de la profondeur des piqûres ([DBB94]). Lors d'essais potentiodynamiques (voir chapitre 1), la dispersion des potentiels E_p fut longtemps attribuée aux conditions expérimentales. Cette dispersion

1. Centre de Corrosion Marine et Biologique : www.corrodys.com

persiste malgré un contrôle rigoureux des expériences. Shibata et al. ([ST76, ST77]) émirent ainsi l'hypothèse que cette dispersion serait due à l'aspect intrinsèquement stochastique du phénomène de piqûration. Il en résulte un traitement statistique et un développement de modèles probabilistes pour étudier le phénomène. Le critère retenu dans ce travail est la profondeur des piqûres. L'une des questions majeures directement liée à ce critère est de prédire la profondeur maximale des piqûres pour une surface objective de taille S en ayant uniquement analysé des échantillons de surface s . L'origine et l'importance de cette question réside dans le fait qu'en pratique, la surface objective est très grande (aile d'avion, coque de navire, ...), ce qui rend impossible le relevé de toutes les mesures (profondeur des piqûres) sans oublier le coût qui en résulte. L'objectif est alors de réduire les incertitudes du point de vue statistique lors de l'extrapolation spatiale et de donner des intervalles de confiance des estimations des profondeurs maximales. Pour cela, la théorie des valeurs extrêmes offre un traitement adapté à ce type de problème. Initialement appliquée en hydrologie, la théorie des valeurs extrêmes (TVE) décrit le comportement asymptotique du maximum d'un échantillon et propose des estimations des événements rares qui sont souvent la cause de catastrophes de grande ampleur (crues, avalanches, fissures critiques, ...) entraînant une rupture brutale. Vu la complexité du phénomène de corrosion qui dépend du couple matériau/milieu, il est important de signaler que les études menées dans le cadre de ce travail ont été appliquées sur des tôles d'aluminium avec dans un milieu corrosif décrit dans le chapitre dédié à l'expérimentation.

L'expérimentation présente une étape de première importance dans ce travail. En effet, elle permet d'alimenter les simulations numériques et confronter les résultats de ces simulations avec celles de l'expérimentation. Les conditions fixées pour ces expérimentations (voir section 4.1) n'ont pas été facile à satisfaire dans la mesure où il n'y a pas beaucoup de couple milieu/matériau qui peuvent aboutir à des échantillons vérifiant nos conditions. Une recherche bibliographique a conclu à retenir l'aluminium comme matériau et un milieu dont la composition est détaillée au chapitre 4. Notons que des courbes de potentiel libre ont aussi été tracé dans l'objectif d'expliquer le comportement à la corrosion par piqûres des échantillons utilisés.

L'organisation de ce travail se présente comme suit :

Au premier chapitre, nous présentons le phénomène de corrosion par piqûres et détaillons les différents facteurs qui le favorise. Ensuite, un aperçu sur les mécanismes (amorçage + propagation) est donné avant de se focaliser sur la phase de propagation qui représentera le cadre de ce travail. Puis, le problème est posé mathématiquement et on revient sur les lois classiques des profondeurs des piqûres rencontrées dans la littérature ainsi que sur les tests statistiques utilisés pour valider de telles lois.

Au deuxième chapitre, nous introduisons la théorie des valeurs extrêmes et présentons les trois lois limites du maximum d'un échantillon en insistant sur leurs propriétés mathématiques et les conditions de leur obtention et application. Ensuite, on distinguera les lois des profondeurs en fonction de leur queues de distributions. La loi des valeurs extrêmes généralisée (Generalized Extreme Value) GEV est présentée, comme étant la forme généralisée des 3 lois limites, en illustrant son utilité du point de vue estimation. Enfin, on introduit la méthode de dépassement de seuil POT (Peaks Over Threshold) qui est une alternative à l'approche des maxima par blocs (Gumbel et GEV). Pour ces approches, différentes méthodes d'estimation des paramètres sont présentées (régression, moments, moments de probabilités pondérés, maximum de vraisemblance).

Au troisième chapitre, les méthodes d'analyse des données spatiales sont présentées. Il s'agit d'étudier l'existence d'interactions entre les piqûres ce qui permet ainsi de vérifier l'hypothèse d'in-

dépendance nécessaire lors de l'application de TVE. Ce chapitre est divisé en deux parties. La première s'intéresse à ces interactions en tenant compte de la position des piqûres et la deuxième en tenant compte à la fois de la position et de la profondeur associée à chaque piqûre.

Au quatrième chapitre, nous présentons l'expérimentation effectuée durant ce travail en fournissant une description détaillée du matériel utilisé, les différentes manipulations pour corroder des plaques d'aluminium ainsi que l'acquisition des données en terme de mesure des profondeurs par microscopie et profilométrie optique. Les données relevées sur des plaques d'aluminium sont utilisées pour détecter les interactions et mettre en oeuvre la méthode de Gumbel. Le recours à la simulation numérique couplée à l'expérimentation est justifié dans l'optique de clarifier certaines interrogations autour de cette méthode.

Au cinquième chapitre, les méthodes alternatives à l'approche classique de Gumbel à savoir les approches GEV et POT sont présentées. Une comparaison entre ces différentes approches en terme d'estimation de quantiles est effectuée. Enfin, nous terminons ce travail par un commentaire sur les différents résultats obtenus et par une mise en exergue de l'apport de ce travail par rapport à la littérature.

Chapitre 1

Corrosion par piqûres

1.1 Généralités

La corrosion par piqûres est la forme de corrosion localisée la plus souvent rencontrée. De l'industrie maritime et pétrolière (navire, pipeline, ...) en passant par la construction et le transport (câbles de suspension, canalisations, citernes, ...), elle peut toucher tous les secteurs et s'avérer très coûteuse. La figure (1.1) montre la répartition des différentes formes de corrosion pour les aciers inoxydables.

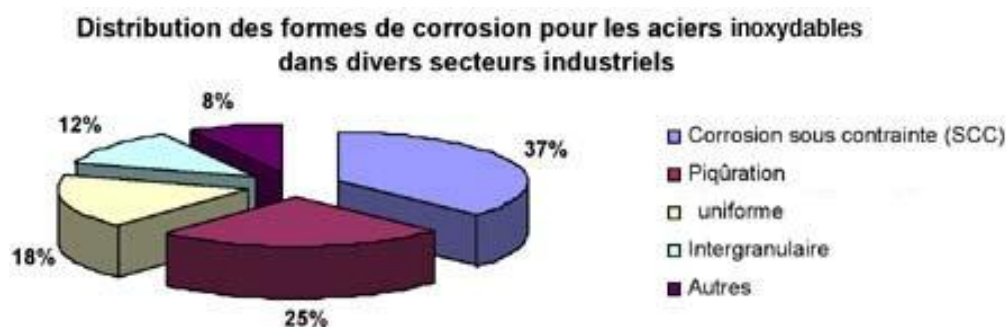


FIG. 1.1: Bilan des différentes formes de corrosion [X-PER-X]

Vis à vis de son milieu corrosif, un matériau métallique peut se trouver dans l'un des trois états possibles : l'immunité, la passivité et l'activité.

1. **L'immunité** : Dans cet état, la corrosion du matériau est impossible. Cela implique que le couple matériau/milieu se trouve dans un état de stabilité, et cette absence de corrosion n'est pas due à l'existence physique d'une « couche protectrice » entre le matériau et le milieu mais à l'absence de réaction électrochimique. Cette situation est rencontrée dans le cas des matériaux nobles comme l'or, le platine dans les milieux aqueux de pH neutre.
2. **La passivité** : Le matériau est protégé naturellement par une « couche protectrice » appelée film passif d'une épaisseur de l'ordre du nanomètre. En jouant le rôle de barrière entre le matériau et son milieu, ce film passif protège le matériau contre la corrosion. Ce comportement électrochimique est celui des matériaux comme l'aluminium, le titane, l'acier inoxydable dans les milieux aqueux de pH neutre.

3. **L'activité** : Le matériau réagit directement avec son milieu et se corrode généralement de manière uniforme sur toute la surface. C'est le cas par exemple du fer en milieu acide.

Contrairement à la corrosion généralisée, où toute la surface est attaquée suite à la répartition spatio-temporelle aléatoire des sites anodiques et cathodiques, la corrosion par piqûre se développe localement en formant une pile galvanique de corrosion entre la surface extérieure, passive et cathodique, et l'intérieur de la piqûre, actif et anodique (voir figure (1.2)) entraînant une cavité qui peut rapidement devenir profonde bien qu'une petite quantité de métal se dissolve globalement.

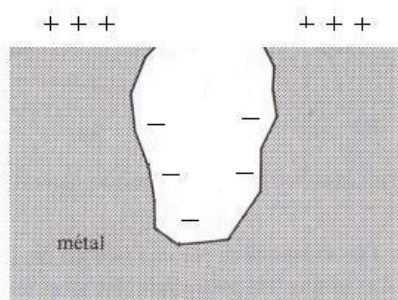


FIG. 1.2: Pile galvanique de corrosion : + Cathode, - Anode ([Lan03] modifiée)

Cela montre le caractère insidieux et dangereux du phénomène qui touche aussi les matériaux passivables présentant donc une bonne résistance à la corrosion généralisée. Il en résulte deux mécanismes d'attaque différents. La figure (1.3) schématise le développement dans le temps de l'attaque de la corrosion par piqûres et de la corrosion généralisée.

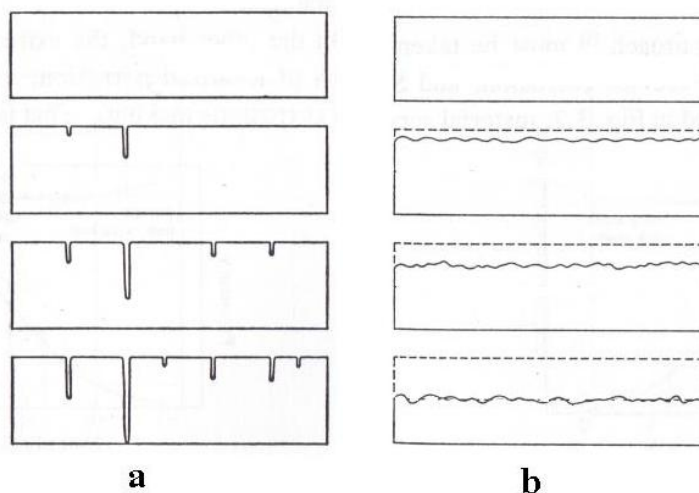


FIG. 1.3: Attaque par corrosion : a- Localisée, b- Généralisée ([Kow94] modifiée)

Finalement, la morphologie des piqûres de corrosion peut prendre plusieurs aspects selon les conditions expérimentales. La figure (1.4) illustre des exemples de formes de corrosion ([Bar04]). Dans le chapitre 3 nous montrerons la morphologie de quelques piqûres issues de notre expérimentation.

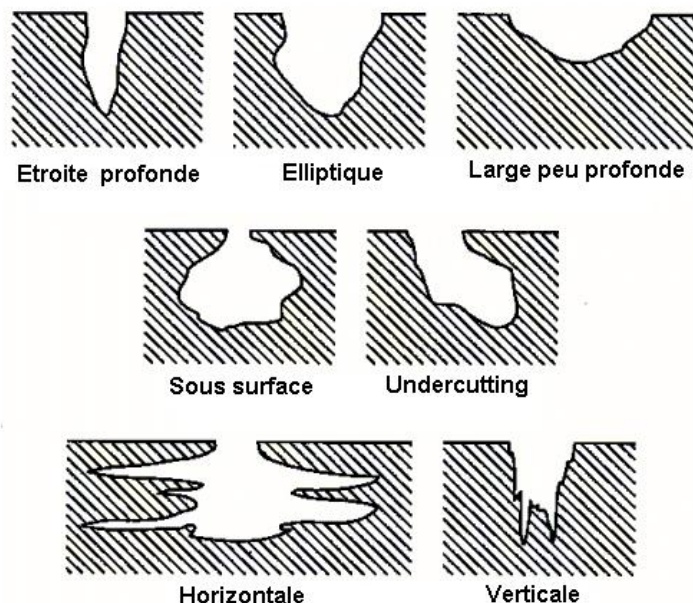


FIG. 1.4: Exemples de morphologie des piqûres de corrosion ([Bar04])

Les mécanismes et les causes de la corrosion par piqûres ont été largement étudiés dans la littérature ([SS86, Mar02]). L'objectif de cette partie est d'en présenter les aspects généraux.

1.2 Caractérisation de la sensibilité à la corrosion par piqûres

1.2.1 Potentiel de piqûres

L'une des principales caractéristiques pour étudier la sensibilité à la corrosion par piqûres est l'existence d'une valeur critique du potentiel du matériau dans son milieu corrosif pour laquelle ce type d'attaque est susceptible de se manifester. Ce concept a été introduit pour la première fois par Brenner ([DBB94]) mais c'est à Pourbaix en 1963 que revient la distinction entre deux potentiels celui de piqûration et celui de repassivation.

Il existe plusieurs méthodes pour déterminer ces deux potentiels caractéristiques. La plus utilisée est la méthode potentiodynamique qui consiste à tracer des courbes intensité (i)- potentiel (E) par balayage du potentiel dans le sens croissant et décroissant. Le potentiel de piqûration E_p traduit la rupture stable du film passif alors que le potentiel de repassivation E_r (ou de protection) est associé à l'arrêt des piqûres. Le potentiel de piqûration, est obtenu lors du balayage dans le sens croissant, il correspond à l'augmentation brutale du courant. Quant au potentiel de repassivation, il est obtenu lors de l'inversion du sens du balayage. Il correspond à l'intersection du palier de passivation avec la courbe retour. A noter que plus E_p est grand et plus la résistance à la corrosion par piqûres est élevée. La figure (1.5) montre un schéma simplifié de ce type de courbes.

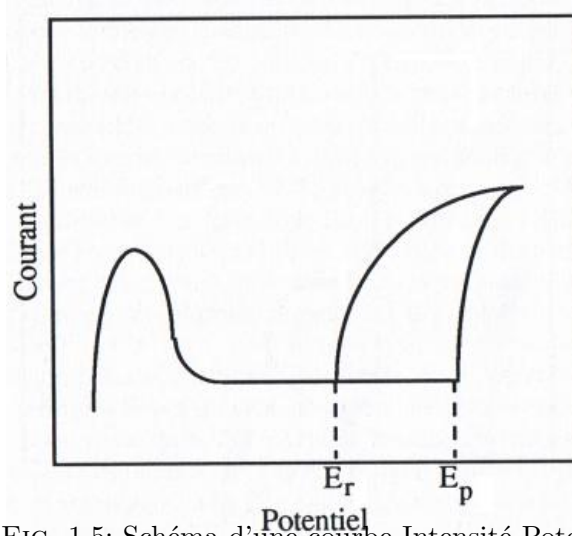


FIG. 1.5: Schéma d'une courbe Intensité-Potentiel

Ces deux potentiels peuvent aussi être déterminés par la méthode galvanostatique qui consiste à appliquer un courant constant. Le potentiel de piqûration correspond à la stabilisation du potentiel suite à la formation de piqûres stables (voir figure (1.6)).

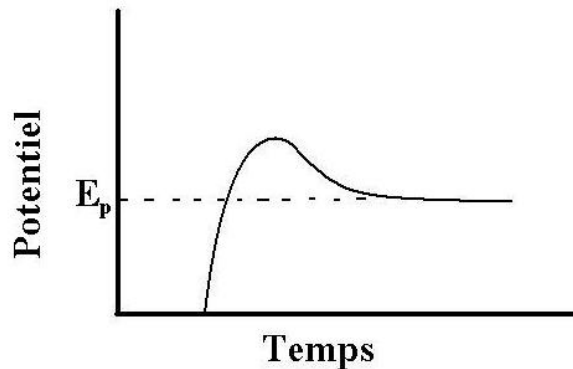


FIG. 1.6: Schéma de détermination du potentiel de piqûration par la méthode galvanostatique

Il est important de rappeler que la valeur du potentiel de piqûres n'est pas une caractéristique intrinsèque au matériau. Il s'agit d'une variable aléatoire qui dépend fortement de la microstructure du métal, de l'état de surface, de la composition chimique du milieu, de la température et aussi de la méthode de mesure utilisée.

1.2.2 Temps d'amorçage

Certains auteurs étudient la sensibilité à la corrosion par piqûres en mesurant leur temps d'amorçage. Il existe différentes interprétations pour définir ce temps d'amorçage ([DBB94]). Il peut être avancé comme étant :

- Le temps nécessaire à la piqûre pour atteindre une taille pour laquelle le courant de piqûration est supérieur au courant de passivité sur le reste de la surface.
- Le temps nécessaire pour que les réactions locales causent une décroissance du pH jusqu'à une valeur permettant un maintien de la dissolution du métal entraînant la propagation des piqûres.

- Le temps nécessaire aux ions agressifs pour pénétrer dans le film passif.

Face à ces différentes interprétations, le mécanisme d’amorçage reste très complexe bien que plusieurs expériences ont mis en évidence des relations reliant le temps d’attaque et l’épaisseur du film passif. Par conséquent des modèles probabilistes ont été élaborés pour caractériser ce temps et tenir compte des aléas du phénomène.

1.2.3 Approche statistique : Amorçage + Propagation

Amorçage : La dispersion du potentiel de piqûration lors des essais potentiodynamique a principalement été reliée au caractère intrinsèquement stochastique de l’amorçage des piqûres. Ainsi, le traitement de cette phase doit se faire via des approches probabilistes et des méthodes statistiques ([ST76, ST77]). Le modèle proposé par Shibata et Takeyama suppose que le processus d’amorçage des piqûres est un processus markovien. Cela signifie que la probabilité d’amorçage d’une piqûre à l’instant t ne dépend pas du nombre de piqûres amorcées auparavant. Autrement dit, il y a indépendance des amorçages des piqûres sur la surface des échantillons.

Propagation : En plus des approches basées sur le potentiel et la cinétique des piqûres, l’aspect propagation peut aussi caractériser la sensibilité à la corrosion par piqûres. Dans cette approche, on s’intéresse à la distribution statistique des profondeurs des piqûres et le but est de prévoir la perforation pour une surface objective S en ayant analysé des surfaces échantillons de taille s . C’est dans ce cadre que se place le travail de cette thèse ; nous reviendrons en détail sur cette approche dans les chapitres suivants.

Enfin, l’évolution dans le temps de la moyenne des profondeurs x est souvent donnée par la loi empirique ([Kow94]) :

$$x - x_a = k (t - t_a)^n \quad (1.1)$$

où :

- x_a la profondeur à partir de laquelle les piqûres continuent à croître.
- t_a le temps d’amorçage des piqûres.
- k et n des coefficients déterminés expérimentalement.

1.3 Facteurs favorisant la corrosion par piqûres

Les facteurs qui influencent la corrosion par piqûres sont multiples, mais on peut les classer en deux catégories : ceux liés au matériau et ceux liés au milieu.

1.3.1 Matériau

- **Composition :** l’incorporation d’éléments d’alliages aux matériaux peut augmenter la résistance à la corrosion. C’est le cas par exemple de la présence du chrome pour les aciers inoxydables dans des proportions bien précises. Dans le cas où la concentration en chrome devient supérieure à 12%, un film passif se forme à la surface de l’alliage. Dans le cas de l’aluminium et de ses alliages, ce film se forme spontanément. Les caractéristiques de ce film peuvent dépendre de plusieurs paramètres. Nous citons :

1. La composition chimique et l’épaisseur : La composition chimique du film passif dépend à la fois du substrat métallique, des traitements de surface subis par le matériau et

du milieu dans lequel il est immergé (molécules d'eau, anions OH^- , cations de métaux oxydables, ...). Quant à l'épaisseur, elle est en général de l'ordre du nanomètre et il apparaît dans de nombreux travaux ([DBB94]) que le temps d'amorçage est corrélé positivement avec l'augmentation de l'épaisseur du film passif.

2. Conductivité : Le processus électrochimique de la corrosion entraîne des transferts de charges électrique ; ainsi on a une bonne résistance à la corrosion par piqûres si le film se comporte comme un isolant.
- **Hétérogénéités métallurgiques** : La préexistence de toute sorte d'hétérogénéité peut constituer de véritables sites d'amorçages de piqûres. Parmi ces sites, on peut distinguer ceux relatifs aux défauts cristallins tels que les dislocations, les joints de grain, des ségrégations souvent rencontrées pour les alliages d'aluminium ainsi que des inclusions non métalliques tel que les sulfures de manganèse pour les aciers inoxydables ([DBB94]).

1.3.2 Milieu

Comme le phénomène de corrosion fait intervenir le couple matériau/milieu, l'activité du milieu peut influencer la corrosion par piqûres autant que le matériau. En effet la présence d'espèces agressives telles que les ions halogénures peut facilement déstabiliser le film passif. La température est aussi un facteur important puisqu'elle augmente la vitesse de propagation des piqûres et diminue les temps d'amorçage et les potentiels de piqûration ([DBB94]). Finalement, dans certains milieux, des bactéries peuvent aussi intervenir en modifiant les conditions locales d'aération ou en créant des produits de métabolisme agressifs.

1.4 Mécanisme de la corrosion par piqûres

Comme indiqué précédemment, le processus de corrosion par piqûres est complexe et nécessite un traitement prudent pour mieux interpréter les caractéristiques du phénomène. En général, on distingue les trois étapes : la rupture de passivité en condition métastable, la croissance de piqûres stables et la repassivation.

1.4.1 Rupture de la passivité

Les films passifs ne sont pas inertes. Ils sont en permanence le siège de différentes réactions qui déterminent leur aptitude à protéger contre la corrosion par piqûres. En général, la plupart des auteurs distinguent trois mécanismes de rupture de la passivité à savoir les mécanismes de pénétration, de rupture de film et d'adsorption.

1. **Mécanisme par pénétration** : Ce mécanisme met en jeu le transfert d'anions agressifs à travers le film passif jusqu'à la surface du métal où débute leur mise en réaction. Ce transfert peut être expliqué ([DBB94]) par la présence d'un champ électrique élevé et une concentration importante de défauts dans la couche d'oxyde. En milieu chloruré, cela suppose que la piqûre se produit lorsque les ions Cl^- arrivent à l'interface métal-film. La figure (1.7) montre un schéma de ce type de mécanisme.

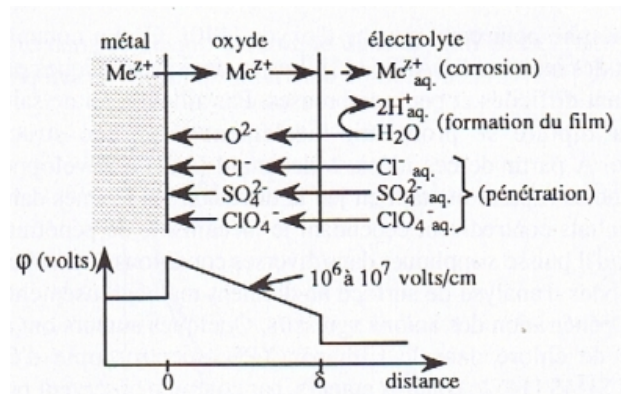


FIG. 1.7: Schéma du mécanisme d'amorçage par pénétration ([DBB94])

2. **Mécanisme par rupture de film** : Vetter et Strehblow (cité dans [DBB94]) ont proposé un mécanisme qui nécessite des ruptures dans le film permettant l'accès direct des anions à la surface métallique non protégée. La figure (1.8) montre un schéma de ce type de mécanisme.

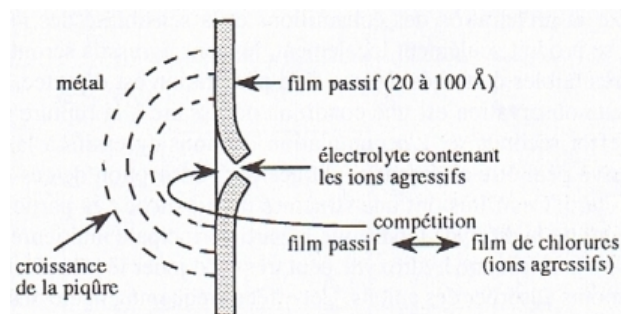


FIG. 1.8: Schéma du mécanisme d'amorçage par rupture du film ([DBB94])

3. **Mécanisme par adsorption** : Il commence avec la formation en surface de complexes qui peuvent mettre en jeu des anions agressifs et causer l'amincissement local du film passif et sa destruction en formant des piqûres.

Ces mécanismes ne se produisent pas forcément de manière exclusive les uns par rapport aux autres. Un mécanisme d'adsorption peut par exemple précéder la pénétration d'ions agressifs. Cependant ils ont tous un effet accélérateur de la différence de potentiel métal/milieu et de la concentration des ions agressifs en solution.

1.4.2 Croissance des piqûres

Une fois les piqûres amorçées, la condition générale pour assurer leur croissance est d'avoir une instabilité du film passif. Cela peut se produire si les concentrations locales et les teneurs en espèces agressives sont suffisantes pour empêcher toute repassivation. Du point de vue expérimental, le suivi de la propagation des piqûres est principalement effectué par l'étude des densités de courant déduites des mesures de courant de dissolution et de la géométrie des piqûres.

1.4.3 Repassivation

Comme indiqué ci-dessus, le comportement des piqûres est assez complexe. Après son amorçage, la piqûre peut continuer à croître, mais ce processus peut s'arrêter pour différentes raisons. C'est la repassivation. Parmi les facteurs qui peuvent arrêter la croissance des piqûres on peut citer la chute ohmique. Par contre elle ne concerne que les piqûres qui sont dans un stade avancé avec un rayon de quelques mm ([DBB94]), et ne peut servir d'explication pour les piqûres de petites dimensions (de l'ordre du μm).

Face à cette complexité du phénomène de piqûration, nous avons choisi d'étudier le phénomène du point de vue statistique en se plaçant dans la phase de propagation sans se préoccuper des mécanismes. Ainsi, le problème se ramène à la quantification du risque de perforation des structures en estimant la profondeur maximale des piqûres pour une surface objective S . Pour aborder ce type de problème, la théorie des valeurs extrêmes offre un traitement statistique. Cela fera l'objet du chapitre suivant. Ainsi, pour mettre en oeuvre les différentes méthodes statistiques et les simulations numériques, nous avons besoin de connaître la distribution en loi de probabilité des profondeurs des piqûres. Ce chapitre se termine par un rappel des différentes lois classiques rencontrées en corrosion ainsi que leur propriétés, et nous présentons les trois tests d'ajustement utilisés dans ce travail à savoir le test du Chi-deux, le test de Kolmogorov et le test d'Anderson Darling.

1.5 Lois de probabilités en corrosion par piqûres

1.5.1 Loi de Poisson

C'est une loi discrète utilisée pour décrire la variable aléatoire N_p correspondant au nombre de piqûres dans le temps et dans l'espace ([Fie92, Gum58]). Si λ désigne le nombre moyen de piqûres par unité de surface, alors la probabilité d'observer k piqûres sur une surface A (d'aire $|A|$) est donnée par :

$$\mathbb{P}(N_p = k) = \frac{(\lambda |A|)^k}{k!} \exp(-\lambda |A|) \quad (1.2)$$

L'espérance et l'écart type de cette variable sont respectivement donnés par :

$$\mathbb{E}[N_p] = \lambda |A| \quad \sigma[N_p] = \sqrt{\lambda |A|} \quad (1.3)$$

où $\mathbb{E}[\cdot]$ désigne l'opérateur de l'espérance et $\sigma[\cdot]$ l'opérateur de l'écart-type.

1.5.2 Loi Exponentielle

La loi exponentielle peut modéliser la variable aléatoire correspondant au temps d'amorçage des piqûres notée T_a . Si μ désigne le nombre moyen d'apparitions de piqûres par unité de temps, alors la probabilité qu'au moins une piqûre s'amorce à l'instant t est donnée par :

$$\mathbb{P}(t) = 1 - \exp(-\mu t) \quad (1.4)$$

L'espérance et l'écart-type sont donnés par :

$$\mathbb{E}[T_a] = \frac{1}{\mu} \quad \text{et} \quad \sigma[T_a] = \frac{1}{\mu^2} \quad (1.5)$$

1.5.3 Loi Normale

Appelée aussi la loi de Gauss, c'est la loi la plus utilisée par excellence. En corrosion par piqûres, elle peut représenter la variable aléatoire se référant à la profondeur des piqûres X ([Lei54]). La probabilité d'avoir une profondeur inférieure à une valeur x est entièrement définie avec la donnée de deux paramètres μ et σ correspondant respectivement à la moyenne et l'écart-type de la variable X . Son expression est donnée par :

$$\mathbb{P}(X \leq x) = \frac{1}{\sqrt{2\pi}\sigma} \int_{-\infty}^x \exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) dt \quad (1.6)$$

L'espérance et l'écart-type sont donnés par :

$$\mathbb{E}[X] = \mu \quad \text{et} \quad \sigma[X] = \sigma \quad (1.7)$$

1.5.4 Loi log-normale

Une variable aléatoire Y strictement positive suit la loi log-normale de paramètres μ et σ si son logarithme suit une loi normale de paramètres μ et σ . Ce type de distribution peut modéliser la profondeur des piqûres ([NLB⁺02]) et aussi le temps de défaillance dans la corrosion sous contrainte ([Kow94]). La probabilité pour que Y prenne des valeurs inférieures à y est donnée par :

$$\mathbb{P}(Y \leq y) = \int_0^y \frac{1}{\sqrt{2\pi}\sigma} \frac{1}{t} \exp\left(-\frac{1}{2} \left(\frac{\log t - \mu}{\sigma}\right)^2\right) dt \quad (1.8)$$

L'espérance et l'écart-type sont donnés par :

$$\mathbb{E}[Y] = \exp\left(\mu + \frac{\sigma^2}{2}\right) \quad \text{et} \quad \sigma[Y] = \sqrt{\exp(2\mu + \sigma^2) (\exp(\sigma^2) - 1)} \quad (1.9)$$

1.5.5 Loi de lambda distribution généralisée

La loi de lambda distribution généralisée connue aussi sous le nom de Tukey-Lambda est connue par la donnée de sa fonction quantile (l'inverse de la fonction de répartition) définie par :

$$Q(u, \lambda_1, \lambda_2, \lambda_3, \lambda_4) = \lambda_1 + \frac{u^{\lambda_3} - (1-u)^{\lambda_4}}{\lambda_2} \quad (1.10)$$

où λ_1, λ_2 représentent respectivement le paramètre de localisation et d'échelle et les paramètres λ_3 et λ_4 sont déterminés respectivement par le coefficient de dissymétrie (skewness) et le coefficient d'aplatissement (kurtosis).

Ce type de distribution a été utilisé pour estimer les quantiles des profondeurs maximales des piqûres de corrosion ([NBLI03]).

1.6 Tests classiques d'ajustement et d'analyse de données

1.6.1 Tests classiques d'ajustement

1.6.1.1 Test du Chi-deux

C'est sans doute le test le plus utilisé. Il tire son nom du fait que l'on lit la valeur critique dans la table de la loi de Chi-deux. Il comporte plusieurs variantes, essentiellement le test d'adéquation qui compare la distribution observée dans un échantillon à une distribution théorique, le test

d'indépendance qui vérifie si deux caractères d'une population sont indépendants et le test d'homogénéité qui teste si les échantillons observés sont issus d'une même population. Dans ce travail, on s'intéresse à la première variante. Ainsi, si l'on dispose d'un échantillon de valeurs représentant des profondeurs de piqûres ; dans un premier temps, ces valeurs sont réparties en r classes C_1, \dots, C_r . Les probabilités théoriques dans chacune des classes sont notées p_1, \dots, p_r , les effectifs observés n_1, \dots, n_r et la taille totale de l'échantillon N .

L'hypothèse testée (*nulle*) H_0 est : « L'échantillon observé est issu de la distribution théorique » avec un risque d'erreur α .

Les étapes du test sont les suivantes :

1. On calcule les effectifs théoriques Np_i pour $i = 1, \dots, r$
2. On calcule la valeur observée de la variable du test :

$$\chi^2 = \sum_{i=1}^r \frac{(n_i - Np_i)^2}{Np_i} \quad (1.11)$$

3. On cherche la valeur critique χ_α^2 dans la table de la loi de Chi-deux à $r - 1$ degrés de liberté.
4. Si $\chi^2 < \chi_\alpha^2$, on ne peut rejeter l'hypothèse nulle, sinon on la rejette.

Une condition d'application du test à vérifier a priori est que $Np_i \geq 5$, sinon on procède à un regroupement de classes.

1.6.1.2 Test de Kolmogorov

Il est souvent utilisé de préférence au test de Chi-deux lorsque le caractère observé est continu. Soit X ce caractère de fonction de répartition théorique connue F_0 . Etant donné un échantillon (x_1, \dots, x_N) de N observations, l'hypothèse nulle à tester est : « La fonction de répartition F_0 de X est égale à la fonction de répartition observée F ».

On peut résumer les étapes du test comme suit :

1. On ordonne les valeurs observées $x_1 \leq x_2 \leq \dots \leq x_N$
2. La fonction de répartition de l'échantillon dite empirique se calcule en posant :

$$F(x_1) = \frac{1}{N}, \quad F(x_2) = \frac{2}{N}, \dots, F(x_N) = 1 \quad (1.12)$$

3. On calcule :

$$K = \max_{1 \leq i \leq n} \left(\frac{i}{n} - F_0(x_i), F_0(x_i) - \frac{i-1}{n} \right) \quad (1.13)$$

4. On lit la valeur critique D_n dans la table de la loi de Kolmogorov. Si $K < D_n$, on ne rejette pas l'hypothèse nulle, sinon on la rejette.

1.6.1.3 Test d'Anderson-Darling

Comme dans le test de Kolmogorov, le test d'Anderson-Darling consiste à comparer la distribution théorique $F_0(x)$ à la distribution de l'échantillon (empirique) $F(x)$ mais en calculant la statistique suivante :

$$\int_{-\infty}^{+\infty} [F(x) - F_0(x)]^2 w(x) F'(x) dx \quad (1.14)$$

où $w(x)$ est une fonction de pondération, et $F'(x)$ la fonction dérivée de $F(x)$.

Le cas standard du test d'Anderson-Darling correspond au choix de la fonction de pondération suivante :

$$w(x) = \frac{1}{F_0(x)[1 - F_0(x)]} \quad (1.15)$$

Ce choix permet de tenir compte des faibles et des fortes fréquences. Le test devient plus sensible au comportement des fréquences rares pour un choix de la fonction de pondération égale à

$$w(x) = \frac{1}{1 - F_0(x)} \quad (1.16)$$

ce qui peut s'avérer utile dans le cadre des valeurs extrêmes. Le test n'est pas rejeté à un risque α si la statistique du test est inférieure à la valeur critique tabulée.

1.6.2 Tests d'analyse de données

L'analyse des données est une branche des statistiques qui s'intéresse à la description des données conjointes. Il s'agit de chercher les liens qui peuvent exister entre les différentes variables étudiées ou de les classer en sous groupes plus homogènes. Les principales méthodes d'analyse de données sont :

- L'analyse de variance (ANOVA) : Pour tester s'il existe une différence significative (en moyenne) entre les différents échantillons.
- L'analyse en composante principale (ACP) : Utilisée pour trouver les variables les plus influentes dans le phénomène étudié.
- L'analyse factorielle discriminante (AFD) : Permet d'identifier des groupes homogènes au sein de la population.

Dans ce travail, c'est le test ANOVA qui va être utilisé. En effet, on va s'intéresser à tester s'il existe une différence significative entre les mesures et les estimations qu'on détaillera au quatrième chapitre.

1.6.2.1 Tests d'analyse de variance ANOVA à un facteur

Etant donné k groupes (ou échantillons), le test ANOVA consiste à tester l'égalité des moyennes des groupes. Le terme « un facteur » signifie que le test porte sur un seul paramètre qui est la moyenne. Supposons que les données puissent être résumées par la variable (x_{ij}) avec $1 \leq i \leq n$ et $1 \leq j \leq k$ où n est le nombre d'observations de chaque groupe, et k le nombre de groupes. Le tableau (1.1) schématise ces données :

Facteur	Groupe 1	Groupe 2	...	Groupe k
Effectif	n_1	n_2	...	n_k
Mesure	x_{11}	x_{12}	...	x_{1k}
Mesure	x_{21}	x_{22}	...	x_{2k}
⋮	⋮	⋮	⋮	⋮
Moyennes	\bar{x}_1	\bar{x}_2	...	\bar{x}_k

TAB. 1.1: Exemple de tableau de données

On cherche à tester :

$H_0 : \bar{x}_1 = \bar{x}_2 \cdots = \bar{x}_k = \bar{x}$ contre

H_1 : Il existe au moins deux groupes qui ont deux moyennes significativement différentes.

Le principe du test ANOVA est basé sur la décomposition de la variation totale des données en deux composantes :

$$Var_T = Var_{inter} + Var_{intra} \quad (1.17)$$

où la variable Var_T correspond à la variation totale autour de la moyenne. Elle est définie par :

$$Var_T = \sum_{i,j} (x_{ij} - \bar{x})^2 \quad (1.18)$$

La variable Var_{inter} est la variation due au facteur (intergroupe). Elle est définie par :

$$Var_{inter} = \sum_{j=1}^k n_j (\bar{x}_j - \bar{x})^2 \quad (1.19)$$

et la variable Var_{intra} correspond à la variation due à l'erreur (intragroupe). Elle est donnée par :

$$Var_{intra} = \sum_{i,j} (x_{ij} - \bar{x}_j)^2 \quad (1.20)$$

La statistique du test, notée F , est donnée par :

$$F = \frac{Var_{inter}/(k-1)}{Var_{intra}/(n-k)} \quad (1.21)$$

Sous l'hypothèse nulle H_0 , la statistique F suit la loi de Fisher à $(k-1, n-k)$ degrés de liberté. Pour un seuil α donné, le test permet de rejeter H_0 si la valeur de F est supérieure à la valeur critique (ou théorique) calculée à partir de la loi de Fisher. En pratique, la plupart des logiciels de statistique renvoient la p-valeur du test qu'on compare avec le seuil α fixé auparavant. Cette valeur est définie comme étant le niveau significatif le plus bas où l'hypothèse nulle peut-être rejetée. Ainsi, si la p-valeur est inférieure à α , l'hypothèse nulle est rejetée, et elle acceptée dans le cas contraire.

Pour appliquer le test ANOVA, il faut vérifier les hypothèses suivantes :

- L'indépendance des observations.
- La normalité des résidus. Le vecteur résidu est défini comme étant la différence entre les composantes du vecteur et leur moyenne.
- L'homoscédasticité : c'est l'hypothèse qui concerne l'égalité des variances des échantillons.

La vérification de l'indépendance des échantillons est laissée à l'expérimentateur car elle dépend de la façon dont les données ont été relevées. Quant à la normalité des résidus, il est possible de la vérifier graphiquement par la méthode des quantile-quantile plot en mettant en abscisse les quantiles de la loi normale et en ordonnée les observations. Ainsi, la normalité est acceptée si les points des observations sont répartis linéairement par rapport à la droite des quantiles théoriques. Dans le cas où la linéarité est peu évidente à cause de l'existence de points s'écartant de la droite théorique, l'un des tests d'ajustement présentés précédemment peut être envisagé. Enfin, pour l'homoscédasticité, l'un des tests couramment utilisés est celui de Bartlett. Il consiste à calculer la statistique T_B donnée par :

$$T_B = \frac{(N - k) \ln (S_p^2) - \sum_{i=1}^k (n_i - 1) \ln (S_i^2)}{1 + \frac{1}{3(k-1)} \left(\sum_{i=1}^k \left(\frac{1}{n_i - 1} \right) - \frac{1}{N-k} \right)} \quad (1.22)$$

où k représente le nombre de classes, n_i le nombre d'observations dans le groupe i , $N = \sum_{i=1}^k n_i$, $S_p^2 = \frac{1}{N-k} \sum_{i=1}^k (n_i - 1) S_i^2$ et S_i^2 la variance du groupe i .

La règle de décision consiste à rejeter l'hypothèse d'égalité des variances si $T_B > \chi_{\alpha, k-1}^2$ où $\chi_{\alpha, k-1}^2$ est la valeur critique donnée par la loi de Chi-deux au seuil α et à $k - 1$ degrés de liberté.

Les tests paramétriques sont contraignants en terme de conditions d'applications. Mais une fois que toutes ces conditions sont réunies, ces test sont plus puissants que leurs analogues en approche non-paramétrique. Dans le cas où les conditions d'application du test ANOVA ne sont pas réunies, l'analyse de variance peut se faire à l'aide de sa version non-paramétrique. Il s'agit du test de Kruskal-Wallis qui s'applique sur les rangs des observations. Ce test s'effectue de la manière suivante :

Premièrement, on détermine le rang des observations de la plus petite (rang 1) à la plus grande (rang $N = \sum_{i=1}^k n_i$). Dans le cas d'observations égales, des rangs moyens sont affectés. Soit r_{ij} le rang de l'observation x_{ij} . La statistique du test est donnée par :

$$T_{KW} = (N - 1) \frac{S_t^2 - C}{S_r^2 - C} \quad (1.23)$$

où $S_t^2 = \sum_{i=1}^k s_i^2/n_i$, s_i la somme des rangs du groupe i , $S_r^2 = \sum_{i,j} r_{ij}^2$ et $C = \frac{1}{4}N(N + 1)^2$

La règle de décision est la même que pour la statistique T_B .

Conclusion : Dans ce chapitre, une introduction au phénomène de la corrosion par piqûres a été développée. Cela consistait à rappeler les aspects généraux du phénomène en terme de mécanismes. Dans ce travail, l'étude de ce phénomène se fera du point de vue statistique, c'est ainsi que les différentes distributions et tests statistiques utilisés pour ce propos ont été présentés à la fin de ce chapitre. Ces outils seront appliqués à des cas basés sur des simulations numériques et sur l'expérimentation. Cela fera l'objet des chapitres 4 et 5.

Chapitre 2

Théorie des Valeurs Extrêmes

2.1 Vers la théorie des valeurs extrêmes

L'étude de la fréquence des événements rares qui peuvent résulter de différents phénomènes (pluviométrie, crue, tremblement de terre, fissures critiques, ...) constitue un problème majeur et de grand intérêt, car la prise en compte du risque dû à ces événements peut efficacement aider à éviter de réelles catastrophes. A titre d'exemple, notons qu'une estimation correcte du risque de crue, fournit des informations indispensables pour le dimensionnement des digues et de leur importance. Dans le cadre de ce travail, une bonne estimation de la valeur maximale des profondeurs des piqûres de corrosion dans une structure aide à anticiper le risque de perforation de celle-ci ou celui de sites potentiels d'amorçage de fissures de taille critique pouvant causer une rupture brutale dans le cas d'une structure sollicitée mécaniquement au contact d'un milieu corrosif. On citera par exemple la fatigue-corrosion. Dans un tel contexte, l'objectif de ce travail, est de prévoir la valeur maximale des profondeurs de piqûres sur une surface donnée, ainsi que l'incertitude sur son estimation. Du point de vue statistique, cela se ramène à l'analyse des queues de distributions en cherchant à estimer les quantiles extrêmes dépassés avec une probabilité p très faible.

Pour aborder ce type de problème, les approches classiques basées sur les fréquences empiriques et les tests d'adéquation montrent rapidement leurs limites quand il s'agit des queues de distribution. Deux problèmes apparaissent :

a) Fréquences empiriques : Pour une variable aléatoire X donnée, qui admet comme réalisation l'échantillon de taille n , $x_1 < x_2 < \dots < x_n$, une approximation de la fonction de répartition de cet échantillon peut être exprimée via la fonction de répartition empirique suivante :

$$\hat{F}_n(x) = \frac{i}{n} \text{ si } x \in [x_i, x_{i+1}[\quad (2.1)$$

L'estimation par exemple de la probabilité $P(X > x_i)$ pour $i = 1, \dots, n$ est donnée par $1 - \hat{F}_n(x_i)$. Le problème se pose lorsqu'on souhaite estimer $P(X > x)$ avec $x > x_n$. Avec la construction ci-dessus (eq. 2.1), cette probabilité est nulle. En d'autres termes, on ne peut pas attribuer de probabilité pour des événements non observés. Ce problème a été mis en évidence par ([BV64]) et ([BGST04]).

b) Tests d'adéquation : Les tests classiques (c.f (1.6.1) par exemple), mesurent l'ajustement des données dans la partie centrale de l'intervalle où est observé l'échantillon ([Gar02]). Ainsi, dans le contexte où on s'intéresse aux queues des distributions, l'extrapolation aux quantiles ex-

trêmes peut conduire à des estimations grossières. Cela intervient en particulier lorsque $p < \frac{1}{n}$ ([EAD02, HM82]). Dieblot et al. ([DGG03]) proposent le test ET (Exponential Tail) pour vérifier si la fonction de répartition du modèle testé est bien adaptée aux queues de distribution. Cela se justifie par le fait que les parties centrales de plusieurs distributions peuvent présenter un comportement similaire, alors que dans les queues des écarts importants apparaissent engendrant des estimations pouvant varier sensiblement, surtout pour les très faibles probabilités p .

Face à ces difficultés, la littérature s'est intéressée au comportement asymptotique de la loi du maximum, avec l'idée que la connaissance de cette loi permet la détermination du comportement des valeurs extrêmes situées en queues de distribution. Pour cela, il existe deux approches classiques. La première est basée sur les maxima par bloc alors que la seconde est basée sur la notion de dépassement de seuil.

Dans la littérature, c'est la première approche qui a été étudiée. Les premiers résultats théoriques sont essentiellement dûs à Fisher et Tippett (1928), puis à Gnedenko (1943). Mais c'est à partir des années 50 que cette branche des statistiques s'est développée avec notamment les travaux de Jenkinson ([Jen55]) et l'ouvrage de Gumbel ([Gum58]). Ce dernier constitue une référence classique. Plus tard, la littérature s'est enrichie en particulier avec les ouvrages de Leadbetter et al. ([LLR83]), de Resnick ([Res87]) ainsi que des travaux de Galambos ([Gal95, Gal87]). Ce dernier ayant traité et proposé un historique des principaux résultats. Récemment, la modélisation des valeurs extrêmes s'est accrue, en particulier dans le domaine des finances avec comme ouvrage de référence, celui de Embrechts et al. ([EKM97]).

Quant à la deuxième approche, elle a été développée dans l'optique de proposer une alternative à l'approche des maxima par bloc, dont la principale critique est la perte d'information due au fait qu'on ne conserve que la valeur maximale pour chaque bloc. Un des résultats fondamentaux dans la théorie des valeurs extrêmes est le théorème qui stipule l'existence d'une loi asymptotique des dépassements de seuil (les excès). Celui-ci a été démontré par Pickands ([Pic75]) ainsi que par Balekema et al. ([BdH74, BdH78a, BdH78b]). Parmi les travaux ayant contribué au développement de cette approche, on cite ceux de Davidson et al. ([DS90]). Dans ce chapitre, nous présenterons les principaux résultats théoriques liés aux deux approches en illustrant leurs différences.

2.2 L'approche des maxima par blocs

2.2.1 Théorème fondamental des valeurs extrêmes

Dans ce qui suit, on désigne par X la variable aléatoire (v.a) qui représente la profondeur et de fonction de distribution F . On suppose que les données observées sont des réalisations d'une suite de variables aléatoires X_1, X_2, \dots, X_n indépendantes et identiquement distribuées (i.i.d) de même loi F . On désigne par $X_{1,n} \leq \dots \leq X_{n,n}$ la statistique d'ordre associée. Le théorème fondamental des valeurs extrêmes est analogue au théorème central limite (TCL) rapporté à la somme d'une suite de variables aléatoires de même espérance et de variance finie ; qui convergent vers une loi normale. Le résultat du théorème TVE est le suivant :

Théorème 2.1 (Fisher-Tippett). *Soit $M_n = \max(X_1, \dots, X_n)$. S'il existe deux suites normali-*

santes réelles $(a_n)_{n \geq 1} > 0$ et $(b_n)_{n \geq 1}$ et une distribution non-dégénérée¹ G telles que

$$\mathbb{P} \left(\frac{M_n - b_n}{a_n} \leq x \right) = F^n(a_n x + b_n) \xrightarrow{\text{loi}} G(x), \quad n \rightarrow +\infty \quad (2.2)$$

alors G est nécessairement de l'un des trois types suivant :

$$\textbf{Type Fréchet} : G(x) \equiv \phi_\alpha(x) = \begin{cases} 0 & \text{si } x \leq 0 \\ \exp(-x^{-\alpha}) & \text{si } x > 0 \end{cases} \quad (2.3)$$

$$\textbf{Type Weibull} : G(x) \equiv \psi_\alpha(x) = \begin{cases} \exp(-(-x)^\alpha) & \text{si } x \leq 0 \\ 1 & \text{si } x > 0 \end{cases} \quad (2.4)$$

$$\textbf{Type Gumbel} : G(x) \equiv \Lambda(x) = \exp(-\exp(-x)) \quad x \in \mathbb{R} \quad (2.5)$$

avec α un paramètre strictement positif.

La variable $\frac{M_n - b_n}{a_n}$ est appelée maximum normalisé.

Remarque : On note que les deux suites normalisantes ne peuvent être ignorées car la loi du maximum serait dégénérée dans la mesure où $\mathbb{P}(M_n \leq x) = F^n(x)$ tend vers 0 ou 1 puisque $F(x) \in [0, 1]$.

Exemples :

- **Loi de Cauchy (Type Fréchet) :** Soit X_1 une variable qui suit la loi de Cauchy de paramètre 1. La densité de cette loi est $f(x) = \frac{1}{\pi(1+x^2)}$, $x \in \mathbb{R}$. On montre que :

$$\mathbb{P} \left(\frac{\pi M_n}{n} \leq x \right) = \mathbb{P} \left(X_1 \leq \frac{nx}{\pi} \right)^n \xrightarrow{n \rightarrow +\infty} \exp(-x^{-1}), \quad x > 0 \quad (2.6)$$

- **Loi uniforme (Type Weibull) :** Soit X_2 une variable aléatoire suivant la loi uniforme sur $[0, \theta]$, $\theta > 0$. La fonction de répartition de cette loi est $F(x) = x/\theta$ pour $x \in [0, \theta]$. Soit F_n la fonction de répartition de $n \left(\frac{M_n}{\theta} - 1 \right)$. On montre que F_n est de type Weibull. En effet, si $x \geq 0$ on a $F_n(x) = 1$ car $M_n < \theta$. Pour $x < 0$ on a :

$$\mathbb{P} \left(n \left(\frac{M_n}{\theta} - 1 \right) \leq x \right) = \mathbb{P} \left(X_2 \leq \theta + \theta \frac{x}{n} \right)^n \xrightarrow{n \rightarrow +\infty} \exp(x), \quad x < 0 \quad (2.7)$$

- **Loi exponentielle (Type Gumbel) :** Soit X_3 une variable aléatoire qui suit la loi exponentielle de paramètre $\lambda > 0$. La fonction de répartition de cette variable est $F(x) = 1 - \exp(-\lambda x)$ pour $x > 0$. On montre que :

$$\mathbb{P}(\lambda M_n - \log(n) \leq x) = \mathbb{P} \left(X_3 \leq \frac{x + \log(n)}{\lambda} \right)^n \xrightarrow{n \rightarrow +\infty} \exp(-\exp(-x)), \quad x \in \mathbb{R} \quad (2.8)$$

¹Non réduite à un point

2.2.2 Loi des valeurs extrêmes généralisée GEV

Jenkison ([Jen55]) a proposé une seule forme paramétrique qui regroupe les trois lois limites, appelée loi des valeurs extrêmes généralisée (GEV), et qui dépend d'un seul paramètre ξ . Elle est donnée par :

$$\mathbf{GEV} : G_\xi(x) = \begin{cases} \exp\left(- (1 + \xi x)^{-\frac{1}{\xi}}\right) & \text{si } \xi \neq 0, \forall x \text{ tel que } 1 + \xi x > 0 \\ \exp(-\exp(-x)) & \text{si } \xi = 0 \end{cases} \quad (2.9)$$

Le paramètre ξ est appelé indice extrême. Son signe renseigne sur le type de la loi asymptotique du maximum et on a pour ξ non nul, $|\xi| = \frac{1}{\alpha}$. On distingue alors entre :

- $\xi = 0 \Rightarrow$ Type Gumbel et on a $\Lambda(x) = G_0(x)$
- $\xi > 0 \Rightarrow$ Type Fréchet avec $\phi_\alpha(x) = G_{\frac{1}{\alpha}}(\alpha(x-1))$.
- $\xi < 0 \Rightarrow$ Type Weibull avec $\psi_\alpha(x) = G_{-\frac{1}{\alpha}}(\alpha(x+1))$.

La figure (2.1) montre trois exemples de la fonction densité de la loi GEV en fonction de la variable réduite x

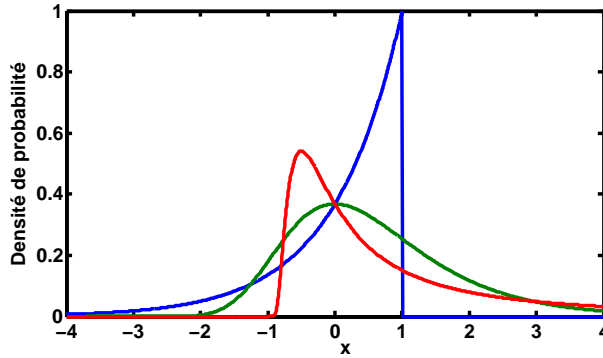


FIG. 2.1: Exemple de fonctions de densité de probabilité avec différentes valeurs de ξ

La distinction entre les trois lois limites fait appel à la notion du domaine d'attraction du maximum. Cela fait l'objet de la partie suivante.

2.2.3 Domaine d'attraction du maximum

Dire qu'une distribution F appartient au domaine d'attraction du maximum de G_ξ que l'on note $MDA(G_\xi)$, signifie que la loi asymptotique du maximum normalisé d'un échantillon de v.a. i.i.d est G_ξ . L'objectif de cette partie est de présenter les conditions d'appartenance à un MDA , dans le cas où la loi F est connue. Dans le cas où celle-ci est inconnue, ce qui est souvent le cas en pratique, des techniques graphiques de détermination de ce MDA existent et seront présentées. Pour plus de détails sur les démonstrations des résultats suivants, le lecteur pourra consulter les références ([Res87, Gal87, BGST04]).

Les deux caractéristiques classiques d'appartenance à un domaine d'attraction sont souvent données à partir de la fonction du hasard et des fonctions à variations lentes qui sont définies dans

la suite de ce paragraphe. Les notions suivantes seront nécessaires dans ce qui va suivre :

- Le point terminal d'une distribution F est défini par :

$$w(F) = \sup \{x \in \mathbb{R} : F(x) < 1\} \quad (2.10)$$

- L'inverse généralisée d'une fonction de répartition est donnée par :

$$F^{\leftarrow}(y) = \inf \{x : F(x) \geq y\} = Q(y) \quad (2.11)$$

où $Q(\cdot)$ représente la fonction quantile. On note que l'inverse généralisée coïncide avec l'inverse classique F^{-1} lorsque la fonction F est strictement croissante.

- La fonction de survie d'une distribution est définie par :

$$\bar{F}(x) = 1 - F(x) \quad (2.12)$$

- La fonction du hasard d'une distribution F de densité f est définie par :

$$r(x) = \frac{f(x)}{1 - F(x)} \quad (2.13)$$

- On dit qu'une fonction ℓ de la variable t est à variation lente si :

$$\forall x > 0 \quad \lim_{t \rightarrow +\infty} \frac{\ell(tx)}{\ell(t)} = 1 \quad \text{et} \quad \ell(t) > 0 \quad (2.14)$$

2.2.3.1 Domaine d'attraction de Fréchet $\xi > 0$

Les lois appartenant à ce domaine d'attraction sont caractérisées par une queue à décroissance lente (polynomiale) à l'infini, et un point terminal $w(F) = +\infty$. Elles sont dites aussi distributions à queues lourdes. Une caractérisation de ce domaine d'attraction noté $MDA(\phi_\alpha)$ est donnée par le théorème suivant :

Théorème 2.2. *Une fonction de répartition F appartient au $MDA(\phi_\alpha)$ si et seulement si*

$$\bar{F}(x) = x^{-\alpha} \ell(x) \quad (2.15)$$

où ℓ est une fonction à variation lente. Dans ce cas les deux suites normalisantes sont données par :

$$a_n = F^{\leftarrow} \left(1 - \frac{1}{n} \right) \quad \text{et} \quad b_n = 0 \quad (2.16)$$

La deuxième caractérisation qui est basée sur la fonction du hasard est donnée par le théorème suivant :

Théorème 2.3 (Von Mises). *si $w(F) = +\infty$ et si*

$$\lim_{x \rightarrow +\infty} x r(x) = \alpha > 0 \quad (2.17)$$

alors $F \in MDA(\phi_\alpha)$

Le tableau (2.1) montre quelques exemples de distributions appartenant au domaine d'attraction de Fréchet.

Distribution	$1 - F(x)$	ξ	$\ell(x)$
<i>Pareto</i> (α) $\alpha > 0$	$x^{-\alpha}$	$\frac{1}{\alpha}$	1
<i>Burr</i> (η, τ, λ) $\eta, \tau, \lambda > 0$	$\left(\frac{\eta}{\eta+x^\tau}\right)^\lambda$ $x > 0$	$\frac{1}{\lambda\tau}$	$\left(\frac{\eta}{1+\frac{\eta}{x^\tau}}\right)^\lambda$
<i>Pareto-Généralisé</i> (σ, ξ) $\sigma, \xi > 0$	$\left(1 + \frac{\xi x}{\sigma}\right)^{-\frac{1}{\xi}}$ $x > 0$	ξ	$\left(\frac{\sigma}{\xi}\right)^{\frac{1}{\xi}} \left(1 + \frac{\sigma}{\xi x}\right)^{-\frac{1}{\xi}}$
<i>Fréchet</i> (α) $\alpha > 0$	$1 - \exp(-x^{-\alpha})$ $x > 0$	$\frac{1}{\alpha}$	$1 - \frac{x^{-\alpha}}{2} + o(x^{-\alpha})$

TAB. 2.1: Exemples de distributions de type Fréchet

2.2.3.2 Domaine d'attraction de Weibull $\xi < 0$

Les lois de ce domaine d'attraction sont bornées à droite, ainsi le point terminal $w(F)$ est fini. Elles peuvent modéliser des phénomènes dont la variable étudiée reste inférieure à une valeur donnée, quelle que soit l'évolution du phénomène. Une caractérisation d'appartenance à ce domaine est donnée par le théorème suivant :

Théorème 2.4. *Une distribution F appartient au $MDA(\psi_\alpha)$, si et seulement si $w(F) < +\infty$ et*

$$\bar{F}\left(w(F) - \frac{1}{x}\right) = x^{-\alpha}\ell(x) \quad (2.18)$$

les suites normalisantes sont données par :

$$a_n = w(F) - F^{\leftarrow}\left(1 - \frac{1}{n}\right) \quad \text{et} \quad b_n = w(F) \quad (2.19)$$

La caractérisation via la fonction du hasard est donnée par :

Théorème 2.5 (Von Mises). *Si $w(F) < +\infty$ et*

$$\lim_{x \rightarrow w(F)} (w(F) - x) r(x) = \alpha > 0 \quad (2.20)$$

alors $F \in MDA(\psi_\alpha)$

Le tableau (2.2) montre quelques exemples de distributions appartenant au domaine d'attraction de Weibull.

Distribution	$1 - F\left(w(F) - \frac{1}{x}\right)$	ξ	$\ell(x)$
<i>Uniforme</i>	$\frac{1}{x}$, $x > 1$	-1	1
<i>Weibull</i>	$1 - \exp(-x^{-\alpha})$	$-\frac{1}{\alpha}$	$1 - \frac{x^{-\alpha}}{2} + o(x^{-\alpha})$
<i>Reverse Burr</i> (λ, β, τ)	$\left(\frac{\beta}{\beta+x^\tau}\right)^\lambda$ $x > 0$	$-\frac{1}{\lambda\tau}$	$\beta^\lambda (1 - \lambda\beta x^{-\tau} + o(x^{-\tau}))$

TAB. 2.2: Exemples de distributions de type Weibull

2.2.3.3 Domaine d'attraction de Gumbel $\xi = 0$

Les lois appartenant à ce domaine d'attraction ont une queue à décroissance exponentielle à l'infini. C'est le cas par exemple des lois normale, log-normale et exponentielle. Le cas de ce domaine est plus délicat à traiter car il n'y a pas de lien entre la queue de distribution et les fonctions à variations lentes comme c'est le cas pour les deux autres domaines d'attraction. Dans la littérature il n'y a pas de conditions nécessaires et suffisantes faciles à mettre en oeuvre. Cependant Von Mises (1936) a donné une caractérisation plus simple, illustrée dans le théorème suivant.

Théorème 2.6. *S'il existe une fonction mesurable R de la variable t appelée fonction auxiliaire, et telle que :*

$$\lim_{x \rightarrow w(F)} \frac{1 - F(t + xR(t))}{1 - F(t)} = \exp(-x) \quad (2.21)$$

alors $F \in MDA(\Lambda)$, et de plus un choix de $R(t)$, a_n et b_n , peut être donné comme suit :

$$R(t) = \frac{1}{1 - F(t)} \int_t^{w(F)} (1 - F(y)) dy \quad (2.22)$$

$$b_n = F \leftarrow \left(1 - \frac{1}{n}\right) \quad \text{et} \quad a_n = R(b_n) \quad (2.23)$$

La caractérisation par la fonction du hasard est par ailleurs donnée par le théorème suivant :

Théorème 2.7 (Von Mises). *si $r(x)$ est strictement positive au voisinage de $w(F)$, différentiable et vérifiant :*

$$\lim_{x \rightarrow w(F)} \frac{dr(x)}{dx} = 0 \quad (2.24)$$

alors, $F \in MDA(\Lambda)$

Le tableau (2.3) donne quelques exemples de distributions dans le domaine d'attraction de Gumbel.

Distribution	$1 - F(x)$
<i>BenktanderII</i> , $\alpha, \beta > 0$	$x^{-(1-\beta)} \exp\left(-\frac{\alpha}{\beta} x^\beta\right)$
<i>xi</i> (m, λ), $m, \lambda > 0$	$\frac{\lambda^m}{\xi(m)} \int_x^{+\infty} u^{m-1} \exp(-\lambda u) du, x > 0$
<i>Logistic</i>	$\frac{1}{1 + \exp(x)}$
<i>Weibull</i> (λ, τ), $\lambda, \tau > 0$	$\exp(-\lambda x^\tau)$

TAB. 2.3: Exemples de distributions de type Gumbel

Remarques :

- Si une fonction de répartition F appartient à deux domaines d'attractions différents U et V , alors on peut montrer qu'il existe des constantes $a > 0$ et b tels que ([Res87]) :

$$V(x) = U(ax + b) \quad (2.25)$$

On dit alors que U et V sont de même type.

- On peut vérifier que [EKM97] :

$$(X \in MDA(G_\xi), \xi > 0) \Leftrightarrow \left(\log\left(X^{1/\xi}\right) \in MDA(G_0)\right) \Leftrightarrow (-X^{-1} \in MDA(G_\xi), \xi < 0) \quad (2.26)$$

- Une distribution peut ne pas appartenir à un de ces trois domaines d'attraction ; c'est le cas de la distribution définie par : $F(x) = 1 - \frac{1}{\log(x)}$, $x \geq e$.

- L'absence des deux suites normalisantes n'est pas contraignante, d'autant plus que la vraie loi des observations est souvent inconnue. En pratique si l'on dispose de k réalisations du maximum, il suffit d'écrire

$$\mathbb{P}(M_k \leq x) \approx G_\xi \left(\frac{x - b_k}{a_k} \right) \equiv G_{\xi, \mu, \sigma}(x) = G_\xi \left(\frac{x - \mu}{\sigma} \right) \quad (2.27)$$

Ce qui ramène le problème de la détermination du domaine d'attraction à l'estimation des 3 paramètres suivants : ξ l'indice extrême, dit aussi paramètre de forme, μ le paramètre de localisation et σ le paramètre d'échelle. Pour une série de valeurs (y_1, \dots, y_n) observées sur un échantillon aléatoire (Y_1, \dots, Y_n) de la variable Y , on introduit les notations suivantes :

- Moyenne empirique (de l'échantillon) : $\bar{y}_m = \sum_{i=1}^m y_i$
- Variance empirique : $s_Y^2 = \frac{1}{m} \sum_{i=1}^m (y_i - \bar{y})^2$
- Variance empirique corrigée : $s_y^{*2} = \frac{1}{m-1} \sum_{i=1}^m (y_i - \bar{y})^2$
- Quantile de la loi de Student d'ordre α à m degrés de liberté : $t_{\alpha, m}$
- Quantile de la loi normale d'ordre α : z_α
- Vecteur des paramètres de la loi GEV : $\Theta = (\sigma, \xi, \mu)$
- Loi Normale de moyenne 0 et de matrice de variance-covariance V : $\mathcal{N}(0, V)$
- Vecteur des paramètres de la loi de Gumbel : $\Theta_0 = (\mu_0, \sigma_0)$
- Moment centré d'ordre k , d'une v.a de densité f définie, sur un domaine D :

$$m_k = \int_D (y - \mathbb{E}[Y])^k f(y) dy \quad (2.28)$$

En particulier, on a : $m_1 = \mathbb{E}[Y]$ et $m_2 = \mathbb{V}[Y]$ où $\mathbb{V}[\cdot]$ représente l'opérateur de la variance.

2.2.4 Méthodes d'estimation des paramètres et des quantiles extrêmes

2.2.4.1 Estimation dans le domaine de Gumbel

Dans ce domaine, l'indice extrême est nul et la loi du maximum est donnée par la fonction de répartition de Gumbel G_0 :

$$G_0 \left(\frac{x - \mu_0}{\sigma_0} \right) = \exp \left(- \exp \left(- \frac{x - \mu_0}{\sigma_0} \right) \right) \equiv G_0(y) \quad (2.29)$$

avec $y = \frac{x - \mu_0}{\sigma_0}$. μ_0 et σ_0 sont respectivement, le paramètre de localisation et le paramètre d'échelle, dans le cas où $\xi = 0$.

Si X désigne la variable aléatoire étudiée (la profondeur des piqûres dans le cadre de cette thèse), on dispose en pratique de n réalisations X_1, \dots, X_n de cette variable. Mais le maximum de cet échantillon ne constitue qu'une seule observation. L'idée alors proposée par Gumbel (1958), est de subdiviser l'échantillon de taille n en m blocs de taille n/m . On retient les valeurs maximales respectives de chaque bloc que l'on note Y_1, \dots, Y_m et qui vont être utilisées par la suite pour estimer les paramètres de la loi de Gumbel. Cette démarche sera présentée plus en détail au quatrième chapitre. Dans ce travail, on comparera différentes méthodes d'estimations ainsi que leurs impacts sur les quantiles extrêmes. On donnera ici les expressions générales des méthodes classiques d'estimations ainsi que d'autres méthodes rarement utilisées en corrosion par piqûres. Les intervalles de confiance correspondants seront aussi précisés.

Méthode de la régression linéaire C'est la méthode la plus simple à mettre en oeuvre ; méthode très souvent utilisée en corrosion par piqûres. Etant donnés Y_1, \dots, Y_m les maxima par blocs, et $Y_{1,m} \leq \dots \leq Y_{m,m}$ la statistique d'ordre associée. La méthode de régression linéaire consiste à ajuster le nuage de points

$$(-\log(-\log(G_0(Y_{i,m}))), Y_{i,m}) \quad (2.30)$$

à une droite d'équation $Y_{i,m} = ax + b$ en calculant les coefficients a et b par la méthode des moindres carrés. On a par ailleurs $x = -\log(-\log(G_0(Y_{i,m})))$. En pratique, la fonction de répartition $G_0(Y_{i,m})$ est souvent estimée par :

$$\hat{F}(Y_{i,m}) = \frac{i}{m+1} \quad (2.31)$$

où \hat{F} est la fonction de répartition empirique.

Pour estimer les quantiles extrêmes dépassés avec une faible probabilité p , on introduit la notion de la période de retour T définie par :

$$T = \frac{1}{p} = \frac{1}{1 - G_0(y)} \quad (2.32)$$

L'interprétation de cette notion dans le contexte de la corrosion par piqûres sera présenté au quatrième chapitre.

Le coefficient directeur et l'ordonnée à l'origine de la droite ajustée, nous permettent de calculer les estimations des paramètres, et on obtient : $\hat{\mu}_0^{reg} = b$ et $\hat{\sigma}_0^{reg} = a$. L'expression générale du quantile extrême, pour une période de retour T pour la loi de Gumbel, est donnée par :

$$q_0 = \mu_0 + \sigma_0 \left(-\log \left(-\log \left(1 - \frac{1}{T} \right) \right) \right) \quad (2.33)$$

L'estimation du quantile extrême est ainsi obtenue en remplaçant les paramètres μ et σ par leurs estimations calculées par régression linéaire :

$$q_0^{reg} = \hat{\mu}_0^{reg} + \hat{\sigma}_0^{reg} \left(-\log \left(-\log \left(1 - \frac{1}{T} \right) \right) \right) \quad (2.34)$$

Sous l'hypothèse de normalité des résidus e_i , l'intervalle de confiance de la droite d'ajustement par la méthode de la régression linéaire, est donné par ([Sap06]) :

$$\left[\hat{y}_0 \pm t_{\alpha/2, m-2} R \left(1 + \frac{1}{m} + \frac{(x_0 - \bar{x})^2}{(m-1) s_x^2} \right)^{1/2} \right] \quad (2.35)$$

où

$$\hat{y}_0 = ax_0 + b, \quad x_0 = -\log \left(-\log \left(1 - \frac{1}{T} \right) \right), \quad R = \frac{1}{m-2} \sum_{i=1}^m e_i^2, \quad e_i = y_i - \hat{y}_i,$$

$$y_i = Y_{i,m}, \quad \hat{y}_i = ax_i + b, \quad x_i = -\log(-\log(F(Y_{i,m}))), \quad x = (x_i)_{i=1, \dots, m}$$

Méthode du maximum de vraisemblance Il s'agit de chercher les paramètres qui maximisent la fonction de vraisemblance d'une fonction densité f , définie par :

$$L(\mu_0, \sigma_0) = \prod_{i=1}^m f(y_i; \mu_0, \sigma_0) \quad (2.36)$$

En pratique, pour des raisons de simplification, on maximise le logarithme de la fonction de vraisemblance. Cela ne change en rien la solution puisque $[\log(f(x))]' = 0 \Rightarrow [f(x)]' = 0$. On considère ainsi la fonction définie comme suit :

$$\log L(\mu_0, \sigma_0) = -m \log(\sigma_0) - \sum_{i=1}^m \exp\left(-\frac{Y_i - \mu_0}{\sigma_0}\right) - \sum_{i=1}^m \frac{Y_i - \mu_0}{\sigma_0} \quad (2.37)$$

Cela conduit à résoudre le système suivant :

$$\begin{cases} \frac{\partial \log L(\Theta_0)}{\partial \mu_0} = 0 \\ \frac{\partial \log L(\Theta_0)}{\partial \sigma_0} = 0 \end{cases} \quad (2.38)$$

En résolvant ce système, on trouve les valeurs estimées respectives de μ_0 et σ_0 . Pour certaines lois, on peut trouver une solution analytique explicite. Mais en général, il faut passer par une résolution numérique, via par exemple, la méthode de Newton-Raphson. Les propriétés asymptotiques de l'estimateur du maximum de vraisemblance, sont données par le théorème général suivant :

Théorème 2.8. : Soient y_1, \dots, y_m des réalisations indépendantes d'une loi de probabilité $f(., \theta)$ avec $\theta = (\theta_i)_{1 \leq i \leq d}$ désigne le vecteur des d paramètres de f à estimer et $\hat{\Theta}$ son estimateur du maximum de vraisemblance. Alors on a :

$$\sqrt{m}(\hat{\Theta} - \theta) \xrightarrow{loi} \mathcal{N}(0, V_0^{-1}) \text{ pour } m \rightarrow +\infty \quad (2.39)$$

où V_0^{-1} représente la matrice de variance-covariance des paramètres qui est égale à l'inverse de la matrice d'information de Fisher définie par :

$$(V_0)_{1 \leq i, j \leq d} = -\mathbb{E} \left[\frac{\partial^2 \log L(\theta)}{\partial \theta_i \partial \theta_j} \right] \quad (2.40)$$

De plus, la construction de l'intervalle de confiance pour un quantile $q = g(\theta)$ qui s'exprime en fonction des paramètres à estimer $\theta_1, \dots, \theta_d$ est donnée par la méthode delta formulée comme suit :

$$\sqrt{m} \left(g(\hat{\theta}) - g(\theta) \right) = \sqrt{m} (\hat{q} - q) \xrightarrow{loi} \mathcal{N}(0, (\nabla g)^t V_0^{-1} (\nabla g)) \text{ pour } m \rightarrow +\infty \quad (2.41)$$

où ∇g représente le gradient de g et $(.)^t$ l'opérateur de transposition.

Dans le cas particulier de Gumbel, on a $\theta = \Theta_0 = (\mu_0, \sigma_0)$ et l'intervalle de confiance du quantile extrême q_0^{mv} peut être exprimé comme suit :

$$\left[\hat{q}_0^{mv} \pm z_{\alpha/2} \sqrt{\frac{(\nabla g)^t V_0^{-1} (\nabla g)}{m}} \right] \quad (2.42)$$

où \hat{q}_0^{mv} est donnée par l'équation (2.33) en remplaçant les estimations des paramètres μ_0 et σ_0 par ceux du maximum de vraisemblance et l'on a :

$$(\nabla g)^t = \left[\frac{\partial q_0}{\partial \mu_0}, \frac{\partial q_0}{\partial \sigma_0} \right] = \left[1, -\log \left(-\log \left(1 - \frac{1}{T} \right) \right) \right] \quad (2.43)$$

D'après la formule (2.40), pour obtenir la matrice de Fisher, il faudrait calculer l'espérance des dérivées d'ordre 2 de la log-vraisemblance, ce qui nécessiterait en pratique, d'avoir plusieurs réalisations de cette fonction aléatoire, ce qui est rarement, voire jamais le cas. En pratique, la matrice d'information de Fisher V est donc remplacée par la matrice d'information observée \hat{V} qui en représente une réalisation particulière. Dans ce cas la matrice V coïncide avec la matrice Hessienne et, pour la distribution de Gumbel, on peut écrire :

$$\hat{V}_0 = \begin{pmatrix} \frac{\partial^2 \log L(\mu_0, \sigma_0)}{\partial \mu_0^2} & \frac{\partial^2 \log L(\mu_0, \sigma_0)}{\partial \mu_0 \partial \sigma_0} \\ \frac{\partial^2 \log L(\mu_0, \sigma_0)}{\partial \mu_0 \partial \sigma_0} & \frac{\partial^2 \log L(\mu_0, \sigma_0)}{\partial \sigma_0^2} \end{pmatrix} \Big|_{\mu_0 = \hat{\mu}_0^{mv}, \sigma_0 = \hat{\sigma}_0^{mv}} \quad (2.44)$$

Méthode des moments : La méthode consiste à évaluer les deux premiers moments théoriques m_1 et m_2 avec les deux premiers moments observés (moyenne et variance empiriques). On aboutit ainsi à deux équations reliant les paramètres de la loi de Gumbel. On a, en effet :

$$\begin{cases} \mathbb{E}[Y] &= \mu_0 + \gamma_* \sigma_0 &= \bar{y}_m \\ \mathbb{V}[Y] &= \frac{1}{6} \pi^2 \sigma_0^2 &= s_Y^2 \end{cases} \quad (2.45)$$

où $\gamma_* \approx 0,57721$ représente la constante d'Euler. La résolution du système 2.45 donne les estimateurs suivants :

$$\hat{\sigma}_0^{mom} = \frac{\sqrt{6} s_Y}{\pi} \quad \text{et} \quad \hat{\mu}_0^{mom} = \bar{y}_m - \gamma_* \hat{\sigma}_0^{mom} \quad (2.46)$$

Pour estimer les quantiles extrêmes, on remplace dans la formule 2.33, les estimations de μ_0 et σ_0 , par celles estimées par la méthode des moments (Eq. 2.46). Cette méthode a été critiquée dans le sens où elle peut renvoyer des estimations biaisées si la distribution empirique des observations s'écarte sensiblement de la droite de Gumbel. Ceci est dû à l'utilisation du moment d'ordre 2 qui donne plus de poids aux valeurs faibles de l'échantillon. Ce problème peut être corrigé en utilisant par exemple la méthode des moments de probabilité pondérés traitée dans la section 2.2.4.2.

On note que la distribution asymptotique conjointe de $(\hat{\mu}_0^{mom}, \hat{\sigma}_0^{mom})$ suit une loi Normale de moyenne (μ_0, σ_0) ([KN00]) avec des variances et coefficient de corrélation $\rho_{\hat{\mu}_0^{mom}, \hat{\sigma}_0^{mom}}$ donnés par :

$$\mathbb{V}(\hat{\mu}_0^{mom}) \approx \frac{1.1678 \sigma_0^2}{m} \quad (2.47)$$

$$\mathbb{V}(\hat{\sigma}_0^{mom}) \approx \frac{1.1 \sigma_0^2}{m} \quad (2.48)$$

$$\rho_{\hat{\mu}_0^{mom}, \hat{\sigma}_0^{mom}} = \frac{\pi^2 [\sqrt{\beta_1} - 3\gamma_* (\beta_2 - 1) / 2\pi] / 6}{[\{\pi^2 / 6 + \gamma_*^2 (\beta_2 - 1) / 4 - \pi (\gamma_* \sqrt{\beta_1})\} (\beta_2 - 1)]^{1/2}} \quad (2.49)$$

où β_1 et β_2 représentent respectivement le skewness et le kurtosis de l'échantillon.

Pour obtenir l'expression de l'intervalle de confiance des quantiles pour la méthode des moments, il faut calculer la matrice de variance-covariance V_{mom} des paramètres. Cette dernière joue le même rôle que la matrice V_0 calculée pour la méthode du maximum de vraisemblance. En d'autres termes, l'intervalle de confiance est déterminé par la méthode delta. Il est donné par l'équation (2.42) avec g donnée par l'équation (2.43) et V_0^{-1} remplacée par V_{mom} . A noter qu'à partir des variances et du coefficient de corrélation donnés par les équations (2.47), (2.48) et (2.49), la matrice V_{mom} est entièrement connue. En effet le coefficient de corrélation $\rho(X, Y)$ de deux variables X et Y est liée à la covariance par la relation :

$$\rho(X, Y) = \frac{cov(X, Y)}{S_X S_Y} \quad (2.50)$$

où S_X et S_Y représentent respectivement les écarts-types de X et Y .

2.2.4.2 Estimation dans les domaines de Fréchet et de Weibull $\xi \neq 0$

Pour $\xi \neq 0$, la loi du maximum est donnée par la fonction de répartition *GEV* :

$$G_\xi\left(\frac{x-\mu}{\sigma}\right) = \exp\left(-\left(1 + \xi\frac{x-\mu}{\sigma}\right)^{-\frac{1}{\xi}}\right) \equiv G_\xi(y), \quad 1 + \xi\frac{x-\mu}{\sigma} > 0 \quad (2.51)$$

Nous présentons pour ces deux domaines, les méthodes d'estimation du maximum de vraisemblance et des moments de probabilité pondérés.

Notons que le quantile de la loi *GEV*, pour une période de retour T , est donné par la formule suivante :

$$q = \mu + \frac{\sigma}{\xi} \left[-\left(-\log\left(1 - \frac{1}{T}\right)\right)^{-\xi} \right] \quad \xi \neq 0 \quad (2.52)$$

Méthode du maximum de vraisemblance : Comme dans le cas où $\xi = 0$, la méthode consiste à chercher les paramètres qui maximisent la fonction de log-vraisemblance donnée par :

$$\begin{aligned} \log L(\sigma, \xi, \mu) = & -m \log(\sigma) - \left(1 + \frac{1}{\xi}\right) \sum_{i=1}^m \log\left(1 + \xi\frac{Y_i - \mu}{\sigma}\right) \\ & - \sum_{i=1}^m \left(1 + \xi\frac{Y_i - \mu}{\sigma}\right)^{-\frac{1}{\xi}}, \quad 1 + \xi\frac{Y_i - \mu}{\sigma} > 0, \quad i = 1, \dots, m \end{aligned} \quad (2.53)$$

On est aussi conduit à résoudre le système suivant :

$$\begin{cases} \frac{\partial \log L(\Theta)}{\partial \sigma} = 0 \\ \frac{\partial \log L(\Theta)}{\partial \xi} = 0 \\ \frac{\partial \log L(\Theta)}{\partial \mu} = 0 \end{cases} \quad (2.54)$$

On obtient ainsi les estimations $\hat{\sigma}^{mv}$, $\hat{\xi}^{mv}$ et $\hat{\mu}^{mv}$ des paramètres σ , ξ et μ . Les conditions de régularité et de convergence de cet estimateur ont été étudié en détail dans ([Smi85]). Dans le cas

où $\xi > -0.5$, les propriétés usuelles de consistance et de normalité sont vérifiées. En effet pour $m \rightarrow +\infty$,

$$\sqrt{m} \left((\hat{\sigma}^{mv}, \hat{\xi}^{mv}, \hat{\mu}^{mv}) - (\sigma, \xi, \mu) \right) \xrightarrow{loi} \mathcal{N}(0, V) \quad , \quad \xi > -0.5 \quad (2.55)$$

où V représente la matrice d'information de Fisher, dont les composantes sont

([BGST04]) :

$$\left\{ \begin{array}{l} V_{1,1} = \frac{1}{\sigma^2 \xi^2} (1 - \Gamma(2 + \xi) + d) \\ V_{1,2} = V_{2,1} = -\frac{1}{\sigma \xi^2} \left(1 - \gamma_* - c + \frac{1 - \Gamma(2 + \xi)}{\xi} + \frac{d}{\xi} \right) \\ V_{1,3} = V_{3,1} = -\frac{1}{\sigma^2 \xi} (d - \Gamma(2 + \xi)) \\ V_{2,2} = \frac{1}{\xi^2} \left[\frac{\pi^2}{6} + \left(1 - \gamma_* + \frac{1}{\xi} \right)^2 - \frac{2c}{\xi} + \frac{d}{\xi^2} \right] \\ V_{2,3} = V_{3,2} = -\frac{1}{\sigma \xi} \left(c - \frac{d}{\xi} \right) \\ V_{3,3} = \frac{d}{\sigma^2} \end{array} \right. \quad (2.56)$$

où $d = (1 + \xi)^2 \Gamma(1 + 2\xi)$,

$c = \Gamma(2 + \xi) \left(\psi(1 + \xi) + \frac{1 + \xi}{\xi} \right)$, $\psi(x) = \frac{d \log \Gamma(x)}{dx}$ et $\Gamma(\cdot)$ la fonction Gamma d'Euler.

Comme dans le domaine de Gumbel, l'intervalle de confiance du quantile extrême se construit à l'aide de la méthode delta, et on a :

$$\left[\hat{q}^{mv} \pm z_{\alpha/2} \sqrt{\frac{(\nabla g)^t V^{-1} (\nabla g)}{m}} \right] \quad (2.57)$$

où \hat{q}^{mv} est donné par l'équation (2.52) en remplaçant les paramètres par leur estimations obtenues par le maximum de vraisemblance. On a également $(\nabla g)^t = \left[\frac{\partial q}{\partial \sigma}, \frac{\partial q}{\partial \xi}, \frac{\partial q}{\partial \mu} \right]$ avec :

$$\left\{ \begin{array}{l} \frac{\partial q}{\partial \sigma} = \frac{1}{\xi} \left(-\log\left(1 - \frac{1}{T}\right) \right)^{-\xi} - 1 \\ \frac{\partial q}{\partial \xi} = -\frac{\sigma}{\xi^2} \left[\left(-\log\left(1 - \frac{1}{T}\right) \right)^{-\xi} - 1 \right] - \frac{\sigma}{\xi} \left(-\log\left(1 - \frac{1}{T}\right) \right)^{-\xi} \log\left(-\log\left(1 - \frac{1}{T}\right) \right) \\ \frac{\partial q}{\partial \mu} = 1 \end{array} \right. \quad (2.58)$$

Moments de probabilité pondérés PWM Introduite par Greenwood et al. ([GLMW79]), les moments de probabilités pondérés sont définis par :

$$M_{p,r,s} = \mathbb{E} [Y^p (F(Y))^r (1 - F(Y))^s] \quad (2.59)$$

où p, r, s sont des réels. Pour $\xi \neq 0$, un choix possible pour calculer les moments $M_{p,r,s}$ est de prendre $p = 1, s = 0$ et $r = 0, 1, 2$ [BGST04]. On obtient ainsi pour la loi GEV :

$$M_{1,r,0} = \frac{1}{r+1} \left\{ \mu - \frac{\sigma}{\xi} \left[1 - (r+1)^\xi \Gamma(1-\xi) \right] \right\} \quad \xi < 1 \quad (2.60)$$

Les estimations PWM $\hat{\sigma}^{pwm}, \hat{\xi}^{pwm}, \hat{\mu}^{pwm}$ des paramètres de la loi *GEV* constituent la solution du système d'équations, obtenu pour $r = 0, 1, 2$ ([HWW85]) :

$$M_{1,0,0} = \mu - \frac{\sigma}{\xi} (1 - \Gamma(1-\xi)) \quad (2.61)$$

$$2M_{1,1,0} - M_{1,0,0} = \frac{\sigma}{\xi} \Gamma(1-\xi) (2^\xi - 1) \quad (2.62)$$

$$\frac{3M_{1,2,0} - M_{1,0,0}}{2M_{1,1,0} - M_{1,0,0}} = \frac{3^\xi - 1}{2^\xi - 1} \quad (2.63)$$

Il existe plusieurs estimateurs de $M_{1,r,0}$ pour r entier. Hosking et al. [HWW85] préconisent dans la pratique, l'utilisation d'un estimateur asymptotiquement consistant donné par :

$$\hat{M}_{1,r,0} = \frac{1}{m} \sum_{i=1}^m \left(\hat{F}(Y_{i,m}) \right)^r Y_{j,m} \quad (2.64)$$

où un choix de la fonction de répartition empirique $\hat{F}(Y_{i,m})$ peut être fait parmi les deux expressions suivantes :

$$\hat{F}(y_{i,m}) = \frac{i-a}{m} \quad 0 < a < 1 \quad (2.65)$$

$$\hat{F}(y_{i,m}) = \frac{i-a}{m+1-2a} \quad -\frac{1}{2} < a < \frac{1}{2} \quad (2.66)$$

Hosking et al. ([HWW85]) recommandent l'utilisation de la formule (2.65) avec $a = 0.35$. Après avoir évalué les moments $M_{1,r,0}$ pour $r = 0, 1, 2$, l'équation (2.63) peut être résolu numériquement permettant ainsi l'obtention de $\hat{\xi}^{pwm}$. Ce résultat est ensuite utilisé dans l'équation (2.62) pour calculer $\hat{\sigma}^{pwm}$. Enfin $\hat{\mu}^{pwm}$ est obtenu à partir de l'équation (2.61). Le résultat asymptotique est alors obtenu :

$$\sqrt{m} \left((\hat{\sigma}^{pwm}, \hat{\xi}^{pwm}, \hat{\mu}^{pwm}) - (\sigma, \xi, \mu) \right) \xrightarrow{loi} \mathcal{N}(0, W) \quad \xi < 0.5 \quad (2.67)$$

où les composantes de la matrice symétrique W sont données comme suit :

$$W_{k,k} = \left[\frac{\sigma}{\xi} (k+1)^\xi \right]^2 \left(\Gamma(1-2\xi) K(k/(k+1)) - \Gamma^2(1-\xi) \right) \quad (2.68)$$

$$W_{k,k+1} = \frac{1}{2} \left(\frac{\sigma}{\xi} \right)^2 \left((k+2)^{2\xi} \Gamma(1-2\xi) K(k/(k+2)) + (k+1)^\xi \left[(k+1)^\xi - 2(k+2)^\xi \right] \Gamma^2(1-\xi) \right) \quad (2.69)$$

$$\begin{aligned}
W_{k,k+k'} &= \frac{1}{2} \left(\frac{\sigma}{\xi} \right)^2 \left((k+k'+1)^{2\xi} \Gamma(1-2\xi) K\left(k/(k+k'+1)\right) \right. \\
&\quad - (k+k')^\xi \Gamma(1-2\xi) K\left((k+1)/(k+k')\right) \\
&\quad \left. + 2(k+1)^\xi \left[(k+k')^\xi - (k+k'+1)^\xi \right] \Gamma^2(1-\xi) \right) \quad k' \geq 2
\end{aligned} \tag{2.70}$$

où $K(x) = H_y(-\xi, -2\xi, 1-\xi, -x)$ avec H_y représentant la fonction hypergéométrique.

Ainsi, de l'équation (2.68), on peut déterminer $W_{1,1}$, $W_{2,2}$ et $W_{3,3}$ pour $k = 1, 2, 3$. L'équation (2.69) permet l'obtention de $W_{1,2}$ et $W_{2,3}$ pour $k = 1, 2$. Enfin l'équation 2.70 fournit le terme $W_{1,3}$ pour $k = 1$ et $k' = 2$.

L'intervalle de confiance du quantile extrême est aussi construit à l'aide de la méthode delta par :

$$\left[\hat{q}^{pwm} \pm z_{\alpha/2} \sqrt{\frac{(\nabla g)^t W(\nabla g)}{m}} \right] \tag{2.71}$$

où \hat{q}^{pwm} est donné par l'équation (2.52), en remplaçant les paramètres σ , ξ , et μ par leurs estimations *PWM* respectives. ∇g est donnée par l'équation (2.58).

2.3 L'approche des dépassements de seuil

La caractérisation des événements extrêmes peut aussi se faire à l'aide de l'approche des dépassements de seuil. Soient X_1, \dots, X_n des v.a de même loi F . Cette approche s'intéresse à l'analyse du comportement asymptotique des excès qui sont définis par $Z_i = X_i - u | X_i > u$ où u représente un seuil bien choisi. La fonction de répartition des excès au-delà du seuil u est définie par :

$$F_u(z) = \mathbb{P}(X - u \leq z | X > u) = \begin{cases} \frac{F(u+z) - F(u)}{1 - F(u)} & \text{si } z \geq 0 \\ 0 & \text{si } z < 0 \end{cases} \tag{2.72}$$

La loi asymptotique des excès est de plus donnée par le théorème suivant :

Théorème 2.9 (Pickands (1975)). *Si F appartient à l'un des trois domaines d'attraction ($MDA(\text{Fréchet})$, $MDA(\text{Gumbel})$, et $MDA(\text{Weibull})$), alors il existe une fonction $\sigma(u)$ positive, telle que :*

$$\lim_{u \rightarrow w(F)} \sup_{0 < z < w(F) - u} |F_u(z) - H_{\xi, \sigma(u)}(z)| = 0 \tag{2.73}$$

$H_{\xi, \sigma(u)}(z)$ est la fonction de répartition de la loi de Pareto Généralisée (*GPD*, *Generalized Pareto Distribution*), définie par :

$$H_{\xi, \sigma(u)}(z) = \begin{cases} 1 - \left(1 + \frac{\xi z}{\sigma(u)}\right)^{-1/\xi} & \text{si } z \in [0, +\infty[\quad \text{et } \xi > 0 \\ 1 - \exp\left(-\frac{z}{\sigma(u)}\right) & \text{si } z \in [0, +\infty[\quad \text{et } \xi = 0 \\ 1 - \left(1 + \frac{\xi z}{\sigma(u)}\right)^{-1/\xi} & \text{si } z \in [0, -\frac{\sigma(u)}{\xi}[\quad \text{et } \xi < 0 \end{cases} \tag{2.74}$$

La convergence en loi du maximum vers une loi GEV et la convergence en loi des excès vers la loi GPD sont donc équivalentes. En d'autres termes, pour un seuil suffisamment élevé qui doit être choisi par l'expérimentateur, la loi des excès peut être approchée par une loi GPD dont l'indice extrême ξ est le même que celui de la loi GEV.

2.3.1 Modèle de renouvellement (ou modèle POT : Peaks Over Threshold)

Le modèle de renouvellement pour les v.a X_1, \dots, X_n est une bonne illustration de l'approche des dépassements de seuil. Il s'appuie sur les deux assertions suivantes :

1. Le nombre de dépassements par X_1, \dots, X_n d'un seuil u suit une loi de Poisson de paramètre τ , notée $P(\tau)$
2. La loi des dépassements au delà du seuil u est une loi GPD.

L'assertion (2) est justifiée par le théorème de Pickands. En effet, pour un seuil suffisamment élevé, la loi des dépassements peut être approchée par la loi GPD, Quant à l'assertion (1), elle se justifie via l'approximation de la loi de Poisson par la loi binomiale. En effet, si l'on désigne par $\{X_i > u_n\}$ l'événement « Dépassement d'un seuil u_n » et sa fonction indicatrice $I_{\{X_i > u_n\}}$ définie par :

$$I_{\{X_i > u_n\}} = \begin{cases} 1 & \text{si } X_i > u_n \\ 0 & \text{sinon} \end{cases} \quad (2.75)$$

alors le nombre de dépassements de u_n , égal à $N_{u_n} = \sum_{i=1}^n I_{\{X_i > u_n\}}$, suit une loi binomiale de paramètres $(n, 1 - F(u_n))$ qui peut être approchée par une loi de Poisson lorsque $n \rightarrow +\infty$, $1 - F(u_n) \rightarrow 0$ et $n(1 - F(u_n)) \rightarrow \tau$. Dans ce modèle la période de retour T_u est liée à la fonction de distribution $H_{\xi, \sigma(u)}$ et à la probabilité de dépassement par la relation suivante [Ros85] :

$$T_u = \frac{1}{\tau p} = \frac{1}{\tau (1 - H_{\xi, \sigma(u)}(z))} \quad (2.76)$$

Dans la suite, on pose $\sigma(u) = \sigma_u$ pour désigner le paramètre d'échelle de la loi GPD.

Les principales difficultés de ce modèle sont le choix du seuil et la méthode d'estimation des paramètres σ_u et ξ . Dans la suite, nous présenterons quelques méthodes classiques souvent utilisées à ce propos.

2.4 Méthodes pour le choix du seuil

2.4.1 Méthodes graphiques

Fonction de la moyenne des excès : (Mean Excess Function (MEF)) : Pour une loi GPD avec $\xi < 1$, la fonction moyenne des excès $e(u)$ est définie par :

$$e(u) = \mathbb{E}[X - u \mid X > u] = \frac{\sigma_u + \xi u}{1 - \xi} \quad (2.77)$$

C'est une fonction linéaire en u . Le choix du seuil correspond alors au début de la linéarité observée sur le graphe défini par $(u, e(u))$. En pratique, la fonction $e(\cdot)$ est approchée par son approximation empirique $\hat{e}_n(\cdot)$ définie par :

$$\hat{e}_n(u) = \frac{\sum_{i=1}^n (X_i - u)}{\sum_{i=1}^n I_{\{X_i > u\}}} \quad (2.78)$$

Cette méthode permet aussi de distinguer les différents domaines d'attraction. En effet, si la linéarité est caractérisée par une pente positive, alors les données appartiennent au MDA(Fréchet), une pente nulle correspond au MDA(Gumbel), et une pente négative indique l'appartenance au MDA(Weibull). La figure (2.2) montre un exemple de trois échantillons issus des trois domaines d'attraction.

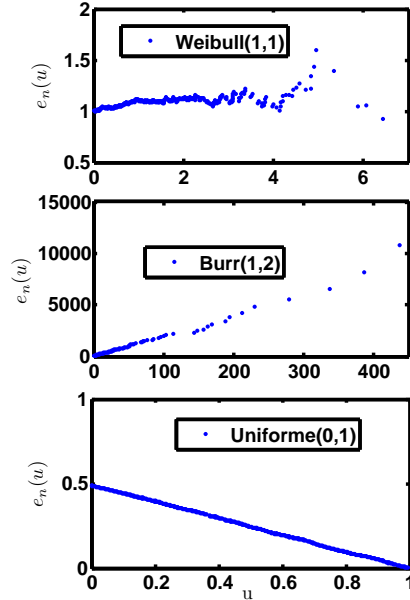


FIG. 2.2: (u, \hat{e}_n) pour 1000 valeurs issues des lois : Weibull(1,1) \in MDA(Gumbel), Burr(1,2) \in MDA(Fréchet) et Uniforme(0,1) \in MDA(Weibull)

2.4.2 Méthodes numériques

Parmi les méthodes numériques couramment utilisées pour le choix du seuil, il y a la méthode du choix automatique proposée par Reiss et al.([RT97]), et la méthode de Pickand ([Pic75]). On note qu'il existe d'autres méthodes numériques pour lesquelles le choix optimal du seuil est celui qui minimise la variance et le biais de l'estimateur utilisé pour les paramètres.

Dans la suite, nous désignons par k le nombre de valeurs extrêmes retenues au delà du seuil u . En pratique, on prend comme estimateur du seuil u la $(n - k)$ ème statistique d'ordre $X_{n-k,n}$, permettant ainsi de définir les excès $Z_{j,k}$ par $Z_{j,k} = X_{n-k+j,n} - X_{n-k,n}$ pour $j = 1, \dots, k$

Choix automatique du seuil : Reiss et Thomas ont proposé une méthode heuristique pour choisir le nombre de valeurs extrêmes k^* utilisées pour estimer les paramètres de la loi ([RT97]). Celui-ci est donné par :

$$k^* = \min_k \frac{1}{k} \sum_{i \leq k} i^\beta \left| \hat{\xi}_i - \text{med} \left(\hat{\xi}_1, \dots, \hat{\xi}_k \right) \right| \quad k = 2, \dots, n-1 \quad \text{et} \quad 0 \leq \beta \leq \frac{1}{2} \quad (2.79)$$

où $\hat{\xi}_j$ (pour $j \leq k$) est l'estimation de l'indice extrême en utilisant les j plus grandes valeurs et med désigne la médiane.

Méthode de Pickands : Cette méthode est basée sur la minimisation de la distance de Kolmogorov-Smirnov entre H_{ξ, σ_u} et la fonction de distribution empirique \hat{H}_n des excès qui peut être construite comme pour l'équation (2.31). Le nombre optimal de valeurs extrêmes k^* est donné par ([Pic75]) :

$$k^* = \min_{k=2, \dots, n-1} \left\| H_{\hat{\xi}, \hat{\sigma}_u}(z_i) - \hat{H}_n(z_i) \right\|_{\infty} \quad i = 1, \dots, k \quad (2.80)$$

où z_i représentent les excès au delà du seuil u et $\|f\|_{\infty}$ désigne la valeur supérieure de la fonction f .

2.5 Méthodes d'estimation des paramètres et des quantiles extrêmes

Trois méthodes d'estimation des paramètres et des quantiles extrêmes sont principalement étudiées dans cette section.

Méthode des moments : Comme dans le cas de la distribution de Gumbel, la méthode consiste à évaluer les moments théoriques et les moments empiriques de façon à obtenir :

$$\begin{cases} \mathbb{E}(Z) = \frac{\sigma_u}{1-\xi} = \bar{Z} \\ \mathbb{V}(Z) = \frac{\sigma_u^2}{(1-\xi)^2(1-2\xi)} = S_Z^{*2} \end{cases} \quad (2.81)$$

où \bar{Z} et S_Z^{*2} représentent respectivement la moyenne et la variance empirique des excès. La résolution du système (2.81) conduit aux estimateurs des moments suivants :

$$\hat{\xi}_u^{mom} = \frac{1}{2} \left(1 - \frac{\bar{Z}^2}{S_Z^{*2}} \right) \quad (2.82)$$

$$\hat{\sigma}_u^{mom} = \frac{\bar{Z}}{2} \left(1 + \frac{\bar{Z}^2}{S_Z^{*2}} \right) \quad (2.83)$$

Ces estimateurs vérifient le résultat asymptotique ci-dessous [BGST04] :

$$\sqrt{k} \left(\left(\hat{\xi}_u^{mom}, \hat{\sigma}_u^{mom} \right) - (\xi, \sigma_u) \right) \xrightarrow{loi} \mathcal{N}(0, C) \quad \text{pour } k \rightarrow +\infty \quad (2.84)$$

où la matrice C est égale à :

$$C = c \begin{pmatrix} (1-2\xi)^2(1-\xi+6\xi^2) & -\sigma_u(1-2\xi)(1-4\xi+12\xi^2) \\ -\sigma_u(1-2\xi)(1-4\xi+12\xi^2) & 2\sigma_u^2(1-6\xi+12\xi^2) \end{pmatrix} \quad (2.85)$$

avec $c = \frac{(1-\xi^2)}{(1-2\xi)(1-3\xi)(1-4\xi)}$ pour $\xi < \frac{1}{4}$

Le quantile extrême des excès dépassé par une faible probabilité p s'obtient ainsi en inversant la fonction de répartition H_{ξ, σ_u} , et l'on a :

$$q^{POT} = \begin{cases} \frac{\sigma_u}{\xi} (p^{-\xi} - 1) & \text{si } \xi \neq 0 \\ -\sigma_u \log(p) & \text{si } \xi = 0 \end{cases} \quad (2.86)$$

Le quantile extrême pour les observations q_u^{POT} est donné par l'équation :

$$q_u^{POT} = u + q^{POT} \quad (2.87)$$

En remplaçant le seuil et les paramètres par leurs estimateurs respectifs, on obtient alors l'estimateur des moments suivant :

$$\hat{q}_u^{mom} = \begin{cases} X_{n-k,n} + \frac{\hat{\sigma}_u^{mom}}{\hat{\xi}_u^{mom}} \left(p^{-\hat{\xi}_u^{mom}} - 1 \right) & \text{si } \hat{\xi}_u^{mom} \neq 0 \\ -\hat{\sigma}_u^{mom} \log(p) & \text{si } \hat{\xi}_u^{mom} = 0 \end{cases} \quad (2.88)$$

L'intervalle de confiance pour le quantile extrême se construit à l'aide de la méthode delta de la manière suivante :

Dans le cas où $\xi \neq 0$, on a l'intervalle suivant :

$$\left[\hat{q}_u^{mom} \pm z_{\alpha/2} \sqrt{\frac{(\nabla g)^t C^{-1} (\nabla g)}{k}} \right] \quad (2.89)$$

où

$$(\nabla g)^t = \left[\frac{\partial q^{POT}}{\partial \xi}, \frac{\partial q^{POT}}{\partial \sigma_u} \right] = \left[-\frac{\sigma_u}{\xi^2} (p^\xi - 1) - \frac{\sigma_u}{\xi} p^{-\xi} \log(p), \frac{1}{p} (p^{-\xi} - 1) \right] \quad (2.90)$$

Dans le cas où $\xi = 0$, la loi GPD se réduit à la loi exponentielle, et les quantiles ainsi que les intervalles de confiance, peuvent être facilement obtenus à l'aide de cette loi.

Méthode du maximum de vraisemblance : La fonction de log-vraisemblance des excès est donnée par :

$$\log(\sigma_u, \xi) = -k \log(\sigma_u) - \left(\frac{1}{\xi} + 1 \right) \sum_{i=1}^k \log \left(1 + \frac{\xi Z_i}{\sigma_u} \right) \quad \xi \neq 0 \quad (2.91)$$

$$\log(\sigma_u, 0) = -k \log(\sigma_u) - \frac{1}{\sigma_u} \sum_{i=1}^k Z_i \quad (2.92)$$

La résolution numérique du système composé des équations (2.91) et (2.92) conduit aux estimateurs du maximum de vraisemblance $\hat{\sigma}_u^{mv}$ et $\hat{\xi}_u^{mv}$ vérifiant pour $\xi > -1/2$ le résultat asymptotique suivant :

$$\sqrt{k} \left(\left(\hat{\xi}_u^{mv}, \hat{\sigma}_u^{mv} \right) - (\xi, \sigma_u) \right) \xrightarrow{loi} \mathcal{N}(0, D) \quad \text{pour } k \rightarrow +\infty \quad (2.93)$$

où la matrice D est donnée par [BGST04] :

$$D = (1 + \xi) \begin{pmatrix} 1 + \xi & -\sigma_u \\ -\sigma_u & 2\sigma^2 \end{pmatrix} \quad (2.94)$$

Le quantile extrême s'obtient à partir de la formule (2.88), en remplaçant les paramètres par leurs estimateurs du maximum de vraisemblance. Ceci donne un intervalle de confiance de q_u^{mv} construit à l'aide de la méthode delta et défini par :

$$\left[\hat{q}_u^{mv} \pm z_{\alpha/2} \sqrt{\frac{(\nabla g)^t D^{-1} (\nabla g)}{k}} \right] \quad (2.95)$$

avec $(\nabla g)^t$ donnée par la formule (2.90)

Méthode des moments de probabilité pondérés : On reprend la définition des moments de probabilité pondérés (2.59), avec $p = 1$, $r = 0$ et $s = 0, 1$. Pour la loi GPD, on a :

$$M_{1,0,s} = \frac{\sigma_u}{(s+1)(s+1-\xi)} \quad \xi < 1 \quad (2.96)$$

En remplaçant $M_{1,0,s}$ par leurs estimateurs empiriques :

$$\hat{M}_{1,0,s} = \frac{1}{k} \sum_{i=1}^k \left(1 - \frac{i}{k+1} \right) Z_{i,k} \quad (2.97)$$

la résolution de l'équation (2.96), pour $s = 0$ et $s = 1$ conduit aux estimateurs PWM suivants :

$$\hat{\xi}_u^{pwm} = 2 - \frac{\hat{M}_{1,0,0}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}} \quad (2.98)$$

$$\hat{\sigma}_u^{pwm} = \frac{2\hat{M}_{1,0,0}\hat{M}_{1,0,1}}{\hat{M}_{1,0,0} - 2\hat{M}_{1,0,1}} \quad (2.99)$$

Pour $\xi < 1/2$, la convergence asymptotique de ces estimateurs est exprimée par :

$$\sqrt{k} \left(\left(\hat{\xi}_u^{pwm}, \hat{\sigma}_u^{pwm} \right) - (\xi, \sigma_u) \right) \xrightarrow{loi} \mathcal{N}(0, E) \quad \text{pour } k \rightarrow +\infty \quad (2.100)$$

où la matrice E a pour expression :

$$E = e \begin{pmatrix} (1-\xi)(2-\xi)^2(1-\xi+2\xi^2) & -\sigma_u(2-\xi)(2-6\xi+7\xi^2-2\xi^3) \\ -\sigma_u(2-\xi)(2-6\xi+7\xi^2-2\xi^3) & \sigma^2(7-18\xi+11\xi^2-2\xi^3) \end{pmatrix} \quad (2.101)$$

où $e = \frac{1}{(1-2\xi)(3-2\xi)}$

Comme précédemment, l'intervalle de confiance du quantile extrême pour la méthode PWM est donné par :

$$\left[\hat{q}_u^{pwm} \pm z_{\alpha/2} \sqrt{\frac{(\nabla g)^t E^{-1} (\nabla g)}{k}} \right] \quad (2.102)$$

2.6 Autres méthodes d'estimation du paramètre ξ

2.6.1 Estimateur de Hill

C'est l'un des estimateurs les plus répandus de l'indice extrême. Introduit par Hill ([Hil75]), il est défini de la façon suivante :

$$\hat{\xi}_{k,n}^{(H)} = \frac{1}{k} \sum_{j=1}^k \log X_{n-j+1,n} - \log X_{n-k,n} \quad (2.103)$$

Cet estimateur est valable seulement dans le domaine d'attraction de Fréchet c'est à dire si $\xi > 0$. Il est interprété graphiquement comme étant la pente du graphe :

$$\left(\log \left(\frac{n+1}{j} \right), \log X_{n-j+1,n} \right) \quad (2.104)$$

appelé "Pareto quantile plot". Dans le MDA(Fréchet), ce graphe présente une linéarité au voisinage des points extrêmes avec une certaine pente ξ . Ainsi l'estimateur de Hill est une approximation simple de cette pente. Pour généraliser aux autres domaines d'attractions, différents estimateurs ont été proposés, entre autres l'estimateur des moments de Dekkers et l'estimateur UH de Beirlant.

2.6.2 Estimateur des moments de Dekkers

Introduit par ([DEdH89]), cet estimateur est présenté comme une extension de l'estimateur de Hill pour $\xi \in \mathbb{R}$. Il est défini comme suit :

$$\hat{\xi}_{k,n}^{(M)} = \hat{\xi}_{k,n}^{(H)} + 1 - \frac{1}{2} \left(1 - \frac{\left(\hat{\xi}_{k,n}^{(H)} \right)^2}{S_{k,n}} \right)^{-1} \quad (2.105)$$

où

$$S_{k,n} = \frac{1}{k} \sum_{j=1}^k (\log X_{n-j+1,n} - \log X_{n-k,n})^2 \quad (2.106)$$

Sous certaines conditions de régularité ([DEdH89]), la normalité asymptotique de l'estimateur est établie, et on a pour $k \rightarrow +\infty$, $n \rightarrow +\infty$ et $n/k \rightarrow 0$:

$$\sqrt{k} \left(\hat{\xi}_{k,n}^{(M)} - \xi \right) \xrightarrow{loi} \begin{cases} \mathcal{N}(0, 1 + \xi^2) & \text{si } \xi \geq 0 \\ \mathcal{N}\left(0, (1 - \xi^2) (1 - 2\xi) \left\{ 4 - 8 \frac{1-2\xi}{1-3\xi} + \frac{(5-11\xi)(1-2\xi)}{(1-3\xi)(1-4\xi)} \right\} \right) & \text{si } \xi < 0 \end{cases} \quad (2.107)$$

Contrairement à l'estimateur de Hill, cet estimateur ne peut pas être interprété graphiquement. Une autre généralisation a été proposée par Beirlant ([BVT96]) qui a donné lieu à l'estimateur UH , vu comme un « un quantile plot généralisé ».

2.6.3 Estimateur UH

il est défini par :

$$UH_{k,n} = X_{n-k,n} \left(\frac{1}{k} \sum_{j=1}^k \log X_{n-j+1,n} - \log X_{n-k,n} \right) = X_{n-k,n} \hat{\xi}_{k,n}^{(H)} \quad (2.108)$$

Le « quantile plot généralisé » correspond au nuage de points suivant :

$$\left(\log \left(\frac{n+1}{j} \right), \log UH_{j,n} \right) \quad j = 1, \dots, n \quad (2.109)$$

$$\hat{\xi}_{k,n}^{(UH)} = \frac{1}{k} \sum_{j=1}^k \log UH_{j,n} - \log UH_{k+1,n} \quad (2.110)$$

Ainsi, selon que la pente est positive, nulle ou négative, les données sont dans le MDA(Fréchet), MDA(Gumbel) ou MDA(Weibull).

Sous certaines conditions de régularité la normalité asymptotique de l'estimateur UH est donnée par :

$$\sqrt{k} \left(\hat{\xi}_{k,n}^{(UH)} - \xi \right) \xrightarrow{loi} \begin{cases} \mathcal{N} \left(0, A + \xi^2 \right) & si \quad \xi \geq 0 \\ \mathcal{N} \left(0, \frac{(1-\xi)(1+\xi+2\xi^2)}{1-2\xi} \right) & si \quad \xi < 0 \end{cases} \quad (2.111)$$

La figure (2.3), montre un exemple d'application de cet outil.

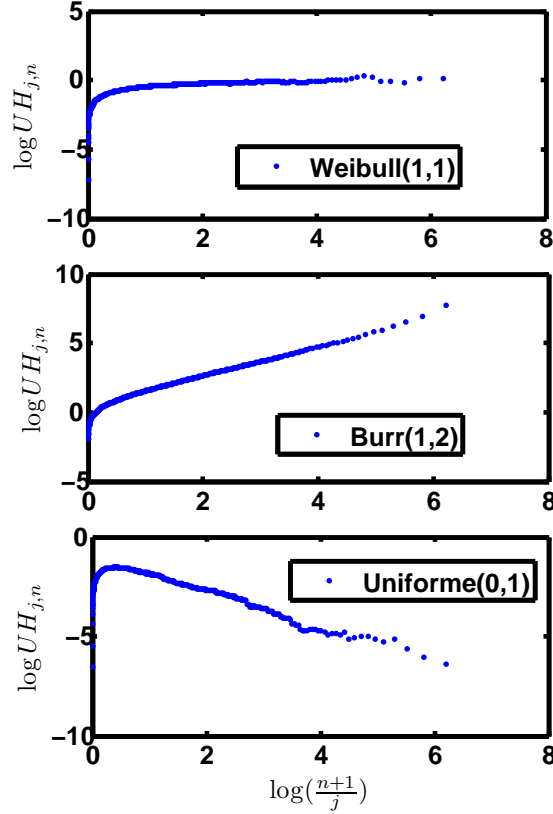


FIG. 2.3: $UH_{k,n}(u)$ pour 1000 valeurs issues des lois : Weibull(1,1) \in MDA(Gumbel), Burr(1,2) \in MDA(Fréchet) et Uniforme(0,1) \in MDA(Weibull)

Les estimateurs $\hat{\xi}_{k,n}^{(M)}$ et $\hat{\xi}_{k,n}^{(UH)}$ sont insuffisants pour donner des estimations des quantiles extrêmes. Il faudra en plus la connaissance d'un estimateur du paramètre d'échelle. Nous utilisons dans ce travail l'estimateur proposé par [dHR93].

$$\hat{\sigma}_{k,n}^{(\cdot)} = \frac{X_{n-k,n} \sqrt{3 \left(\hat{\xi}_{k,n}^{(H)} \right)^2 - S_{k,n}}}{\sqrt{3 \left[\rho_1 \left(\hat{\xi}_{k,n}^{(\cdot)} \right) \right]^2 - \rho_2 \left(\hat{\xi}_{k,n}^{(\cdot)} \right)}} \quad (2.112)$$

où $S_{k,n}$ est donnée par l'équation (2.106), avec

$$\rho_1(\xi) = \begin{cases} 1 & \text{si } \xi \geq 0 \\ \frac{1}{1-\xi} & \text{si } \xi < 0 \end{cases} \quad (2.113)$$

$$\rho_2(\xi) = \begin{cases} 2 & \text{si } \xi \geq 0 \\ \frac{2}{(1-\xi)(1-2\xi)} & \text{si } \xi < 0 \end{cases} \quad (2.114)$$

où $\hat{\xi}_{k,n}^{(\cdot)}$ est remplacé par $\hat{\xi}_{k,n}^{(M)}$ ou $\hat{\xi}_{k,n}^{(UH)}$

Conclusion : Dans ce chapitre, plusieurs méthodes d'estimation des paramètres ont été présentées et cela pour les différentes approches à savoir les approches de Gumbel, GEV et POT. Dans la suite, plusieurs simulations numériques ont été mises en oeuvre afin de distinguer la pertinence de chaque méthode et son influence sur l'estimation des profondeurs maximales. Ce travail est présenté dans les chapitres 4 et 5.

Chapitre 3

Analyse et Statistique Spatiales des Données

3.1 Introduction

Les processus générateurs des données spatiales, sont des modèles mathématiques qui permettent de décrire comment des objets sont distribués dans le plan ou dans l'espace et qui servent à identifier d'éventuelles dépendance entre ces objets. Dans l'étude de tels processus, on peut faire la distinction entre deux niveaux :

- Le premier niveau étant basique. Il consiste à décrire la structure en identifiant les objets à leurs coordonnées. A titre d'exemple, ces objets, peuvent être des arbres, des fissures, des étoiles, etc.
- Le deuxième niveau permet d'inclure les caractéristiques des objets appelées « marques » tels que la hauteur des arbres, la longueur des fissures, la luminosité des étoiles, etc.

A partir de ces deux niveaux, deux notions sont distinguées ([Gra06]) :

1. **L'analyse spatiale** si l'étude des processus se fait via le premier niveau.
2. **La statistique spatiale** si l'étude se fait via le deuxième niveau.

L'un des objectifs majeur de l'analyse et la statistique spatiale, est de quantifier la dépendance entre les objets. Ainsi, on parle de :

1. **Interaction** pour l'analyse spatiale où on distingue entre trois types de distributions spatiales : agrégée, régulière et aléatoire.
2. **Autocorrélation** pour la statistique spatiale qui mesure la tendance des objets voisins à se ressembler ou au contraire à se différencier du point de vue marque.

Du point de vue mathématique, un processus spatial est défini comme suit :

$$Z_s = \{Z(s) : s \in D \subset \mathbb{R}^2\} \quad (3.1)$$

où D est un ensemble d'indices et $Z(s)$ la valeur de l'objet localisée en s .

Cressie ([Cre00]) distingue trois types de données spatiales :

1. **Données géostatistiques** : D est fixé et s varie continuellement sur D .
2. **Données sur réseau ou latticiel** : D est fixé et discret.
3. **Données ponctuelles** : Les objets sont observés en des positions aléatoires, c'est à dire que D est un ensemble aléatoires d'indices.

Les processus spatiaux sont utilisés dans plusieurs domaines, en particulier en foresterie et en écologie ([MUM98, GPC95]), en astronomie ([HCH⁺06]), en épidémiologie ([Law01]) ou encore en pluviométrie ([MV04]).

S'agissant de la corrosion par piqûres, les processus générateurs de données ponctuelles (processus ponctuels) apparaissent les mieux adaptés pour étudier la distribution spatiale des piqûres et pour quantifier leurs dépendances. En effet, les piqûres apparaissent à des endroits aléatoires et sont considérées comme étant une réalisation d'un processus stochastique. L'attribut de la piqûre est identifié, dans ce travail, à sa profondeur. Récemment, ces processus ponctuels ont été adaptés au problème de la corrosion par piqûres. Parmi les travaux qui illustrent l'application de ces modèles, on cite les références suivants : ([CLKG07, CG08]), ([BOHS05]) et ([CH96]).

Dans la suite, une réalisation d'un processus ponctuel X_s est un ensemble d'objets (ou événements) s_1, \dots, s_n dans une région d'étude \mathbf{R} . Les coordonnées de s_i sont composées de l'abscisse et l'ordonnée de l'événement i , et on écrit $s_i = (s_{i1}, s_{i2})$.

Pour décrire une structure spatiale, les notions de propriétés du premier et du second ordre sont très utiles.

Propriété du premier ordre : Etant donné un processus spatial X_s , la propriété du premier ordre est décrite par l'intensité $\lambda_X(s)$. Elle est définie comme étant le nombre moyen d'événements par unité de surface au point s . Son expression est donnée par ([Dig03]) :

$$\lambda_X(s) = \lim_{|ds| \rightarrow 0} \left\{ \frac{\mathbb{E}[N_X(ds)]}{|ds|} \right\} \quad (3.2)$$

où ds est une petite région autour du point s , $|ds|$ l'aire de ds et $N_X(ds)$ le nombre d'événements dans ds .

Propriété du second ordre : Elle est décrite par l'intensité du second ordre $\gamma_X(s_i, s_j)$ et contient des informations sur la dépendance spatiale entre événements. Elle est définie par ([Dig03]) :

$$\gamma_X(s_i, s_j) = \lim_{|ds_i|, |ds_j| \rightarrow 0} \left\{ \frac{\mathbb{E}[\{N_X(ds_i) - \lambda_X(s_i)|ds_i|\} \{N_X(ds_j) - \lambda_X(s_j)|ds_j|\}]}{|ds_i||ds_j|} \right\} \quad (3.3)$$

Parmi les caractéristiques explorées dans un processus spatial, il y a la stationnarité et l'isotropie :

- Un processus est dit stationnaire s'il est invariant par translation. Dans ce cas, sa propriété du premier ordre est constante, et on pose $\lambda_X(s) = \lambda_X$.
- Un processus est dit isotrope s'il est invariant par rotation. Dans ce cas sa propriété du second ordre dépend seulement de la distance h entre deux événements s_i et s_j , et on écrit $\gamma_X(s_i, s_j) = \gamma_X(h)$

3.2 Analyse spatiale des données : Cas stationnaire

L'objectif de cette partie est de présenter quelques aspects théoriques de l'analyse spatiale des données par les processus ponctuels stationnaires. Une étude similaire dans le cas non stationnaire sera présentée par la suite.

En général, on distingue entre trois types de structures spatiales : aléatoire, agrégée et régulière. Une présentation des processus générateurs de telles structures fera aussi l'objet de cette partie.

3.2.1 Structure aléatoire

Une structure aléatoire connue sous le nom de structure spatiale complètement aléatoire (complete spatial randomness) et notée par la suite CSR, vérifie les deux conditions suivantes ([IPSS08]) :

1. L'intensité du premier ordre est constante, ainsi le nombre d'événements dans une région $A \subset \mathbf{R}$ d'aire $|A|$ suit la loi de Poisson de paramètre $\lambda_X |A|$ et on écrit :

$$\mathbb{P}(N_X(A) = k) = \frac{(\lambda_X |A|)^k}{k!} \exp(-\lambda_X |A|) \quad k \in \mathbb{N} \quad (3.4)$$

qui représente la probabilité de trouver k événements dans la région A .

2. Il n'y a pas d'interactions entre les événements qui sont uniformément et indépendamment distribués dans la région A .

Dans une structure CSR, les événements ont la même probabilité d'occuper un endroit dans la région A , et les événements n'ont tendance ni à s'attirer ni à se repousser. Cette structure sert d'hypothèse nulle lors de l'investigation de la dépendance entre événements. En d'autres termes, on mesure la déviation de la structure CSR pour identifier les deux autres types de structures. Le processus ponctuel qui génère les structures CSR est le processus de Poisson homogène noté HPP ; processus qui a la propriété d'être stationnaire et isotrope.

Le nombre d'événements générés par la simulation d'un processus HPP avec une intensité λ_X , peut être différent d'une simulation à une autre, mais avec une intensité moyenne égale à λ_X . La simulation de tels processus dans une région A peut être décomposée en deux étapes :

1. Générer un nombre n suivant la loi de Poisson de paramètre $\lambda_X |A|$ qu'on note $\mathcal{P}(\lambda_X |A|)$. Il est obtenu comme étant le plus petit entier n tel que ([IPSS08]) :

$$\prod_{i=1}^{n+1} u_i < \exp(-\lambda_X |A|) \quad (3.5)$$

où u_i sont des réalisations de la loi uniforme dans l'intervalle $[0, 1]$ qu'on note $\mathcal{U}_{[0,1]}$.

2. Générer indépendamment n coordonnées de points selon la loi \mathcal{U}_A qui représente la loi uniforme dans la région A . Si la région A est rectangulaire, par exemple $A = [a, b] \times [c, d]$, alors les coordonnées des points sont données par :

$$((b - a)U_1 + a, (d - c)U_2 + c) \quad (3.6)$$

où U_1 et U_2 des v.a suivant la loi $\mathcal{U}_{[0,1]}$.

Si la région A est quelconque, on peut procéder à la méthode de rejet ([MM02]) en générant des points dans un rectangle contenant la région A selon le procédé ci-dessus. Cette fois-ci, les points qui n'appartiennent pas à A sont rejetés.

Pour générer une structure CSR avec exactement n événements, il faut faire appel au processus binomial (spatial) qui n'est d'autre que le processus HPP conditionné au nombre d'événements souhaité. Il détient ce nom du fait que la variable aléatoire $N_X(B) | N_X(A) = n$ représentant le nombre d'événements dans un ensemble B contenu dans A suit la loi binomiale $\mathcal{B}(n, |B| / |A|)$. En effet on montre que ([IPSS08]) :

$$\mathbb{P}(N_X(B) = k | N_X(A) = n) = \binom{n}{k} \left(\frac{|B|}{|A|} \right)^k \left(1 - \frac{|B|}{|A|} \right)^{n-k} \quad (3.7)$$

où $\binom{n}{k} = \frac{n!}{k!(n-k)!}$ représente le nombre de combinaisons de k éléments parmi n . La figure 53.1° montre un exemple de réalisation d'un processus HPP dans le carré unité.

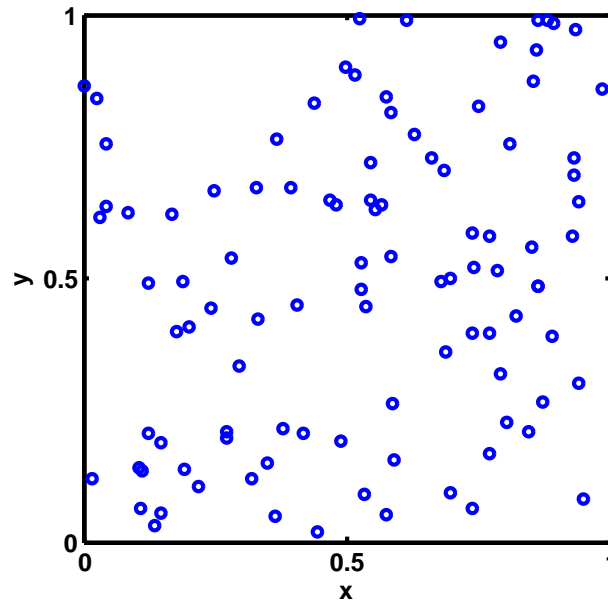


FIG. 3.1: Exemple de réalisation d'un processus HPP

3.2.2 Structure Agrégée

En général, une structure agrégée est caractérisée par une densité spatiale qui varie sensiblement. En effet, sur certains endroits, la densité est très élevée, et autour peut exister une densité très faible, voire nulle. La définition des agrégats (cluster) est parfois subjective ([IPSS08]). Ils peuvent être présentés comme étant un groupement de points avec une inter-distance (moyenne des distances entre points) inférieure à l'inter-distance dans toute la structure. La distinction entre le phénomène d'agrégation et d'hétérogénéité spatiale n'est pas un problème facile. L'ambiguïté entre les deux notions réside dans la définition qu'on leur donne. L'hétérogénéité correspond à une variation locale de l'intensité $\lambda_x(s)$, alors que l'agrégation peut être vue comme le résultat de dépendance entre événements. Souvent, l'apparition d'agrégats dans une structure est attribuée au processus physique du phénomène étudié, à l'interaction entre les événements ou tout simplement au caractère aléatoire intrinsèque au phénomène. L'intérêt de ce type de structure est de mettre en évidence, pour certains processus physiques difficiles à interpréter, la dépendance entre les événements générés et dans le contexte de corrosion par piqûres, cela permettra de tester l'indépendance spatiale entre les piqûres ; hypothèse nécessaire pour une bonne application de la théorie des valeurs extrêmes.

Parmi les processus qui permettent la génération de structures agrégées, les processus de Neyman-Scott sont un outil classique pour atteindre cet objectif ([IPSS08]). Ils consistent à générer les centres des clusters appelés « parents » suivant un processus HPP d'intensité λ_p , ensuite, des « descendants » sont générés autour de ces centres et leurs nombres sont issus d'une variable aléatoire

N_c , et leurs coordonnées sont indépendamment distribuées selon la même loi f_c . Si μ_p représente le nombre moyen de « descendants » par cluster, alors l'intensité finale du processus est constante et est égale à $\lambda_p \mu_c$. Notons que les « parents » sont seulement utilisés comme intermédiaires lors de la simulation et sont exclus de la structure agrégée finale. L'un des exemples simple des processus de Neyman-Scott, est le processus modifié de Thomas, noté MTCP (Modified Thomas Cluster Process), qui consiste à générer la v.a N_c suivant une loi de Poisson $\mathcal{P}(\mu_p)$, et les coordonnées selon une loi Normale bivariée d'écart-type σ_c et centrées sur les parents. La figure (3.2) montre des exemples de réalisations d'un tel processus dans le carré unité.

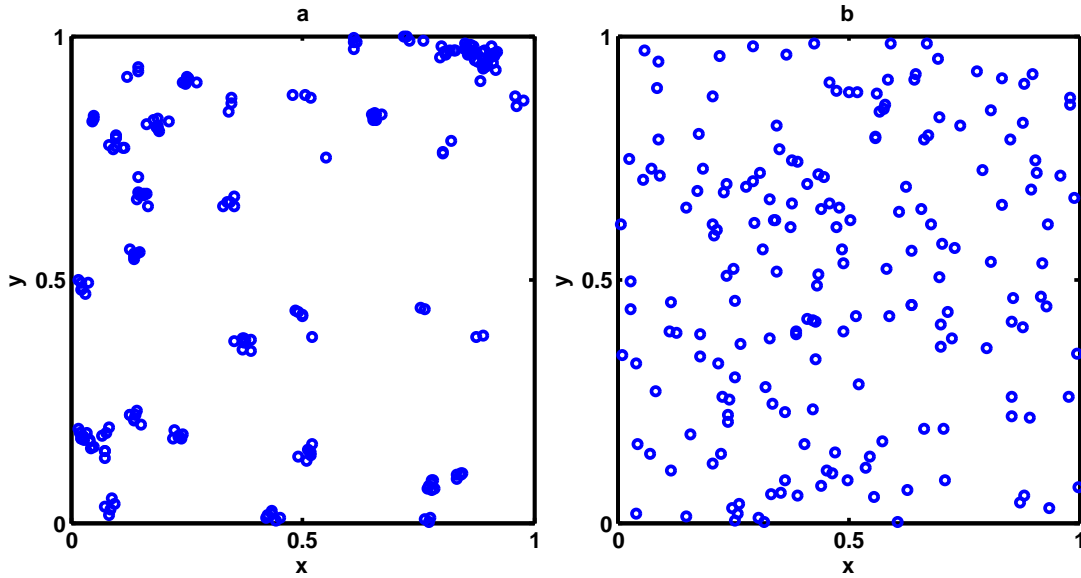


FIG. 3.2: Exemples de réalisations d'un processus MTCP avec $\lambda_p = 40$, $\mu_c = 4$ et : (a) $\sigma_c = 0,01$, (b) $\sigma_c = 0,1$

3.2.3 Structure Régulière

Dans ce type de structure, il n'y a pas de points situés à une distance inférieure à une distance minimale notée δ . Selon la valeur de δ , les points sont distribués plus ou moins régulièrement dans la région étudiée. Plus précisément, la distinction se fait entre structure régulière et structure dispersée, et le processus qui les génère s'appelle processus à noyau dur. Il peut représenter des objets qui restent à une distance minimale entre eux ou des objets impénétrables.

Parmi les méthodes classiques pour simuler ce type de structure, il y a le processus d'effacement de points et le processus d'inhibition séquentiel :

1. **Processus d'effacement** : Il sert à régulariser une configuration spatiale en éliminant les paires de points trop proches ([DLS06]). Si X_s est une réalisation d'un HPP d'intensité λ_x , alors pour une distance δ fixée, on efface toutes les paires de points $s_i \neq s_j$ pour lesquelles $\|s_i - s_j\| \leq 2\delta$. La configuration finale permet à chaque point s_i d'avoir son espace propre qui est la boule de centre s_i et de rayon δ .
2. **Processus d'inhibition séquentiel PIS** : En partant d'une surface vide et en fixant une inter-distance δ , le remplissage se fait par tirage aléatoire de points selon un processus HPP. Un point est rejeté s'il est à une distance inférieure à δ des autres points.

La figure (3.3) montre des exemples d'un tel processus dans le carré unité.

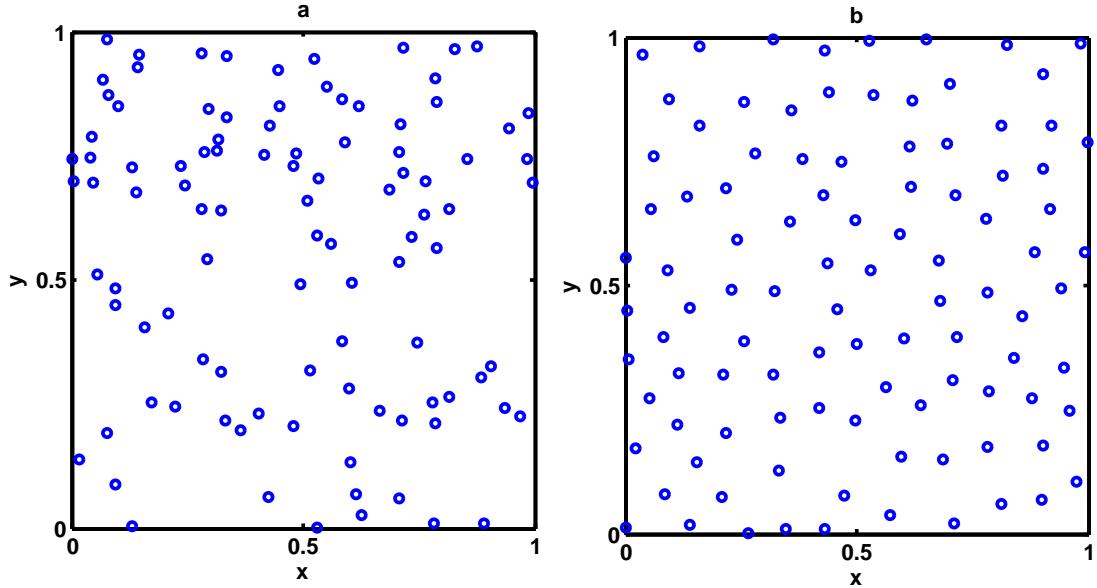


FIG. 3.3: Exemple de réalisations d'un processus d'inhibition séquentiel avec : (a) $\delta = 0,02$, (b) $\delta = 0,08$

3.2.4 Méthodes d'investigation des structures spatiales

Pour distinguer entre les trois types de structures spatiales citées précédemment, plusieurs approches sont possibles. Dans ce travail, trois techniques ont été considérées, à savoir, la méthode de Ripley issue de la classe des méthodes basées sur les distances, la méthode des quadrats utile dans le cadre de la méthode de Gumbel (voir Chapitre 4), et la méthode basée sur l'analyse spectrale, jusqu'à maintenant jamais adaptée au problème de la corrosion par piqûres.

3.2.4.1 Méthode de Ripley

Parmi les techniques classiques d'investigation des interactions entre objets, il y a les méthodes basées sur la mesure de distances. En général, les variables qui peuvent être considérées pour cet objectif sont :

1. **Distances des paires de points** : $s_{ij} = \|s_i - s_j\|$ qui représente toutes les distances entre les points s_i et s_j ($i \neq j$) dans la région étudiée.
2. **Distances aux voisins les plus proches** : qui représente la distance minimale entre un point s_i et les voisins les plus proches.
3. **Distances des espaces vides** : $d(u) = \min_i \|u - s_i\|$ qui représente la distance entre un point fixé de référence u dans la région étudiée aux points observés s_i .

La méthode de Ripley est l'une des méthodes basées sur les distances les plus utilisées. Introduite par Ripley ([Rip76, Rip77]), elle repose sur le calcul d'une fonction $K(r)$ (dite fonction de Ripley) qui mesure le nombre de voisins à distance r d'un point quelconque de la structure. Elle est définie par :

$$K(r) = \frac{\mathbb{E}(N_r(r))}{\lambda_X} \quad (3.8)$$

où N_r représente la v.a correspondante au nombre de voisins à distance r d'un point s_i et λ_X l'intensité du processus

. On note que la méthode suppose que le processus est stationnaire. Son expression dans le cas non-stationnaire est présentée dans la section (3.3).

Pour une structure CSR, qui sert d'hypothèse nulle, la fonction $K(r)$ est égale à πr^2 . Une structure agrégée est détectée si $K(r) > \pi r^2$ car les points ont en moyenne plus de voisins que dans une structure CSR. Inversement, $K(r) < \pi r^2$ pour une structure régulière où les points ont moins de voisins en moyenne.

En pratique, la fonction $K(r)$ est inconnue, elle est remplacée par l'estimateur suivant :

$$\hat{K}(r) = \frac{1}{\hat{\lambda}_X} \frac{1}{N} \sum_{i=1}^n \sum_{j=1}^n 1(d_{ij} \leq r) \quad (3.9)$$

où n le nombre de points observés, $\hat{\lambda}_X$ l'estimateur de la densité donné par $n/|A|$ avec A la région étudiée et

$$1(d_{ij} \leq r) = \begin{cases} 1 & \text{if } d_{ij} \leq r \\ 0 & \text{sinon} \end{cases} \quad (3.10)$$

La figure (3.4) montre un exemple d'estimation des trois structures par la méthode de Ripley.

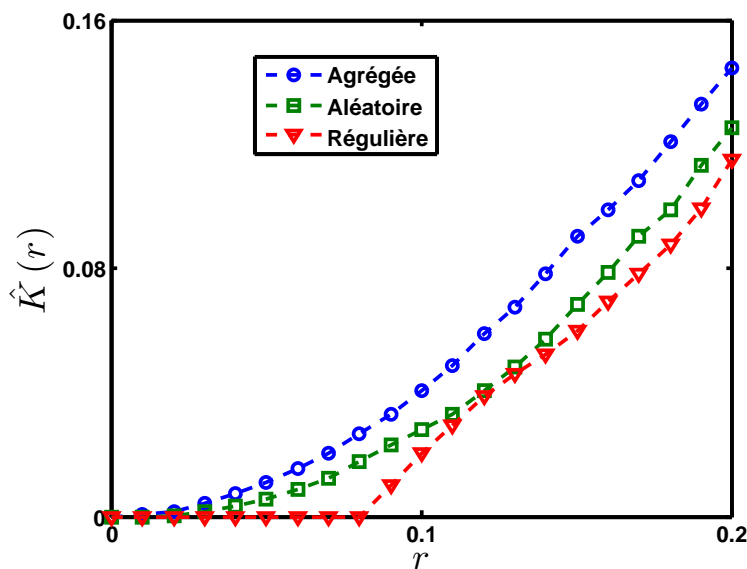


FIG. 3.4: Exemple de fonction Ripley pour les trois types de structures

A cause d'effets de bord, l'estimateur donné par l'équation (3.9) est biaisé. En effet, les valeurs de $K(r)$ sont sous estimées à cause des points proches de la bordure, qui ont moins de voisins que les autres points. Pour cela, il existe plusieurs méthodes de correction de bord. Celle utilisée dans ce travail est la méthode de correction de Ripley ([Rip77]). Dans ce cas, l'estimation de la fonction $K(r)$ est donnée par :

$$\hat{K}_c(r) = \frac{1}{\hat{\lambda}} \frac{1}{N} \sum_{i=1}^N \sum_{j=1}^N \frac{1(d_{ij} \leq r)}{w_{ij}} \quad (3.11)$$

où w_{ij} représente la proportion du cercle de centre s_i passant par s_j et qui se trouve à l'intérieur de la région étudiée. Il est à noter que le calcul de $\hat{K}_c(r)$ se fait souvent jusqu'à des valeurs de r au plus inférieure ou égale à la moitié de la longueur de la zone étudiée (rectangulaire dans ce travail), car pour les grandes valeurs de r l'interprétation de $\hat{K}_c(r)$ n'a plus beaucoup de sens vu que dans ce cas, peu de points sont utilisés pour son calcul. Parfois, l'ordre du quart de la longueur suffit par souci de prudence ([Gor00]). Les figures (3.5) et (3.6) illustrent respectivement le problème des effets de bord et la différence entre les estimations avec et sans correction de bord.

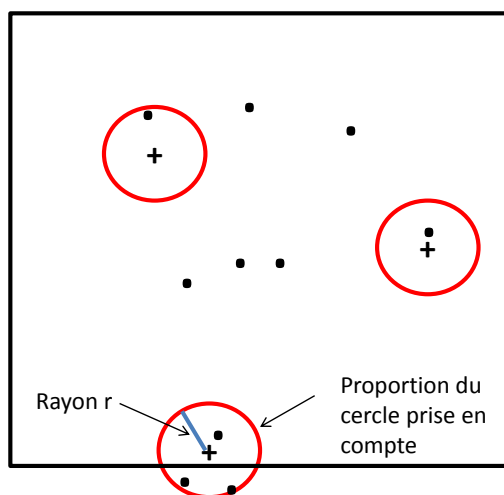


FIGURE 3.5 – Illustration du problème des effets de bord

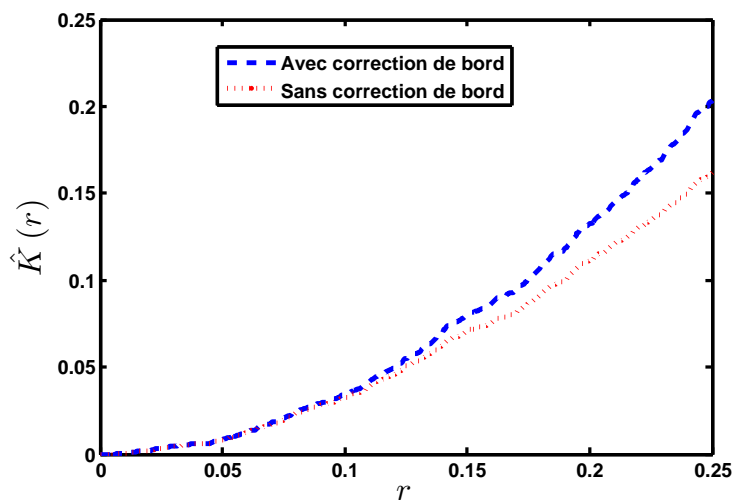


FIGURE 3.6 – Comparaison des estimations de $K(r)$ avec et sans correction de bord pour cent points générés avec un process HPP sur le carré unité

En pratique, pour une structure observée, la fonction de Ripley est estimée par l'équation (3.11). Ensuite, un intervalle de confiance est calculé par la méthode de Monte-Carlo. Il est basé sur la simulation de N_{simul} réalisations du processus HPP ; la fonction de Ripley étant calculée pour chaque simulation. **Pour une valeur de r fixée, les bornes de l'intervalle de confiance sont données par les statistiques d'ordre $N_{simul} \alpha/2$ et $N_{simul} (1 - \alpha/2)$. Si la fonction de Ripley reste à l'intérieur de l'enveloppe, alors la structure observée est une CSR. Si la fonction de Ripley est au dessus de l'enveloppe, la structure sera alors supposée agrégée. Inversement, si cette fonction est au dessous de l'enveloppe, elle sera supposée régulière.**

3.2.4.2 Méthode des quadrats

L'une des alternatives aux approches basées sur la mesure des distances, est la méthode des quadrats ([Gle20]). Le principe de la méthode consiste à subdiviser la région d'étude A (contenant n points) en m domaines (carré ou rectangle) de mêmes aires appelés quadrats. Le nombre de points n_i dans chaque quadrat est calculé et comparé au nombre de points théoriques attendus \bar{n} qui, dans le cas d'une configuration uniforme est égale à $\bar{n} = n/m$. Ensuite, la statistique de Pearson (test de Chi-deux) est calculée. **Sous l'hypothèse nulle (CSR) la v.a suit la loi de Chi-deux à $m - 1$ degrés de liberté. Ainsi, une structure CSR n'est pas rejetée si la statistique de Pearson observée est dans la zone d'acceptation pour un seuil α donné.**

Un deuxième indice déduit de la statistique de Pearson peut être établi pour distinguer les structures agrégées et régulières. C'est l'indice de dispersion noté I et défini par :

$$I = \frac{(m - 1) S^2}{\bar{n}} \quad (3.12)$$

avec S^2 la variance du nombre de points par quadrat.

Ainsi, si I appartient à l'intervalle de confiance défini à partir des quantiles de la loi de Chi-Deux $[\chi_{m-1;1-\alpha}^2, \chi_{m-1;\alpha}^2]$, alors la structure est CSR, si $I > \chi_{m-1;\alpha}^2$ alors la structure est agrégée et si $I < \chi_{m-1;1-\alpha}^2$ alors la structure est régulière.

La figure (3.7) montre un découpage en 4×4 quadrats, où le nombre de points sur chacun de ces quadrats est indiqué. La structure indiquée étant une réalisation d'un HPP dans le carré unité.

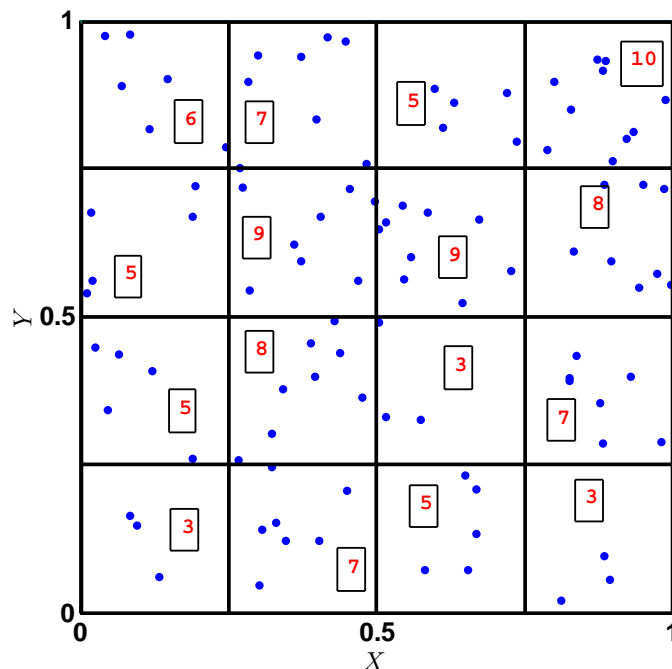


FIG. 3.7: Illustration de la méthode des quadrats : découpage 4×4 pour un processus HPP à cent points dans le carré unité

Cette méthode a l'avantage d'être simple à mettre en oeuvre et ne nécessite pas la connaissance exacte des coordonnées des points car seul leur nombre par quadrat compte. Néanmoins, l'une des principales critiques envers cette méthode, est la dépendance de la statistique de Pearson à la taille des quadrats. C'est le problème qui consiste à déterminer l'échelle à laquelle la structure peut être considérée comme aléatoire ou non.

3.2.4.3 Analyse spectrale

L'analyse spectrale des processus spatiaux était ignorée des praticiens en faveur des méthodes basées sur la mesure des distances, et cela suite à sa relative complexité et la non-vulgarisation scientifique de son utilisation. Les travaux de Bartlett ([Bar64]) étaient les premiers à ouvrir le champ vers une large application de l'approche spectrale dans différents domaines. L'objectif de cette section est d'en présenter les aspects théoriques généraux ainsi que l'illustration de son application aux trois types de structures étudiées précédemment.

3.2.4.4 Fonction de densité spectrale

Pour définir la fonction de densité spectrale, il est nécessaire d'introduire la notion de fonction de covariance spectrale définie par :

$$\kappa_X(a, b) = \lambda_X(a) \delta(a - b) + \gamma_X(a, b) \quad a, b \in \mathbb{R}^2 \quad (3.13)$$

où $\delta(\cdot)$ est la fonction de Dirac qui vérifie pour toute fonction intégrable :

$$\int_{\mathbb{R}^2} f(x) \delta(x - x') dx = f(x') \quad (3.14)$$

La fonction de densité spectrale f_X est définie comme étant la transformée de Fourier de la fonction de covariance spectrale. Ainsi, pour un processus stationnaire, f_X s'écrit comme suit ([MR96]) :

$$f_X(w_1, w_2) = \lambda_X + \int_{\mathbb{R}^2} \gamma_X(c_1, c_2) \exp(-i(w_1 c_1 + w_2 c_2)) dc_1 dc_2 \quad (3.15)$$

Soient $\Omega = [0, L_x] \times [0, L_y]$ contenant n points et $\{(x_i, y_i), i = 1, \dots, n\}$ leurs coordonnées. A cause de la complexité de la fonction de densité spectrale, Bartlett ([Bar64]) a proposé une approximation \hat{f}_X appelée périodogramme et basée sur la transformée discrète de Fourier. Son expression est donnée par ([MR96]) :

$$\hat{f}_X(p, q) = F_X(p, q) \overline{F_X(p, q)} = \{A(p, q)\}^2 + \{B(p, q)\}^2 \quad (3.16)$$

avec

$$F_X(p, q) = \sum_{i=1}^n \exp\left(-2i\pi \left(p \frac{x_i}{L_x} + q \frac{y_i}{L_y}\right)\right) = A(p, q) + iB(p, q), \quad (p, q) \in \mathbb{Z}^2 \quad (3.17)$$

Pour simplifier cette expression, la structure est souvent ramenée à la surface unité via une transformation. Dans ce cas la formule (3.18) devient :

$$F_X(p, q) = \sum_{i=1}^n \exp\left(-2i\pi \left(px'_i + qy'_i\right)\right), \quad (x'_i, y'_i) \in [0, 1]^2 \quad (3.18)$$

La propriété principale du périodogramme est la symétrie, c'est à dire :

$$\hat{f}_X(-p, -q) = \hat{f}_X(p, q) \quad (3.19)$$

Cette propriété est très utile car elle permet de réduire le calcul du périodogramme. En effet, il est calculé pour les valeurs positives et négatives de q (par exemple), et seulement pour les valeurs positives de p . Mugglestone et al.([MR96]) proposent une discrétisation avec $p = 0, \dots, 16$ et $q = -16, \dots, 15$. Dans ce travail, la discrétisation choisie est telle que $p = 0, \dots, p_{max}$ et $q = -q_{max}, \dots, q_{max}$, où $p_{max} = q_{max} = 64$ afin de permettre l'exploration de la structure à des échelles encore plus fines.

3.2.4.5 Interprétation du périodogramme

Les valeurs du périodogramme représentent la contribution de chaque valeur du couple de fréquences défini par $(w_p, w_q) = (2\pi p, 2\pi q)$ à la variance totale du processus spatial. Ainsi, pour une structure CSR, aucune valeur du périodogramme ne domine le reste des valeurs et pour une structure agrégée (respectivement régulière), le périodogramme présente des grandes valeurs (respectivement des faibles valeurs) pour les petites fréquences. Les figures (3.8,3.9,3.10) montrent des exemples de périodogrammes calculés à partir des processus HPP, MTCP et PIS.

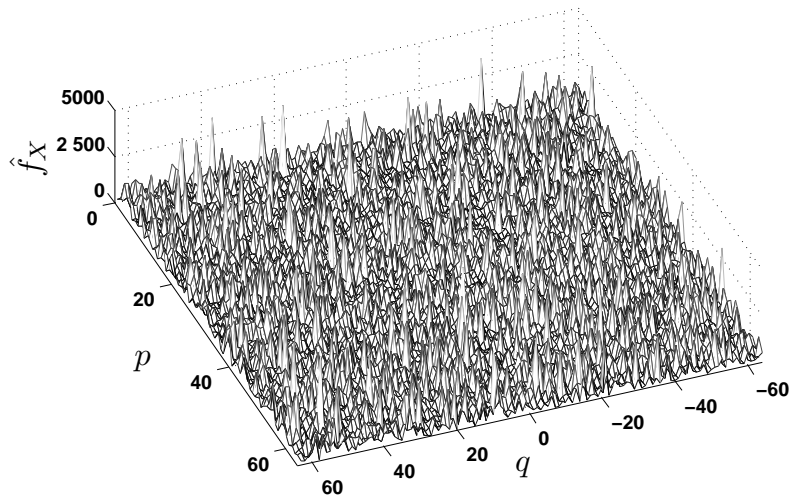


FIG. 3.8: Périodogramme d'un processus HPP à 500 points dans un carré unité

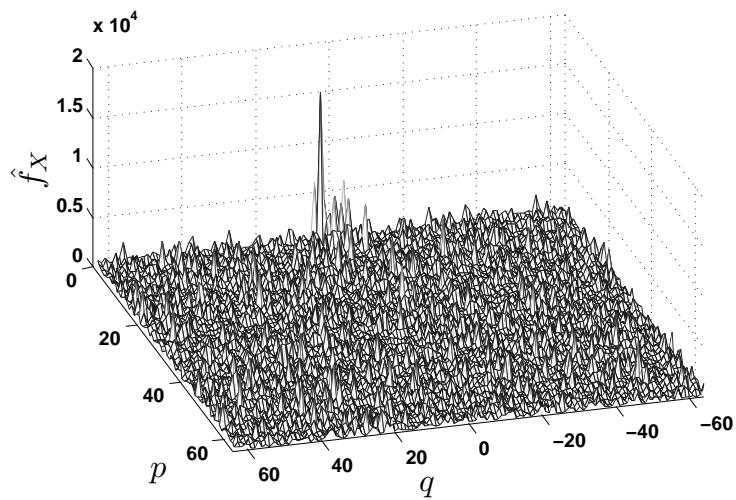


FIG. 3.9: Périodogramme d'un processus MTCP à 497 points dans un carré unité avec : $\lambda_p = 50$, $\mu_p = 10$ et $\sigma_c = 0,05$

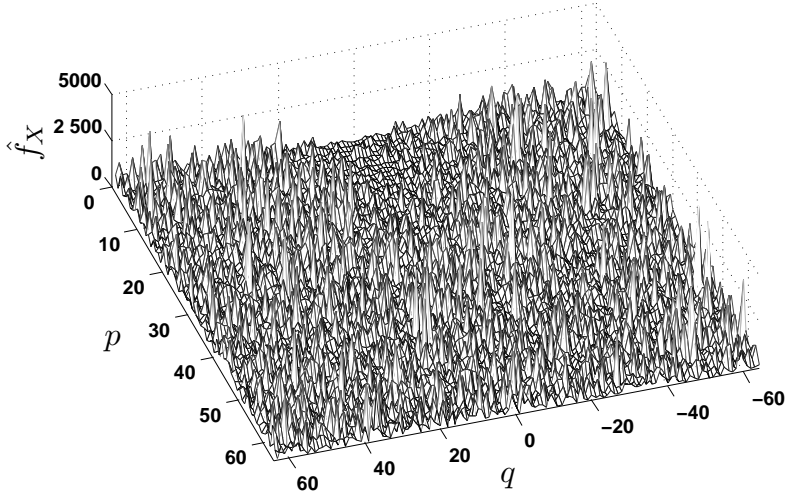


FIG. 3.10: Périodogramme d'un processus PSI à 500 points dans un carré unité avec $\delta = 0,025$

Pour une meilleure interprétation du périodogramme, ([RF84, RF83]) ont proposé une représentation polaire décrite par le $R - Spectre$ qui caractérise les échelles spatiales de la structure étudiée, et par le $\theta - Spectre$ qui vérifie la présence de direction privilégiée.

3.2.5 $R - Spectre$ et $\theta - Spectre$

Le $R - Spectre$ noté par $\hat{f}_R(r)$ est défini par :

$$\hat{f}_R(r) = \frac{1}{n_r} \sum_{r'} \sum_{\theta} \hat{f}_X(r', \theta), \quad 1 \leq r \leq \sqrt{p_{max}^2 + q_{max}^2} \quad (3.20)$$

$\hat{f}_R(r)$ représente la moyenne des n_r valeurs $\hat{f}_X(r', \theta)$ du périodogramme pour lesquelles $r-1 < r' \leq r$. Dans cette représentation polaire, $\hat{f}_X(r', \theta)$ correspond à $\hat{f}_X(p, q)$ où le couple de valeurs (p, q) est défini par :

$$r' = \sqrt{p^2 + q^2}, \quad \theta = \arctan\left(\frac{q}{p}\right) \quad (3.21)$$

Par exemple, pour calculer $\hat{f}_R(6)$, on prend la moyenne de toutes les n_r valeurs du périodogramme pour lesquelles $5 < \sqrt{p^2 + q^2} \leq 6$. Le choix d'un pas égal à 1 est recommandé en pratique ([MR96]).

D'une façon similaire, le $\theta - Spectre$ noté par $\hat{f}_\Theta(\theta)$ est construit en moyennant les valeurs du périodogramme pour lesquelles $\theta - 5^\circ < \theta' \leq \theta + 5^\circ$. Un pas égal à 10° est également recommandé en pratique ([MR96]). le $\theta - Spectre$ est défini par :

$$\hat{f}_\Theta(\theta) = \frac{1}{n_\theta} \sum_r \sum_{\theta'} \hat{f}_X(r, \theta'), \quad \theta = -90^\circ, \dots, 90^\circ \quad (3.22)$$

Finalement, il est à noter que $\hat{f}_X(0, 0)$ est exclu de la représentation polaire car sa grande valeur égale à n^2 domine la forme du périodogramme.

En calculant ces deux spectres, la distinction entre les trois types de structures est basée sur les lois de probabilités de ces deux spectres. En effet, sous l'hypothèse nulle (CSR), Mugglestone et al. ([MR96, MR01]) montrent que :

$$\frac{2\hat{f}_X(w_p, w_q)}{n} \sim \chi_2^2, \quad (w_p, w_q) \neq (0, 0) \quad (3.23)$$

et

$$\frac{2\{\hat{f}_X(0, 0) - n\}}{n} \sim \chi_1^2 \quad (3.24)$$

où χ_n^2 représente la loi de Chi-Deux à n degrés de liberté.

Pour le $R - Spectre$, on a :

$$\frac{\hat{f}_R(r)}{n} \sim \frac{1}{2n_r} \chi_{n_r}^2 \quad 1 \leq r \leq \sqrt{p_{max}^2 + q_{max}^2} \quad (3.25)$$

avec

$$\mathbb{E} \left[\frac{\hat{f}_R(r)}{n} \right] = 1 \quad (3.26)$$

Et pour le $\theta - Spectre$, on a :

$$\frac{\hat{f}_\theta(\theta)}{n} \sim \frac{1}{2n_\theta} \chi_{n_\theta}^2 \quad \theta \in [-90^\circ, 90^\circ] \quad (3.27)$$

avec

$$\mathbb{E} \left[\frac{\hat{f}_\theta(\theta)}{n} \right] = 1 \quad (3.28)$$

Par conséquent, sous l'hypothèse CSR, $\frac{\hat{f}_R(r)}{n}$ reste à l'intérieur de l'intervalle de confiance $\frac{1}{2n_r} \left[\chi_{2n_r, \frac{\alpha}{2}}^2, \chi_{2n_r, 1-\frac{\alpha}{2}}^2 \right]$ où α représente le risque du test.

Pour les faibles valeurs de r , une structure est supposée agrégée si $\frac{\hat{f}_R(r)}{n}$ est au dessus de la bande supérieure de l'intervalle de confiance alors qu'une structure est supposée régulière si $\frac{\hat{f}_R(r)}{n}$ est au dessous de la bande inférieure de l'intervalle de confiance. Le $R - Spectre$ donne des informations sur la configuration spatiale de la structure étudiée et les échelles pour lesquelles cette configuration peut être considérée comme aléatoire, agrégée ou régulière. Par ailleurs, le $\theta - Spectre$ teste l'isotropie de la structure, mais son interprétation directe, ($\frac{\hat{f}_R(r)}{n}$ qui reste à l'intérieur de l'intervalle de confiance $\frac{1}{2n_r} \left[\chi_{2n_r, \frac{\alpha}{2}}^2, \chi_{2n_r, 1-\frac{\alpha}{2}}^2 \right]$ sous l'hypothèse CSR) nécessite une analyse complémentaire (Voir commentaire figure (3.12)).

Pour illustrer ce propos, les deux spectres ont été calculés pour les trois types de structures. Les figures (3.11), (3.12) et (3.13) montrent ces résultats pour $\alpha = 10\%$.

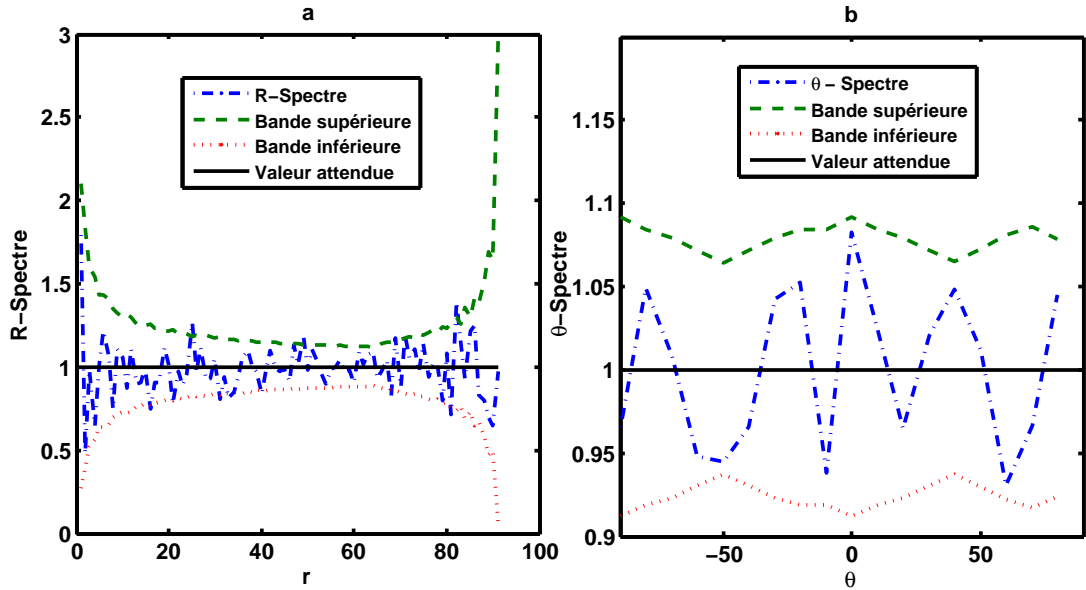


FIG. 3.11: Spectres d'un processus HPP à 500 points dans un carré unité : (a) $R - Spectre$, (b) $\theta - Spectre$. La valeur attendue est égale à 1.

Le processus HPP étant stationnaire et isotrope, les deux spectres restent à l'intérieur de l'intervalle de confiance confirmant ces deux propriétés.

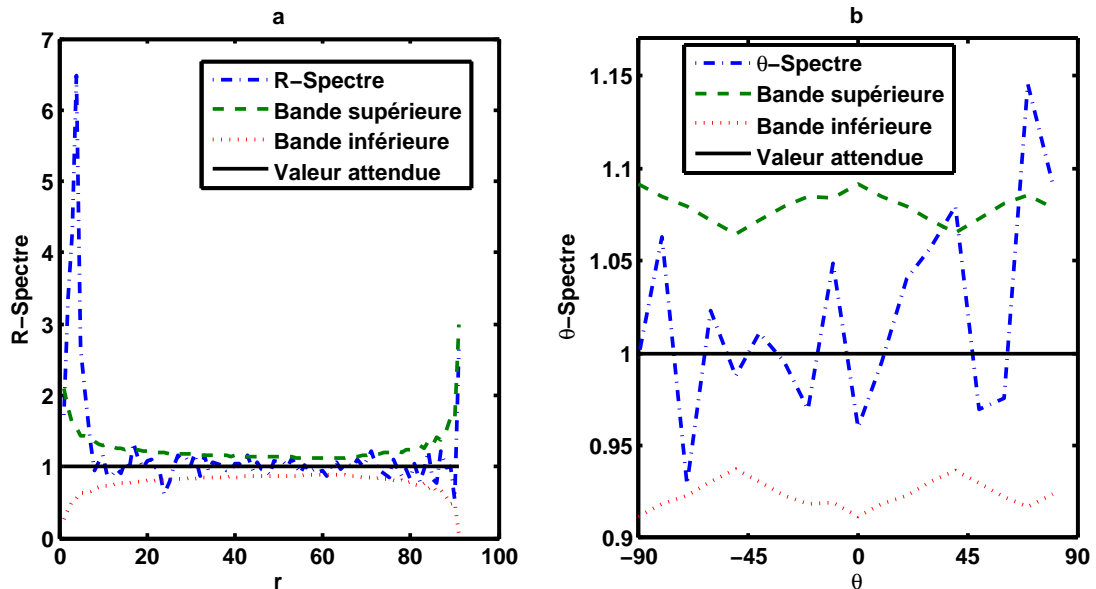


FIG. 3.12: Spectres d'un processus MTCP à 487 points dans un carré unité avec $\lambda_p = 50$, $\mu_p = 10$ et $\sigma_c = 0,05$: (a) $R - Spectre$, (b) $\theta - Spectre$

Le $R - Spectre$ du processus MTCP simulé présente pour les faibles valeurs, des grandes valeurs r dépassant la borne supérieure de l'intervalle de confiance. Cela intervient pour toutes les valeurs

$r \leq 7$ ce qui correspond à une échelle d'agrégation jusqu'à $1/7$ de la surface unité. D'autre part, le θ -Spectre reste dans l'intervalle de confiance sauf pour des angles proches de 90° . Cela peut être dû aux fluctuations de l'échantillonnage car le processus MTCP est isotrope. Cependant, lorsque le phénomène d'agrégation est accentué, le θ -Spectre du processus MTCP peut sortir de l'intervalle de confiance, mais cela n'indique pas une anisotropie. Schabenger ([SG05]) a évoqué ce problème et pour y remédier, une analyse complémentaire consiste à calculer la statistique $\frac{\hat{f}_\Theta(\theta)}{\hat{f}_\Theta(-\theta)}$, qui sous l'hypothèse d'isotropie, suit une distribution de Fisher. Ce calcul n'est pas présenté ici, car l'intérêt de ce travail est porté sur l'exploration de la structure et non pas sur l'orientation des objets y appartenant.

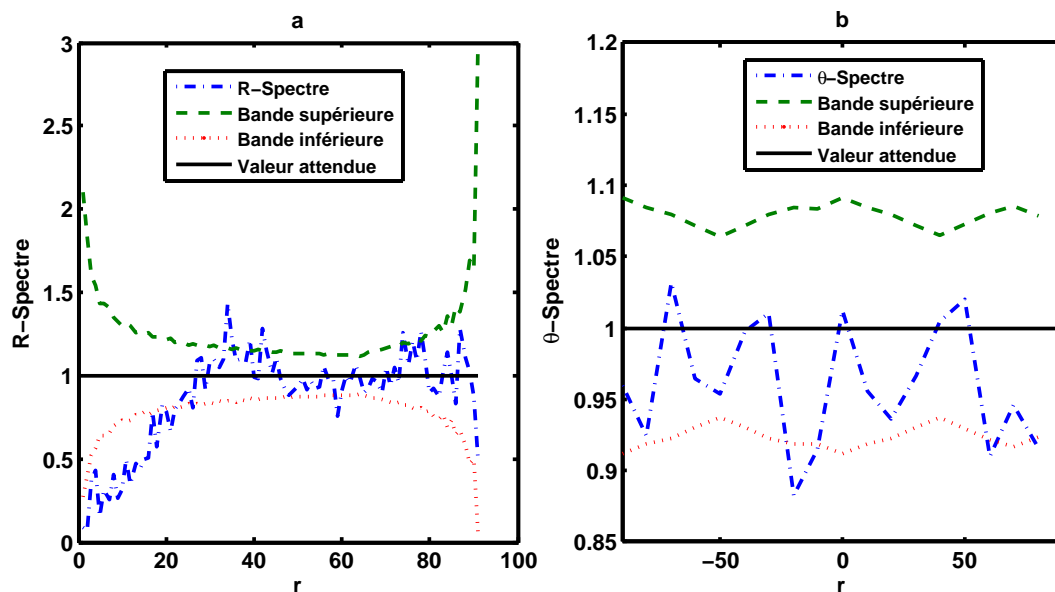


FIG. 3.13: Spectres d'un processus PSI à 500 points dans un carré unité avec $\delta = 0,025$: (a) R -Spectre, (b) θ -Spectre

S'agissant de la structure régulière, le R -Spectre présente des valeurs inférieures à la bande inférieure de l'intervalle de confiance pour des petites valeurs de r confirmant ainsi la structure régulière du processus PSI simulé. Le pic observé à $r = 34$ correspond à une échelle de $1/34 \approx 0,029$ qui est proche de $\delta = 0,025$ la distance inter-points retenue pour la simulation. L'écart entre les deux valeurs peut être due aux fluctuations de l'échantillonnage. En calculant la moyenne des pics observés lors de la génération de 100 structures PSI avec les mêmes paramètres, la valeur moyenne de la position du pic est approximativement égale à $0,026$.

Remarques Les méthodes citées précédemment ont été développées dans le cadre d'un processus stationnaire (homogène). Ainsi, la validité de ces méthodes est restreinte au cas des processus à intensité constante. Il faut rappeler que le terme de **stationnaire** est souvent utilisé au lieu du terme (homogène) ([IPSS08]). La notion d'interaction apparaît lorsqu'il y a dépendance entre les objets situés dans la région étudiée. Par conséquent avec une structure CSR il y a une totale indépendance entre les points, alors qu'on attribue l'observation d'une structure agrégée ou régulière à l'existence d'une interaction entre les points due au phénomène physique (ou autre) qui les génère. Comme évoqué au paragraphe (3.2.2) il y a une réelle ambiguïté entre la notion d'interaction et

non-homogénéité. La deuxième notion signifie que l'intensité du processus dépend de la position mais cela n'implique pas forcément l'existence d'une interaction entre les points c'est le cas par exemple du processus de Poisson non homogène qui est caractérisé par l'indépendance entre les positions des points malgré leur distribution spatiale non uniforme. La caractérisation des processus non homogène fera l'objet de la section suivante. Enfin, le terme d'interaction est lié à l'existence d'agrégation ou d'inhibition dans la structure indépendamment de la distribution spatiale des points et l'objectif de l'analyse spatiale est de mettre en évidence cette interaction dans le cas stationnaire (homogène) et non stationnaire (hétérogène).

3.3 Analyse spatiale des données : Cas non-stationnaire

les processus non-stationnaires sont caractérisés par une intensité $\lambda_X(s)$ qui dépend de la position (x, y) des points. Cette intensité va être notée dans la suite par $\lambda_X(x, y)$. Le processus de Poisson non-homogène noté *NHPP* est la forme la plus simple des processus non-homogènes. Il repose sur les deux assertions suivantes :

1. Le nombre de points dans une région A suit une loi de Poisson de paramètre $\int_A \lambda_X(x, y) dx dy$.
2. les variables aléatoires correspondantes aux nombres de points dans k sous-ensembles disjoints de A sont indépendantes, pour toute valeur de k .

Les méthodes présentées précédemment peuvent s'adapter au cas non-stationnaire. En général, pour étudier une structure, la première étape consiste à ajuster la configuration observée à un modèle puis la deuxième étape consiste à tester la validité de cet ajustement et enfin, en cas d'acceptation du modèle, à le simuler. La classe des processus non-stationnaires est très élargie et l'implémentation des méthodes pour quantifier les interactions ne sont pas toujours faciles à mettre en oeuvre. Dans ce travail, seul les processus NHPP et les processus agrégés avec une intensité des descendants non constante, seront étudiés.

3.3.1 Méthodes d'ajustement d'une structure spatiale

Il existe plusieurs méthodes pour ajuster une structure observée A . Pour cela, deux approches vont être présentées dans ce travail. La première consiste à estimer directement l'intensité $\lambda_X(x, y)$ de A , la deuxième est la méthode du minimum de contraste qui consiste à utiliser une statistique connue de la structure. La minimisation de son écart avec la statistique observée donnera les paramètres caractérisant cette structure.

3.3.1.1 Estimation de l'intensité $\lambda_X(x, y)$

Soit n le nombre de points observés dans la région A . Pour estimer l'intensité $\lambda_X(x, y)$ d'un processus NHPP, il est fréquent d'utiliser la méthode du maximum de vraisemblance. Par contre, si l'indépendance entre les points n'est pas satisfaite, alors pour certaines classes de processus (processus de Gibbs par exemple) cette méthode n'est pas facile à mettre en oeuvre ([BT00]), elle est remplacée par la méthode du maximum de pseudo-vraisemblance qui fait appel à la notion d'intensité conditionnelle ([BT00]). Dans le cas d'un processus NHPP, les deux méthodes du maximum de vraisemblance et du maximum de pseudo-vraisemblance coïncident. Comme en théorie des valeurs extrêmes, le log-vraisemblance est plus pratique à résoudre. Ainsi, si on suppose que l'intensité $\lambda_X(x, y)$ dépend d'un vecteur de paramètres Θ qu'on note $\lambda_\Theta(x, y)$, alors la fonction de log-vraisemblance est définie par :

$$\log L(\Theta; (x, y)) = \sum_{i=1}^n \log \lambda_{\Theta}(x_i, y_i) - \int_A \lambda_{\Theta}(s) ds \quad (3.29)$$

Si, par exemple, $\log \lambda_{\Theta}(x, y)$ est linéaire, alors le système de maximisation de la fonction log-vraisemblance admet une solution unique. Ce travail sera réalisé en conservant cette hypothèse.

$$\lambda_{\Theta}(x, y) = \exp(\theta_0 + \theta_1 x + \theta_2 y) \quad , \quad \Theta = (\theta_0, \theta_1, \theta_2) \quad (3.30)$$

3.3.1.2 Méthode du minimum de contraste

Soit $T_{\eta}(s)$ une statistique connue qui résume les propriétés d'une structure et qui dépend d'un vecteur de paramètres notée η , et $T(s)$ la statistique observée. La méthode du minimum de contraste consiste à chercher $\hat{\eta}$, l'estimateur de η , tel que ([Bad07]) :

$$\hat{\eta} = \min_{\eta} D(T(s), \mathbb{E}[T_{\eta}(s)]) \quad (3.31)$$

où D représente une métrique. Diggle ([Dig03]) a proposé d'utiliser comme statistique, la fonction de Ripley avec la métrique suivante :

$$\int_a^b |K(r) - \hat{K}(r)|^p dr \quad (3.32)$$

Exemples

Cas stationnaire : Dans le cas d'un processus MTCP, le vecteur des paramètres est $\Theta = (\lambda_p, \mu_c, \sigma_c)$. Sa fonction de Ripley est connue analytiquement, et elle est donnée par ([Bad07]) :

$$D_p = K(r) = \pi r^2 + \frac{1}{\lambda_p} \left(1 - \exp\left(-\frac{r^2}{4\sigma_c^2}\right) \right) \quad (3.33)$$

Ainsi, avec la statistique (3.32), les estimateurs des paramètres de MTCP sont calculés.

Cas non-stationnaire : Dans ce cas, l'estimation de la fonction de Ripley \hat{K}_{NH} est donnée par ([Dig03]) :

$$\hat{K}_{NH}(r) = \frac{1}{|A|} \sum_{i=1}^n \sum_{j \neq i} \frac{1 \{ \|s_i - s_j\| \leq r \}}{w_{ij} \hat{\lambda}_X(s_i) \hat{\lambda}_X(s_j)} \quad (3.34)$$

où $\hat{\lambda}_X(s)$ est une estimation de l'intensité et w_{ij} la proportion du cercle définie de la même façon que dans le cas stationnaire.

L'une des méthodes pour générer une structure agrégée non-stationnaire est de simuler les parents selon un processus HPP d'intensité λ_p , ce qui a été réalisé ici, et les descendants selon un processus NHPP d'intensité $\mu_{NH}(x, y)$ (voir section 3.3.3). Par conséquent, l'intensité finale $\lambda_X(x, y)$ est égale à $\lambda_p \mu_{NH}(x, y)$ et l'estimation de ses paramètres via la méthode du minimum de contraste peut se faire comme suit :

1. Estimer l'intensité finale $\lambda_X(x, y)$ (par la méthode du maximum de vraisemblance par exemple).
2. Calculer $\hat{K}_{NH}(r)$.
3. Estimer les paramètres en minimisant la statistique donnée par l'équation (3.32).

3.3.2 Test de la validité de l'ajustement

Dans ce travail, deux méthodes seront utilisées. Il s'agit de la méthode des enveloppes et la méthode des quadrats.

La méthode des enveloppes Le principe est le même que pour la méthode de Ripley dans le cas stationnaire. Si on souhaite tester l'hypothèse que le processus ajusté est un *NHPP*, alors via les simulations de Monte-Carlo, la fonction de Ripley $\hat{K}_{NH}(r)$ est calculée pour chaque réalisation. Ainsi, le modèle ajusté sera accepté si la fonction de Ripley observée reste dans l'enveloppe, et il est rejeté dans le cas contraire.

La méthode des quadrats C'est le même principe que dans le cas stationnaire. La statistique de Pearson est calculée. Cette fois-ci, \bar{n} qui représente le nombre de points théorique (attendu) n'est plus constant. Si $\hat{\lambda}_X(x, y)$ est l'intensité estimée, alors le nombre de points attendus dans le i ème quadrat C_i et qu'on note $n(C_i)$, est égale à $\int_{C_i} \hat{\lambda}_X(x, y) dx dy$. Ainsi, la statistique du test χ_{nh}^2 est donnée par :

$$\chi_{nh}^2 = \sum_{i=1}^m \frac{(n_i - n(C_i))^2}{n(C_i)} \quad (3.35)$$

où n_i est le nombre de points observés dans le i ème quadrat. Enfin, la statistique χ_{nh}^2 suit la loi de Chi-Deux à $m - 1$ degrés de liberté, et la règle de décision est la même que dans le cas stationnaire.

3.3.3 Simulation d'un processus non-stationnaire

Il existe plusieurs algorithmes pour simuler un processus spatial d'intensité $\lambda_X(x, y)$. Les plus connus sont les méthodes de Monte Carlo par chaînes de Markov pour simuler les structures à interaction (dépendance entre points) et l'algorithme de Metropolis-Hastings pour le cas sans interaction (indépendance entre points) ([DLS06]). S'agissant dans ce travail de simuler que les processus *NHPP*, la méthode simple d'effacement de points apparaît répondre à ce besoin. Elle est basée sur les étapes suivantes ([DLS06]) :

1. Simuler $X = \{s_i\}$ un processus HPP avec une intensité λ_{max} tel que :

$$\lambda_{max} = \max_{x, y \in A} \lambda_X(x, y) \quad (3.36)$$

2. On efface indépendamment chaque point s_i avec la probabilité $\left(1 - \frac{\lambda_X(x, y)}{\lambda_{max}}\right)$.

Ainsi, le processus résultant a l'intensité souhaitée. Avec cette méthode, et pour générer une structure agrégée non-stationnaire, il suffit de générer les descendants selon cette procédure, les paramètres λ_p et σ_c étant choisi de la même façon que dans le cas stationnaire.

La figure (3.14) montre une réalisation d'un processus *NHPP* d'intensité $1 + 100(x + y)$ dans le carré unité par la méthode d'effacement de points.

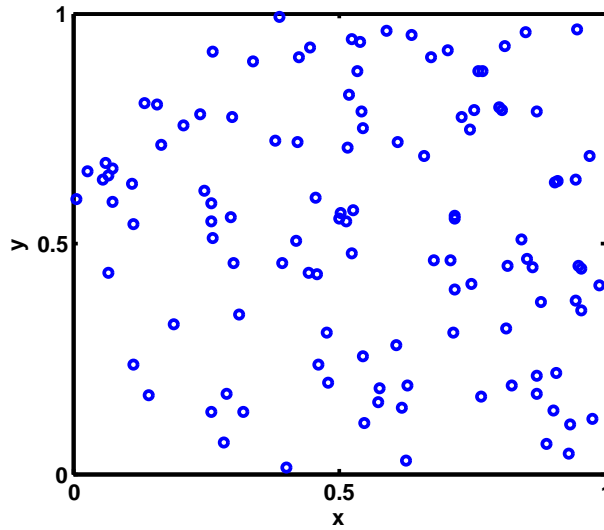


FIG. 3.14: Exemple de réalisation d'un processus *NHPP* d'intensité $1 + 100(x + y)$ dans le carré unité

3.4 Statistique spatiale des données

Comme indiqué au début de ce chapitre, la statistique spatiale des données s'intéresse à mettre en évidence les corrélations entre les objets qui sont associés à des attributs. Les processus marqués présente l'outil mathématique pour étudier ce type de problème et l'objectif de cette partie est d'en présenter les aspects théoriques généraux qui vont être adaptés à notre problème de corrosion par piqûres au chapitre suivant.

3.4.1 Fonction de corrélation des marques K_{mm}

La propriété du second ordre pour un processus marqué met en évidence les différences entre les valeurs des marques par rapports à la distance qui sépare les points. Soit $\{s_i, m_i\}_{i=1, \dots, n}$ un processus marqué. $\{s_i\}$ étant les positions générées par un processus spatial X_s et $\{m_i\}_{i=1, \dots, n}$ représentent les marques qui sont une réalisation d'une variable aléatoire M de densité de probabilité f_M et d'espérance μ_m . Pour les points voisins, on parle d'inhibition (ou compétition) si les valeurs des marques ont tendance à être plus grandes que la moyenne μ_m et on parle de stimulation dans le cas contraire.

Pour caractériser cette propriété du second ordre, l'espérance des produits des marques C_{mm} est nécessaire :

$$C_{mm}(r) = \mathbb{E}[M(s_i)M(s_j), \|s_i - s_j\| = r] \quad (3.37)$$

Sous l'hypothèse d'indépendance des marques, on a $C_{mm}(r) = \mu_m^2$. Une normalisation de cette variable conduit à la définition de la fonction de corrélation des marques K_{mm} introduite par Stoyan ([Sto84]) :

$$K_{mm}(r) = \frac{C_{mm}(r)}{\mu_m^2} \quad (3.38)$$

Ainsi, le cas où $K_{mm}(r) > 1$ est interprété comme un phénomène de stimulation alors que le cas où $K_{mm}(r) < 1$ est interprété comme un phénomène d'inhibition. Dans le cas général, le calcul de la fonction $K_{mm}(r) > 1$ est souvent basé sur l'introduction de fonctions tests $t(m, m')$ ([IPSS08]). Dans ce cas, la définition générale de la fonction produit des marques noté C_t est :

$$C_t(r) = \mathbb{E}[t(M(s_i), M(s_j)), \|s_i - s_j\| = r] \quad (3.39)$$

Dans ce cas, la fonction de corrélation des marques généralisée K_t est donnée par :

$$K_t(r) = \frac{C_t(r)}{\mu_t} \quad (3.40)$$

où μ_t est le coefficient de normalisation défini par :

$$\mu_t = \int_0^{+\infty} \int_0^{+\infty} t(m, m') f_M(m) f_M(m') dm dm' \quad (3.41)$$

Parmi les fonctions tests les plus utilisées, il y a :

- $t_1(m, m') = mm'$.
- $t_2(m, m') = \frac{1}{2}(m - m')^2$ appelée variogramme et souvent utilisée pour les données géostatistiques.
- $t_3(m, m') = (m - \mu_{t_3})(m' - \mu_{t_3})$

Dans ce travail, la fonction test utilisée est t_1 . Avec cette fonction, on retrouve les expressions de C_{mm} et K_{mm} introduite pour caractériser la propriété du deuxième ordre. Chaque choix de fonction test conduit à une interprétation spécifique de la fonction de corrélation des marques. Le choix de t_1 est souvent commode surtout pour les v.a (marques) continues.

3.4.2 Estimation de la Fonction de corrélation des marques

Dans ce travail, l'estimation de la fonction K_{mm} est donnée seulement dans le cas de processus stationnaire et isotrope.

Soit A la région étudiée et A_s la région translatée de A . En d'autres termes, si $A = [0, a] \times [0, b]$ alors $A_s = [s_x, a + s_x] \times [s_y, b + s_y]$.

Une estimation de la fonction $K_{mm}(r)$ est donnée par la formule suivante ([IPSS08]) :

$$\hat{K}_{mm} = \frac{1}{\hat{\lambda}_X^2} \frac{1}{\hat{C}_{mm}} \sum_{\substack{s_i, s_j \in A \\ s_i \neq s_j}}^n \frac{m_i m_j 1(\|s_i - s_j\| \leq r)}{|A_{s_i} \cap A_{s_j}|} \quad (3.42)$$

où $\hat{\lambda}_X^2$ est l'estimation du carré de l'intensité λ_X qui est donné par ([IPSS08]) :

$$\hat{\lambda}_X^2 = \frac{n(n-1)}{|A|^2} \quad (3.43)$$

et C_{mm} est simplement estimé par $\hat{C}_{mm} = \hat{\mu}_m^2$ avec $\hat{\mu}_m$ la moyenne arithmétique des $\{m_i\}_{i=1, \dots, n}$.

Conclusion : Dans ce chapitre, l'analyse et la statistique des données spatiales ont été présentées. Dans le contexte de ce travail, l'utilité d'introduire ces notions réside dans la possibilité de mettre en évidence la dépendance entre les piqûres de corrosion en terme d'interaction et en terme de corrélation entre les profondeurs en fonction de leurs positions. La présence de telles dépendances affaiblit les conditions théoriques nécessaires à l'application de la méthode de Gumbel pour l'estimation de la profondeur maximale. Les processus spatiaux (avec ou sans marques) représentent l'outil mathématique pour étudier ce type de problème. La distinction entre les processus stationnaires et non stationnaires est primordiale. Les premiers sont caractérisés par une intensité constante par translation alors que les seconds ont une intensité qui dépend de la position. En pratique, on ne possède qu'une région étudiée, appelée souvent « fenêtre », extraite d'une structure spatiale. En étudiant cette fenêtre, il est difficile d'affirmer si le processus global est stationnaire ou pas et l'hétérogénéité observée dans cette fenêtre n'est pas synonyme de la présence de dépendance. C'est pour cela que l'hypothèse de stationnarité est souvent émise dans la littérature pour étudier les processus spatiaux. Par contre, si on arrive à montrer la non-stationnarité du processus (en possédant plusieurs réalisations (fenêtres)), il faut adapter les méthodes appliquées au cas stationnaire. Enfin, l'hétérogénéité n'implique pas l'existence d'une dépendance spatiale (cas de NHPP), d'où la nécessité, avant de commencer l'étude, de savoir ou supposer que le processus étudié est stationnaire ou pas.

Chapitre 4

Analyse des données expérimentales et Méthode de Gumbel

Ce chapitre est divisé en deux parties :

1. La première est consacrée à l'expérimentation. Il s'agit de présenter les différentes manipulations réalisées pour obtenir des échantillons ayant subi une corrosion par piqûres. L'exploitation de ces échantillons est faite en terme de mesures de densité de piqûres et de profondeur. Les outils mathématiques présentés au troisième chapitre sont aussi utilisés.
2. La deuxième s'intéresse à la méthode de Gumbel (développée au deuxième chapitre). Un travail de simulation numérique est mis en œuvre dans l'objectif d'élucider certaines questions autour de l'application de cette méthode. Les paramètres de cette simulation numérique sont issus de l'expérimentation.

4.1 Analyse expérimentale des données

L'expérimentation présente une étape de première importance dans ce travail. En effet, elle permet d'alimenter les simulations numériques et confronter les résultats de ces simulations avec celles de l'expérimentation. Pour mener cette comparaison, on a fixé quelques conditions à vérifier et qui peuvent se résumer dans ce qui suit :

1. Piqûres facilement mesurables (absence de produits de corrosion, absence de coalescence, non déviantes, ...).
2. Observation du phénomène de corrosion par piqûres en un temps raisonnable (quelques heures).
3. Expérimentation simple sur échantillons de grandes tailles (essais à potentiel libre à température ambiante).

Trouver le couple qui vérifie ces conditions s'est avéré très délicat. Une recherche bibliographique [BG95] nous a conduit à retenir l'aluminium comme matériau, et le milieu dont les compositions sont : 2,5 g de NaCl, 20 g de FeSO₄, 125 cm³ de H₂SO₄ et 1 L de H₂O.

A noter que les concentrations indiquées ici ont été adaptées selon nos besoins.

Ainsi, les plaques échantillons sont immergées dans différents bacs (2 plaques par bac) contenant cette solution agressive pendant quatre heures. Ces plaques n'ont subi aucun traitement préalable mis à part un nettoyage de la surface à l'acétone. Les échantillons utilisés pour cette expérimentation sont issus de trois tôles d'aluminium de dimension $1 \times 2 \text{ m}^2$. Dans leur ordre chronologique

d'utilisation, la première et la deuxième tôles ont une épaisseur de 0,8 mm, alors que la troisième à une épaisseur de 1 mm. Sous les mêmes conditions d'attaque (présentées dans la suite), le comportement à la corrosion par piqûres évolue significativement. Pour illustrer ces différents résultats, les échantillons de la première tôle sont analysés. Ensuite, une étude est basée sur les courbes de potentiel libre tracées pour comparer le comportement à la corrosion par piqûres de la première et de la deuxième tôle. Enfin, une étude basée sur les processus spatiaux a été mise en oeuvre pour les échantillons de la troisième tôle.

4.1.1 Préparation des échantillons de la première tôle

Pour cette première série d'expériences, l'objectif est d'essayer d'obtenir une reproductibilité de la densité de piqûres des échantillons en les soumettant aux mêmes conditions d'attaque. Pour cela, la tôle a été découpée en trois bandes notées A,B et C de largeur 5 cm et dans le sens de la largeur de la tôle. Chaque bande (de longueur 1 m) étant découpée en 6 plaques de dimension $5 \times 15 \text{ cm}^2$ en éliminant 2,5 cm de chaque côté et en laissant un espace de 1 cm entre les plaques. Le schéma (4.1) illustre cette préparation.

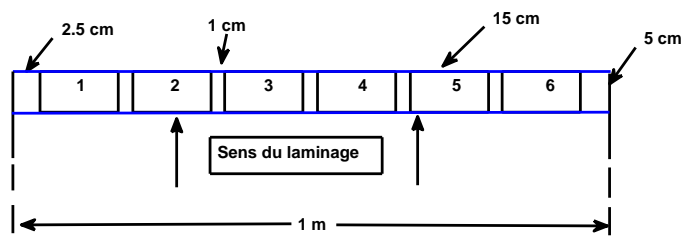


FIGURE 4.1 – Illustration du découpage de la tôle en échantillons

4.1.1.1 Densité de piqûres des échantillons

La densité de piqûres des échantillons se calcule en général via un logiciel d'analyse d'images (ici Image J). Mais pour cette première série expérimentale, cela n'a pas été possible à cause de l'état de surface des échantillons après attaque. En fait, un phénomène de coalescence de piqûres a été observé rendant l'analyse d'images très délicate. La figure (4.2) montre l'état de surface d'une partie extraite de l'une des plaques attaquées.



FIGURE 4.2 – Relevé d'un échantillon d'environ 10 cm^2 sur une plaque montrant l'état de surface et le phénomène de coalescence des piqûres

Par conséquent, un calcul précis de la densité de piqûres n'a pas été possible pour les échantillons issus de la première tôle. Cependant, cela a été possible avec les échantillons issus de la troisième

tôle. Cette densité a été estimée en appliquant un masque (voir figure 4.3) de cinq carrés de 1 cm^2 dans lesquels le nombre de piqûres a été déterminé.

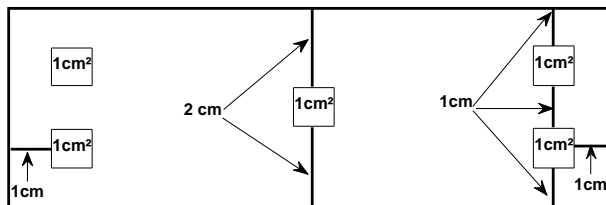


FIG. 4.3: Schéma du masque à cinq carrés de 1 cm^2

Une analyse de variance (voir chapitre 1) a été appliquée pour tester s'il existe une différence significative entre les densités observées. Il s'agit de calculer les densités sur chaque bande en fonction de la position de la plaque, par exemple pour la bande A , on compare les densités de A_1, A_2, \dots, A_6 . Ensuite, on compare les densités des bandes A, B et C pour les positions 1, 2, \dots , 6.

Pour une plaque donnée, on peut considérer que la densité est estimée à partir de cinq réalisations qui sont le nombre de piqûres calculé sur chacun des cinq carrés. Ici, un groupe (voir chapitre 1) est représenté par les cinq valeurs de la densité. Vu la faible taille des groupes, c'est le test de Kruskal-Wallis qui sera privilégié au test ANOVA, car il est difficile de vérifier les conditions de normalité et d'homoscédasticité nécessaires pour le test ANOVA.

La figure (4.4) illustre les boîtes à moustaches avec les résultats du test de Kruskal-Wallis au seuil de $\alpha = 0,05$. Une description de ces boîtes à moustaches est donnée en annexe (A.1).

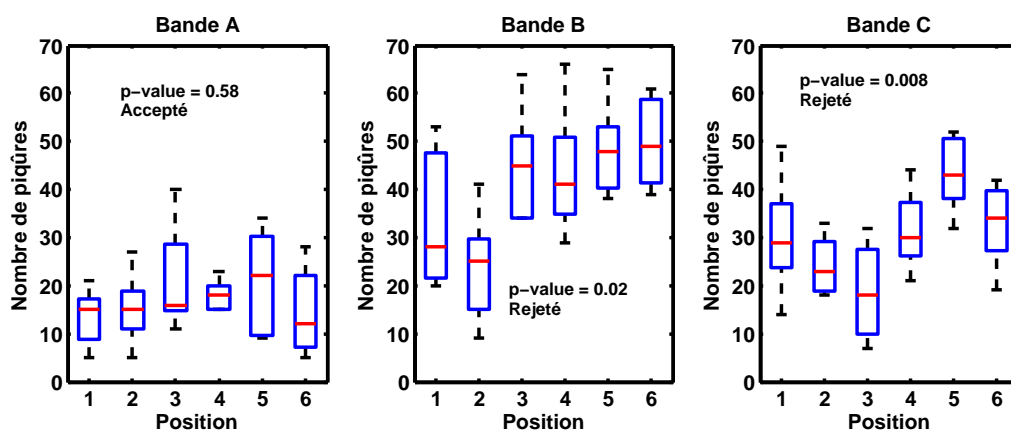


FIG. 4.4: Boîtes à moustaches et résultats du test de Kruskal-Wallis pour la comparaison du nombre de piqûres par carrés sur chacune des bandes A, B et C

Les boîtes à moustaches montrent une certaine dispersion en terme de nombre de piqûres au sein de chaque bande. Le test de Kruskal-Wallis a rejeté l'hypothèse nulle pour les bandes B et C. En d'autres termes, il existe une différence significative (en moyenne) entre les densités de piqûres des plaques issues des bandes B et C. Toutefois, cette conclusion est à prendre avec prudence. En effet, le fait de ne pas pouvoir analyser entièrement les plaques et se restreindre aux masques à cinq carrés,

peut conduire à de mauvaises interprétations des variations de la densité. Cette dernière dépend fortement de la position, du nombre et de la taille des carrés. Il faut noter aussi que cette densité de piqûres est liée aux erreurs expérimentales, aux propriétés du matériau et à la distribution sous-jacente des profondeurs de piqûres. Cette dernière est liée en particulier au caractère stochastique intrinsèque du phénomène de corrosion par piqûres. Quant au test qui porte sur la comparaison du nombre de piqûres selon la position, il est résumé dans la figure (4.5).

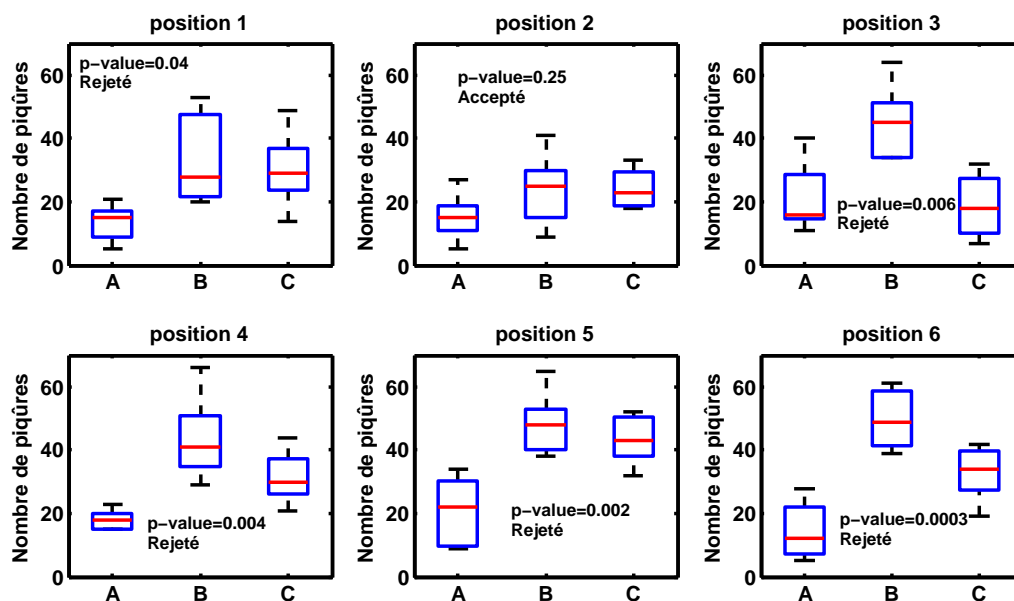


FIG. 4.5: Boîtes à moustaches et résultats du test de Kruskal-Wallis pour la comparaison du nombre de piqûres par carrés sur chacune des positions 1 à 6

Hormis la position 2, le test a rejeté l'hypothèse nulle pour les autres positions. Cela veut dire, qu'il est difficile de retrouver la même densité de piqûres pour une position donnée en parcourant la tôle (passage d'une bande à une autre). Ce resultat peut être dû aux mêmes facteurs évoqués ci-dessus en y ajoutant l'état de surface de la tôle qui peut être légèrement modifié (par exemple, la tôle peut être laminée différemment aux extrémités par rapport à sa partie centrale) ce qui peut avoir un impact direct sur la densité de piqûres.

Cette partie expérimentale montre bien la difficulté d'obtenir, sous les mêmes conditions expérimentales, la même densité de piqûres. Cette forte variabilité, impose une réflexion vis à vis du travail de simulation numérique mis en œuvre plus tard et montre l'intérêt des processus spatiaux étudiés au chapitre précédent qui permettent de générer des densités ajustées aux observations.

4.1.2 Courbes à potentiel libre pour la première et la deuxième tôle

Comme indiqué au début de ce chapitre, les plaques de la deuxième tôle n'ont pas été attaquées. Pour analyser et comparer le comportement à la corrosion par piqûres, les courbes montrant les variations du potentiel libre en fonction du temps ont été tracées au moyen d'un potentiostat-galvanostat reliée à une cellule électrochimique standard à trois électrodes. La figure (4.6) montre le schéma de ce dispositif expérimental.

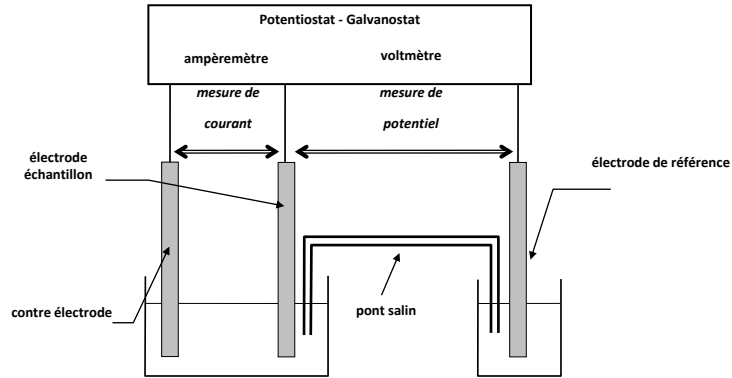


FIG. 4.6: Schéma du dispositif à trois électrodes utilisé pour la réalisation des courbes à potentiel libre

Dans une cellule électrochimique standard, l'électrode échantillon est une pastille de 12 mm de diamètre constitué du matériau à étudier. La mesure du potentiel de la surface de l'électrode échantillon plongé dans le milieu corrosif se fait par rapport au potentiel connu d'une électrode de référence. Dans le cadre de notre étude, une électrode au calomel saturé (couple $\text{Hg}/\text{Hg}_2\text{Cl}_2$) a été utilisé comme électrode de référence. Cette électrode a un potentiel de 0,241 V par rapport à l'électrode normale à hydrogène. La troisième électrode, appelée contre électrode, est en platine et permet de faire circuler un courant au travers de l'échantillon. Si ce courant est positif, les réactions anodiques sont privilégiées à la surface de l'échantillon. Si le courant est négatif, ce sont les réactions de réduction qui sont privilégiées. Quand aucun courant ne circule, les réactions anodiques et cathodiques se compensent en condition de potentiel libre. La figure (4.7) montre le comportement électrochimique de différentes pastilles d'aluminium prélevées dans les deux tôles et plongées en condition de potentiel libre dans la solution corrosive décrite précédemment après avoir subi un nettoyage aux ultrasons dans un bain d'acétone.

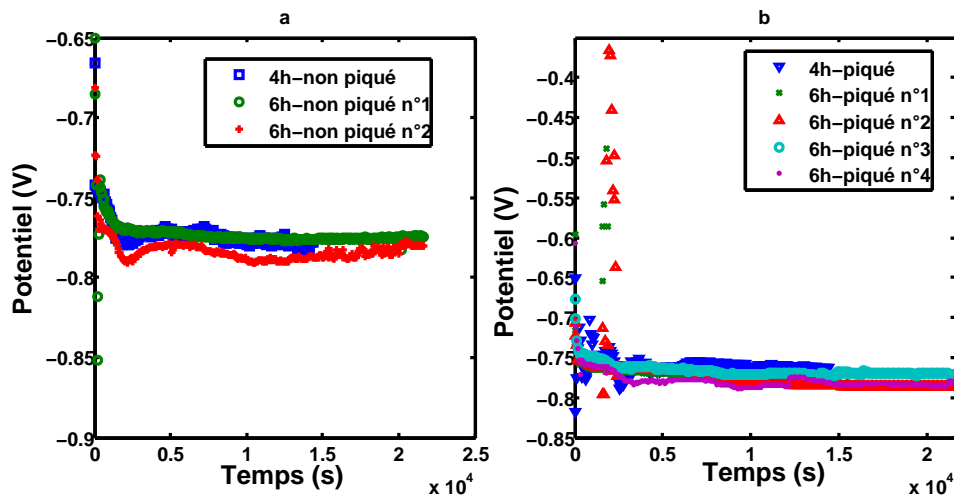


FIG. 4.7: Variations du potentiel en fonction du temps pour les deux tôles : a-plaque non-piquée pour la deuxième tôle, b-plaque piquée pour la première tôle

Pour les plaques issues de la deuxième tôle qui n'ont pas été piquées (figure (4.7)-a), on observe toujours un comportement stable du potentiel pour toutes les pastilles testées (potentiel qui converge vers une valeur de $-0,76\text{ V}$). Pour les plaques issues de la première tôle qui ont été piquées (figure (4.7)-b), on observe soit un comportement stable du potentiel pour certaines pastilles, soit des courbes avec des pics en potentiel pour d'autres après un temps d'environ 1800 s . A signaler que cette valeur (du temps) est proche de celle pour laquelle on a constaté visuellement la formation des piqûres de corrosion sur les plaques. Enfin, cette expérimentation n'a pas permis une explication définitive de la différence de comportement entre les deux tôles, d'autres expériences seront envisagées dans le futur.

4.1.3 Analyse spatiale pour la troisième tôle

Pour cette tôle, une étude par analyse spatiale a été effectuée afin de comparer la distribution spatiale des piqûres. Cette fois-ci, la dimension d'une plaque est de $15 \times 10\text{ cm}^2$. Trois bandes ont été analysées qu'on note bande 1, bande 2 et bande 3, puis à l'aide d'un logiciel d'analyse d'images, les coordonnées de chaque piqûre sont ensuite extraites. La figure (4.8) montre la variation du nombre des piqûres en fonction de la position de la plaque et cela pour chaque bande.

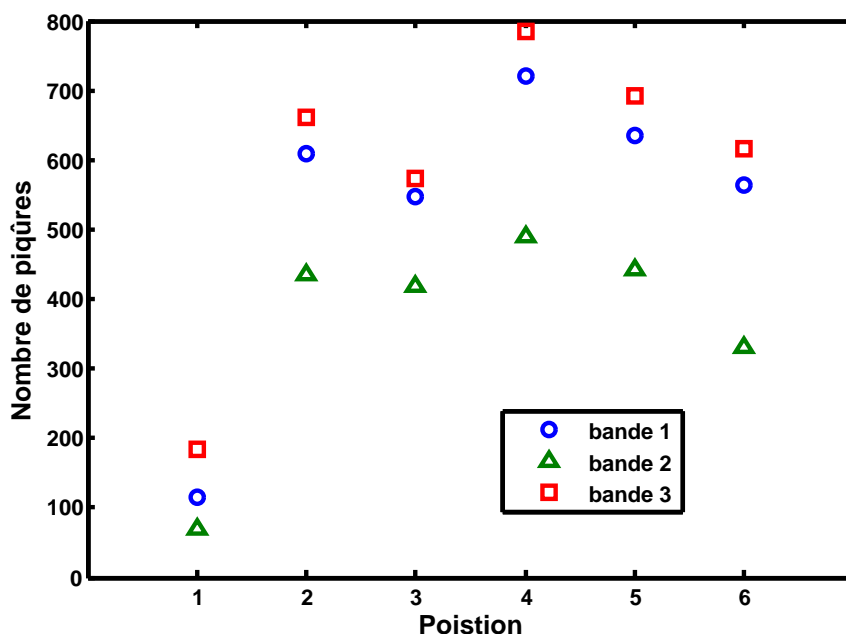


FIG. 4.8: Variations du nombre de piqûres en fonction de la position de la plaque pour les trois bandes

On remarque que le nombre de piqûres suit le même sens de variation pour les trois bandes en fonction de la position. Les bandes 1 et 3 sont proches en terme de nombre de piqûres alors que la deuxième tôle présente une densité de piqûration plus faible. On remarque aussi que les trois plaques de position 4 correspondent au nombre de piqûres le plus élevé. La figure (4.9) montre l'image binaire de la plaque de position 4 issue de la bande 3.

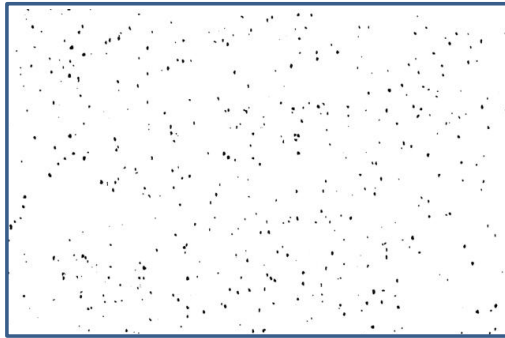


FIG. 4.9: Image binaire de la plaque de position 4 issue de la bande 3

La comparaison de la distribution spatiale des piqûres dans les différentes plaques est effectuée en calculant les deux spectres (voir (3.2.4.3)). Pour illustrer ces résultats on se restreint aux plaques de positions 3 et 4 issues des différentes bandes. Les figures (4.10) et (4.11) montrent le résultat du calcul pour les deux plaques de positions 3 et 4 issues de la bande 1. Pour les autres plaques, ces mêmes graphes sont présentés en annexe.

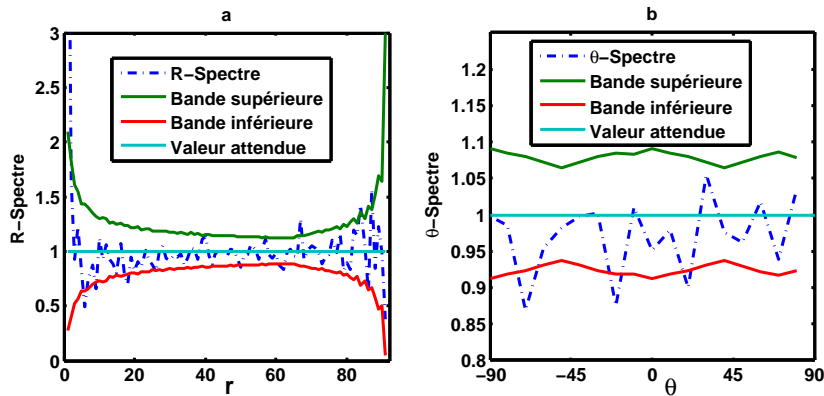


FIG. 4.10: Spectres pour la plaque de position 3 issue de la bande 1 : a- R -Spectre, b- θ -Spectre

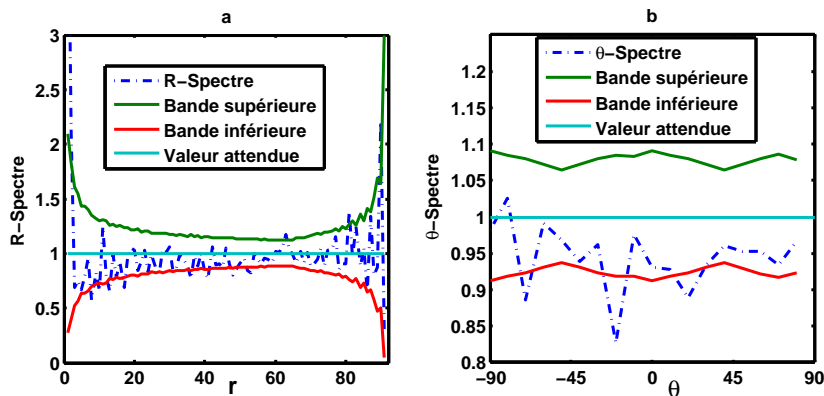


FIG. 4.11: Spectres pour la plaque de position 4 issue de la bande 1 : a- R -Spectre, b- θ -Spectre

A l'aide de l'analyse spatiale des données, on peut distinguer plusieurs structures spatiales des piqûres. Pour les plaques de positions 3 et 4 issues de la bande 1 (figures (4.10) et (4.11)), on

remarque une certaine similitude entre les deux R – *spectres* pour la plupart des échelles explorées (valeurs de $r \geq 2$) mais avec des orientations différentes en se référant aux θ – *Spectre*. Les mêmes conclusions peuvent être tirées Pour les autres plaques (voir annexe A). La différence en terme de nombre de piqûres et d'orientation confirme la difficulté d'obtenir une reproductibilité de la structure spatiale des piqûres.

L'intérêt de l'utilisation des processus spatiaux réside dans la possibilité de révéler certains détails de la structure spatiale telle que l'indépendance entre les piqûres ou encore la présence de directions privilégiées. Trouver l'échelle pour laquelle les piqûres peuvent être considérées comme indépendantes peut constituer une information intéressante dans le sens où cette échelle peut être considérée comme la taille minimale des blocs d'analyses utilisés dans la méthode de Gumbel (voir section (4.2)).

4.1.4 Mesure des profondeurs

Pour mesurer les profondeurs des piqûres, deux instruments ont été utilisés, à savoir, le microscope optique et le profilomètre optique. Leur description est résumée dans ce qui suit :

- **Microscope optique** : Il mesure la différence de hauteur qui correspond à une image nette (surface de la piqûre analysée correspondant à la surface de la tôle) et une image floue (fond de la piqûre). Il a l'avantage d'être simple à utiliser et permet des mesures rapides des profondeurs. Le grossissement utilisé dans ce travail était de $\times 50$.
- **Profilomètre optique** : Le capteur optique de l'appareil permet de mesurer l'altitude d'un point de l'échantillon. Le capteur, qui est ici une caméra, permet de mesurer un ensemble de points sur un échantillon surfacique correspondant à la zone observée. Contrairement à un profilomètre tactile où l'acquisition se fait point par point, on mesure ici l'ensemble des points, correspondants aux pixels caméra 640x480, quasi-simultanément. Le profilomètre optique permet aussi l'acquisition de profils et le calcul des paramètres standard de rugosité. Il permettra dans ce travail de mesurer les profondeurs et d'analyser la forme des piqûres.

La figure (4.12) montre le profilomètre utilisé dans ce travail. Le modèle est un VEECO WYKO NT9300.

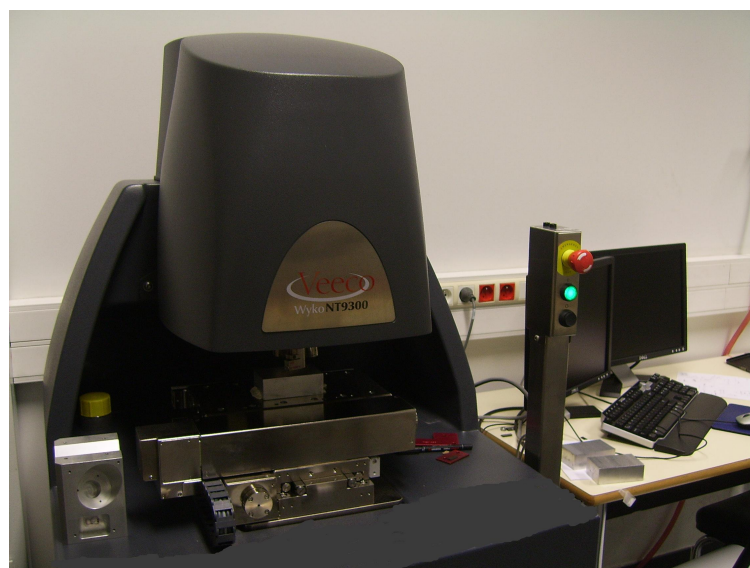


FIG. 4.12: Profilomètre optique VEECO WYKO NT9300

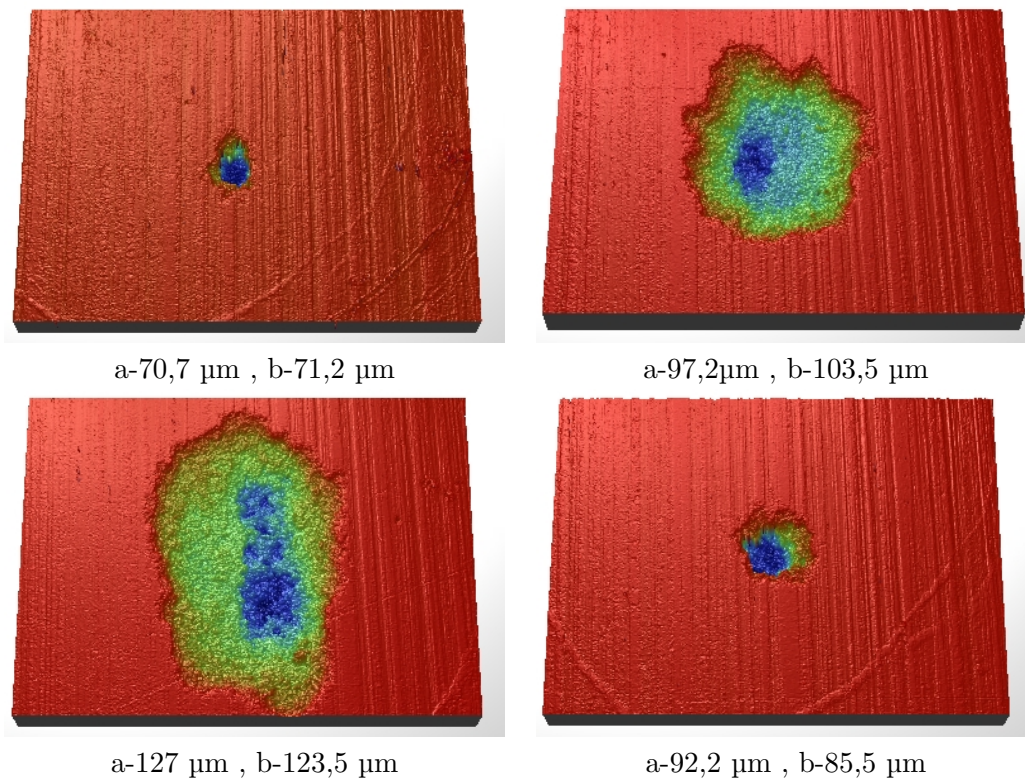


FIG. 4.13: Exemples de morphologie de piqûres (issues de la tôle 1) caractérisée à l'aide du profilomètre optique VEECO avec leurs mesures de profondeurs : a-mesure par profilomètre, b-mesure par microscope

L'objectif de cette section est de comparer la précision de ces deux instruments. Pour cela, plusieurs mesures ont été effectuées. L'histogramme suivant (fig. (4.14)), montre les différentes mesures (sur des plaques issues de la tôle 1) enregistrées sur les deux appareils sur plusieurs piqûres plus ou moins profondes.

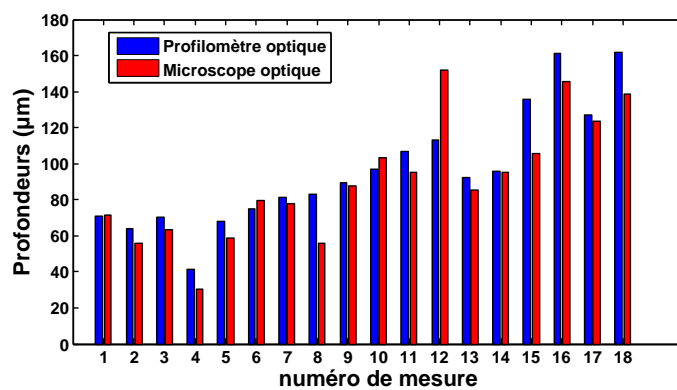


FIG. 4.14: Histogramme des profondeurs mesurées par les deux appareils

Pour tester s'il y a une différence significative entre les deux mesures, un test ANOVA a été mis en œuvre. Mais avant de procéder à ce test, la vérification des conditions de son application est nécessaire, à savoir, les tests de normalité et d'égalité des variances des résidus.

Par rapport à la normalité, un quantile-quantile plot a été tracé pour visualiser l'adéquation des résidus à la loi normale. La figure (4.15) montre les résultats obtenus pour chaque appareil.

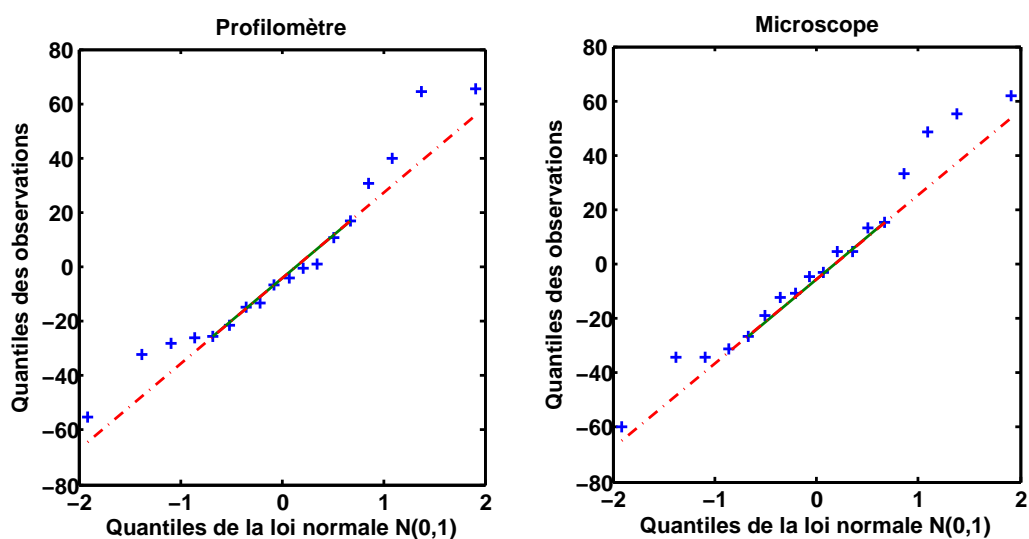


FIG. 4.15: Quantile-quantile plot pour les résidus des mesures issues des deux appareils

Graphiquement, les résidus des deux mesures semblent suivre une loi normale dans la partie centrale des observations. Cependant un écart apparaît aux extrémités. C'est pour cette raison que le test de Kolmogorov a été effectué pour confirmer l'hypothèse de normalité des résidus au seuil $\alpha = 0,05$. En effet, la p-valeur est égale à 0,96 pour le microscope optique, et pour le profilomètre optique, elle est égale à 0,71.

Pour l'égalité des variances, le test de Bartlett a été appliqué, et l'hypothèse d'égalité n'a pas été rejetée au seuil $\alpha = 0,05$. La p-valeur du test étant égale à 0,95.

Avec les conclusions de ces deux tests, le test ANOVA peut être effectué. La figure (4.16) montre les boîtes à moustaches des deux mesures. La p-valeur du test ANOVA est de 0,58 et l'hypothèse d'égalité des moyennes ne peut être alors rejetée.

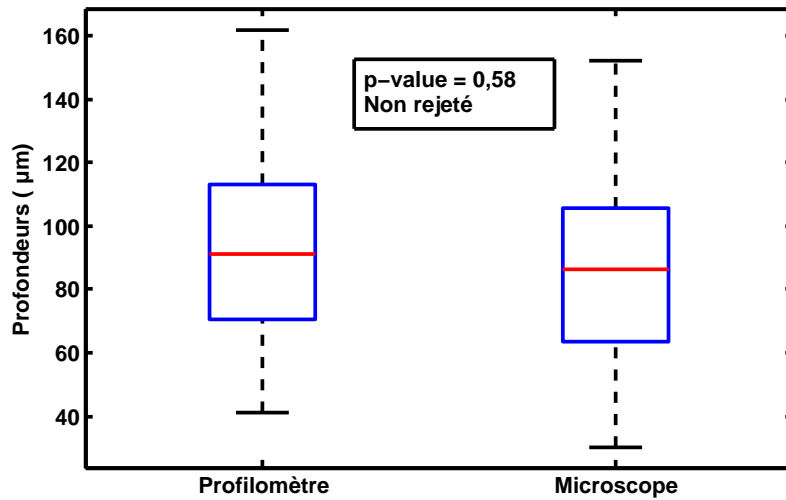


FIG. 4.16: Boîtes à moustaches et résultat du test ANOVA pour les deux mesures

Cette étude statistique a permis de conclure qu'on ne perdra pas (d'une façon significative) de précision sur les mesures de profondeurs des piqûres en utilisant un simple microscope optique. **En comparaison avec le profilomètre optique, cet appareil permettra d'effectuer rapidement de nombreuses mesures de profondeurs.**

4.1.5 Etude complète sur une plaque de 75 cm²

L'étude complète consiste à mesurer la profondeur de chaque piqûre à la surface de la plaque d'aluminium de 75 cm² et à en connaître ses coordonnées. Cette étude permet d'introduire les processus spatiaux simples et marqués, objet de ce paragraphe, et de comparer les différentes méthodes d'estimation de la profondeur maximale avec l'avantage de connaître la vraie valeur maximale. Ce dernier point sera détaillé dans la section suivante (4.2).

La plaque de 75 cm² est issue de la tôle 1. Elle a subi un traitement thermique de recuit ¹ pendant 1 h à 300 °C avant d'être immergée dans la même solution agressive que celle décrite auparavant. La figure (4.17) montre cette plaque d'aluminium.

¹Le recuit a permis d'obtenir un état de surface de qualité meilleure facilitant le traitement par analyse d'image

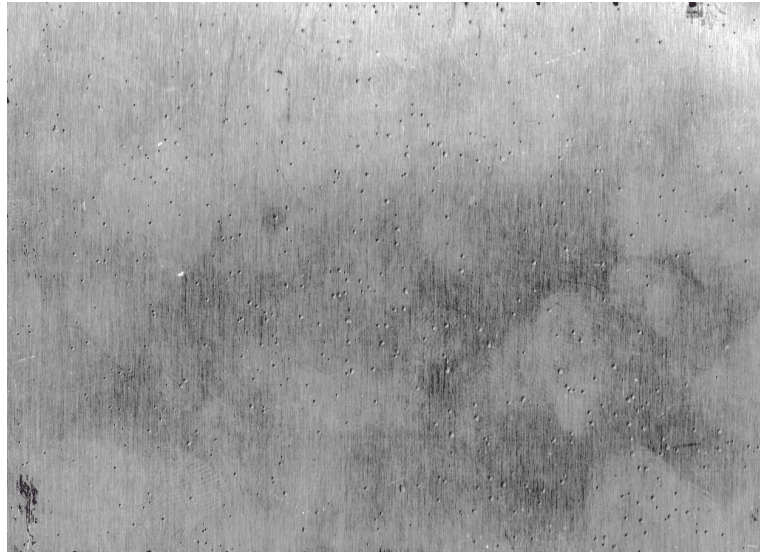


FIG. 4.17: Plaque de 75 cm²

Les coordonnées des piqûres sont extraites à l'aide du logiciel d'analyse d'images. Un traitement numérique consistant à isoler les contours des piqûres est nécessaire, ensuite, ce sont les centres de masse de ces piqûres qui sont identifiés aux coordonnées. La figure (4.18) montre la répartition spatiale de ces centres de masse.

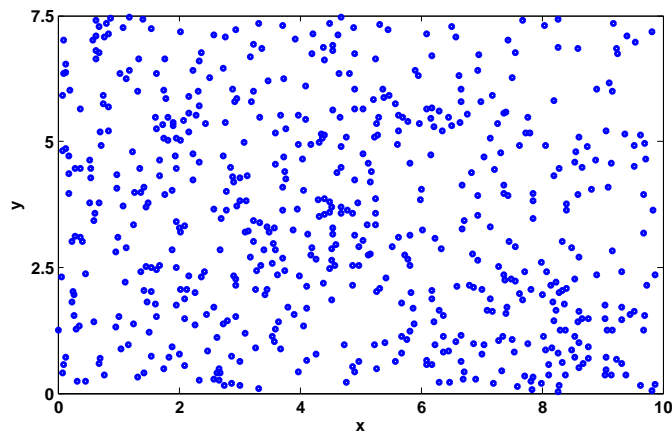


FIG. 4.18: Répartition spatiale des centres de masse des piqûres

La distribution des 603 valeurs de profondeur mesurées avec le microscope optique est illustrée par l'histogramme de la figure (4.19). La valeur minimale est de 70 μm et la valeur maximale de 163 μm . Enfin, la figure (4.20) montre la répartition des profondeurs en fonction des positions des piqûres.

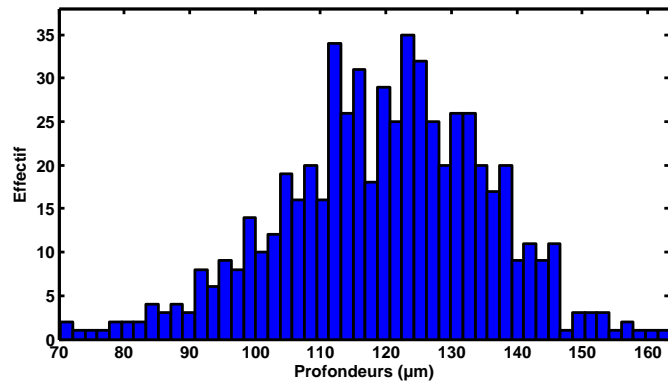


FIG. 4.19: Histogramme des 603 mesures de profondeur de piqûre

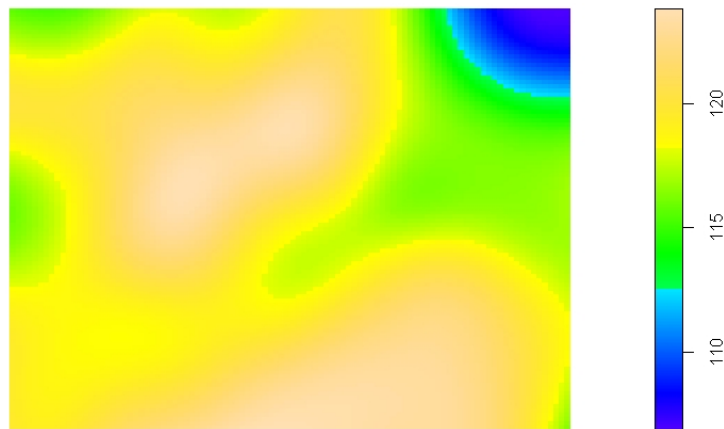


FIG. 4.20: Carte des profondeurs (en μm) en fonction des positions des piqûres

Les trois tests classiques présentés au premier chapitre, à savoir les tests du Chi-deux, de Kolmogorov et d'Anderson-Darling ont tous rejeté l'hypothèse que les 603 profondeurs peuvent être issues d'une loi normale, Weibull et log-normale.

La fonction de corrélation des marques, est illustrée dans la figure (4.21).

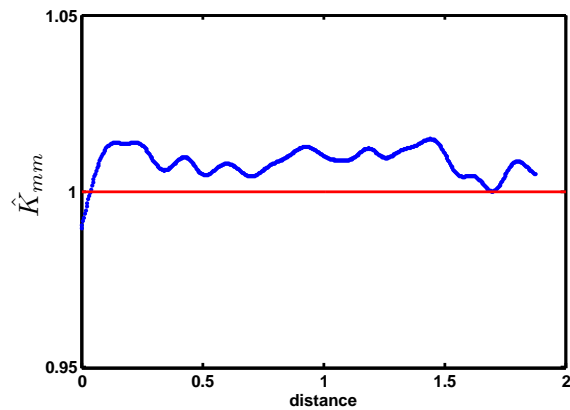


FIG. 4.21: Fonction de corrélation des marques pour la plaque de 75 cm^2

La fonction de corrélation des marques est très proche de 1 et le rapport entre la valeur maximale et la valeur minimale est inférieure à 1%. L'hypothèse d'indépendance entre les profondeurs et leurs positions peut alors être retenue dans ces conditions.

4.2 Méthode de Gumbel

Dans cette partie, l'objectif est de répondre à certaines questions autour de l'application de la méthode de Gumbel présentée au deuxième chapitre. Les étapes d'application de cette méthode se résument dans ce qui suit :

- Décomposer la surface étudiée en n blocs d'analyses de même taille et d'en choisir aléatoirement m . Sur chacun de ces m blocs, la valeur de la profondeur maximale est retenue.
- Les m valeurs maximales retenues sont utilisées pour estimer les paramètres de la loi de Gumbel par une des méthodes d'estimation.

La figure (4.22) illustre les étapes de cette méthode.

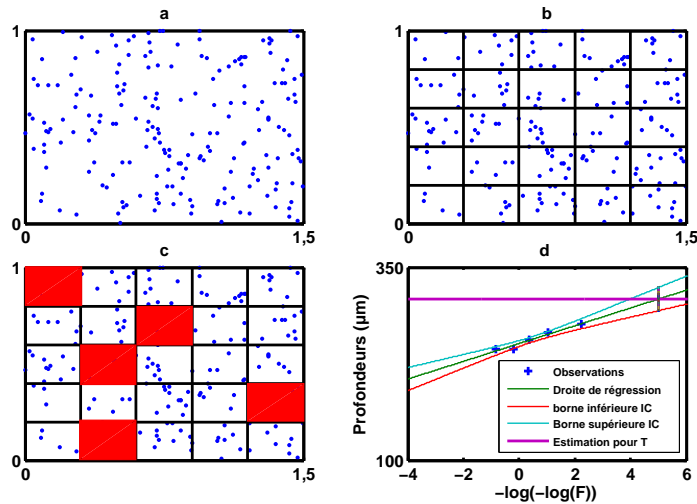


FIG. 4.22: Illustration des étapes de la méthode de Gumbel avec comme méthode d'estimation des paramètres la régression linéaire : a- Surface étudiée, b-Découpage en n blocs, c-Choix aléatoire de m blocs parmi n ($m=5$), d- Estimation pour T souhaitée avec intervalle de confiance (IC)

Pour cette méthode, les questions qui se posent portent sur :

1. La taille des blocs d'analyse et leur nombre nécessaire pour une meilleure estimation de la profondeur maximale pour la période de retour souhaitée.
2. La meilleure méthode d'estimation des paramètres de localisation et d'échelle de la loi de Gumbel en terme de biais et de variance.
3. Au cas où la méthode d'estimation des paramètres est la régression linéaire, quelle est la meilleure approximation de la distribution empirique ?

Le traitement de la troisième question donne des indications quant à l'impact de la taille des blocs sur l'estimation de la profondeur maximale. Ainsi la troisième question sera traitée en premier, ce qui conduira à tester la pertinence de plusieurs approximations de la fonction de la distribution empirique. Ensuite, une discussion autour de la taille des blocs d'analyses et leur nombre est ouverte dans le but de proposer un critère de découpage. Enfin, une comparaison entre les différentes méthodes d'estimation des paramètres proposées au deuxième chapitre (voir 2.2.4.1) est mise en œuvre.

Pour répondre à ces questions, la simulation numérique constitue un outil très important. Pour s'approcher des données réelles, les paramètres des simulations numériques effectuées dans ce travail sont issus de l'expérimentation. Ces paramètres sont essentiellement la loi des profondeurs et la densité spatiale des piqûres.

4.2.1 Paramètres des simulations numériques

4.2.1.1 Loi des profondeurs

L'application de la méthode de Gumbel suppose que la loi sous-jacente des profondeurs appartient au $MDA(Gumbel)$. Deux lois appartenant à ce domaine d'attraction seront testées dans ce travail : La loi normale et la loi log-normale.

La plaque de 75 cm² étudiée au paragraphe (4.1.5) constitue la base de calcul des paramètres des simulations numériques. La nécessité de faire des simulations et répondre aux questions posées a

conduit à considérer les distributions les plus courantes. Pour fixer les paramètres de façon réaliste, on les estime à partir de tout l'échantillon.

Le tableau (4.1) résume l'estimation des paramètres de ces deux lois.

Loi de probabilité	paramètres	Estimation (μm)	95% IC (μm)
normale	Localisation	119,6	[118, 4; 120, 9]
	Echelle	15,8	[14, 9; 16, 7]
log-normale	Localisation	4,77	[4, 76; 4, 78]
	Echelle	0,13	[0, 12; 0, 14]

TAB. 4.1: Estimation et indice de confiance à 95% des paramètres de localisation et d'échelle pour les lois normale et log-normale

4.2.1.2 Densité de piqûres

Le paramètre de la simulation numérique lié à la densité de piqûration pose beaucoup plus de difficultés que celui des profondeurs. En effet, différentes densités ont été observées et l'ajustement via les techniques développées au troisième chapitre par le calcul de l'intensité spatiale $\lambda_X(x, y)$ ne sera valable que pour la surface étudiée, ainsi l'extrapolation pour une surface plus grande n'aura aucun sens. A rappeler que l'objectif de la simulation numérique est de tester la pertinence de la méthode de Gumbel en simulant une grande surface S avec une densité de piqûres et des valeurs de profondeurs proches de ce qu'on peut observer en pratique.

La difficulté réside aussi dans le fait que ce problème n'est pas assez discuté dans la littérature. Hormis le rappel de l'aspect stochastique de l'amorçage des piqûres et la distinction entre les trois types de structures (CSR, agrégée, régulière), on ne trouve pas de lois de probabilités couramment utilisées pour la densité spatiale. L'idée adoptée alors dans ce travail, est de simuler une surface selon un processus CSR, avec une densité de piqûres égale au nombre de piqûres observées divisé par la taille de la surface. Pour la plaque de 75 cm^2 , cela donne une densité de $603/75 \approx 8$ piqûres par cm^2 .

L'intérêt de la simulation d'une structure CSR réside dans le fait qu'elle permet de se placer dans les conditions théoriques de l'application de la méthode de Gumbel. A savoir, l'indépendance entre les piqûres (acquise par définition d'une CSR), et la possibilité d'avoir le même nombre de piqûres par bloc lors du découpage. Quant à l'homogénéité en terme de profondeur, c'est la même loi de probabilité qui est utilisée pour générer toutes les piqûres. Ainsi, avec cette méthodologie, on peut comparer les différentes méthodes d'estimation sans se préoccuper des conditions d'application de la méthode de Gumbel. Enfin, la conclusion qui concerne l'indépendance entre les profondeurs et les positions lors du calcul de la fonction de corrélation des marques (voir 4.1.5) est respectée dans les simulations numériques.

4.2.2 Méthode de Gumbel et régression linéaire

Avant de traiter le problème du choix de l'approximation empirique, une discussion autour de l'application de la méthode de Gumbel est présentée. Cela concerne le choix des axes de la régression linéaire et l'utilisation des quantiles comme valeur pour donner des estimations des profondeurs maximales.

4.2.2.1 Choix des axes pour la régression linéaire

L'estimation des paramètres de la loi de Gumbel à l'aide de la régression linéaire peut se faire de deux façons. Cette différence concerne le choix des axes de la régression, ou en d'autres termes, le choix de la variable à expliquer et la variable explicative. Rappelons que le modèle de la régression linéaire s'écrit sous la forme $y = ax + b + \epsilon$, où y représente la variable à expliquer, x la variable explicative et ϵ le résidu qui suit une loi normale de moyenne nulle.

La distinction entre les deux variables x et y n'est pas toujours facile et dépend souvent de ce qu'on souhaite faire ([DS98]). En général, la variable explicative est supposée déterministe, alors que la variable à expliquer est supposée aléatoire ([CL07]). Dans la littérature ([Shi96, Shi91]), cette méthode est appliquée avec un choix de x assimilable aux profondeurs observées et y égale à $-\log(-\log(1 - \frac{1}{T}))$, alors que dans ce travail (voir 2.2.4.1), c'est le choix inverse qui a été fait. Pour la présentation des résultats, nous appelons « axes classiques », l'approche présentée dans la littérature, et « axes inversés » l'approche présentée dans ce travail.

Avant d'évoquer les raisons de ce choix, revenons sur la notion de période de retour en désignant par S la surface objective pour laquelle on souhaite connaître une estimation de la profondeur maximale, et par s la surface échantillon (ou bloc d'analyse). Soient F l'une des trois distributions des valeurs extrêmes et Y la v.a des profondeurs. Si on note N_e la v.a qui correspond au nombre d'échantillons (blocs) nécessaires pour dépasser pour la première fois la profondeur z , alors :

$$\mathbb{P}(N_e = k) = (1 - F(y)) F(y)^{k-1} \quad (4.1)$$

En d'autres termes, $\mathbb{P}(N_e = k)$ correspond à la probabilité que les profondeurs des piqûres restent inférieures à la valeur z après l'observation de k échantillons. N_e est une v.a suivant la loi géométrique de paramètre $1 - F(y)$, ainsi l'espérance de la variable aléatoire N_e est égale à l'inverse du paramètre, et on écrit :

$$\mathbb{E}[N_e] = \frac{1}{1 - F(y)} = T(y) \quad (4.2)$$

Ainsi, la période de retour définie au deuxième chapitre par l'équation (2.32) correspond au nombre moyen de blocs nécessaires pour dépasser pour la première fois la valeur y . Par conséquent, on retrouve la définition de cette notion comme étant le rapport entre la surface objective et le bloc d'analyse s et on écrit $T = S/s$. Enfin, il faut rappeler que cette notion de période de retour suppose une homogénéité du point de vue environnement dans lequel la surface objective est exposée, et une homogénéité spatiale quant à la distribution des piqûres ([Kow94]).

En tenant compte de la définition de la période de retour et l'aspect théorique de la régression linéaire, le choix des axes dans les deux approches peut se présenter de la façon suivante :

- **Axes classiques** : La variable des profondeurs est la variable explicative. Dans ce cas l'extrapolation se fait par rapport à la surface objective. En d'autres termes, pour une profondeur donnée, quelle est la surface pour laquelle cette profondeur sera dépassée pour la première fois? (déterminée à partir de la connaissance de la droite de régression).
- **Axes inversés** : La variable explicative s'exprime en fonction de la période de retour (plus exactement $-\log(-\log(1 - \frac{1}{T}))$), dans ce cas l'extrapolation se fait en terme de profondeur, plus précisément, pour une surface objective S , on estime la profondeur maximale des piqûres.

La deuxième approche apparaît la plus adaptée au problème de l'extrapolation, en effet on fixe la période de retour et on estime la profondeur correspondante. Dans la littérature, malgré l'utilisation des axes classiques, la période de retour est fixée, puis la profondeur correspondante est déterminée à partir de la connaissance de la droite de régression ([Shi96, Shi91]). En outre, pour tester la qualité de l'ajustement, il est nécessaire de vérifier la normalité des résidus. Un résidu est défini comme étant la différence entre la valeur observée et la valeur ajustée, cette vérification est plutôt adaptée aux résidus concernant les profondeurs que ceux concernant les périodes de retour.

4.2.2.2 Utilisation des quantiles

L'intérêt de l'application de la théorie des valeurs extrêmes en corrosion par piqûres est d'estimer la profondeur maximale. En pratique ce sont les quantiles extrêmes qu'on estime. Ainsi, une profondeur critique dépassée pour une surface objective sera toujours associée à une probabilité et cela est dû à la définition même de la période de retour. Néanmoins, on peut s'intéresser à l'espérance des valeurs maximales en admettant l'approximation que la loi du maximum sur chaque bloc de taille s suit une loi des valeurs extrêmes. Ainsi, si on désigne par M_T le maximum pour une période de retour T , alors l'espérance de cette variable est donnée par :

$$\mathbb{E}[M_T] = \int_{-\infty}^{+\infty} y (G_0^T(y))' dy = \gamma_* \sigma_0 + \mu_0 + \sigma_0 \log(T) \quad (4.3)$$

où $(G_0(y))'$ est la dérivée de la distribution de Gumbel définie par l'équation (2.29).

4.2.3 Choix de l'approximation empirique

Cette approximation est utilisée seulement lorsque la méthode d'estimation est la régression linéaire. L'objectif de cette section est de comparer l'approximation de Weibull, utilisée dans les premiers travaux de Gumbel [Gum58], avec d'autres approximations afin de choisir la mieux adaptée à la méthode de Gumbel. Les distributions empiriques testées dans ce travail sont résumées dans le tableau (4.2).

Approximation	Référence
$\frac{i}{m+1}$	Weibull ([Wei39])
$\frac{i-0.5}{m}$	Hazen ([Haz14])
$\frac{i-0.44}{m+0.12}$	Gringorten ([Gri63])
$\frac{i-0.4}{m+0.2}$	Cunnane ([Cun78])

TAB. 4.2: Expression des quatre distributions empiriques utilisées

La simulation numérique consiste à implémenter la méthode de Gumbel en utilisant les quatre approximations empiriques. Il s'agit de simuler une surface de $100 \times 100 \text{ cm}^2$. Ainsi 80000 piqûres seront générées selon une structure CSR et avec des profondeurs suivant une loi normale et lognormale (voir tableau (4.1)). La surface S est ensuite découpée en n blocs de même taille, n variant entre 100 et 5000 ce qui correspond à un calcul de quantiles avec des probabilités de dépassement variant entre $1/100$ et $1/5000$. Pour un découpage donné en n blocs, 10, 20, \dots , 50 blocs ont été choisis aléatoirement parmi les n blocs qui couvrent S , et pour lesquels on extrait la profondeur maximale. On présentera seulement les résultats pour le choix 10 blocs. Pour les autres choix, les conclusions restent les mêmes et elle sont présentées en annexe (pour la loi normale). La procédure de calcul est répétée 1000 fois. En d'autres termes, 1000 surfaces vont être simulées conduisant

à 1000 estimations des quantiles pour les différentes périodes de retour. On signale que pour les découpages fins, certains blocs ne contiennent aucune piqûres. Ainsi, si l'un de ces blocs est choisi pour en extraire la valeur maximale, un nouveau tirage (choix aléatoire des m blocs) est fait. Les figures (4.23) (a) et (b) illustrent respectivement pour les moyennes et les écarts-types des quantiles pour l'approche des axes inversés.

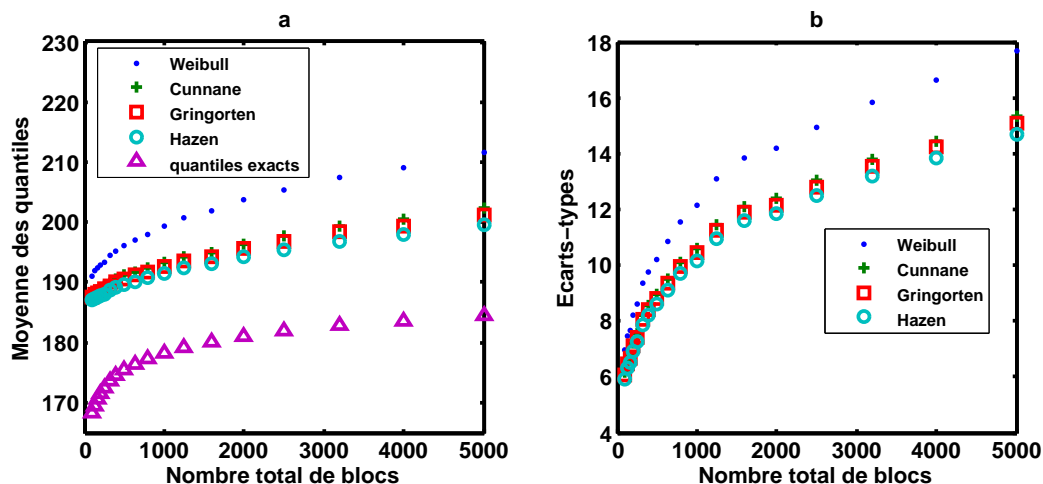


FIG. 4.23: Moyennes des 1000 quantiles estimés à partir des profondeurs générées selon une loi normale avec $m = 10$ blocs : a-Moyenne des quantiles, b-Ecarts-types des quantiles

Pour les découpages présentés, c'est l'approximation de Weibull qui s'écarte le plus des trois autres approximations et de la valeur exacte du quantile. Pour les deux figures (a) et (b), l'écart par rapport à la valeur exacte est autant plus important que le nombre de total de blocs qui couvre la surface objective S est grand, ce qui correspond à des blocs d'analyses de petites tailles. En prenant 10 blocs pour chaque découpage, l'allure des courbes s'explique par le fait qu'en travaillant avec des grands blocs d'analyses, la surface totale analysée ($10 \times$ la taille du bloc) est plus importante, ce qui augmente ainsi la probabilité de récupérer des valeurs maximales contenant plus d'informations sur la queue de distribution ce qui conduit à de meilleures estimations. En prenant un nombre de blocs plus important, comme c'est illustré dans les figures (A.6) et (A.7) présentées en annexe A avec respectivement 30 et 50 blocs, l'amplitude des quantiles diminue réduisant ainsi le biais des estimations mais gardant le même comportement pour chaque approximation empirique.

S'intéressant au choix de la meilleure approximation de la distribution empirique, celle de Weibull semble la moins adaptée. Pour analyser cette différence, le test d'analyse de variance a été appliqué sur les différentes estimations et cela pour chaque découpage. Ce test permet de conclure qu'il existe une différence significative entre les quatre approximations. Pour les trois autres, elles renvoient des estimations très proches les unes des autres.

S'agissant de la comparaison entre les quantiles estimés à partir des axes classiques et inversés, le test ANOVA permet de conclure à l'existence d'une différence significative entre ces deux approximations et cela pour toute valeur de découpage. Quant à la figure (4.24), elle montre les

variations des estimations de Hazen et de Gringorten ainsi que leurs coefficients de variation noté CV pour les deux choix des axes. On s'est limité à ces deux approximations car ce sont celles qui s'écartent le moins de la valeur exacte. Rappelons que le CV est défini comme étant le rapport entre l'écart-type et la moyenne. Notons aussi que le choix du CV comme indicateur de dispersion est préféré à l'écart-type lorsque les séries comparées n'ont pas les mêmes unités ou lorsqu'il existe une différence significative entre les moyennes (ou d'amplitudes) entre ces séries ; ce qui est le cas ici.

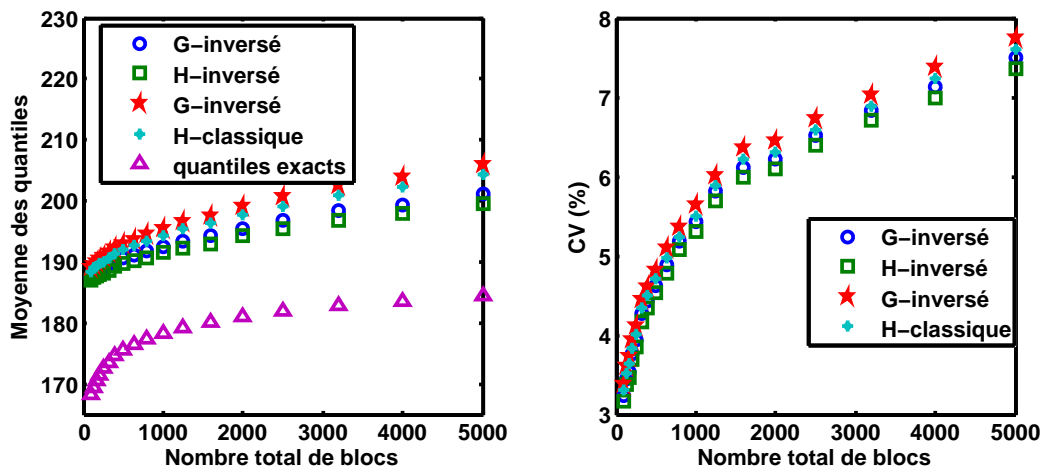


FIG. 4.24: Moyenne des quantiles et coefficients de variations (CV) pour les approximations de Hazen et Gringorten : a-moyenne des quantiles, b- Coefficient de variation

Les quantiles estimés à partir des deux approches d'axes classiques et axes inversés restent proches en terme de variance. On constate la même chose pour un nombre de bloc plus grand. Néanmoins, les estimations avec l'approche des axes inversés présentent un biais moins important. Cela pourra avoir son importance dans le choix de l'approche si l'on souhaite obtenir des estimations plus précises ou plus larges (l'écart maximal entre les deux approches étant de $7 \mu\text{m}$ en prenant $m = 10$ blocs et de $4 \mu\text{m}$ en prenant $m = 50$ blocs).

Pour l'utilisation des espérances du maximum comme estimation (voir eq.(4.3)), la figure (4.25) montre les variations de cette variable en fonction du découpage.

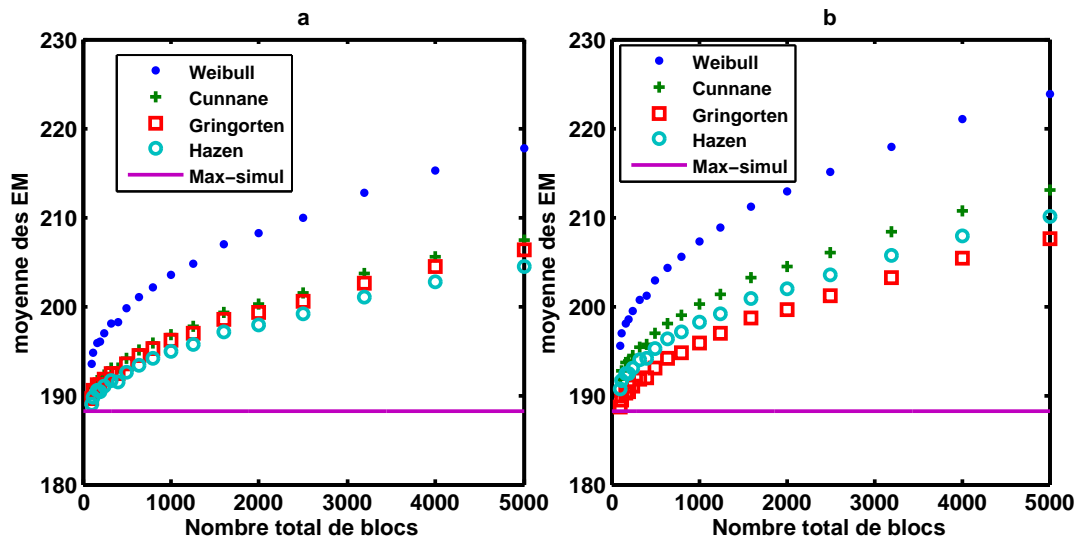


FIG. 4.25: Moyennes des 1000 espérances du maximum (notées EM) estimées à partir des profondeurs générées selon une loi normale : a-axes inversés, b-axes classiques

On remarque aussi que c'est l'approximation de Weibull qui s'écarte le plus de la valeur maximale simulée et cela pour les deux choix des axes. Une étude simulatoire à l'image de ce qui a été présenté précédemment, a conduit aux mêmes conclusions. En plus l'analyse de variance concernant la comparaison des estimations données par les quantiles et les espérances du maximum a aussi permis de conclure à l'existence d'une différence significative.

On note que les objectifs des estimations en utilisant les quantiles et les espérances du maximum sont différents. En effet, pour la première, on estime les quantiles extrêmes qui sont dépassés avec une faible probabilité, alors que pour la deuxième, on estime la moyenne des profondeurs maximales. Dans la suite, on continue le travail en utilisant les quantiles car c'est l'approche la utilisée dans la littérature et la plus développée du point de vue théorique. Néanmoins, l'utilisation des espérances du maximum mérite tout notre intérêt dans le sens où elle donne une estimation directe de la profondeur maximale contrairement aux quantiles qui sont associés à des probabilités de dépassement.

Dans toutes les simulations précédentes, le nombre de piqûres par bloc suit une loi de Poisson $\mathcal{P}(\lambda s)$ avec $\lambda = 8$ représentant le nombre de piqûres par cm^2 et s la taille du bloc d'analyse. Cela conduit à un écart type du nombre de piqûres égal à $\sqrt{\lambda s}$. Sachant que l'application de la méthode de Gumbel nécessite l'égalité du nombre de piqûres dans les blocs d'analyse, nous proposons de procéder aux simulations précédentes mais avec des blocs contenant le même nombre de piqûres et comparer les résultats avec ceux de l'approche précédente. On appellera dans la suite « bloc-égal » les blocs d'analyse pour ces nouvelles simulations, et « bloc-différent » pour ceux étudiés avec les précédentes. La figure (4.26) montre les résultats de ces simulations dans l'approche des axes classiques et inversés et cela en prenant 10 blocs pour l'estimation des paramètres.

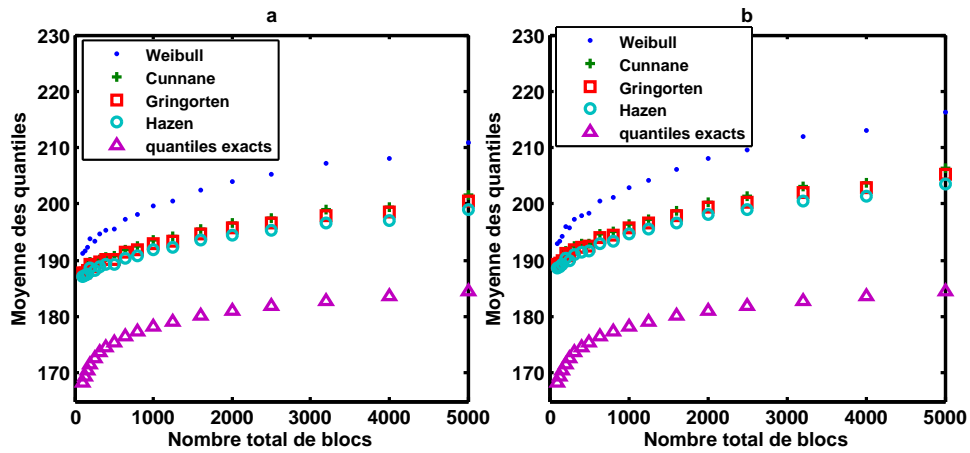


FIG. 4.26: Moyenne des quantiles dans le cas de « bloc-égal » : a- axes inversés, b-axes classiques

Une analyse de variance similaire a conduit aux mêmes conclusions ; à savoir l'existence d'une différence significative entre les estimations calculées à partir des quatre approximations empiriques.

Enfin, pour comparer les estimations calculées avec les deux simulations « bloc-égal » et « bloc-différent », les figures (4.27 a et b) montrent les variations de la p-valeur pour les différents découpages et cela pour l'approximation de Hazen avec un choix de $m = 10$ et $m = 50$ blocs. Le même résultat est constaté pour les autres approximations.

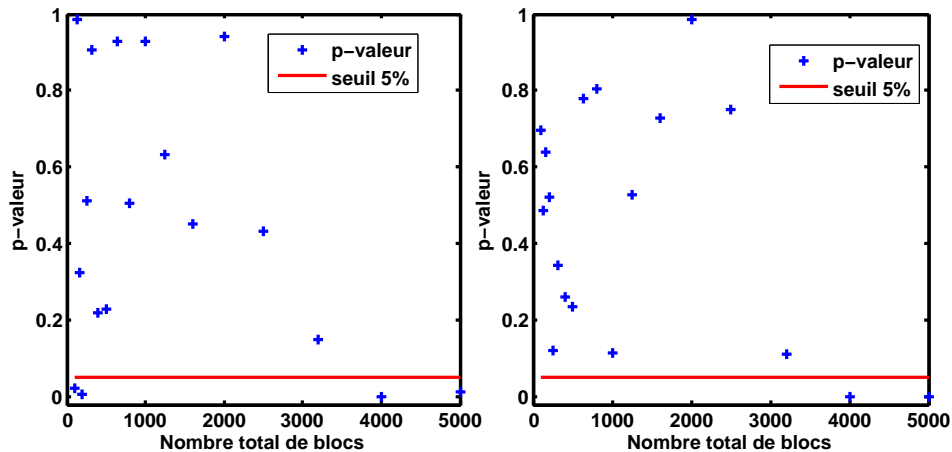


FIG. 4.27: Test ANOVA pour comparer les estimations calculées à partir de « bloc égal » et « bloc différent » : a- $m = 10$ blocs, b- $m = 50$ blocs

Pour $m = 10$ blocs, on remarque que pour la plupart des découpages (hormis ceux en 100, 200, 4000 et 5000 blocs), le test ANOVA permet de conclure qu'il n'y a pas de différence significative

en travaillant avec des blocs contenant le même nombre de piqûres et des blocs dont le nombre suit la loi de Poisson. Pour $m = 50$, le test est rejeté seulement pour les découpages en 4000 et 5000 blocs. Cela montre, que dans la méthode de Gumbel, un découpage pour lequel le nombre de piqûres par bloc suit la loi de Poisson, permet d'avoir des estimations proches de celles calculées à partir de blocs contenant le même nombre de piqûres. Ainsi, une méthode de découpage selon une structure CSR permet de se rapprocher des conditions théoriques de l'application de la méthode de Gumbel. Ce point est discuté dans la section 4.2.4.1. En plus, les découpages à partir desquels le test ANOVA est rejeté peuvent déterminer la taille minimale pour laquelle la méthode de Gumbel est applicable. Pour ces simulations, la taille minimale est égale à $10000/4000 = 2,5 \text{ cm}^2$.

Pour terminer cette section, nous présentons les estimations des quantiles en utilisant la loi lognormale. La figure (4.28) illustre ce résultat. Une étude similaire à celle présentée pour la loi normale conduit aux mêmes conclusions.

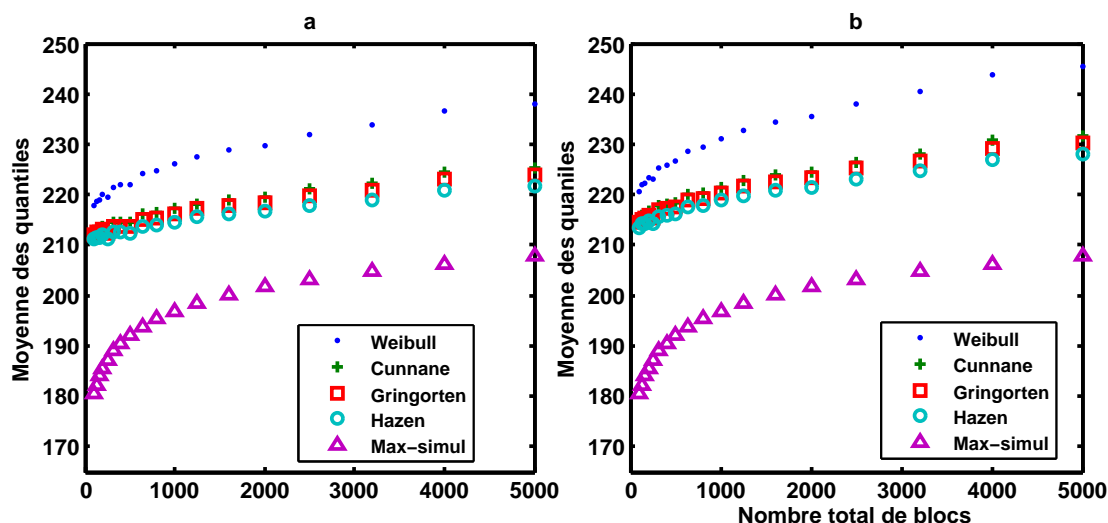


FIG. 4.28: Moyennes des 1000 quantiles estimés à partir des profondeurs générées selon une loi log-normale : a-axes inversés, b-axes classiques

Le même comportement est constaté comme dans le cas de la loi normale. La différence significative réside dans l'ordre de grandeur des quantiles. En effet, la loi log-normale présente une queue plus lourde que celle de la loi normale ; ce qui conduit à des quantiles extrêmes plus importants. Une caractérisation simple des queues lourdes est le calcul du coefficient d'aplatissement (ou kurtosis) de la distribution. La comparaison se fait par rapport à la loi normale qui a un kurtosis égale à 3. Ainsi une distribution a une queue plus lourde que celle de la loi normale si son kurtosis est supérieure à 3. Dans cette étude, le kurtosis calculé à partir de la loi lognormale selon les paramètres présentés au tableau (4.1), a donné une valeur de 3,27. Cela justifie la différence des quantiles extrêmes estimés à partir des deux lois.

Dans cette section, les quatre approximations empiriques ont été testées via des simulations numériques et cela selon plusieurs approches. En résumé, **deux types de simulations** ont été mises en œuvre, la première appelée « **bloc-égal** » et la deuxième « **bloc-différent** ». Pour ces

deux simulations, les quatre approximations empiriques ont été utilisées dans **deux approches** à savoir : l'utilisation des **axes classiques et inversés**, pour estimer les profondeurs maximales à l'aide des quantiles et de l'espérance du maximum. Ensuite une comparaison entre toutes ces estimations a permis de conclure que :

- Pour les deux types de simulations et dans toutes les approches, il existe une différence significative entre les quatre approximations.
- C'est l'approximation de Weibull qui s'écarte le plus de la valeur exacte (maximum simulé noté Max-simul dans les graphiques) en utilisant les espérances du maximum et quantile exact en utilisant les quantiles), alors que les approximations de Cunnane, de Gringorten et de Hazen renvoient des meilleures estimations très proches en termes de biais et de variance, et cela pour les deux types de simulations et pour toutes les approches.
- Pour les deux types de simulations, il existe une différence significative entre les estimations (quantiles et espérance du maximum) calculées à partir de l'approche des axes classiques et l'approche des axes inversés.
- Le calcul des quantiles selon l'approche des axes classiques renvoient des estimations légèrement supérieures (en moyenne) à celles données par l'approche des axes inversés.
- La comparaison des estimations calculées selon les deux types de simulations présente une ressemblance dans le sens où l'analyse de variance a permis de rejeter l'existence d'une différence significative. En d'autres termes il est possible de se rapprocher des conditions optimales de l'application de la méthode de Gumbel en travaillant avec un découpage pour lequel le nombre de piqûres sur chaque bloc suit une loi de Poisson.

4.2.4 Choix de la surface d'analyse

La méthode de Gumbel nécessite la décomposition en blocs d'analyse. Le choix de la taille et le nombre de ces blocs est un problème rarement discuté dans la littérature. Shibata ([Shi96]) a discuté ce problème en concluant que le bloc d'analyse doit contenir un nombre minimal de piqûres, et la théorie des valeurs extrêmes n'est plus valable à partir d'une taille critique du bloc d'analyse. Melsher ([Mel08]) a posé le problème de la « surface représentative » et a conclu que, pour récolter les données, il faut adapter la taille des blocs en fonction des profondeurs des piqûres. Ainsi, un grand bloc est nécessaire dans le cas des grandes profondeurs, et un petit bloc dans le cas des petites profondeurs. Il propose un autre critère du choix de la surface d'analyse, en se basant sur la remarque que l'hétérogénéité, en terme de profondeur, est la conséquence du phénomène de métastabilité des piqûres. Ainsi, l'auteur propose comme critère de choix, des surfaces pour lesquelles le pourcentage des piqûres métastables ne dépasse pas 10%.

Dans cette section, un critère de découpage est proposé. Une fois que la taille des blocs d'analyses est déterminée, le nombre de ces blocs peut être choisi de telle sorte qu'ils stabilisent les estimations. Cette stabilisation est quantifiée en terme de variation de l'estimation lors du passage de p à $p + 1$ blocs. En d'autres termes, on va considérer que les estimations se sont stabilisées si le rapport entre l'estimation à $p+1$ blocs et celle à p blocs est inférieure à un certain seuil qu'on peut fixer à l'avance.

4.2.4.1 Discussion d'un critère de découpage en blocs d'analyse

Les deux approches de simulations numériques présentées dans la section précédente consistent à comparer les estimations dans le cas de « bloc-différent » et le cas de « bloc-égal ». La conclusion, est que les deux estimations sont proches en terme de biais et de variance. Sachant que dans la première approche, les coordonnées des piqûres sont générées selon une CSR, la méthode des

quadrats présentée au troisième chapitre est la plus appropriée pour découper la surface d'étude en blocs d'analyse. En effet, elle permet de trouver l'échelle pour laquelle la structure spatiale des piqûres est une CSR. En outre, cette échelle permettra d'avoir des blocs d'analyse pour lesquels l'hypothèse d'égalité des nombres de piqûres n'est pas rejetée. Ainsi, la méthode des quadrats permet de s'approcher des conditions théoriques de l'application de la méthode de Gumbel. A noter que la méthode ne permet pas toujours l'obtention de l'échelle CSR, car en pratique la structure spatiale des piqûres peut être fortement agrégée. Dans ce cas l'estimation de la densité spatiale et son intégration dans des simulations numériques à l'image du calcul présenté dans la section précédente, peut donner des réponses quant au meilleur découpage en terme de biais et de variance. La figure (4.29) montre les résultats de la méthode des quadrats appliquée aux positions 3 et 4 de la bande 1 (issue de la tôle 1) présentée dans la section (4.1.3).

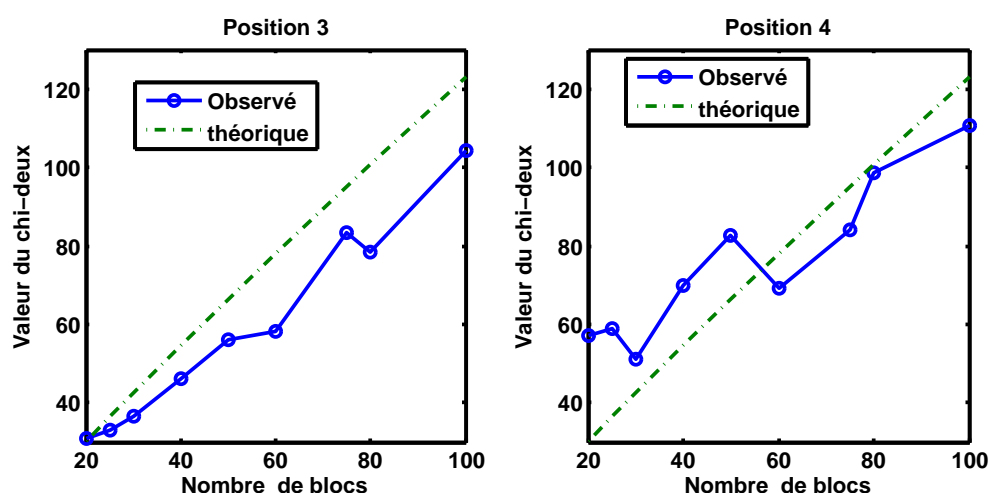


FIG. 4.29: Méthode des quadrats pour les positions 3 et 4 de la bande 1

Le découpage respecte la condition CSR si la valeur observée est inférieure à la valeur théorique du Chi-deux. On remarque qu'en pratique (pour les plaques attaquées), il est possible de trouver le découpage pour lequel l'hypothèse de l'égalité entre les nombres de piqûres par bloc n'est pas rejetée. La difficulté dans cette démarche réside dans le fait qu'il est parfois difficile d'extraire les coordonnées des piqûres.

Désormais, on va s'intéresser à l'influence du choix de la taille des blocs d'analyse ainsi qu'à celui de leur nombre sur les estimations des quantiles extrêmes. Pour cela, on simule une surface de $50 \times 50 \text{ cm}^2$ d'une façon identique à celle de 10000 cm^2 (même densité de 8 piqûres par cm^2 et répartition spatiale), ce qui conduit à la simulation de 20000 piqûres selon une structure CSR. La réduction de la taille de la surface objective simulée a seulement pour objectif la réduction du temps de calcul. Les conclusions restent les mêmes.

Le procédé de calcul consiste à découper la surface objective en plusieurs blocs entre 100 et 1600, et d'y appliquer la méthode de Gumbel. Pour comparer les estimations pour les différents

découpages, le nombre de blocs sera proportionnel à la taille de la surface objective. En d'autres termes, si on découpe toute la surface objective en 100 blocs, et si on en choisit 5 pour estimer les quantiles extrêmes (soit 5% de la surface totale étant analysée) alors, pour le découpage en 200 blocs, on va en analyser 10. Dans ce travail, cette surface analysée varie entre 5% et 30% de la taille totale de la surface objective. Cela permet d'avoir des estimations comparables pour analyser l'impact de la taille des blocs sur les estimations. Pour chaque découpage, 1000 blocs ont été choisis aléatoirement. Vu que toutes les méthodes d'estimations aboutissent aux mêmes conclusions, seule la régression linéaire sera présentée. La figure (4.30) montre les moyennes des 1000 estimations des quantiles associées à chaque découpage ainsi que leurs écarts-types. A noter que dans la suite, c'est l'approximation de Hazen qui sera utilisée pour la méthode de la régression linéaire.

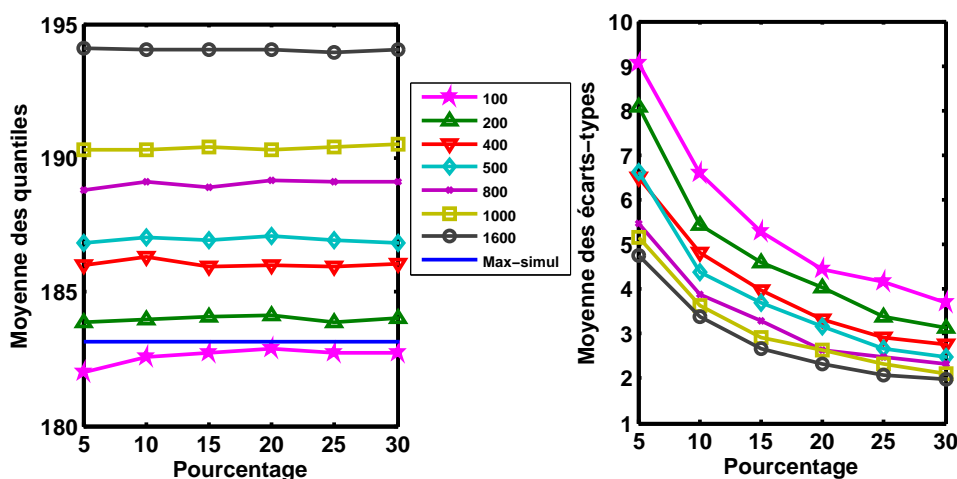


FIG. 4.30: Comparaison des estimations en fonction de la taille des blocs (méthode de la régression) : a-Moyennes des quantiles , b-Moyennes des écarts-types des quantiles

Pour la même taille de surface analysée (égale au nombre de blocs fois la taille d'un bloc), les estimations des quantiles varient. On remarque que plus les blocs ont une grande taille, plus le biais des estimations est faible; en contre partie la variance augmente. Pour un découpage donné, l'augmentation du nombre de blocs n'influe pas sensiblement sur les estimations des quantiles contrairement à leurs tailles. En résumé, en fixant une surface à analyser, la découper en grands blocs aboutira à des estimations à faibles biais mais avec une variance plus importante, et la découper en petits blocs conduira à des surestimations mais avec une variance moins importante. Selon le besoin et l'objectif du travail, on peut privilégier un découpage par rapport à un autre. Par exemple, si on choisit un découpage en petit blocs, on aura une probabilité plus élevée de surestimer l'événement extrême étudié ce qui évitera le risque de la sous-estimation qui peut s'avérer dangereux. En contre partie il faudra évaluer le coût de cette surestimation pour pouvoir trancher.

4.2.5 Comparaison des méthodes d'estimation

Pour comparer les différentes méthodes d'estimations, la simulation numérique mise en œuvre consiste à générer des valeurs selon la loi de Gumbel avec $\mu_0 = 160$ comme paramètre de localisation, et $\sigma_0 = 7$ comme paramètre d'échelle. Ces deux paramètres ont été à partir de la plaque de 75 cm^2 . En connaissant la loi exacte, on va générer des échantillons dont les tailles varient entre 5 et 100. Ensuite ces échantillons sont utilisés pour estimer les paramètres de localisation et d'échelle qu'on comparera avec les valeurs vraies μ_0 et σ_0 . Chaque échantillon étant généré 1000 fois, les moyennes des quantiles ainsi que leurs écarts-types sont présentés dans la figure (4.31).

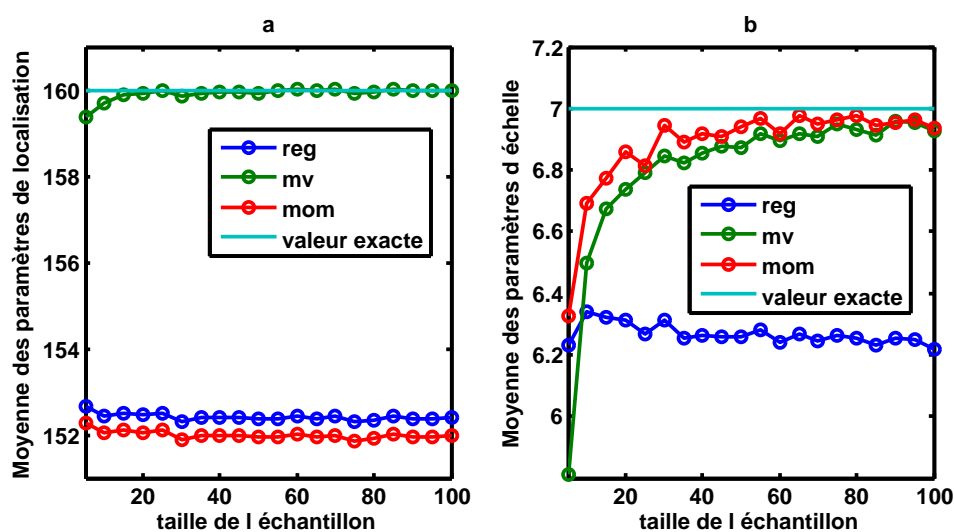


FIG. 4.31: Comparaison des méthodes d'estimations : a-moyennes des paramètres de localisation, b-écarts-types des paramètres d'échelle, reg : régression, mv : maximum de vraisemblance, mom : moment

La simulation mise en œuvre montre bien que la méthode du maximum de vraisemblance présente les meilleures propriétés en terme de biais et de variance. Néanmoins, la résolution numérique du système pour le calcul des paramètres via cette méthode, peut s'avérer délicate dans la mesure où le maximum ne peut être atteint et les intervalles de confiance ne peuvent être calculés avec fiabilité.

4.2.6 Application pour la plaque de 75 cm^2

Dans cette section, on revient sur la plaque de 75 cm^2 , l'objectif est de tester plusieurs découpages et comparer les différentes méthodes d'estimation. La méthodologie est similaire à celle présentée pour la surface simulée de 2500 cm^2 , ainsi 1000 estimations sont calculées en tirant aléatoirement des blocs. La plaque est découpée en 20, 40, 60, puis en 100 blocs. Par exemple, pour le découpage en 20, on choisit aléatoirement 4, 8 puis 12 blocs. Ainsi pour le découpage en 40 blocs, on en choisit 8, 16 puis 24. Pour commencer, la figure (4.32) montre le résultat de la méthode des quadrats appliquée à cette plaque.

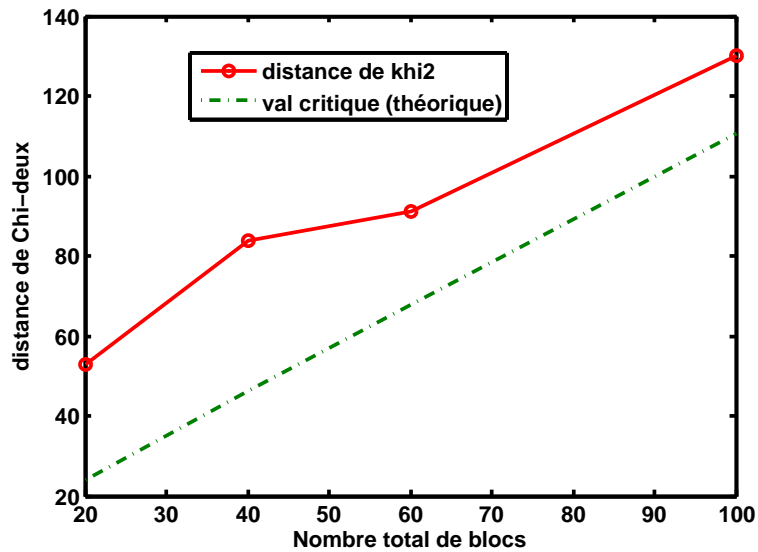


FIG. 4.32: Méthode des quadrats pour la plaque de 75 cm^2

Aucun découpage parmi les quatre utilisés ne respecte le critère de la méthode des quadrats. L'intensité de l'aggrégation n'est pas aussi forte, en effet, pour ces découpages, tous les blocs ont au moins une piqûre. Dans ce cas expérimental, on applique la méthode de Gumbel afin de comparer les résultats avec ceux obtenus dans le cas théorique. La figure (4.33) montre la comparaison des quantiles calculés en utilisant les quatre approximations empiriques en découpant en 100 blocs. Pour les autres découpages, les résultats sont semblables.

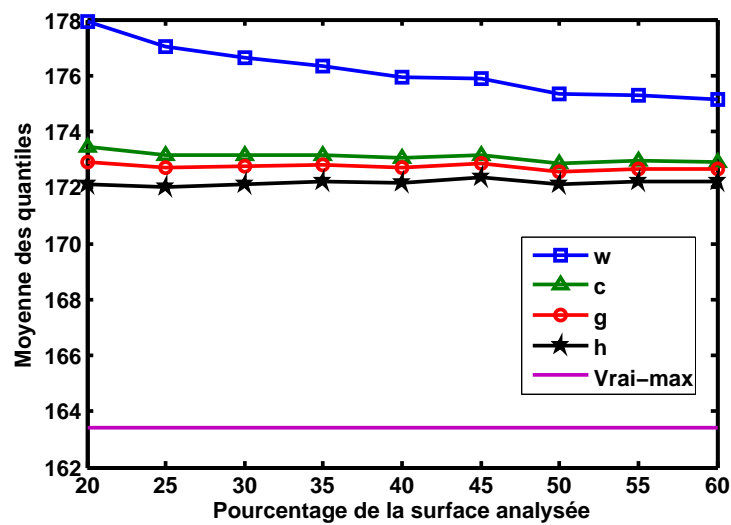


FIG. 4.33: Comparaison entre les quatre approximations empiriques : en abscisse, pourcentage de la surface totale analysée qui est égale à la taille du bloc \times leur nombre

On retrouve le même comportement des estimations vis à vis de chaque approximation empirique que celui constaté avec les simulations numériques. L'approximation de Hazen renvoie le biais le moins important par rapport à la valeur maximale exacte relevée dans la plaque de 75 cm^2 .

L'impact de la taille des blocs d'analyse est présenté sur, la figure (4.34).

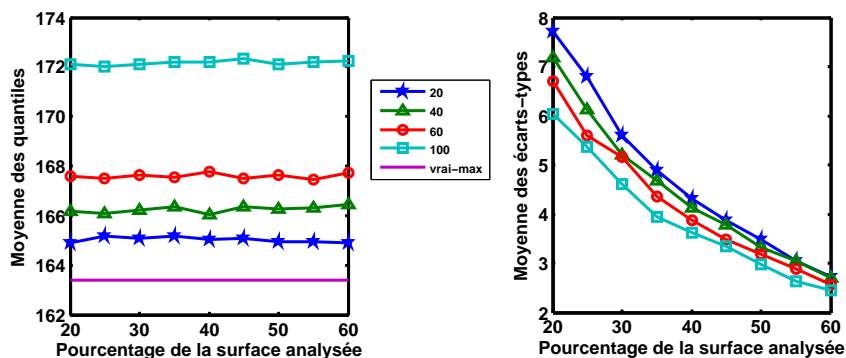


FIG. 4.34: Impact de la taille du bloc d'analyse sur les estimations des quantiles pour la méthode de la régression linéaire

La figure (4.34) montre aussi que les grands blocs renvoient un biais moins important que les petits blocs, et le contraire pour la variance. Pour les autres méthodes d'estimations, les résultats seront présentés en annexe.

L'augmentation du pourcentage de la surface analysée n'améliore pas forcément l'estimation. La figure (4.35) montre la comparaison entre les différentes méthodes d'estimations des paramètres dans la méthode de Gumbel.

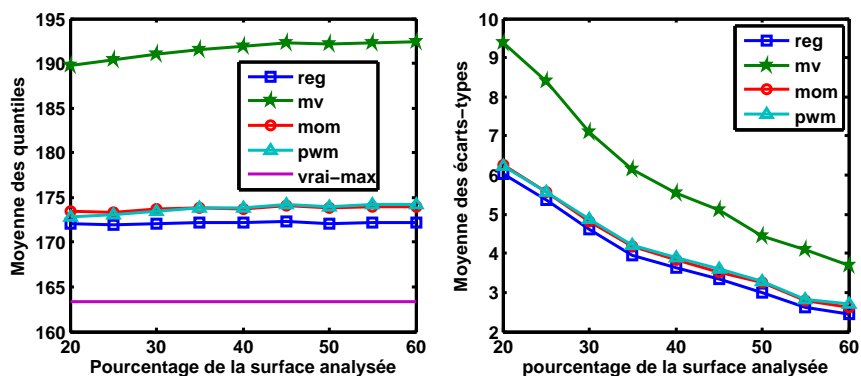


FIG. 4.35: Comparaison des méthodes d'estimations des paramètres pour la plaque de 75 cm^2

Contrairement aux simulations numériques, la méthode du maximum de vraisemblance présente un biais et une variance plus importants que les autres méthodes. Cela est essentiellement dû au fait que les conditions théoriques de l'application de la méthode de Gumbel ne sont pas toutes remplies. Cela montre toute la difficulté à choisir la meilleure méthode quand il s'agit de traiter des cas expérimentaux.

Conclusion : les différentes simulations numériques présentées, ont pour objectif d'élucider certaines questions autour de l'application de la méthode de Gumbel. Pour cela, on s'est placé dans les conditions théoriques d'application de cette méthode. Mais le passage aux cas expérimentaux pose toujours un problème quant au choix optimal de la taille des blocs et de la meilleure méthode d'estimation. Néanmoins, les conclusions de ces simulations numériques consistent à choisir des blocs de grandes tailles pour avoir plus de précisions pour l'estimation des quantiles extrêmes, et des blocs de petites tailles pour les surestimer ce qui permet de prendre plus de sécurité quant aux événements extrêmes. Finalement, le choix de la meilleure méthode d'estimation des paramètres, est difficile quand il s'agit de cas expérimentaux. Par contre, les simulations montrent que la méthode du maximum de vraisemblance présente de meilleures propriétés du point de vue biais et variance.

Chapitre 5

Approches GEV et POT

5.1 Loi des valeurs extrêmes généralisée GEV

Les travaux de Gumbel ([Gum58]) ont très largement contribué à la diffusion de la théorie des valeurs extrêmes, notamment en hydrologie. Le choix de la loi des valeurs extrêmes de type Gumbel était particulièrement privilégié au détriment des deux autres types de lois (Fréchet et Weibull). Ce choix était essentiellement motivé par la simplicité de la loi de Gumbel pour laquelle on cherche à estimer deux paramètres (de localisation et d'échelle) au lieu de trois pour les deux autres types de lois. Néanmoins, il ne faut pas oublier le caractère asymptotique de la théorie des valeurs extrêmes et que les lois de type Fréchet et Weibull sont aussi des candidats pour la modélisation des valeurs maximales. En plus, les estimations des quantiles peuvent être très différentes selon le type de loi utilisée (voir chapitre (2)). Pour illustrer la différence entre les estimations calculées à partir de la loi GEV et la loi de Gumbel, la procédure de calcul utilisée est la même que celle présentée pour l'approche « bloc-égal »(voir chapitre (4)) avec un nombre de blocs égale à 50 et un nombre de simulations égale à 5000. La loi étant utilisée est la loi normale de mêmes paramètres que ceux présentés dans le tableau (4.1). La figure (5.1) montre la moyenne des 5000 estimations des quantiles ainsi que la moyenne des écarts-types. Les deux méthodes d'estimation utilisées dans cet exemple sont le maximum de vraisemblance (MV) pour la loi GEV et la régression linéaire avec l'approximation de Hazen pour la loi de Gumbel.

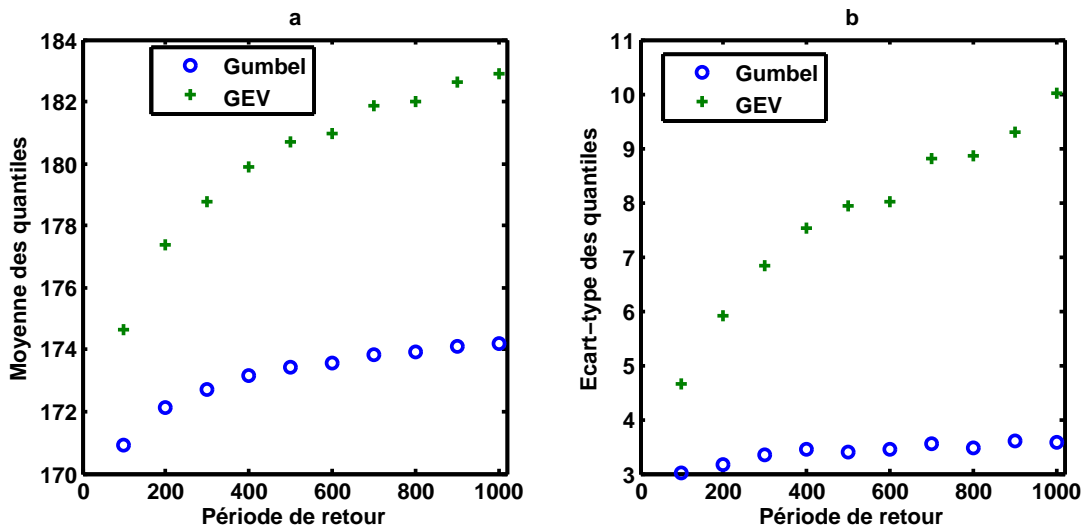


FIG. 5.1: a- Moyenne des quantiles, b- Ecart-types des quantiles

Cet exemple montre bien la différence qui peut exister entre les estimations (et leurs écarts-types) en utilisant la loi GEV ou la loi de Gumbel. L'accroissement des écarts-types lors du passage de la loi de Gumbel à la loi GEV est essentiellement dû au fait que l'augmentation du nombre de paramètres conduit souvent à l'augmentation de l'incertitude liée à l'estimation de chacun d'eux. Ce résultat rejoint les remarques émises par ([PMA01]).

Cette différence illustre le danger d'utilisation systématique de la loi de Gumbel comme distribution des valeurs maximales. Ainsi, il faut déterminer si la distribution de Gumbel est suffisante pour modéliser le comportement des valeurs extrêmes. Pour cela, le test proposé par Hosking ([Hos84]) peut être utilisé. Il consiste à tester l'hypothèse nulle $H_0 : \xi = 0$ contre l'hypothèse alternative $H_1 : \xi \neq 0$. L'auteur montre que sous l'hypothèse nulle, l'estimateur de l'indice extrême par la méthode des moments de probabilité pondérés $\hat{\xi}^{pwm}$ suit asymptotiquement une loi normale $\mathcal{N}(0; 0, 5633/n)$ où n représente la taille de l'échantillon. En pratique, la statistique $Z = \hat{\xi}^{pwm} \sqrt{n/0,5633}$ est calculée et comparée aux valeurs critiques de la loi normale centrée réduite pour accepter ou rejeter l'hypothèse nulle.

5.1.1 Comparaison des méthodes d'estimation PWM et MV pour la loi GEV

Les méthodes PWM et MV (voir chapitre 2) sont les plus utilisées pour estimer les paramètres de la loi GEV. Dans cette section, on propose de comparer ces deux méthodes en terme de biais et de variance. Pour cela, des échantillons issus de la loi GEV normalisée ($\mu = 0$ et $\sigma = 1$) sont générés pour des valeurs de ξ (l'indice extrême) comprises entre $-0,4$ et $0,4$. Le choix de la loi GEV normalisée mène aux mêmes conclusions qu'un choix plus général de ces deux paramètres, car les deux méthodes d'estimations PWM et MV sont invariantes par transformation linéaire. Quant au domaine de variation de l'indice extrême, il est dû au fait que les estimateurs PWM et MV perdent de leur qualité en dehors de cet intervalle ([HWW85]).

Les paramètres de cette simulation numérique sont les suivants :

- Indice extrême ξ : varie entre $-0,4$ et $0,4$ avec un pas de $0,1$.
- Taille des échantillons n : varie entre 20 et 500 avec un pas de 20 .
- Nombre de simulations : 5000 échantillons sont simulés pour chaque valeur de ξ et pour chaque taille d'échantillon.
- Période de retour : les probabilités de dépassements p sont choisies telle que $p < 1/n$, ce qui correspond à des périodes de retour $T > n$. On choisit ici $T = 2000$. Les conclusions restent les mêmes pour les autres choix de T .

Pour la méthode PWM, le calcul de l'indice extrême passe par la résolution numérique de l'équation (2.63) qui peut s'avérer délicate dans certains cas. En pratique, l'indice extrême est souvent compris entre $-0,5$ et $0,5$. Ainsi Hosking ([HWW85]) propose une approximation linéaire de la fonction $(3^\xi - 1)/(2^\xi - 1)$ pour calculer l'indice extrême. L'approximation est basée sur le développement limité à l'ordre 1 de cette fonction. Elle est donnée par :

$$\frac{3^\xi - 1}{2^\xi - 1} \approx \frac{\log(3)}{\log(2)} + \left(\frac{\log(3)}{2} - \frac{\log(3)^2}{2 \log(2)} \right) \xi \quad (5.1)$$

Ainsi, l'indice extrême est calculé en remplaçant cette expression dans l'équation ((2.63)).

Pour la méthode MV, les échantillons générés avec la valeur de $\xi = 0,4$, présente des difficultés numériques. En effet, au voisinage de cette valeur, certains échantillons font que l'algorithme du maximum de vraisemblance ne converge pas ou renvoie des estimations de ξ dépassant $0,8$ ce qui entraîne une augmentation brusque des quantiles. Les figures (5.2) montre les variations de la fonction quantile de la loi GEV normalisée en fonction de l'indice extrême.

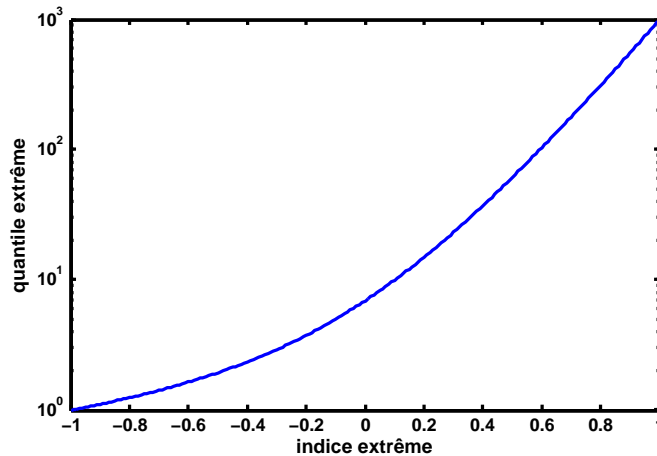


FIG. 5.2: Variation de la fonction quantile de la loi GEV normalisée en fonction de l'indice extrême sur une échelle Logarithmique (en oy)

Pour les échantillons présentant ce type de comportement, le tirage est refait. On montre ici les résultats pour les indices extrêmes $\xi = 0,4$ et $\xi = -0,4$. Dans l'annexe B, on illustre ces résultats pour $\xi = 0,2$, $\xi = -0,2$ et $\xi = 0$. La figure (5.3) montre la moyenne des 5000 estimations de l'indice extrême ξ , du paramètre de localisation μ , du paramètre d'échelle σ et du quantile via les deux méthodes d'estimation MV et PWM.

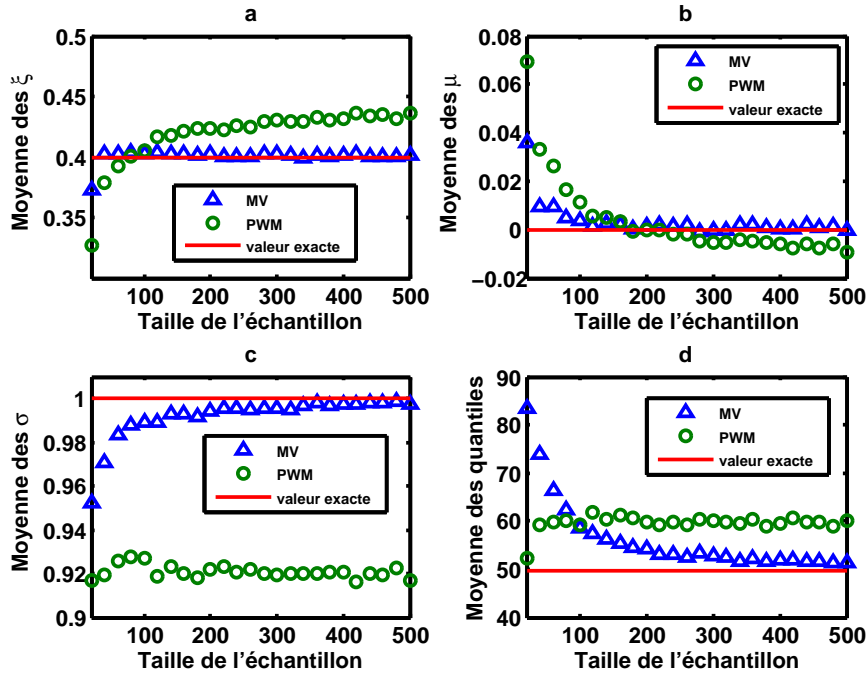


FIG. 5.3: Moyennes des estimations pour les échantillons simulés avec $\xi = 0, 4$ en fonction de leurs tailles : (a)- Moyennes des indices extrêmes, (b)- Moyennes des paramètres de localisation μ , (c)- Moyennes des paramètres d'échelle σ , (d)- Moyennes des quantiles

S'intéressant particulièrement aux estimations des quantiles, la figure (5.4) montre le biais et l'écart-type des estimations.

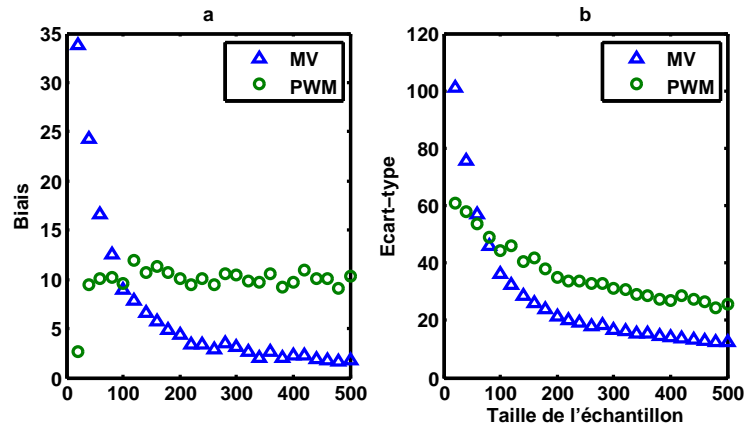


FIG. 5.4: Echantillons pour $\xi = 0, 4$: a- Biais des quantiles, Ecart-types des quantiles

Dans ce cas, c'est la méthode PWM qui présente les meilleures estimations en terme de biais et de variance pour les petites tailles d'échantillon. Ce comportement s'inverse quand il s'agit d'estimer les paramètres à partir d'échantillons de maxima de grandes tailles. Les figures (5.5 et 5.6) montrent les résultats pour les échantillons issus de $\xi = -0, 4$.

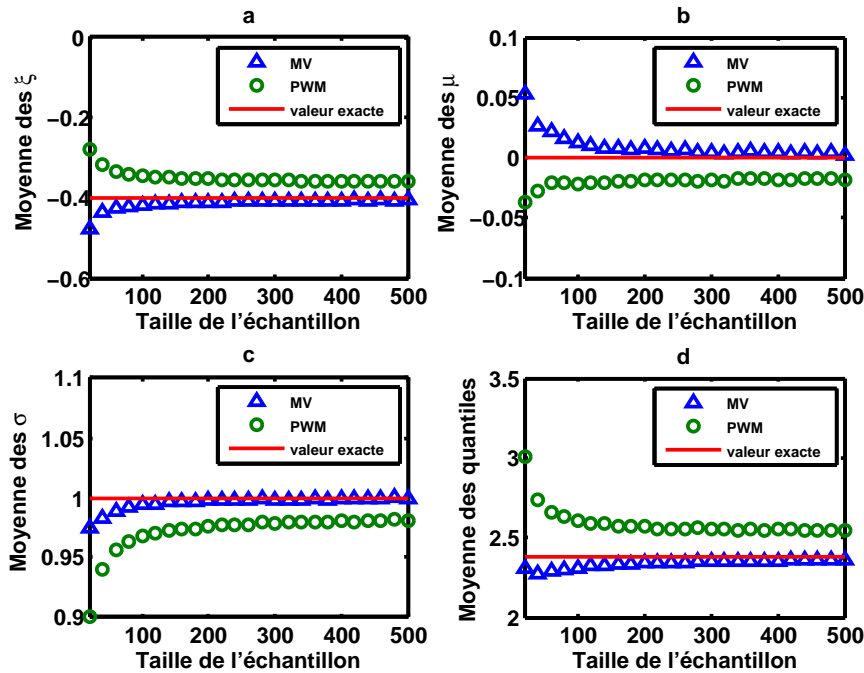


FIG. 5.5: Moyennes des estimations pour les échantillons simulés avec $\xi = -0, 4$ en fonction de leurs tailles : (a)- Moyennes des indices extrêmes, (b)- Moyennes des paramètres de localisation μ , (c)- Moyennes des paramètres d'échelle σ , (d)- Moyennes des quantiles

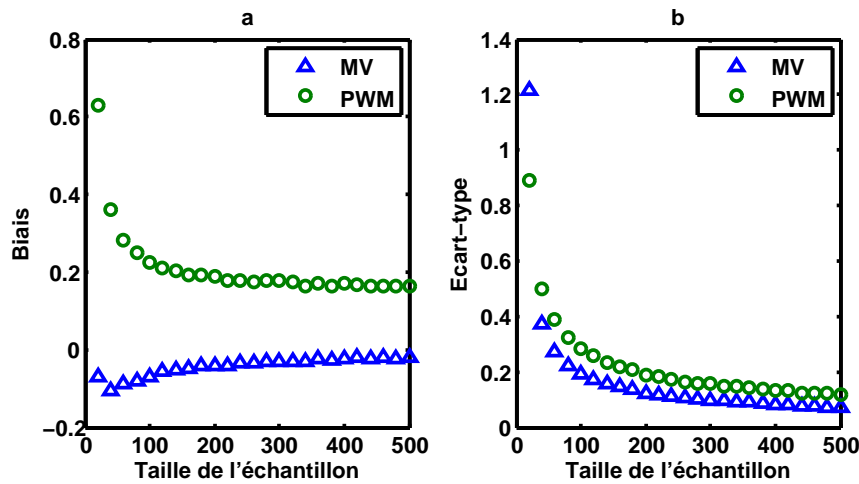


FIG. 5.6: Echantillons pour $\xi = -0, 4$: a- Biais des quantiles, Ecarts-types des quantiles

Pour toutes les valeurs de l'indice extrêmes, et pour les échantillons de grands tailles (≥ 60) la méthode MV présente les meilleurs estimations en terme de biais et de variance. Pour les petites tailles d'échantillons, c'est la méthode PWM qui présente les meilleurs propriétés en particulier en terme d'écart-type. Sachant que les difficultés numériques de la méthode MV disparaissent pour les grandes tailles d'échantillon (problèmes survenus pour ξ proche de 0, 4), on peut privilégier la méthode PWM pour un échantillon de maxima de petite taille et la méthode MV sa taille est grande.

5.2 Loi des dépassements de seuil POT

Pour comparer les deux méthodes d'estimation MV et PWM sous l'approche POT, on suit la même démarche de simulation présentée pour l'approche GEV. Dans ce cas, c'est à partir de la loi de Pareto généralisée normalisée qu'on va générer les échantillons à partir desquels on va calculer les paramètres de forme (indices extrêmes), d'échelle et les quantiles. Quant au paramètre de localisation qui représente le seuil dans l'approche POT, on utilisera sa valeur exacte qui est égale à 0. Ce choix est dû essentiellement au fait que les simulations mises en oeuvre deviennent relativement lourdes lors de la recherche du seuil optimal (via les méthodes numériques (voir 2.4.2)) en particulier pour les échantillons de grandes tailles. Cela ne change en rien les conclusions qui peuvent être tirées en comparant les deux méthodes d'estimations puisque on compare les autres paramètres et les quantiles pour le même seuil. Pour illustrer ces résultats, on se restreint aux quantiles pour comparer les deux méthodes MV et PWM. Les figures (5.7 et B.9) montrent les moyennes des quantiles ainsi que leurs écarts-types pour $\xi = 0,4$ et $\xi = -0,4$. Pour les autres valeurs de ξ , ces résultats sont illustrés dans l'annexe B.

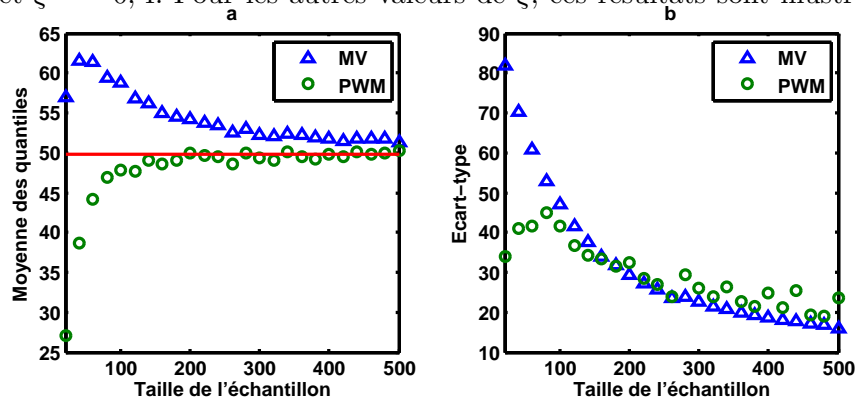


FIG. 5.7: Echantillons pour $\xi = 0,4$: a- Moyennes des quantiles, b-Écarts-types des quantiles

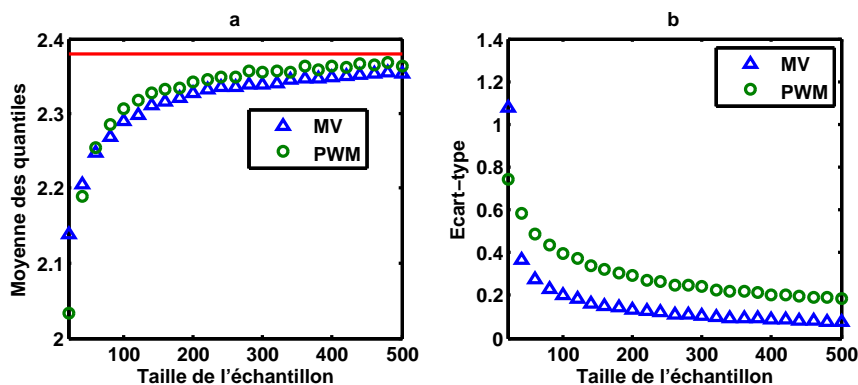


FIG. 5.8: Echantillons pour $\xi = -0,4$: a- Moyennes des quantiles, b- Écarts-types des quantiles

Pour les grands échantillons, les deux méthodes renvoient des estimations comparables en terme de biais et de variance. La différence apparaît notamment quand il s'agit des échan-

tillons de petites tailles en particulier concernant les écarts-types des estimations calculées à partir de la méthode MV. D'une façon générale, on peut émettre les mêmes conclusions quant au choix de la méthode d'estimation par rapport à la méthode GEV, à savoir la méthode PWM pour les échantillons de petites tailles et la méthode MV pour les grandes tailles. Cela évitera les instabilités ou les problèmes de convergences de la méthode MV pour les petits échantillons ($n \leq 60$).

Dans la section qui va suivre, on va s'intéresser à la comparaison entre les quantiles extrêmes déterminés par les deux approches GEV et POT.

5.3 Comparaison entre les approches GEV et POT

Le calcul des quantiles extrêmes dans l'approche POT est fait à partir de la formule (2.87). Ce quantile représente la profondeur dépassée avec une faible probabilité p (la période de retour $T = 1/p$), sachant que cette profondeur est supérieure à un seuil donné. Pour calculer le quantile dépassé avec la probabilité p , on utilise la formule suivante ([MB03]) :

$$\hat{q}_{p,k}^{pot} = X_{n-k,n} + \hat{\sigma}_k \frac{\left(\frac{k}{np}\right)^{\hat{\xi}_k} - 1}{\hat{\xi}_k} \quad \text{pour } p < \frac{k}{n} \quad (5.2)$$

où $X_{n-k,n}$ représente la $(n-k)$ ème statistique d'ordre des observations. $\hat{\sigma}_k$ et $\hat{\xi}_k$ sont respectivement les paramètres d'échelle et de localisation estimés en utilisant k valeurs dépassant le seuil.

Dans cette section, on propose de comparer les différents quantiles sous les deux approches. Pour l'approche des maxima par bloc, on calcule les quantiles de la loi de Gumbel et de la loi GEV par les méthodes MV et PWM. De l'autre côté, pour l'approche de dépassement de seuil, on calcule les quantiles donnés par l'équation (5.2) par les méthodes MV, PWM, moments de Dekkers et l'estimateur UH (voir chapitre (2)). On utilise pour cela, des échantillons issus de la loi normale et la loi lognormale avec les mêmes paramètres présentés auparavant (voir tableau 4.1). Pour cela, on génère 1000 échantillons suivant ces deux lois. En utilisant la même densité spatiale des piqûres (8 piqûres par 10 cm²) que celle estimé au chapitre précédent, cela conduit à la simulation d'échantillons de taille égale à 80 qu'on divise en m blocs selon l'approche « bloc égal »(voir chapitre (4)). Pour illustrer les résultats, on prend $m = 10$, et une période de retour égale à 1000. Le même comportement est observé pour $m > 10$. Les figures ((5.9),(5.10)) montrent respectivement les intervalles de confiances à 10% des estimations des quantiles pour la loi normale et lognormale.

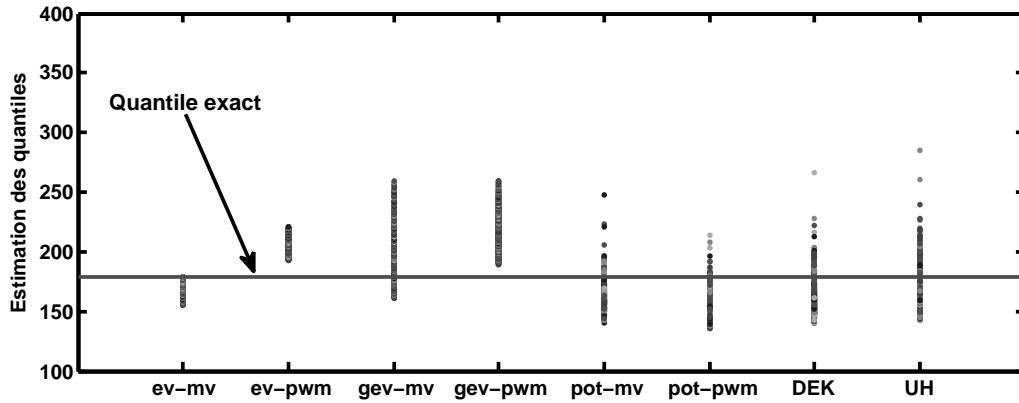


FIG. 5.9: Intervalle de confiance à 10% des estimations des quantiles pour la loi normale. De gauche à droite les quantiles de : Gumbel estimés par mv, Gumbel estimés par pwm, GEV estimés par mv, GEV estimés par pwm, POT estimés par mv, POT estimés par pwm, Dekkers, UH.

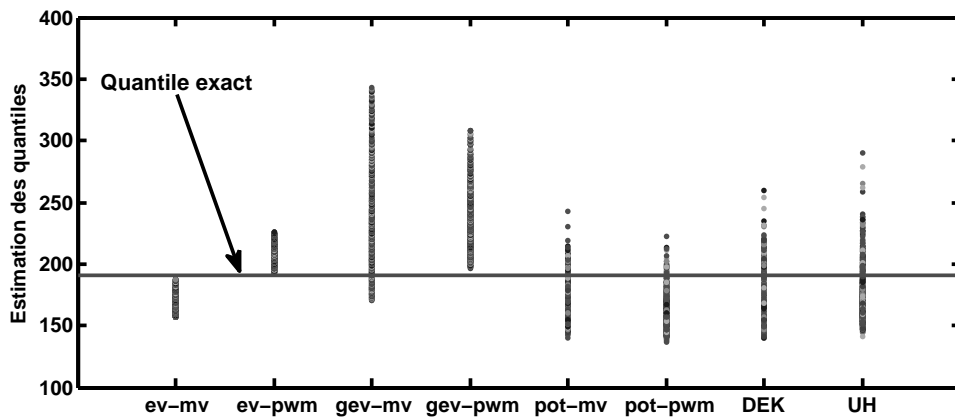


FIG. 5.10: Intervalle de confiance à 10% des estimations des quantiles pour la loi lognormale. De gauche à droite les quantiles de : Gumbel estimés par mv, Gumbel estimés par pwm, GEV estimés par mv, GEV estimés par pwm, POT estimés par mv, POT estimés par pwm, Dekkers, UH.

On note que dans l'approche POT, le seuil a été calculé à l'aide de la méthode de Pickand (voir chapitre (2)). Pour la loi normale, le nombre moyen de valeurs retenues (excès) est de 48, alors que celui de la loi lognormale est de 50. On remarque que la méthode POT présente

des résultats intéressants en terme de biais et de variance par rapport à la méthode GEV et cela pour les deux méthodes d'estimations. Les estimateurs UH et de Dekkers renvoient des estimations comparables. Selon le besoin du praticien, le choix d'une méthode au détriment d'une autre peut être motivé par la volonté de surestimer la profondeur maximale (approche GEV) ou de la connaître avec suffisamment de précision (POT).

La loi normale (utilisée dans cette simulation) rentre dans le domaine d'attraction de Gumbel. On peut remarquer que c'est la distribution de Gumbel qui renvoie les meilleurs résultats en terme de biais. Le fait d'ajouter un troisième paramètre dans les approches GEV et POT (l'indice extrême) a augmenté l'incertitude liée à l'estimation de chaque paramètre, ce qui est observé vue la largeur de l'intervalle de confiance dans ces deux approches.

Dans ce travail, les mesures de profondeurs effectuées appartenaient toutes au domaine d'attraction de Weibull. C'est le cas ici lors de l'étude de la plaque de 75 cm², et aussi le cas d'un travail effectué sur une autre plaque issue de la tôle 1 et résumé dans ([JNI⁺]). Pour cela, nous terminons cette simulation en générant les profondeurs selon une loi des valeurs extrêmes de paramètres $\xi = -0,23$, $\mu = 137,7$ et $\sigma = 8,9$. Ces paramètres ont été estimés à partir de la plaque de 75 cm² à l'aide de la méthode MV. l'objectif, est de voir le comportement des différentes méthodes d'estimations (en particulier dans l'approche de Gumbel) en travaillant avec des valeurs appartenant dans le domaine d'attraction de Weibull. la figure (5.11) illustre ces résultats.

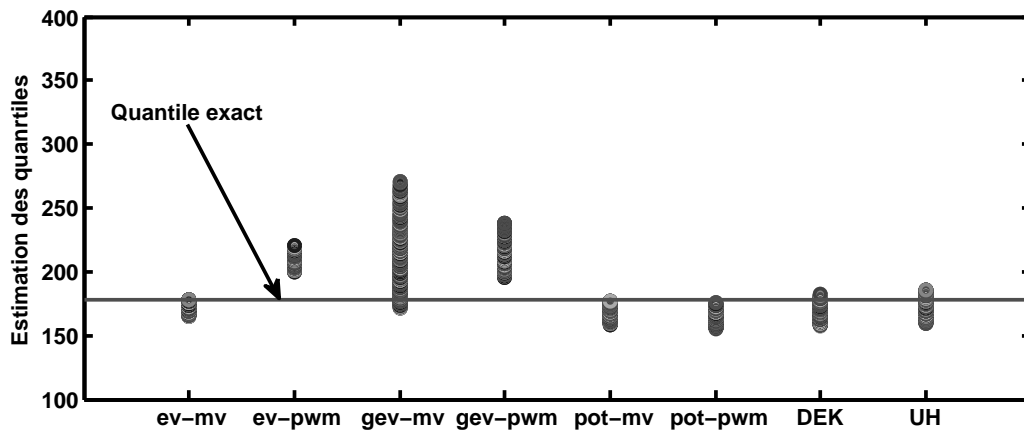


FIG. 5.11: Intervalle de confiance à 10% des estimations des quantiles pour la loi lognormale. De gauche à droite les quantiles de : Gumbel estimés par mv, Gumbel estimés par pwm, GEV estimés par mv, GEV estimés par pwm, POT estimés par mv, POT estimés par pwm, Dekkers, UH.

Les incertitudes constatées lors de l'utilisation de la loi normale et lognormale ont été largement réduites. La loi GEV présente toujours l'intervalle de confiance le plus large, alors que dans l'approche POT, ce dernier a été réduit et renvoie des résultats proches à ceux de

la méthode de Gumbel (avec la méthode d'estimation MV). Dans tous ces étudiés, la loi de Gumbel garde une variance faible par rapports aux autres méthodes. En grande partie cela est dû au nombre de paramètres réduit. Un critère de choix de la méthode d'estimation des paramètres ne peut être basé que sur la variance. Il faut regarder aussi le biais et la façon dont on estime quantiles extrêmes, en d'autres termes, la méthode a tendance à les sous-estimer ou à les surestimer. Tout dépend du coût et le risque de l'étude engagée.

5.4 Application à la plaque de 75 cm²

Dans un premier temps, nous allons nous intéresser au domaine d'attraction des 603 piqûres mesurées sur la plaque de 75 cm². Pour cela, on va utiliser le quantile plot généralisé basé sur l'estimateur UH, et l'estimateur de Dekkers. La figure (5.12) montre les variations de ces deux estimateurs en fonction du nombre des excès avec leurs intervalles de confiances à 10%.

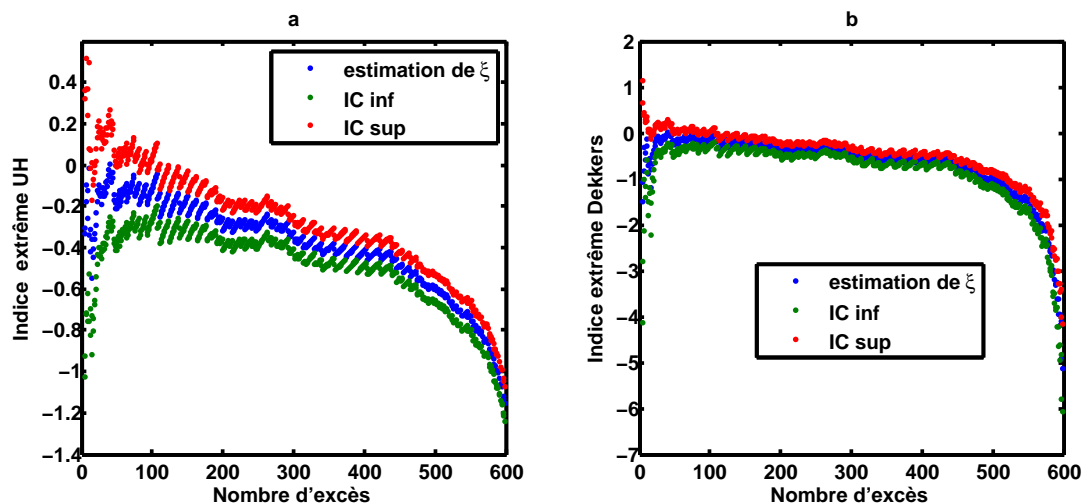


FIG. 5.12: Intervalle de confiance à 10% des estimation de l'indice extrême. IC inf : borne inférieure de l'intervalle de confiance, IC sup : borne supérieure

Ces deux graphiques montrent que les profondeurs des piqûres mesurées sur la plaque de 75 cm² appartiennent au domaine d'attraction de Weibull. En effet, pour les deux estimateurs de l'indice extrêmes, on a des valeurs négatives pour ce paramètre qui caractérise un domaine d'attraction. Cela implique l'existence d'une profondeur limite qu'on peut calculer en estimant les paramètres de la loi GEV ou POT. Cette limite est appelée en théorie des valeur extrêmes, un point terminal.

Désormais, on va s'intéresser à la comparaison des différents estimateurs des quantiles extrêmes à l'image du travail de simulation présentée par les figures (5.9, 5.9). Dans ce cas expérimental, on présente les résultats des estimations en découpant la plaque de 75 cm² en 100 blocs, ensuite on choisit aléatoirement 10 puis 20 blocs pour estimer les quantiles extrêmes pour la période de retour $T = 100$ qu'on compare avec la profondeur maximale exacte relevée

sur la plaque et qui est égale à $163 \mu\text{m}$. Les figures (5.13) et (5.14) illustrent ces résultats.

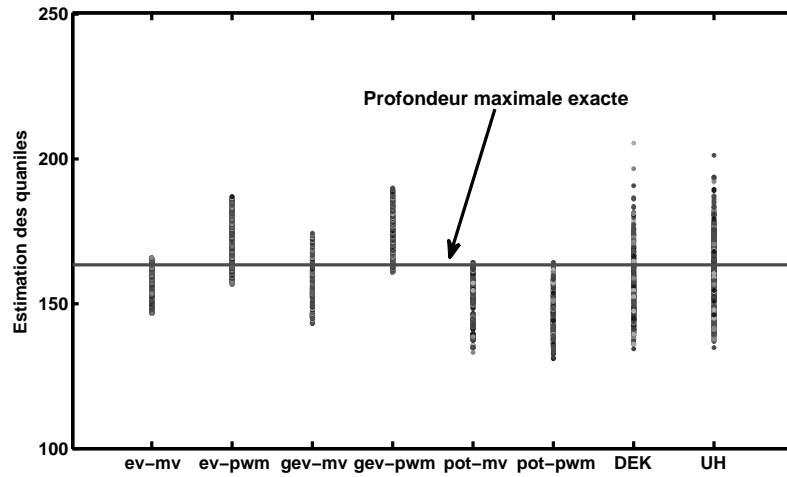


FIG. 5.13: Intervalle de confiance à 10% des estimations des quantiles pour la plaque pour le choix de 10 blocs. De gauche à droite les quantiles de : Gumbel estimés par mv, Gumbel estimés par pwm, GEV estimés par mv, GEV estimés par pwm, POT estimés par mv, POT estimés par pwm, Dekkers, UH.

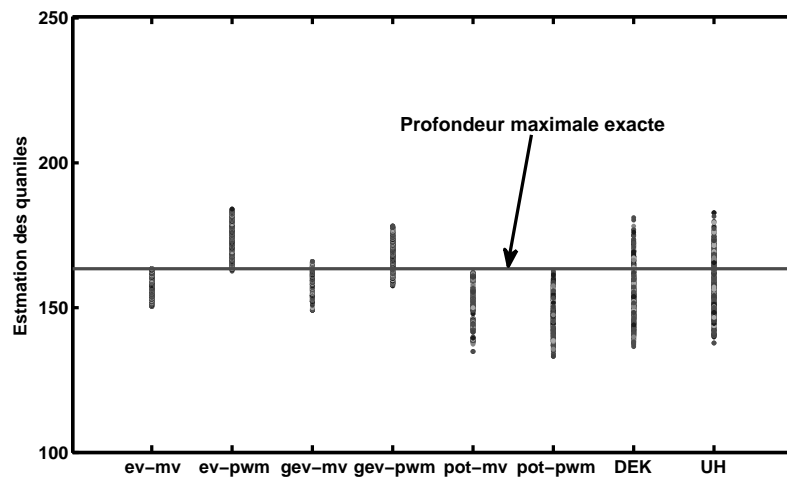


FIG. 5.14: Intervalle de confiance à 10% des estimations des quantiles pour la plaque pour le choix de 20 blocs. De gauche à droite les quantiles de : Gumbel estimés par mv, Gumbel estimés par pwm, GEV estimés par mv, GEV estimés par pwm, POT estimés par mv, POT estimés par pwm, Dekkers, UH.

Dans ce cas expérimental, on retrouve pratiquement le même comportement global des estimations des quantiles sauf pour la loi GEV qui présente cette fois-ci une dispersion moins importante. Son intervalle de confiance est comparable à celui de la loi Gumbel. Cela est dû

essentiellement au fait que la loi GEV est plus adaptée dans ce cas malgré l'estimation de trois paramètres. Enfin, pour l'approche POT, les quantiles sous-estiment la profondeur maximale exacte, cela peut être dû au fait que la méthode POT perde de ses performances en présence de données redondantes ce qui est le cas dans nos mesures de profondeurs. Dans ce travail, le traitement de ces données est fait en prenant la première valeur non répétée. Par exemple si la méthode de Pickands donne k seuil et on a égalité entre la $n - k$ ème et la $n - k + 1$, alors c'est la $n - k + 2$ ème valeur qui est prise comme seuil. Enfin, les deux estimateurs UH et de Dekkers présentent des estimations comparables en terme de biais et de variance.

Conclusion générale

Conclusion générale

La complexité du phénomène de corrosion par piqûres en terme de physique et son caractère stochastique rencontré dans sa phase d'initiation et de propagation, a conduit récemment beaucoup de travaux à s'intéresser à l'étude de ce phénomène du point de vue statistique. L'un des facteurs importants lors de l'étude du phénomène est la profondeur maximale des piqûres. Son estimation est très importante pour prévoir les défaillances des structures car une profondeur critique peut fragiliser les structures en particulier lors du couplage avec des sollicitations mécaniques. Dans ce travail, on s'est intéressé aux méthodes statistiques utilisées pour cet objectif. La théorie des valeurs extrêmes est manifestement l'approche la plus utilisée pour estimer ces profondeurs maximales en particulier la méthode de Gumbel. Cependant, dans le domaine de la corrosion par piqûres, cette méthode est souvent appliquée sans s'intéresser aux conditions de sa validité, et les méthodes d'estimation des paramètres sont peu comparées entre elles. Ce travail a donc pour objectif de revenir sur l'application de la méthode de Gumbel et la comparer avec d'autres approches telles que la méthode des valeurs extrêmes généralisées (généralisation de la loi de Gumbel) et l'approche de dépassement de seuil.

Avant d'entamer ce développement statistique, un travail expérimental était nécessaire. Dans un premier temps, on s'est intéressé à corroder des plaques d'aluminium pour constituer des échantillons afin d'en faire des mesures de profondeurs et calculer les paramètres de simulations numériques utilisées pour répondre à certaines questions autour de la méthode de Gumbel. Pour mener ces expérimentations et pouvoir comparer les résultats de simulations avec ceux de la simulation, on a imposé quelques conditions sur les échantillons. En effet, on a souhaité que les profondeurs des piqûres soient facile à mesurer (absence de produits de corrosion, absence de coalescence, non déviantes, . . .), que le l'observation des piqûres se fait en un temps raisonnable (quelques heures) et que l'expérimentation soit simple. Ainsi, les essais ont été fait à potentiel libre et à température ambiante. Cependant, trouver le couple matériau/milieu qui vérifie ces conditions s'est avéré très délicat. Une recherche bibliographique nous a conduit à choisir l'aluminium comme matériau, et un milieu dont les compositions ont été données dans le chapitre 4.

La première difficulté rencontrée concerne la non-reproductibilité de la densité des piqûres. En effet, sous les mêmes conditions de corrosion, la densité spatiale des piqûres pour la même position des plaques peut être très différente. S'intéressant à la densité spatiale des piqûres, on a introduit les processus spatiaux pour étudier la répétabilité d'une structure spatiale et comment les piqûres sont réparties sur la surface attaquée. Cette méthode stasistique détermine s'il y a une interaction entre piqûres ou pas, ce qui permet de vérifier à l'occasion la condition d'indépendance nécessaire lors de l'application de la théorie des valeurs extrêmes. Ensuite, on a adapté la méthode d'analyse spectrale pour analyser les interactions. Cette méthode, n'a jamais été appliquée en corrosion auparavant, et elle a été appliquée pour étudier la répétabilité de la structure spatiale des piqûres pour la même position (d'une plaque de la tôle). Ces processus spatiaux, ont été utilisé aussi pour

trouver l'échelle pour laquelle on peut considérer que les piqûres sont indépendantes. Cette échelle est proposée comme critère pour trouver la taille de la surface d'analyse pour la méthode de Gumbel. Pour pousser l'étude des processus spatiaux, on a introduit les processus marqués. Il s'agit d'étudier l'interaction entre les piqûres en rajoutant l'information portée sur les profondeurs. En d'autres termes, les processus spatiaux étudient l'interaction entre les piqûres en utilisant seulement les coordonnées des piqûres, par contre les processus marqués fournissent des informations quant à l'attraction ou la répulsion des piqûres ayant des profondeurs données. Ainsi, on peut savoir par exemple si les piqûres les plus profondes ont tendance à s'attirer ou à se repulser. Cette méthode est contraignante car elle nécessite la connaissance à la fois des coordonnées de chaque piqûre et la profondeur y associée. Cependant, sur des échantillons de petites tailles, la méthode est très intéressante car elle peut fournir des informations pour savoir si les profondeurs les plus profondes ont tendance à se regrouper. Cette méthode a été appliquée sur la plaque de 75 cm^2 et a conclu à une totale indépendance entre les positions des piqûres et leurs profondeurs.

Le travail expérimental, a permis d'extraire les paramètres utilisés dans nos simulations numériques. L'objectif de ces simulations est de répondre à certaines questions autour de l'application de la méthode de Gumbel. En particulier le choix de la distribution empirique, la taille et le nombre des surfaces d'analyse et la comparaison des méthodes d'estimation des paramètres. Cela a conduit à conclure que le meilleur choix de la distribution empirique (parmi celles utilisées) est l'approximation de Hazen. Quant au choix de la taille de la surface d'analyse, on remarque que pour les grandes tailles de ces surfaces d'analyse, on a un biais moins important que pour les surfaces de petites tailles mais une variance plus importante. Donc le choix peut s'avérer délicat, néanmoins, selon l'application, on peut privilégier un choix par rapport à l'autre (volonté de surestimer la profondeur maximale par exemple), mais ce choix de la taille peut toujours être donné par les processus spatiaux quand on peut calculer les coordonnées des piqûres. Le nombre des blocs d'analyses peut être choisi de telle sorte qu'on arrive à une stabilisation d'estimation des paramètres.

Pour les approches GEV et POT, elles sont présentées comme des alternatives à la méthode de Gumbel. Cependant elles sont différentes dans leur élaboration. La première est une généralisation de la méthode de Gumbel et utilise les maxima par blocs pour estimer les paramètres, alors que la deuxième est basée sur un choix de seuil, et ce sont les valeurs dépassant ce seuil qui sont utilisés pour estimer les paramètres. Les méthodes d'estimation les plus utilisées dans ces deux approches, sont la méthode du maximum de vraisemblance MV et la méthode des moments de probabilité pondérés. Dans un premier temps, on a comparé les deux méthodes pour les deux approches. Pour les deux approches, la méthode MV présente de meilleures qualités en terme de biais et de variance pour les échantillons de grandes tailles. Ses performances se perdent lorsque l'indice extrême est proche de 0.4 en particulier pour les petites tailles, alors que celles de la méthode PWM restent acceptables dans cette zone. Cela conduit à privilégier la méthode PWM pour les petites tailles d'échantillon et la méthode MV pour les grandes. Enfin, en utilisant la loi normale et lognormale dont les paramètres ont été calculés à partir des profondeurs mesurées sur la plaque de 75 cm^2 , on a pu comparé les quantiles des différentes approches. Notons que ces deux lois appartiennent au domaine d'attraction de Gumbel, donc une comparaison avec les quantiles de la loi de Gumbel est intéressante. A partir de ces comparaisons on remarque que le passage de deux à trois paramètres augmente les incertitudes quand aux estimations des quantiles extrêmes et montre ainsi l'importance de connaître le bon domaine d'attraction. L'approche POT peut présenter une alternative très intéressante à l'approche GEV car en terme de biais et de variance, elle a présenté de meilleurs résultats. La difficulté avec cette méthode est le choix de seuil. Ici, on a utilisé la méthode Pickands, mais il existe d'autres méthodes plus complexes qui aboutissent à un choix meilleur du seuil et

ainsi une meilleure estimation des quantiles. Ces différents estimateurs des quantiles ont été utilisés pour estimer la profondeur maximale sur la plaque de 75 cm^2 (profondeur exacte connue). Cette fois-ci la variance de la méthode GEV est moins importante qu'en simulation. Cela est dû au fait que les observations appartiennent au domaine d'attraction de Weibull et non celui de Gumbel.

Ce travail de thèse a permis, de développer ces approches statistiques et ouvre en même temps la porte à plusieurs perspectives du point de vue expérimentale et mathématique.

Du point de vue expérimentale, on souhaite trouver des conditions expérimentales pour reproduire des densités spatiales comparables et des échantillons facile à analyser via un logiciel d'analyse d'image afin de constituer une base de données composée de coordonnées de piqûres et les profondeurs y associées. Dans ce travail, seule l'extrapolation spatiale a été utilisée. L'extrapolation dans le temps peut être développées mais pour cela il faut caler les conditions de reproduction des densités spatiales avant d'introduire le facteur du temps. Ce travail peut aussi être envisagé sur d'autres couples matériau/milieu.

Du point de vue mathématique, on peut s'intéresser à plusieurs problèmes :

- Traitement des distributions multimodales : Le phénomène de métastabilité empêche certaines piqûres à progresser ce qui peut entraîner l'existence de deux ou plusieurs populations de profondeurs de piqûres. Les méthodes développées dans ce travail ne sont pas adaptées à ce type de problème. Néanmoins, des travaux récents, adaptent cette théorie pour les populations mixtes. Cela peut constituer une perspective intéressante pour traiter ce phénomène de métastabilité.
- Seules les méthodes paramétriques ont été utilisées dans ce travail. Les méthodes non-paramétriques peuvent présenter une alternatives aux méthodes présentés ici. La méthode des noyaux par exemple est l'une des plus utilisés pour ce propos. Par contre, son couplage avec les valeurs extrêmes ne donne pas de bon résultats. Des améliorations de cette méthode pour l'adapter aux valeurs extrêmes ont été développées et peuvent être utilisées. On peut aussi adapter les approches bayésiennes utilisées dans d'autre domaines qui nécessite ce qu'on appelle un « avis d'expert » qui représente une information supplémentaire sur le phénomène et qui peut être incluse dans le traitement.
- La difficulté avec la méthode POT est le choix du seuil. Ici, on a utilisé la méthode de Pickands, mais il existe d'autres méthodes qui peuvent aboutir à un meilleur choix du seuil. Ce sont des méthodes basées sur la minimisation à la fois du biais et de variance.
- Enfin, le couplage du phénomène de la corrosion par piqûres avec la fatigue fait partie des perspectives de ce travail, en considérant que les piqûres les plus profondes peuvent constituer de véritables sites d'amorçage de fissures.

Annexes

Annexe A

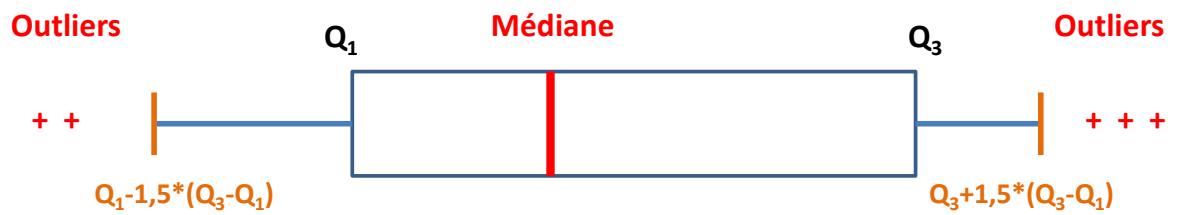


FIG. A.1: Schéma descriptif de la boîte à moustache : Q_1 Q_3 représentent respectivement le premier et le troisième quartile. Les outliers désignent des valeurs considérées comme extrêmes

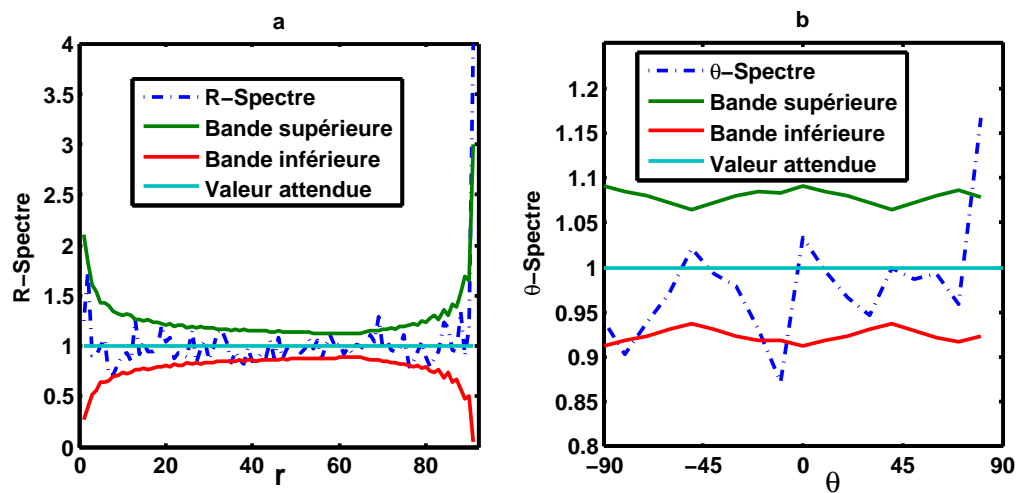


FIG. A.2: Spectres pour la plaque de position 3 issue de la bande 2 : a- $R - Spectre$, b- $\theta - Spectre$

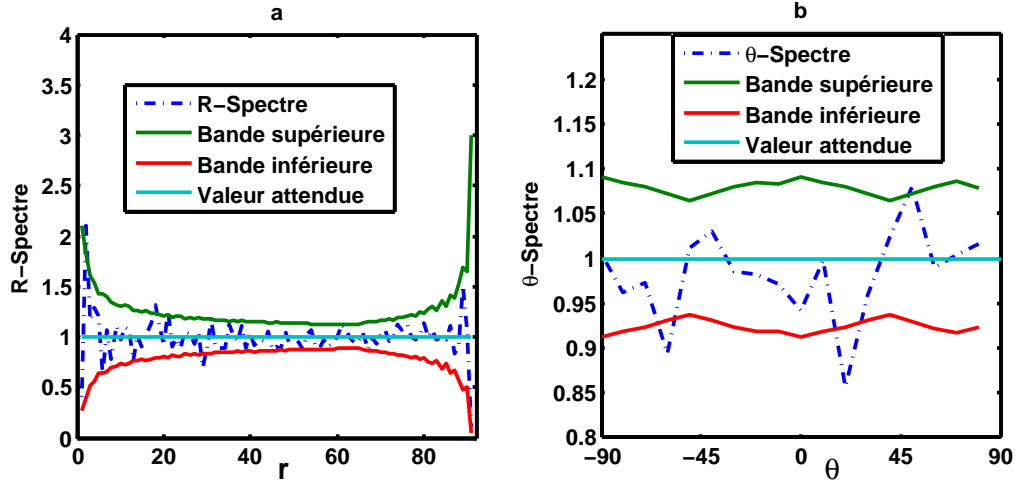


FIG. A.3: Spectres pour la plaque de position 4 issue de la bande 2 : a- $R - Spectre$, b- $\theta - Spectre$

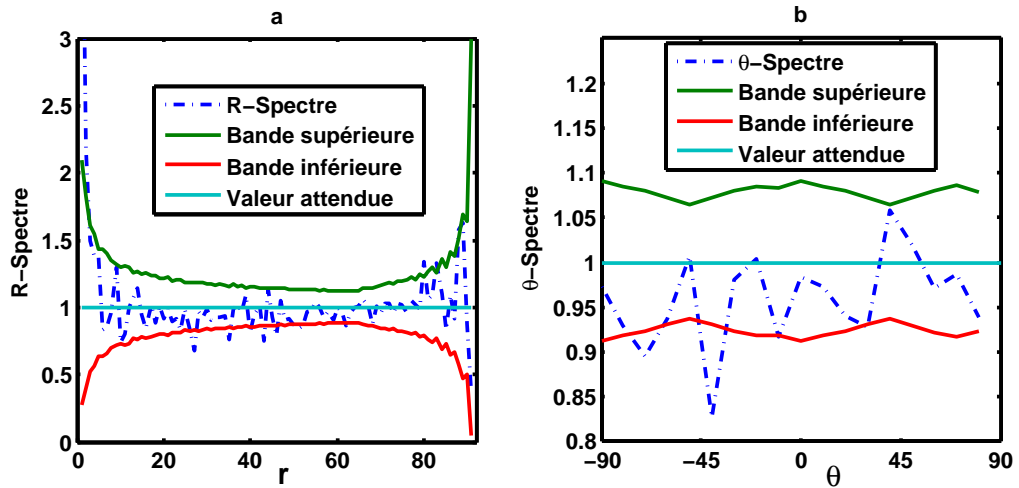


FIG. A.4: Spectres pour la plaque de position 3 issue de la bande 3 : a- $R - Spectre$, b- $\theta - Spectre$

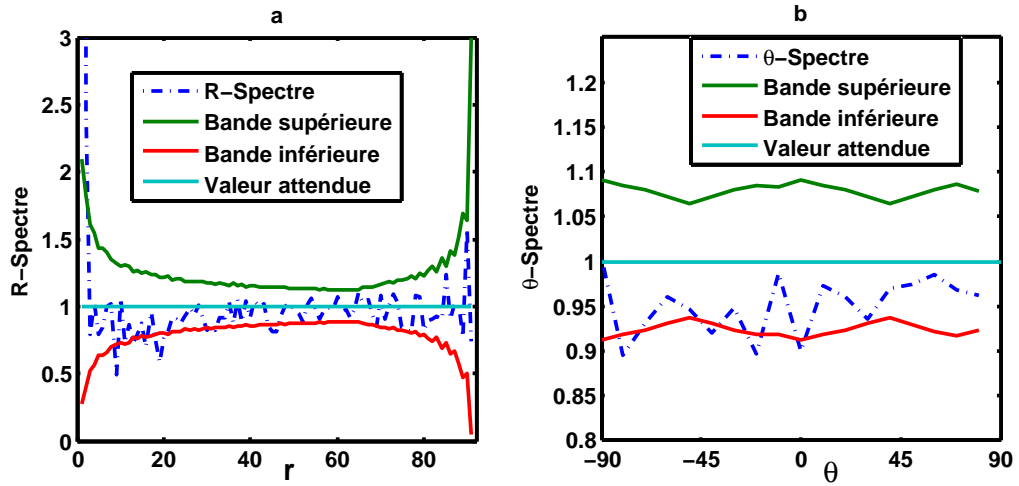


FIG. A.5: Spectres pour la plaque de position 4 issue de la bande 3 : a- R - Spectre, b- θ - Spectre

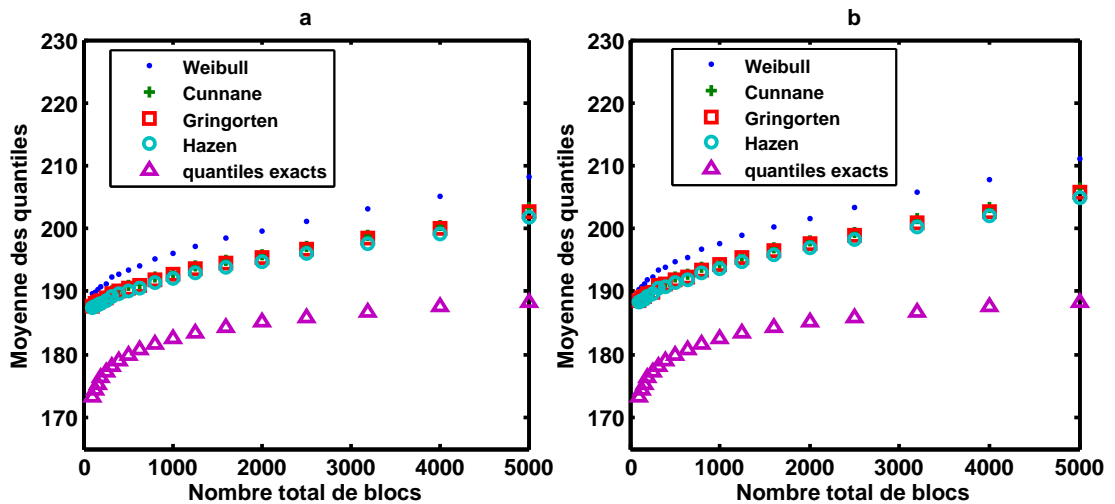


FIG. A.6: Moyennes des 1000 quantiles estimés à partir des profondeurs générées selon une loi normale avec $m = 30$ blocs : a-axes inversés, b-axes classiques

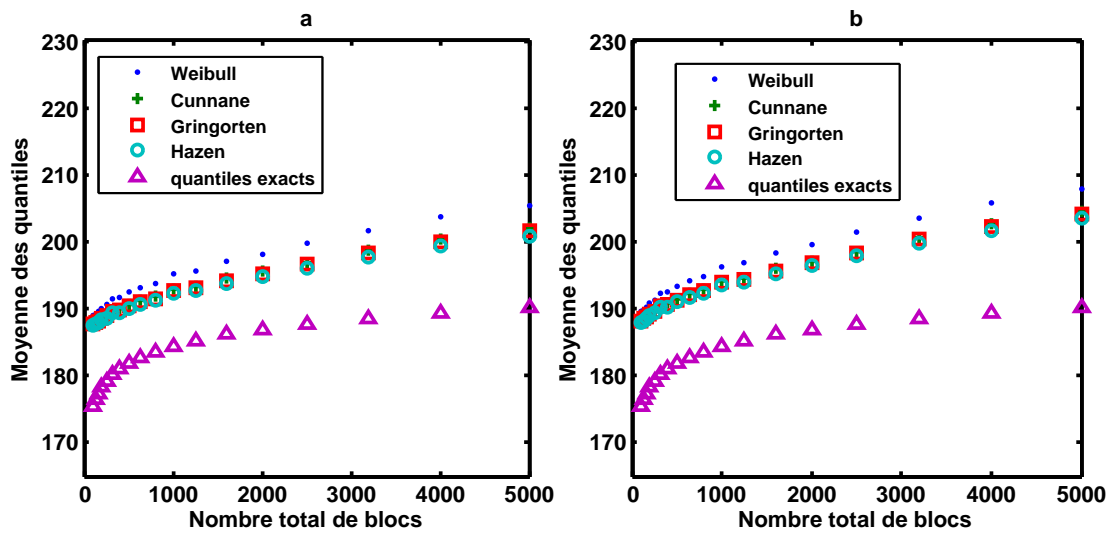


FIG. A.7: Moyennes des 1000 quantiles estimés à partir des profondeurs générées selon une loi normale avec $m = 50$ blocs : a-axes inversés, b-axes classiques

Annexe B

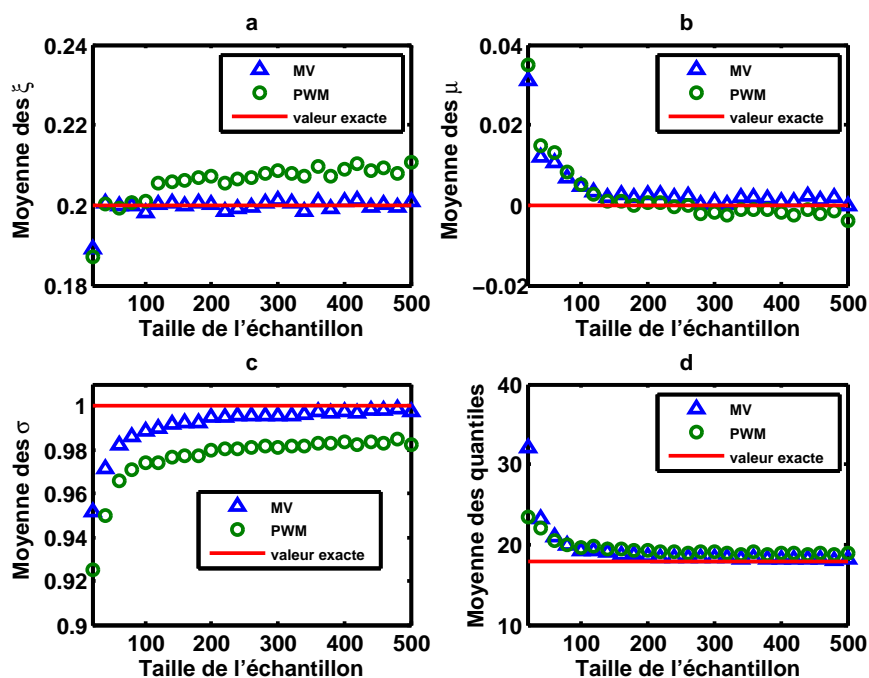


FIG. B.1: Moyennes des estimations pour les échantillons simulés avec $\xi = 0, 2$ en fonction de leurs tailles : (a)- Moyennes des indices extrêmes, (b)- Moyennes des paramètres de localisation μ , (c)- Moyennes des paramètres d'échelle σ , (d)- Moyennes des quantiles

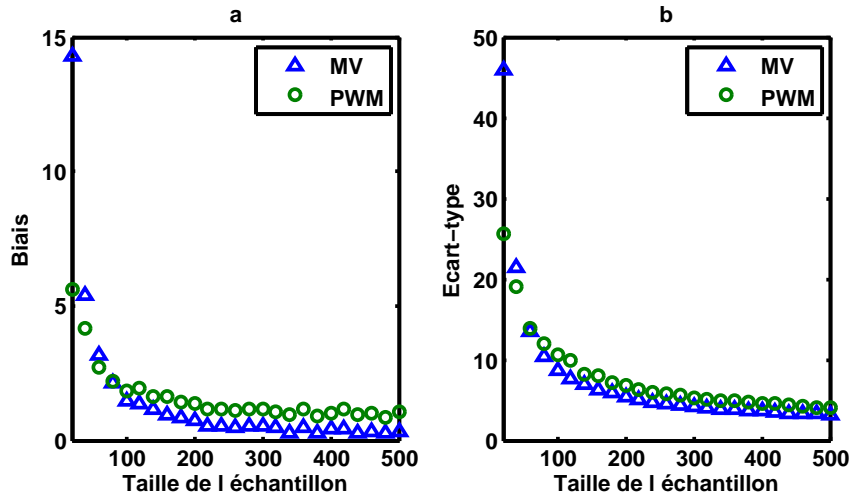


FIG. B.2: Echantillons pour $\xi = 0, 2$: a- Biais des quantiles, Ecart-types des quantiles

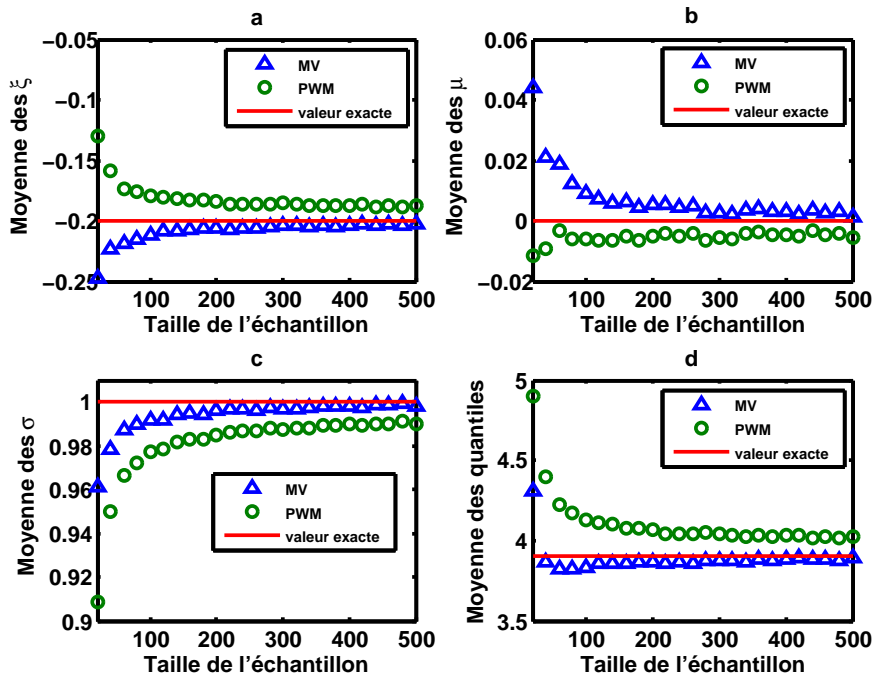


FIG. B.3: Moyennes des estimations pour les échantillons simulés avec $\xi = -0, 2$ en fonction de leurs tailles : (a)- Moyennes des indices extrêmes, (b)- Moyennes des paramètres de localisation μ , (c)- Moyennes des paramètres d'échelle σ , (d)- Moyennes des quantiles

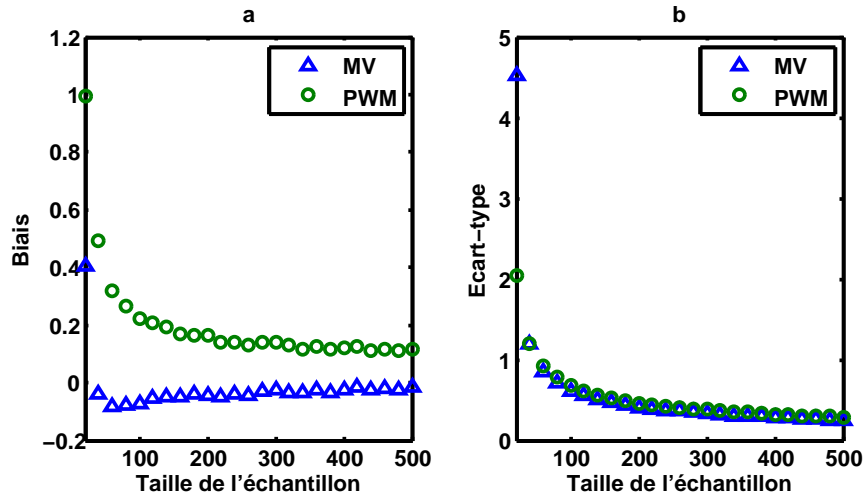


FIG. B.4: Echantillons pour $\xi = -0, 2$: a- Biais des quantiles, Ecart-types des quantiles

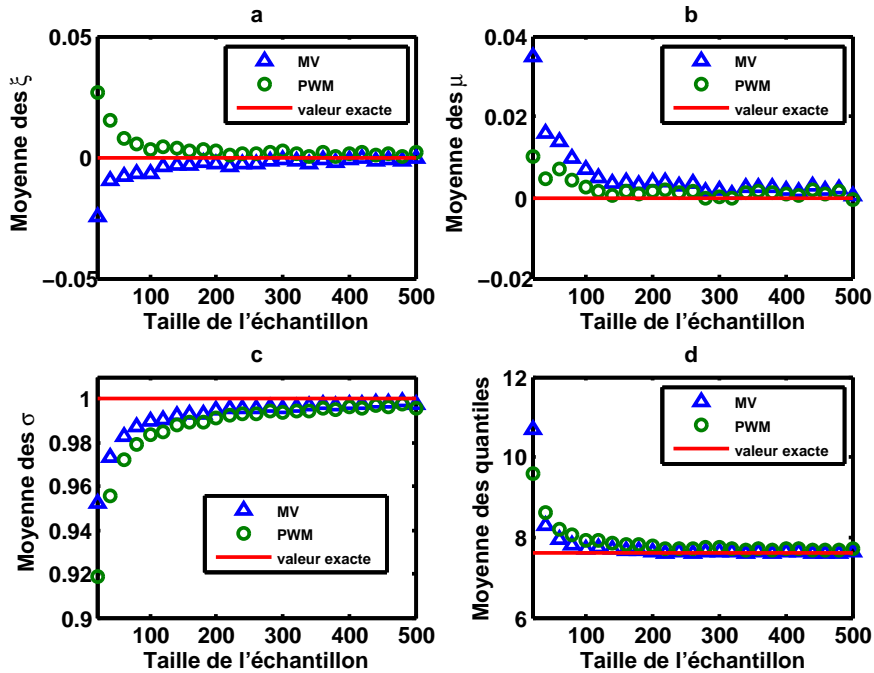


FIG. B.5: Moyennes des estimations pour les échantillons simulés avec $\xi = 0$ en fonction de leurs tailles : (a)- Moyennes des indices extrêmes, (b)- Moyennes des paramètres de localisation μ , (c)- Moyennes des paramètres d'échelle σ , (d)- Moyennes des quantiles

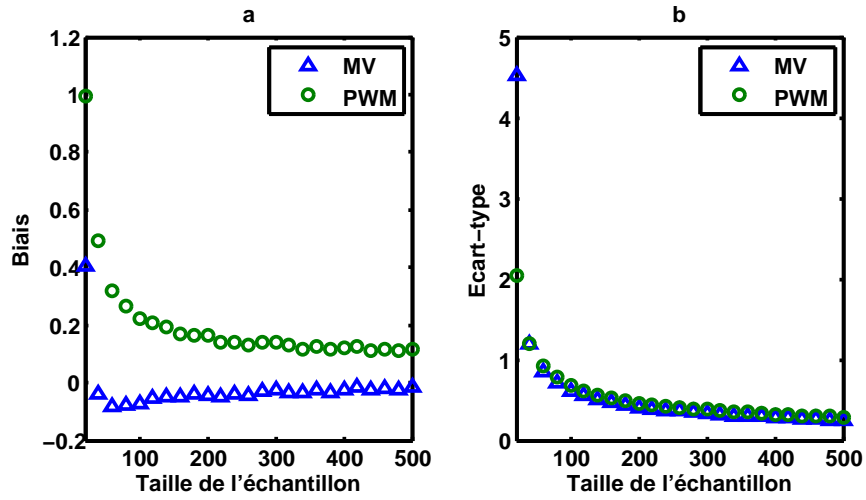


FIG. B.6: Echantillons pour $\xi = 0$: a- Biais des quantiles, Ecart-types des quantiles

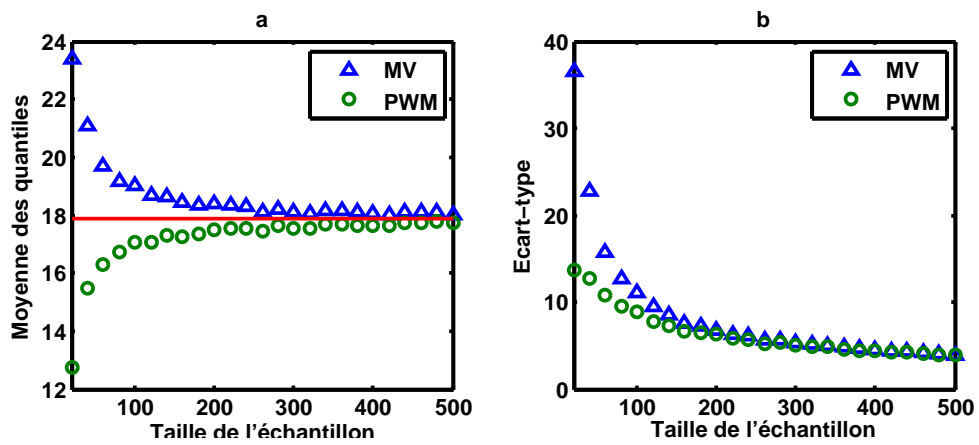


FIG. B.7: Echantillons pour $\xi = 0, 2$: a- Moyennes des quantiles, b- Ecart-types des quantiles

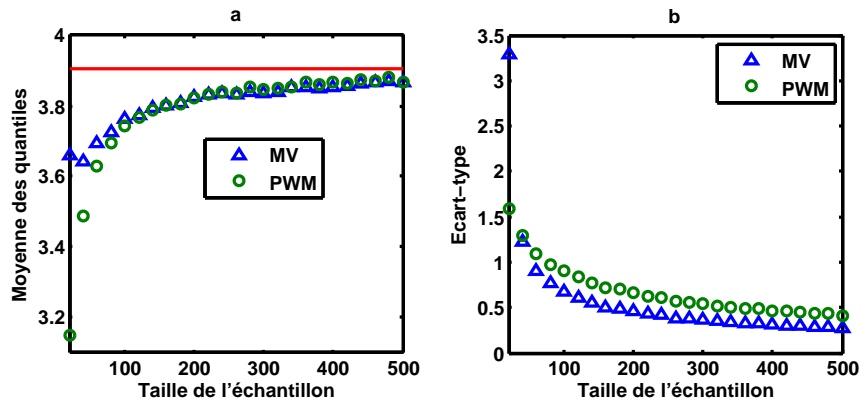


FIG. B.8: Echantillons pour $\xi = -0,2$: a- Moyennes des quantiles, b- Ecart-types des quantiles

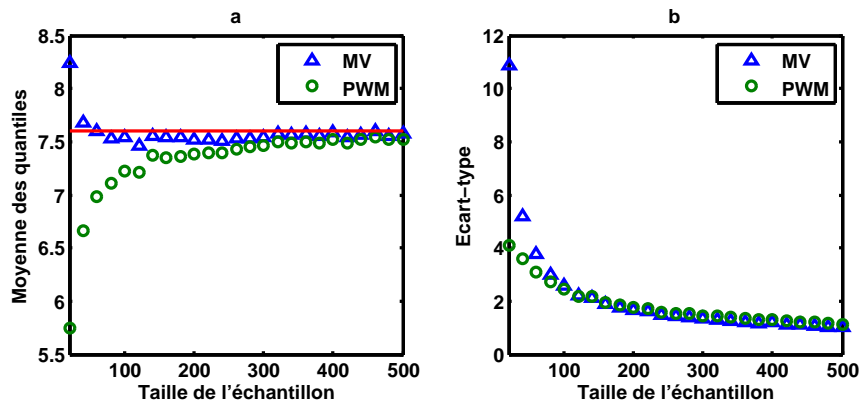


FIG. B.9: Echantillons pour $\xi = 0$: a- Moyennes des quantiles, b- Ecart-types des quantiles

Glossaire

E_r	Potentiel de Repassivation
E_p	Potentiel de Piqûration
S	Surface Objective
s	Surface d'Analyse
ANOVA	Analyse de variance
ACP	Analyse en composante principal
AFD	Analyse factorielle discriminante
AFC	Analyse factorielle des correspondances
GLD	loi généralisée de lambda distribution
F	Statistique de Fisher
T_B	Statistique de Bartlett
T_{KW}	Statistique de Kruskal-Wallis
TVE	Théorie des Valeurs Extrêmes
GEV	Valeurs Extrêmes Généralisées (Generalized Extreme Value)
POT	Dépassement de seuil (Peaks Over Threshold)
N_p	Variable aléatoire correspondant au nombre de piqûres
T_a	Variable aléatoire correspondant au temps d'amorçage
$\mathbb{E}[\cdot]$	Opérateur de l'Espérance
$\sigma[\cdot]$	Opérateur de l'écart-type
p	Probabilité de dépassement
X	La variable aléatoire qui désigne la profondeur des piqûres
X_i	Variable aléatoire de même loi que X
$X_{i:n}$	ième Statistique d'ordre de l'échantillon
Y_i	Valeur maximale au bloc i
$Y_{i:m}$	ième statistique d'ordre des maxima par bloc
F	Fonction de répartition de X
\hat{F}	Fonction de répartition empirique
a_n	Suite normalisante (d'échelle)
b_n	Suite normalisante (de localisation)
<i>i.i.d</i>	Indépendantes et identiquement distribuées
ξ	Indice extrême
$G_\xi(x)$	Loi GEV
ϕ_α	Distribution type Fréchet
ψ_α	Distribution type Weibull
$\Lambda(x)$	Distribution type Gumbel
MDA	Domaine d'Attraction du Maximum
<i>v.a</i>	Variable aléatoire

$w(F)$	Point terminal de la distribution F
$Q(\cdot)$	Fonction quantile
F^{\leftarrow}	Inverse généralisée
F^{-1}	Inverse classique
$r(t)$	Fonction du hasard
$\ell(x)$	Fonctions à variations lentes
$\mathcal{N}(m, \sigma^2)$	Loi normale de moyenne m et d'écart-type σ
μ_0	Paramètre de localisation pour la loi de Gumbel
μ	Paramètre de localisation pour la loi GEV
σ_0	Paramètre d'échelle pour la loi de Gumbel
σ	Paramètre d'échelle pour la loi GEV
n	Taille de l'échantillon des observations
m	Taille de l'échantillon des maxima par blocs
$G_0(y)$	Distribution de Gumbel de paramètres μ et σ
\bar{Y}_m	Moyenne empirique (de l'échantillon)
S_Y^2	Variance empirique
S_Y^{*2}	Variance empirique corrigée
m_k	Moment centré d'ordre k
$\mathbb{V}[\cdot]$	Opérateur de la variance
$\nabla(\cdot)$	Opérateur du gradient
$(\cdot)^t$	Opérateur de transposition
MV	Méthode d'estimation du maximum de vraisemblance
PWM	Méthode d'estimation des moments de probabilité pondérés
$\hat{\mu}_0^{reg}, \hat{\sigma}_0^{reg}$	Paramètres de la loi de Gumbel estimés par régression linéaire
\hat{q}_0^{reg}	Quantile de la loi de Gumbel estimé par régression linéaire
$\hat{\mu}_0^{mv}, \hat{\sigma}_0^{mv}$	Paramètres de la loi de Gumbel estimés par MV
\hat{q}_0^{mv}	Quantile de la loi de Gumbel estimé par MV
$\hat{\xi}^{mv}, \hat{q}^{mv}$	Indice extrême et quantile de la loi GEV estimé par MV
$\hat{\mu}_0^{mom}, \hat{\sigma}_0^{mom}$	Paramètres de la loi de Gumbel estimés par la méthode des moments
\hat{q}_0^{mom}	Quantile de la loi de Gumbel estimé par la méthode des moments
V_{mom}	Matrice de variance-covariance des paramètres pour la méthode des moments
$Cov(X, Y)$	Coefficient de corrélation de X et Y
$\hat{\sigma}^{mv}, \hat{\mu}^{mv}$	Paramètres de la loi GEV estimés par MV
$\hat{\sigma}^{pwm}, \hat{\mu}^{pwm}$	Paramètres de la loi GEV estimés par PWM
$\hat{\xi}^{pwm}, \hat{q}^{mv}$	Indice extrême et quantile de la loi GEV estimés par PWM
$\Gamma(\cdot)$	Fonction Gamma d'Euler
T_u	Période de retour dans le modèle POT
MEF	Fonction de la moyenne des excès (MEF)
u	Seuil pour la méthode POT
k	Nombre de valeurs extrêmes au delà du seuil u
$Z_{j,k}$	jème statistique d'ordre des dépassements des k plus grandes valeurs
k^*	Nombre de valeurs extrêmes optimal au delà du seuil u
$\hat{\xi}_u^{mom}, \hat{\sigma}_u^{mom}$	Indice extrême et paramètre d'échelle Estimés par la méthode des moments pour la loi GPD
\hat{q}_u^{mom}	Quantile de la loi GPD estimé par la méthode des moments
$\hat{\xi}_u^{mv}, \hat{\sigma}_u^{mv}$	Indice extrême et paramètre d'échelle Estimés par MV

\hat{q}_u^{mv}	Quantile de la loi GPD estimé par MV
$\hat{\xi}_u^{pwm}, \hat{\sigma}_u^{pwm}$	Indice extrême et paramètre d'échelle Estimés par PWM
\hat{q}_u^{pwm}	Quantile de la loi GPD estimé par PWM
$\hat{\xi}_{k,n}^{(M)}$	Estimateur des moments de Dekkers
$\hat{\xi}_{k,n}^{(UH)}$	Estimateur de Beirlant
s	Point appartenant au plan
X_s	Processus spatial
$X(s)$	Valeur prise par le processus spatial au point s
\mathbf{R}	Région spatiale d'étude
$\lambda_X(s)$	Intensité du premier ordre du processus spatial X_s
λ_X	Intensité du premier ordre constante
$N_X(ds)$	Nombre d'événements dans ds
$\gamma_X(s_i, s_j)$	Intensité du second ordre aux points s_i et s_j
CSR	Structure complètement aléatoire
HPP	Processus de Poisson Homogène (spatial)
$\mathcal{P}(\lambda_X)$	Loi de Poisson unidimensionnel de paramètre λ_X
$\mathcal{U}_{[0,1]}$	v.a uniforme dans l'intervalle $[0, 1]$
u_i	Réalisation de $\mathcal{U}_{[0,1]}$
$\mathcal{B}(n, p)$	Loi binomiale de paramètres n et p
$\binom{n}{k}$	Nombre de combinaisons de k parmi n
MTCP	Processus de Thomas Modifié pour les Cluster
λ_p	Paramètre du processus <i>HPP</i> pour les parents dans un processus MTCP
μ_c	Paramètre de la loi de Poisson pour le nombre des descendants
σ_c	Ecart type pour la loi normale bivariée des coordonnées des descendants
PIS	Processus d'inhibition séquentiel
δ	Distance minimale entre événements dans une structure régulière
N_r	v.a pour le nombre de voisins à distance r
$K(r)$	Fonction de Ripley
$\hat{K}(r)$	Estimation de la fonction de Ripley sans correction de bord
$\hat{K}_c(r)$	Estimation de la fonction de Ripley avec correction de bord
w_{ij}	Proportion du cercle de centre i et passant par j
N_{simul}	Nombre de simulation Monte-Carlo
κ_X	Fonction de covariance spectrale
f_X	Fonction de densité spectrale
\hat{f}_X	Périodogramme
$\hat{f}_R(r)$	$R - Spectre$
$\hat{f}_\Theta(\theta)$	$\theta - Spectre$
χ_m^2	Loi de Chi-Deux à m degrés de liberté
I	Indice de dispersion pour la méthode des quadrats
$\lambda_X(x, y)$	Intensité d'un processus spatial non homogène
NHPP	Processus de Poisson non homogène
$\hat{\lambda}_X$	Estimation de l'intensité λ_X
Θ	Vecteur de paramètres de l'intensité $\lambda_X(x, y)$
$\hat{K}_{NH}(r)$	Estimation de la fonction de Ripley dans le cas non-stationnaire
$\mu_{NH}(x, y)$	Intensité des descendants dans le cas non-stationnaire
M	v.a des marques

$\{m_i\}_{i=1,\dots,n}$	Réalisation de la v.a M
f_M	Fonction de densité de probabilité des marques
μ_m	Espérance de la v.a des marques
C_{mm}	Espérance du produit de deux marques
K_{mm}	Fonction de corrélation des marques
C_t	Fonction produit de deux marques associée à la fonction test t
K_t	Fonction de corrélation des marques associée à la fonction test t
μ_t	Constante de normalisation associée à la fonction test t
\hat{K}_{mm}	Estimateur de K_{mm}
M^T	v.a du maximum pour $(G_y)^T$
CV	Coefficient de Variation
EM	Espérance du Maximum
$\hat{q}_{p,k}^{pot}$	Quantile de la méthode POT avec k valeurs dépassant le seuil
q_{end}	Point terminal

Bibliographie

- [Bad07] A. Baddeley. *Spatial Point Process and their Applications in Stochastic Geometry p :1-75*. Springer Berlin,Heidelberg, 2007.
- [Bar64] M. S. Bartlett. The spectral analysis of two dimensional point processes. *Biometrika*, 51 :299–311, 1964.
- [Bar04] E. Bardal. *Corrosion and Protection*. Springer, 2004.
- [BdH74] A. Balkema and L. de Haan. Residual life time at great age. *The Annals of probability*, 2(5) :792–804, 1974.
- [BdH78a] A. Balkema and L. de Haan. Limit distributions for order statistics. *i. SIAM Theory of probability and its applications*, 23(1) :77–92, 1978.
- [BdH78b] A. Balkema and L. de Haan. Limit distributions for order statistics. *ii. SIAM Theory of probability and its applications*, 23(2) :341–358, 1978.
- [BG95] L. Balázs and J-F. Gouyet. Two-dimensional pitting corrosion of aluminium thin layers. *Physica A*, 217 :319–338, 1995.
- [BGST04] J. Beirlant, Y. Goegebeur, J. Segers, and J. Teugels. *Statistics of Extremes Thoery and Application*. John Wiley & Sons, Ltd, 2004.
- [BOHS05] N. D. Budiansky, L. Organ, J. L. Hudson, and J. R. Scully. Detection of interactions among localised pitting sites on stainless steel using spatial statistics. *J. Electrochem. Soc*, 152 :B152–B160, 2005.
- [BT00] Adrian Baddeley and Rolf Turner. Practical maximum pseudolikelihood for spatial point patterns. *Australian and New Zealand Journal of Statistics*, 42 :283–322, 2000.
- [BV64] J. Bernier and R. Veron. Sur quelques difficultés rencontrées dans l’estimation d’un débit de crue de probabilité donnée. *revue de statistique appliquée*, 12, 1964.
- [BVT96] J. Beirlant, P. Vynckier, and J. L. Teugels. Tail index estimation, pareto quantile plots, and regression diagnostics. *J. Amer. Statist. Assoc*, 91 :1659–1667, 1996.
- [CG08] J. Lopez De La Cruz and M. A Gutiérrez. Spatial statistics of pitting corrosion patterning : Quadrat counts and the non-homogeneous poisson process. *Corros. Sci*, 50 :1441–1448, 2008.
- [CL07] P. A. Cornillon and E. M. Lober. *Régression théorie et application*. Springer-Verlag, Paris, 2007.

- [CLKG07] J. Lopez De La Cruz, R. H. A. Lindelauf, L. Koene, and M. A. Gutiérrez. Stochastic approach to the spatial analysis of pitting corrosion and pit interaction. *Electrochem. Commun.*, 9 :325–330, 2007.
- [Cre00] N. A. C. Cressie. *Statistics for spatial data*. Wiley, New York, 2000.
- [Cun78] C. Cunnane. Unbiased plotting positions. *J Hydrol*, 37 :205–222, 1978.
- [DBB94] F. Dabosi, G. Beranger, and B. Baroux. *Corrosion localisée*. Les éditions de Physique, 1994.
- [DEdH89] A. L. M. Dekkers, J. H. J. Einmahl, and L. de Haan. A moment estimator for the index of an extreme-value distribution. *Ann. Statist*, 17 :1833–1855, 1989.
- [DGG03] J. Dieblot, M. Garrido, and S. Girard. Asymptotic normality of the et method-application to the et test. *C. R. Acad. Sci. Paris, Ser. I 337*, 2003.
- [dHR93] L. de Haan and H. Rootzén. On the estimation of high quantiles. *J. Statist. Plann. Inference*, 35 :1–13, 1993.
- [Dig03] P. J. Diggle. *Statistical analysis of spatial point patterns*. Oxford University Press Inc., New York, 2003.
- [DLS06] J-J Dreesbeke, M. Lejeune, and G. Saporta. *Analyse Statistique des Données Spatiales*. Editions TECHNIP, Paris, 2006.
- [DS90] A. C. Davison and R. L. Smith. Models for exceedances and high thresholds. *J.R. Statist. Soc, B 52,3* :393–442, 1990.
- [DS98] N. R. Draper and H. Smith. *Applied regression analysis*. Wiley series in probability and statistics, New York, 1998.
- [EAD02] M. A. El-Aroui and J. Dieblot. On the use of the peaks over threshold method for estimating out-of-sample quantiles. *Computational Statistics & Data Analysis*, 39, 2002.
- [EKM97] P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance*. Springer Verlag, Berlin, 1997.
- [Fie92] P. Fievet. *Etude statistique de la corrosion par piqûres, par analyse d'images numérisées, d'aciers inoxydables austénitiques : Effet de l'implantation ionique de molybdène et d'azote*. Thèse de doctorat. Institut National Polytechnique de Toulouse, 1992.
- [Gal87] J. Galambos. *The Asymptotic Theory of Extreme order statistics*. John Wiley & Sons, Inc, 1987.
- [Gal95] J. Galambos. The development of the mathematical theory of extremes in the past half-century. *Theory of the Probability and its Applications*, 39 :234–248, 1995.
- [Gar02] M. Garrido. *Modélisation des événements rares et estimation des quantiles extrêmes, méthodes de sélection de modèles pour les queues de distribution*. Thèse de doctorat. Université Joseph Fourier, 2002.

- [Gle20] H. A. Gleason. Some applications of the quadrat method. *Bulletin of the Torrey Botanical Club*, 47 :21–33, 1920.
- [GLMW79] J. A. Greenwood, J. M. Landwehr, M. C. Matalas, and J. R. Wallis. Probability weighted moments : definition and relation to parameters of several distribution expressible in inverse form. *Water Resources Research*, 15 :1049–1054, 1979.
- [Gor00] F. Goreaud. *Apports de l'analyse de la structure spatiale en forêt tempérée à l'étude et la modélisation des peuplements complexes*. Thèse de doctorat. EN-GREF, Nancy, 2000.
- [GPC95] M. Goulard, L. Pagès, and A. Cabanettes. Marked point process : Using correlation functions to explore a spatial data set. *Biometrical journal*, 37 :837–853, 1995.
- [Gra06] C. Grasland. *Discontinuités statistiques et discontinuités spatiales : L'exemple des inégalités de richesse par habitant en Italie (1951-1991) in Analyse Statistique des Données Spatiales*. Editions TECHNIP, 2006.
- [Gri63] I. I. Gringorten. A plotting rule for extreme probability paper. *J Geophys Res*, 68 :813–814, 1963.
- [Gum58] E.J. Gumbel. *Statistics of Extremes*. Columbia Univ. Press, 1958.
- [Haz14] A. Hazen. Storage to be provided in impounding reservoirs for municipal water supply. *Trans Am Soc Civ Eng*, 77 :1547–1550, 1914.
- [HCH⁺06] G. Harker, S. Cole, J. Helly, C. Frenk, and A. Jenkins. A marked correlation function analysis of halo formation times in the millennium simulation. *Monthly Notices of the Royal Astronomical Society*, 367 :1039–1049, 2006.
- [Hil75] B. M. Hill. A simple general approach to inference about the tail of a distribution. *Ann. Statist*, 3 :1163–1174, 1975.
- [HM82] G. Hahn and W. Meeker. Pitfalls and practical considerations in product life analysis-part i : Basic concepts and dangers of extrapolation. *Journal of Quality Technology*, 14(3), 1982.
- [Hos84] J. R. M. Hosking. Testing whether the shape parameter is zero in the generalized extreme-value distribution. *Biometrika*, 71(2) :367–374, 1984.
- [HWW85] J. R. M. Hosking, J. R. Wallis, and E. F. Wood. Estimation of the generalized extreme-value distribution by the method of probability-weighted moments. *Technometrics*, 27 :251–261, 1985.
- [IPSS08] J. Illian, A. Penttinen, H. Stoyan, and D. Stoyan. *Statistical Analysis and Modelling of Spatial Point Patterns*. John Wiley & Sons Ltd , Chichester, 2008.
- [Jen55] A. F. Jenkinson. The frequency distribution of the annual maxima (or minimum) values of meteorological elements. *Quarterly Journal of the Royal Meteorological Society*, 81, 1955.

- [KBG09] S. Kim, J. T. Burns, and R. P. Gangloff. fatigue crack formation and growth from localized corrosion in al-zn-mg-cu. *Engineering Fracture Mechanics*, 78 :651–667, 2009.
- [KN00] S. Kotz and S. Nadarajah. *Extreme Value Distribution : Theory and Applications*. World Scientific, 2000.
- [Kow94] M. Kowaka. *Introduction to Life Prediction of Industrial Plant Materials : Application of the Extreme Value Statistical Method for Corrosion Analysis*. Allerton Press, INC, 1994.
- [Lan03] D. Landolt. *Corrosion et chimie de surfaces des métaux*. Presses Polytechniques et Universitaires Romandes, 2003.
- [Law01] A. B. Lawson. *Statistical methods in spatial epidemiology*. John Wiley & Sons Ltd , Chichester, 2001.
- [Lei54] J. Leiblein. A new method of analyzing extreme-value data. *NACA, Technical Note 3053*, 1954.
- [LLR83] M.R. Leadbetter, G. Lindgren, and H. Rootzén. *Extremes and related properties of random sequences and series*. Springer Verlag, New York, 1983.
- [Mar02] P. Marcus. *Pitting Corrosion of Metals*. CRC Press, 2002.
- [MB03] G. Matthys and J. Beirlant. Estimating extreme value index and high quantiles with exponential regression models. *Statistica Sinica*, 13 :853–880, 2003.
- [Mel08] R. E. Melshers. Extreme value statistics and long-term marine pitting corrosion of steel. *Probabilistic Engineering Mechanics*, 23 :482–488, 2008.
- [MM02] W. L. Martinez and A. R. Martinez. *Computational Statistics Handbook with Matlab*. Chapman & Hall/CRC, USA, 2002.
- [MR96] M. A. Mugglestone and E. Renshaw. A practical guide to the spectral analysis of spatial point processes. *Comp. Stat. & Data Anal*, 21 :43–65, 1996.
- [MR01] M. A. Mugglestone and E. Renshaw. Spectral tests of randomness for spatial point patterns, environmental and ecological statistics. *Environmental and Ecological Statistics*, 8 :237–251, 2001.
- [MUM98] J. Mateu, J. L. Usó, and F. Montes. The spatial pattern of a forest ecosystem. *Ecological modeling*, 108 :163–174, 1998.
- [MV04] J. Martin-Vide. Spatial distribution of a daily precipitation concentration index in peninsular Spain. *Int. J. Climatol*, 24 :959–971, 2004.
- [NBLI03] D. Najjar, M. Bigerelle, C. Lefebvre, and A. Iost. A new approach to predict the pit depth extreme value of a localized corrosion process. *ISIJ*, 43 :720–725, 2003.
- [NLB⁺02] D. Najjar, C. Lefebvre, L. Bourdeau, M. Bigerelle, and A. Iost. Caractérisation et modélisation topographiques de piqûres de corrosion. *Congrès Matériaux*, 2002.

- [NRC96] D. G. Harlow N. R. Cawley. Spatial statistics of particles and corrosion pits in 2024-t3 aluminium alloy. *J. Mater. Sci*, 31 :5127–5134, 1996.
- [Pic75] J. Pickands. Statistical inference using extreme order statistics. *The Annals of Statistics*, 3 :119–131, 1975.
- [PMA01] C. Perrin, C. Michel, and V. Andréassian. Does a large number of parameters enhance model performance? comparative assessment of common catchment model structures on 429 catchments. *Journal of Hydrology*, 242 :275–301, 2001.
- [PR08] R. M. Pidaparti and A. S. Rao. Analysis of pits induced stresses due to metal corrosion. *Corrosion Science*, 50 :1932–1938, 2008.
- [Res87] S. I. Resnick. *Extreme Values Regular Variation and Point Processes*. Springer Verlag, New York, 1987.
- [RF83] E. Renshaw and E. D. Ford. interpretation of process from pattern using two-dimensional spectral analysis : methods and problems of interpretation. *Appl. Statist*, 32 :51–63, 1983.
- [RF84] E. Renshaw and E. D. Ford. The description of spatial pattern using two-dimensional spectral analysis. *Vegetatio*, 56 :75–85, 1984.
- [Rip76] B. D. Ripley. The second order analysis of stationnary point process. journal of applied probability. *Journal of applied probability*, 13 :255–266, 1976.
- [Rip77] B. D. Ripley. Modelling spatial patterns. journal of the royal statistical society. *Journal of the royal statistical society*, B 39 :172–212, 1977.
- [Ros85] D. Rosbjerg. Estimation in partial duration series with independant ans dependant peak values. *J. Hydrology*, 76 :183–195, 1985.
- [RT97] R. D. Reiss and M. Thomas. *Statistical Analysis of Extreme Values, with Applications to Insurance, Finance, Hydrology and Other Fields*. Birkhäuser, Basel, 1997.
- [Sap06] G. Saporta. *Probabilités analyses des données et Statistiques*. Editions TECHNIP, Paris, 2006.
- [SG05] O. Schabenberger and C. A. Gotway. *Statistical methods for spatial data analysis*. CRC press, USA, 2005.
- [Shi91] T. Shibata. Evaluation of corrosion failure by extreme value statistics. *ISIJ*, 31, No.2 :115–121, 1991.
- [Shi96] T. Shibata. Statistical ans stochastic approaches to localized corrosion. *Corrosion*, 52, No.11 :813–830, 1996.
- [Smi85] R. L. Smith. Maximum likelihood estimation in a class of nonregular cases. *Biometrika*, 72 :67–90, 1985.
- [SS86] Z. Szklarska-Smialowska. *Pitting Corrosion of Metals*. National Association of Corrosion Engineers, 1986.

- [ST76] T. Shibata and T. Takeyama. Pitting corrosion as a stochastic process. *Nature*, 260, 1976.
- [ST77] T. Shibata and T. Takeyama. Stochastic theory of pitting corrosion. *Corrosion*, 33, 1977.
- [Sto84] D. Stoyan. On correlatins of marked point processes. *Math. Nachr*, 116 :197–207, 1984.
- [Wei39] W. Weibull. A statistical theory of strength of materials. *Roy. Swedish Inst. Eng. Res*, 151 :1, 1939.