



HAL
open science

Exploitation du signal pénétrométrique pour l'aide à l'obtention d'un modèle de terrain

Carlos Sastre Jurado

► **To cite this version:**

Carlos Sastre Jurado. Exploitation du signal pénétrométrique pour l'aide à l'obtention d'un modèle de terrain. Génie civil. Université Clermont Auvergne [2017-2020], 2018. Français. NNT : 2018CLFAC003 . tel-01842259

HAL Id: tel-01842259

<https://theses.hal.science/tel-01842259v1>

Submitted on 18 Jul 2018

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.

UNIVERSITÉ CLERMONT AUVERGNE

ÉCOLE DOCTORALE
SCIENCES POUR L'INGÉNIEUR DE CLERMONT-FERRAND

THÈSE

Présentée par

Carlos SASTRE JURADO

Ingénieur Civil – ETSICCP Université de Grenade (UGR)

Pour obtenir le grade de
DOCTEUR D'UNIVERSITÉ
SPECIALITÉ : GÉNIE CIVIL

**EXPLOITATION DU SIGNAL PÉNÉTROMÉTRIQUE POUR L'AIDE À
L'OBTENTION D'UN MODÈLE DE TERRAIN**

Soutenue publiquement le 7 Février 2018 devant le jury :

M. ou Mme	F. Collin	Président
	M. Arroyo	Rapporteur
	G. Auvinet-Richard	Rapporteur
	C. Jacquard	Examineur
	M. Benz-Navarrete	Examineur
	P. Breul	Directeur de thèse
	C. Bacconnet	Co-directeur
	R. Gourvès	Invité

Résumé

Ce travail porte sur la reconnaissance des sols à faible profondeur grâce aux données de résistance de pointe recueillies à l'aide de l'essai de pénétration dynamique à énergie variable, Panda®. L'objectif principal est d'étudier et de proposer un ensemble d'approches dans le cadre d'une méthode globale permettant d'exploiter les mesures issues d'une campagne de sondages Panda afin de bâtir un modèle géotechnique du terrain.

Ce manuscrit est structuré en quatre parties, chacune abordant un objectif spécifique :

- dans un premier temps, on rappelle les principaux moyens de reconnaissance des sols, notamment l'essai de pénétration dynamique Panda. Ensuite on donne un bref aperçu sur le modèle géotechnique et les techniques mathématiques pour décrire l'incertitude dans la caractérisation des propriétés du sol ;
- la deuxième partie porte sur l'identification automatique des unités homogènes du terrain, à partir du signal pénétrométrique Panda. Suite à l'étude réalisée sur l'identification "experte" des couches à partir du signal Panda, des approches statistiques basées sur une fenêtre glissante ont été proposées. Ces techniques ont été étudiées et validées sur la base d'un protocole d'essais en laboratoire et sur des essais effectués en sites naturels et en conditions réelles ;
- la troisième partie porte sur l'identification automatique des matériaux composant les unités homogènes détectées dans le signal Panda à partir des méthodes proposées en partie II. Une méthode de classification automatique basée sur des réseaux de neurones artificiels a été proposée et appliquée aux deux cas d'étude : la caractérisation de sols naturels et la classification d'un matériau granulaire argileux industrialisé (bentonite) ;
- Enfin, la dernière partie est consacrée à la production d'un modèle de terrain basé sur la modélisation et la simulation de la résistance de pointe dynamique au moyen de fonctions aléatoires de l'espace. Cette modélisation est basée sur une approche par champs aléatoires conditionnés par les sondages Panda du terrain. Sa mise en œuvre a été étudiée pour un terrain expérimental situé dans la plaine deltaïque méditerranéenne en Espagne. Des études complémentaires en vue de raffiner cette démarche ont été réalisées pour un deuxième site expérimental dans la plaine de la Limagne en France.

Mots clés : reconnaissance des sols - modèle de terrain - techniques de stratigraphie statistiques - classification automatique - réseaux de neurones - simulation conditionnelle - pénétromètre dynamique Panda.

Abstract

This research focuses on the site characterization of shallow soils using the dynamic cone penetrometer Panda® which uses variable energy. The main purpose is to study and propose several techniques as part of an overall method in order to obtain a ground model through a geotechnical campaign based on the Panda test.

This work is divided into four parts, each of them is focused on a specific topic :

- first of all, we introduce the main site characterization techniques, including the dynamic penetrometer Panda. Then, we present a brief overview of the geotechnical model and the mathematical methods for the characterization of uncertainties in soil properties ;
- the second part deals with the automatic identification of physical homogeneous soil units based on penetration's mechanical response of the soil using the Panda test. Following a study about the soil layers identification based only on expert's judgment, we have proposed statistical moving window procedures for an objective assessment. The application of these statistical methods has been studied for the laboratory and in situ Panda test ;
- the third part focuses on the automatic classification of the penetration curves in the homogeneous soil units identified using the statistical techniques proposed in part II. An automatic methodology to predict the soil grading from the dynamic cone resistance using artificial neural networks has been proposed. The framework has been studied for two different research problems : the classification of natural soils and the classification of several crushed aggregate-bentonite mixtures ;
- finally, the last chapter was devoted to model the spatial variability of the dynamic cone resistance q_d based on random field theory and geostatistics. In order to reduce uncertainty in the field where Panda measurements are carried out, we have proposed the use of conditional simulation in a three dimensional space. This approach has been applied and studied to a real site investigation carried out in an alluvial mediterranean deltaic environment in Spain. Complementary studies in order to improve the proposed framework have been explored based on another geotechnical campaign conducted on a second experimental site in France.

Key words : soil characterization - ground model - statistical methods for identification of layer boundaries - automatic classification - neural networks - conditional simulation - dynamic penetrometer Panda.

Este trabajo trata sobre el reconocimiento de suelos a baja profundidad mediante el uso del penetrómetro a energía variable Panda®. El objetivo principal es el estudio y desarrollo de un conjunto de distintas técnicas en el contexto de una metodología global que permita concebir un modelo geotécnico del terreno a partir de las mediciones efectuadas durante una campaña de ensayos Panda.

Este manuscrito está estructurado en cuatro partes, cada una de las cuales aborda un objetivo específico :

- se comienza por un breve recordatorio sobre los principales ensayos de reconocimiento del terreno con especial énfasis en el ensayo de penetración dinámica Panda. Después se presenta una descripción sucinta del modelo geotécnico y los formalismos matemáticos usados para considerar la incertidumbre asociada a la caracterización de las propiedades del suelo ;
- la segunda parte trata sobre la detección automática de unidades homogéneas del sub-suelo gracias al análisis del penetrograma obtenido durante un ensayo Panda. Tras un estudio acerca de la identificación de las capas del suelo solamente a partir de un penetrograma, se propusieron técnicas estadísticas mediante una ventana móvil para realizar este análisis numéricamente. Un protocolo de ensayos en laboratorio e in situ ha permitido su estudio y validación ;
- la tercera parte se centra en la identificación automática de los geomateriales que componen las unidades homogéneas detectadas gracias a las técnicas estudiadas en la etapa anterior. Una metodología de clasificación automática basada en el uso de redes de neuronas artificiales ha sido propuesta y aplicada a dos casos de estudio : la caracterización de suelos naturales y la clasificación de un material granular arcilloso manufacturado (bentonita) ;
- la última parte está dedicada a la modelización de la resistencia de punta dinámica q_d en el espacio 3D mediante campos aleatorios. Las simulaciones del campo aleatorio son condicionados por los sondeos Panda realizados en el terreno. Este enfoque ha sido estudiado y aplicado a una campaña de reconocimiento de suelos real efectuada en una llanura deltaica del litoral mediterráneo español. Finalmente, se presentan una serie de estudios complementarios realizados en un segundo terreno experimental situado en la llanura de la Limaña en Francia para mejorar la modelización propuesta.

Palabras clave : reconocimiento de suelos - modelo geotécnico - métodos estadísticos de estratigrafía automática - clasificación automática - redes de neuronas - simulaciones condicionales - penetrómetro dinámico Panda.

Sommaire

Table des figures	vii
Liste des tableaux	xv
Notations	xix
Abréviations	xxi
Introduction générale	1
I De la reconnaissance des sols au modèle géotechnique du terrain	5
1 Introduction	6
2 Reconnaissance des sols	7
2.1 Contexte	7
2.2 Cadre normatif	8
2.3 Les principaux moyens de reconnaissance	8
2.3.1 Les essais géophysiques	9
2.3.2 Les forages et carottages	11
2.3.3 Les essais de laboratoire	12
2.3.4 Les sondages in situ	12
2.3.5 Couplage des données	14
2.3.6 Bilan	15
3 L'essai pénétrométrique	16
3.1 L'essai de pénétration au cône statique (CPT et CPTU)	16
3.1.1 Principe et caractéristiques mesurées	16
3.1.2 Équipement et procédures	17
3.1.3 Avantages et inconvénients de l'essai	17
3.2 Les pénétromètres dynamiques	18
3.2.1 Essai de pénétration au carottier (SPT)	18
3.2.2 Essai de pénétration au cône dynamique (DPT) à énergie constante .	19
3.2.3 Essai de pénétration au cône dynamique (DPT) à énergie variable . .	20

3.3	Le pénétromètre dynamique Panda	20
3.3.1	Introduction	20
3.3.2	Principe de la mesure	21
3.3.3	Résultats, exploitation et domaine d'application	23
3.3.4	Le pénétromètre Panda 3	25
3.4	Bilan	26
4	Modèle géotechnique du terrain	28
4.1	Le modèle géotechnique	29
4.1.1	Notion de modèle géotechnique	29
4.1.2	Étapes pour la conception du modèle géotechnique	29
4.1.3	Cadre normatif	30
4.2	Sources d'incertitudes en géotechnique	31
4.2.1	Incertain active ou aléatoire	31
4.2.2	Incertain passive ou épistémique	32
4.3	Prise en compte de l'incertitude	33
4.3.1	Mise en évidence du problème	33
4.3.2	La valeur caractéristique X_k des propriétés géotechniques du sol	35
4.3.3	Évolution de la réglementation internationale : modélisation d'une reconnaissance pour la prise en compte de la variabilité naturelle	37
4.4	Bilan	39
5	Conclusion, problématique et démarche proposée	40
5.1	Conclusion	40
5.2	Problématique traitée et démarche proposée	41
II	Identification automatique des couches de sol homogènes	44
6	Problématique	45
6.1	Contexte	45
6.2	Notion de "couche"	45
6.3	Problème posé par une reconnaissance de couches "experte"	47
7	État de l'art	51
7.1	Méthodes proposées dans la littérature	51
7.1.1	Méthodes de découpage pour l'essai CPT et CPTU	51
7.1.2	Méthodes de découpage pour le pressiomètre	56
7.1.3	Méthodes de découpage pour l'essai Panda	58
7.1.4	Bilan	58
7.2	Méthode de la fenêtre glissante	59
7.2.1	Principe de la méthode	59
7.2.2	Test paramétriques pour la comparaison de deux échantillons	60
7.2.3	Taille de la fenêtre et valeurs seuils	62

8 Démarche proposée	65
8.1 Démarche	65
8.2 Étude sur des essais au laboratoire	66
8.2.1 Protocole expérimental	66
8.2.2 Analyse statistique des pénétrogrammes	67
8.3 Étude sur les tests statistiques par fenêtre glissante	72
8.3.1 Influence du pas de régularisation	72
8.3.2 Influence de la taille de la fenêtre	76
8.3.3 Influence de la transformation logarithmique	78
8.3.4 Influence de la taille de la pointe	80
8.3.5 Bilan	82
8.4 Analyses de l'efficacité des méthodes de découpage statistiques sur des essais in situ en conditions réelles	82
8.4.1 Essais sur le site expérimental de Castello d'Empúries	82
8.4.2 Essais in situ analysés par les experts	86
8.5 Comparatif des stratigraphies issues des essais CPTU et Panda sur le site de Castello d'Empúries	90
8.5.1 Techniques stratigraphiques utilisées pour l'essai CPTU	90
8.5.2 Résultat de la comparaison avec des techniques stratigraphiques pour l'essai CPTU	91
8.5.3 Bilan	93
9 Conclusions	96
III Identification des sols à partir du signal Panda	98
10 Introduction et contexte de l'étude	99
10.1 Classification des sols	99
10.1.1 Principe	99
10.1.2 Classification des sols LPC/USCS	99
10.1.3 Classification des sols pour les terrassements routiers	100
10.2 Systèmes de classification développés à partir des essais in situ	101
10.3 Problématique	102
11 Cadre théorique et démarche proposée	104
11.1 Classification automatique	104
11.1.1 Introduction	104
11.1.2 Algorithmes de classification automatique	105
11.1.3 Les techniques de <i>machine learning</i> en géotechnique	108
11.1.4 Bilan	110
11.2 Les réseaux de neurones artificiels (RNA)	111
11.2.1 Généralités	111
11.2.2 Réseaux de neurones multicouches « feed-forward »	113
11.3 Démarche proposée	119

12 Application à la classification de la nature du sol	121
12.1 Développement du système de classification	121
12.1.1 Élaboration de la base de données	121
12.1.2 Définition des entrées du modèle	123
12.1.3 Définition des sorties du modèle	125
12.1.4 Dimensionnement et optimisation du réseau PMC	126
12.1.5 PNN	130
12.2 Validation : tests sur de nouveaux sondages	132
12.2.1 Démarche	132
12.2.2 Application à des matériaux granulaires en laboratoire	132
12.2.3 Application à des matériaux à l'état hydrique humide	135
12.2.4 Validation sur de nouveaux essais in situ	136
12.3 Bilan global	140
13 Application à la classification d'un matériau granulaire	141
13.1 Contexte	141
13.2 Méthodologie expérimentale	141
13.2.1 Plan général	141
13.2.2 Caractéristiques granulométriques des matériaux	142
13.2.3 Phase de calibration : essais de pénétration Panda	142
13.3 Classification supervisée	143
13.3.1 Démarche générale	143
13.3.2 Approche par Perceptron multicouche (PMC)	143
13.4 Bilan	145
14 Conclusions	146
IV Modélisation de la variabilité spatiale des sols de surface en vue d'obtenir un modèle du terrain	148
15 État de l'art et cadre théorique	149
15.1 Introduction	149
15.2 Modélisation probabiliste de la variabilité spatiale	149
15.2.1 Introduction	149
15.2.2 Théorie des champs aléatoires	149
15.2.3 Théorie des variables régionalisées	153
15.2.4 Méthodes de simulation	157
15.3 Corrélation spatiale des propriétés du sol	160
15.3.1 Introduction	160
15.3.2 Notion d'échelle de fluctuation ou longueur de corrélation	160
15.3.3 Notion de portée	161
15.3.4 Anisotropies	162
15.3.5 Longueurs de corrélation de la résistance de pointe rapportées dans la littérature	164

15.3.6	Stratégie d'échantillonnage et représentativité de la longueur de corrélation	165
15.4	Homogénéité d'un site	168
15.5	Description probabiliste de l'incertitude : la statistique bayésienne	171
15.5.1	Introduction	171
15.5.2	Généralités et techniques numériques	172
15.5.3	Bilan sur l'application de la statistique bayésienne en géotechnique	174
15.6	Bilan d'ouvrages sur la modélisation spatiale des paramètres géotechniques	176
15.7	Conclusions	178
16	Démarche proposée	180
16.1	Introduction	180
16.1.1	Découpage vertical	181
16.1.2	Critère d'homogénéité horizontale	182
16.1.3	Analyse structurale des données pénétrométriques	183
16.1.4	Simulation du champ aléatoire conditionné par les sondages Panda	185
16.2	Cas d'étude : site expérimental de Castello d'Empúries	186
16.2.1	Résultats des essais de pénétration Panda	186
16.2.2	Application des techniques stratigraphiques pour la définition des unités statistiquement homogènes	187
16.2.3	Homogénéité horizontale du site	189
16.2.4	Analyse structurale	194
16.2.5	Simulation de q_d conditionnée par les sondages Panda	201
16.2.6	Post-conditionnement du champ par krigeage. Réduction de l'incertitude et application au plan d'échantillonnage	205
16.2.7	Conclusion	210
17	Apports supplémentaires	212
17.1	Introduction	212
17.2	Apport d'une croix de sondages de type géostatistique	212
17.2.1	Analyse structurale de la première campagne de 8 essais	212
17.2.2	Analyse structurale de la croix de sondages	213
17.2.3	Critère d'homogénéité horizontale	217
17.2.4	Bilan	219
17.3	Variabilité horizontale des valeurs de la résistivité électrique	220
17.4	Application de l'inférence bayésienne	222
17.4.1	Cadre théorique et inférence	222
17.4.2	Loi a priori vague	223
17.4.3	Loi a priori informative	225
17.4.4	Bilan	226
17.5	Conclusion sur les apports supplémentaires	226
18	Conclusion sur la modélisation de la variabilité spatiale	228

Conclusion générale	229
Annexes	258
Annexes	258
A Annexes A	259
A.1 Essais de caractérisation au laboratoire	259
A.2 Essais CPTU sur le site de Castelló d'Empúries	261
A.3 Transformée par ondelettes du profil de résistance dynamique q_d	262
A.4 Courbes statistiques calculées pour les essais in situ analysés par les experts	264
B Annexes B	267
B.1 Le krigeage ordinaire (KO)	267
B.2 Méthodes proposées pour déterminer expérimentalement l'échelle de fluctuation	268
C Annexes C	270
C.1 Le site expérimentale : Le Solar de Castelló d'Empúries	270
C.1.1 Localisation du site	270
C.1.2 Géologie du site	270
C.1.3 Campagne d'essais	271
C.2 Statistiques descriptives de la campagne de sondages	273
C.3 Statistiques de q_d pour chaque unité des modèles de terrain établis	274
C.3.1 Modèle M2	274
C.3.2 Modèle M3	275
C.3.3 Modèle M4	276
C.4 Analyse structurale	277
C.4.1 Modèles de dérives ajustés pour chacun des sondages (modèle M2) . .	277
C.4.2 Estimation des dérives globales pour les modèles (M1, M3, M4) . . .	278
C.5 Conditionnement du champ par krigeage. Réduction de l'incertitude	279
D Annexes D	280
D.1 Le site expérimental d'Aulnat	280
D.1.1 Localisation du site	280
D.1.2 Géologie du site	280
D.1.3 Campagnes d'essai	281
D.1.4 Application des techniques stratigraphiques pour la définition des uni- tés homogènes	284
D.1.5 Tomographie électrique	286

Table des figures

2.1	Prospection électrique : à gauche principe de la mesure utilisant un quadripôle et à droite principaux quadripôles utilisés (Loke (2002))	11
2.2	Essais réalisables en fonction du type de prélèvement (source : Reiffsteck et al. (2012)).	12
2.3	Différents principes d'essais en place : a) exemple d'essai de cisaillement, b) d'expansion, c) de pénétration (source : Nasreddine (2004)).	13
3.1	CPT monté sur un camion (source : Robertson and Cabal (2015))	17
3.2	CPT sur dispositif portable pour des zones d'accès limité (source : Robertson and Cabal (2015))	17
3.3	Première plaquette de diffusion de l'essai Panda (source : (Gourvès (1991)))	21
3.4	Exemple de réalisation d'un sondage Panda	21
3.5	Principe du Panda 1	22
3.6	Principe du Panda 2	22
3.7	(A) Photo de la tête de battage (B) schéma de la tête de battage (C) principe de mesure de la vitesse d'impact	22
3.8	Schéma du mode de mesure Panda 2 de l'énergie apportée au système juste après l'impact du marteau (source : Benz-Navarrete (2009))	23
3.9	Exemple de pénétrogramme obtenu avec l'essai Panda et histogramme des mesures de résistance de pointe	23
3.10	Exemples d'images de matériaux obtenues en laboratoire grâce à l'endoscopie (source : Breul (1999a))	25
3.11	Exemples de la réalisation de l'essai de géondoscopie sur terrain et d'images obtenues	25
3.12	Principe du Panda 3 (source : Benz-Navarrete (2009))	26
3.13	Profils de mesures recueillies par l'essai Panda 3 obtenus en moule d'étalonnage pour un sable d'Allier sec à deux états de densité différents D_1 et D_2 (source : Escobar (2015))	26
4.1	Schéma sur la progression de la reconnaissance géotechnique vers l'élaboration du modèle géotechnique (Magnan (2010))	28
4.2	Étapes pour l'élaboration du modèle géotechnique (source : Reiffsteck et al. (2012))	29
4.3	Organisation des différentes phases d'une étude géotechnique (source : NF P 94-252 (2007))	30

4.4	Échelles de variabilité spatiale des paramètres géotechniques (source : Vanmarcke (1978) et Uzielli et al. (2006))	32
4.5	Exemples de problèmes d'enquête lors d'une reconnaissance géotechnique (source : Favre (2000))	33
4.6	Plan schématique du site qui comportait une vallée fossile (source : Magnan (2000))	33
4.7	Interprétation des ingénieurs de la valeur representative à incorporer dans les calculs pour les mesures de SPT dans les argiles plastiques de Londres et Lambeth (source : Bond and Harris (2008))	34
4.8	Interprétation des ingénieurs de la valeur representative à incorporer dans les calculs pour les mesures de SPT dans les sols graveleux de Thames (source : Bond and Harris (2008))	34
4.9	Approches statistiques pour le calcul des valeurs caractéristiques. Notation : CV_{inh} : CV de la variabilité inhérente, CV_{meas} : CV des erreurs de mesure, CV_{stat} : CV des erreurs statistiques, CV_{trans} : CV des erreurs de transformation du modèle.	36
4.10	Schéma sur les deux approches possibles pour modéliser la variabilité spatiale des paramètres du sol : (a) variable aléatoire et (b) champ aléatoire (source : Papaioannou and Straub (2016))	37
4.11	Incorporation de la variabilité spatiale dans des méthodes de calcul fiabiliste sur la base de la théorie des champs aléatoires (source : Li et al. (2017))	38
5.1	Schéma de la démarche proposée	42
6.1	Nombre de couches décelées pour chaque pénétrogramme de sondage étudié d'après l'analyse des experts	48
6.2	Synthèse comparative du nombre de couches pour chaque essai d'après l'analyse des experts et le nombre de couches identifiées par le géologue. Code couleur : bleu = si le forage géologique a été mené jusqu'à la même profondeur que le sondage pénétrométrique ; rouge = si les profondeurs sont différentes.	48
6.3	Distribution des fréquences	49
6.4	Coupe schématique du sol déduite du sondage à la tarière et découpages proposés par les experts pour l'essai E663s2	50
7.1	Schéma de la procédure par fenêtre glissante	60
8.1	Répartition des matériaux pour les éprouvettes confectionnées	66
8.2	Pénétrogrammes bruts obtenus dans les moules en laboratoire. <i>Noir</i> - pointe 2 cm ² de section fixe, <i>bleu</i> - pointe 2 cm ² de section perdue et <i>rouge</i> - pointe 4 cm ² de section perdue.	68
8.3	Comparaison du q_d moyen en fonction de la taille de la pointe (d_c : diamètre de la pointe)	70
8.4	Comparaison du CV de q_d en fonction de la taille de la pointe (d_c : diamètre de la pointe)	70
8.5	Relation entre le rapport non dimensionnel n_p et la dispersion du pénétrogramme	71
8.6	Histogramme des enfoncements mesurés (pointe de 4 cm ²).	73

8.7	Influence du pas de régularisation ξ (pointe de 4 cm^2). Notation : B - ballast, M - microballast, Sg - sable grossier et Sd - sable de Dunkerque.	73
8.8	Représentation des pénétrogrammes pour différentes valeurs de régularisation ξ (pointe de 4 cm^2). Notation : B - ballast, M - microballast, Sg - sable grossier et Sd - sable de Dunkerque.	73
8.9	Test T ratio : influence du pas de régularisation (ξ) (pointe de 4 cm^2)	74
8.10	Test intra-classe (ρ_I) : influence du pas de régularisation (ξ) (pointe de 4 cm^2)	75
8.11	Test de Bartlett (B_{stat}) : influence du pas de régularisation (ξ) (pointe de 4 cm^2)	75
8.12	T ratio : influence de la longueur (W_d) de la fenêtre glissante (pointe de 4 cm^2)	76
8.13	Intra-classe : influence de la longueur de la fenêtre (W_d) glissante (pointe de 4 cm^2)	77
8.14	Bartlett : influence de la longueur (W_d) de la fenêtre glissante (pointe de 4 cm^2)	77
8.15	T ratio : influence de la transformation logarithmique des valeurs de q_d (pointe de 4 cm^2)	78
8.16	Intra-classe : influence de la transformation logarithmique des valeurs de q_d (pointe de 4 cm^2)	79
8.17	Bartlett : influence de la transformation logarithmique des valeurs de q_d (pointe de 4 cm^2)	79
8.18	T ratio ($\ln(q_d)$) : influence de la taille de la pointe (d_c)	80
8.19	Intra-classe ($\ln(q_d)$) : influence de la taille de la pointe (d_c)	81
8.20	Bartlett (q_d) : influence de la taille de la pointe (d_c)	81
8.21	(a) Implantation des sondages sur le site de Castello d'Empúries. (b) Essais pénétrométriques statiques (CPTU) et dynamiques (Panda). (c) Distribution granulométrique en fonction de la profondeur jusqu'à 5 m (source : Pérez et al. (2013))	83
8.22	Sondage Panda S5 : influence de la longueur (W_d) de la fenêtre glissante	84
8.23	Sondage Panda S3 : influence de la longueur (W_d) de la fenêtre glissante	84
8.24	Bilan sur le nombre de couches détectées grâce aux méthodes statistiques et d'après les experts. La taille de la fenêtre est représentée par un code symbole : $\square = 0.4m$, $\triangleright = 1m$, $\circ = 1.5m$, $\diamond = 2m$	88
8.25	Limites de couches détectées avec les méthodes statistiques. La méthode utilisée est représenté par un code couleur : rouge = T ratio, bleu = Intra-classe et vert = Bartlett. La taille de la fenêtre est représentée par un code symbole : $\square = 0.4m$, $\triangleright = 1m$, $\circ = 1.5m$, $\diamond = 2m$	89
8.26	Comparaison entre les stratigraphies déterminées par le sondage CPTU1 et le sondage Panda S5	94
8.27	Comparaison entre les stratigraphies déterminées par le sondage CPTU2 et le sondage Panda S3	95
10.1	Classification des sols grenus (plus de 50% des éléments $> 80\mu m$)	100
10.2	Classification des sols fins. Abaque de Casagrande	100
10.3	Extrait du tableau synoptique (matériaux « Sols » de $D_{max} < 50 mm$) de la classification GTR des matériaux selon leur nature	101

11.1	Principe de l'hyperplan : pour un ensemble de données linéairement séparables, deux hyperplans séparateurs (a) et (b) possibles. L'hyperplan optimal est celui avec la marge maximale (b) (source : Han et al. (2011))	107
11.2	Schéma de neurone formel (source : Demuth et al. (2014))	111
11.3	Taxonomie des architectures des réseaux de neurones « feed-forward » et « feed-back » (source : Jain et al. (1996))	112
11.4	Structure d'un réseau multicouche <i>feed-forward</i> (source : Han et al. (2011)) .	113
11.5	Interprétation géométrique du rôle du nombre de couches cachées pour un problème de classification dans un espace d'entrées 2D (source : Jain et al. (1996))	115
11.6	Illustration schématique de l'évolution des erreurs pour les ensembles d'apprentissage et validation en fonction de la période d'entraînement τ . L'objectif d'aboutir à la meilleure capacité de généralisation suggère que l'entraînement doit être arrêté à l'itération $\hat{\tau}$ équivalente à la valeur minimale de l'ensemble de validation (source : Bishop (1995)).	115
11.7	Détail du neurone de la couche cachée (<i>pattern unit</i>) (source : Specht (1990))	117
11.8	Architecture du réseau de neurones probabiliste (source : Specht (1990)) . .	117
11.9	Schéma de principe de la méthodologie proposée.	119
12.1	Tamisats à 2 mm et 80 μ m des matériaux de la banque de sondages (BDS) de laboratoire (gauche) et in situ (droite)	122
12.2	Performance du PMC vs nombre de neurones de la deuxième couche de calcul. Pré-traitement des données (min-max)	127
12.3	Performance du PMC vs nombre de neurones de calcul. Pré-traitement des données (min-max)	127
12.4	Performance du PMC vs nombre de neurones de calcul. Pré-traitement des données (z-score)	128
12.5	Matrice de confusion du PMC choisie après la phase d'optimisation	128
12.6	Analyse de sensibilité basée sur l'étude de l'effet d'une modification de chacun des paramètres d'entrée du PMC	130
12.7	Matrices de confusion du PMC avec 12 neurones cachés	130
12.8	Détermination de la valeur optimale du paramètre de lissage σ	131
12.9	Matrices de confusion du PNN avec $\sigma = 0.2$	132
12.10	Matrices de confusion pour les sondages réalisés en pointe de 4 cm^2 de section	133
12.11	Matrices de confusion pour les sondages en pointe de 2 cm^2	134
12.12	Matrices de confusion pour les sondages effectués dans l'argile sableuse. . . .	135
12.13	Matrices de confusion pour les sondages effectués dans le sable de Brive. . . .	136
12.14	Tamisats à 2mm et 80 μ m des nouveaux matériaux in situ	137
12.15	Matrices de confusion	137
12.16	Classification du RNA couplée avec le découpage par fenêtre glissante ($W_d = 1m$). À gauche classification fournie par le modèle PMC. À droite classification fournie par le PNN.	138
12.17	Classification du RNA couplée avec le découpage par fenêtre glissante ($W_d = 2m$). À gauche classification fournie par le modèle PMC. À droite classification fournie par le PNN.	139

13.1	Photo des deux composés de bentonite : concassé à gauche et pellets à droite	142
13.2	Courbes granulométriques des quatre échantillons de bentonite concassée . . .	142
13.3	Équations et courbes de calibration des cinq mélanges de matériaux	143
13.4	Évolution du % moyen de pénétrogrammes bien classés du jeu de données de validation et de test en fonction du nombre de neurones de calcul	144
13.5	Matrices de confusion du PMC	145
15.1	À gauche : exemple de variogramme à palier (source : Bacconnet (1991)). À droite : relation entre variogramme et covariogramme (source : Emery (2001))	162
15.2	(a) Formation isotrope et (b) anisotrope transverse (source : Zhu and Zhang (2013)).	164
15.3	Les trois configurations d'échantillonnage (source : Bacconnet (1991))	167
15.4	À gauche, différence maximale de la moyenne de q_c en fonction de la longueur du profil analysé. À droite, facteur de correction en fonction de la distance relative entre sondages (source : Salgado et al. (2015))	170
15.5	Organigramme de calcul de l'indice de variabilité horizontale du site HVI (source : Salgado et al. (2015))	171
15.6	Aperçu de la mise à jour bayésienne en géotechnique (source : (Straub and Papaioannou (2014)))	175
16.1	Modélisation de la variabilité spatiale d'un profil de sol hétérogène (source : modifié d'après Phoon and Kulhawy (1999))	182
16.2	Schéma du traitement statistique des sondages Panda	184
16.3	Organigramme des différentes étapes de calcul des simulations conditionnelles	185
16.4	Schéma de la procédure de post-conditionnement par krigeage	185
16.5	(a) Pénétrogrammes bruts jusqu'à 5 mètres de profondeur. (b) Histogramme des mesures de résistance. (c) Histogramme des enfoncements enregistrés. . .	186
16.6	Caractéristiques statistiques des huit sondages Panda : (a) Pénétrogrammes régularisés (b) coefficient de variation (c) moyenne et écart-type.	187
16.7	Limites de couches détectées avec les méthodes statistiques. Code couleur : Tratio , Intra-classe et Bartlett . Avec W_d la longueur de la fenêtre testée.	188
16.8	Coupe schématique des 4 modèles de découpage étudiés avec la valeur moyenne et l'écart-type des limites entre couches.	189
16.9	Profils de résistance de pointe statique q_c et stratigraphies interprétées selon l'abaque de classification de Robertson (Robertson (1990))	191
16.10	Pénétrogrammes regularisés au sein de la couche argileuse	192
16.11	Comparaison entre les indices de variabilité issus de l'analyse de q_d avec ceux calculés pour $\ln(q_d)$ en fonction du modèle de découpage	192
16.12	Comparaison entre les indices HVI	193
16.13	À gauche : tendances linéaires issues d'un régression globales des profils de $\ln(q_d)$. À droite : histogrammes des résidus pour les deux unités et ajustement d'une loi normale	195
16.14	Comparaison entre les échelles de fluctuation verticales estimées pour la résistance de pointe q_d en abscisse et pour la transformée logarithmique $\ln(q_d)$.	197
16.15	Autocorrélogrammes moyens et ajustement du modèle exponentiel	197

16.16	Estimation de l'échelle de fluctuation horizontale du logarithme de la résistance de pointe pour les deux unités de terrain considérés	198
16.17	Analyse structurale de la formation sableuse : (a) résidus normalisés après le passage au logarithme et soustraction de la dérive linéaire avec la profondeur (b) autocorrélogrammes empiriques et ajustement des modèles théoriques (c) distribution de résidus et ajustement de la loi normale standard	199
16.18	Analyse structurale de la formation argileuse : (a) résidus normalisés après le passage au logarithme et soustraction de la dérive linéaire avec la profondeur (b) autocorrélogrammes empiriques et ajustement des modèles théoriques (c) distribution des résidus et ajustement de la loi normale standard	200
16.19	Plan de la position relative des sondages. Cercle noir : sondages utilisés pour la simulation du champ. Croix rouge : sondages non pris en compte dans la simulation	201
16.20	En haut moyenne et en bas écart type du modèle 3D suite à 500 tirages . . .	202
16.21	Comparaison entre les pénétrogrammes mesurés et ceux simulés pour différents modèles de découpage	203
16.22	RMSE calculée selon la profondeur pour les pénétrogrammes S3 et S8 après 500 simulations	204
16.23	Réduction de l'incertitude en fonction de θ_H	207
16.24	Apport du conditionnement par les 8 sondages	207
16.25	Influence de la distance de corrélation θ_H et de l'écart Δx sur l'implantation optimale de deux sondages le long d'un profil d'extension L	208
16.26	Influence de la distance de corrélation θ_H et de l'écart Δy sur l'implantation optimale de trois sondages le long d'un profil d'extension L	208
16.27	À gauche, plan d'implantation optimal des 6 sondages. À droite, gain du nouveau schéma par rapport à l'implantation originale en fonction de la distance de corrélation	209
16.28	Apport à la réduction de l'incertitude en fonction de la position d'implantation d'un nouveau sondage pour $\theta_H = 13m$	210
17.1	Résultats des mesures de la première campagne	213
17.2	Pénétrogrammes régularisés et histogrammes des résistances de pointe sur la couche de limons et argiles	213
17.3	Variogrammes verticaux de q_d sur la couche de limons	214
17.4	Fonctions d'autocorrélation verticales de la couche de limons	215
17.5	Analyse structurale de q_d pour les sondages placés dans l'axe Y	215
17.6	Analyse structurale de q_d pour les sondages placés dans l'axe X	216
17.7	Estimation de l'échelle de fluctuation horizontale de q_d pour les deux unités de terrain considérées	217
17.8	Comparaison entre les indices HVI calculés en fonction de la campagne analysée et la modification proposée.	219
17.9	Profils de variation de la résistivité électrique et ajustement de modèles théoriques d'autocorrélation.	220
17.10	Chaînes de Markov simulées	224
17.11	Loi a posteriori marginale de θ_H	224

17.12	Influence de l'écart-type de l'erreur sur l'estimation de la valeur moyenne de θ_H pour une loi a priori informative	225
17.13	Ajustement de la loi lognormale aux valeurs de θ_H estimées à partir de la tomographie électrique	225
17.14	Lois a posteriori marginales de θ_H	225
A.1	Analyse granulométrique par tamisage du ballast	259
A.2	Analyse granulométrique par tamisage du microballast	259
A.3	Analyse granulométrique par tamisage du sable grossier	260
A.4	Analyse granulométrique par tamisage du sable de Dunkerque	260
A.5	Résultats des essais CPTU sur le site de Castello	261
A.6	Classification des sols du site selon l'abaque de (Robertson (1990))	261
A.7	Résultats de l'analyse par la transformée par ondelettes du profil q_d issue du sondage effectué avec une pointe de 4 cm^2 dans le moule HET 1	263
A.8	Courbes des tests statistiques selon la hauteur de la fenêtre utilisée pour les sondages in-situ analysés par les experts	264
A.9	Courbes des tests statistiques selon la hauteur de la fenêtre utilisée pour les sondages in-situ analysés par les experts	265
A.10	Courbes des tests statistiques selon la hauteur de la fenêtre utilisée pour les sondages in-situ analysés par les experts	266
C.1	(a) Situation géographique du terrain. (b) À droite emplacement du Solar de Castello d'Empúries.	270
C.2	Carte géologique (source : Díaz and Ercilla (1993)) et situation géographique du site expérimental (point rouge)	271
C.3	À gauche, le schéma d'implantations des essais in situ. À droite, la position relative des sondages pénétrométriques	271
C.4	Calage de dérives linéaires pour $\ln(q_d)$ au sein des deux unités lithologiques identifiés	277
C.5	Dérives linéaires et histogrammes des résidus sur chacune des unités pour les différentes unités de terrain	278
D.1	(a) Situation géographique du terrain. (b) Emplacement de la parcelle d'Aulnat.	280
D.2	Zone d'étude. Carte géologique de la région de Clermont-Ferrand 1 : 50000. Source : http://planet-terre.ens-lyon.fr/article/BRGM-kml.xml	281
D.3	Première campagne. À gauche, le schéma d'implantations des essais in situ. À droite, la position relative des sondages pénétrométriques	282
D.4	Deuxième campagne. À gauche, le schéma d'implantation des essais in situ. À droite, la position relative des sondages pénétrométriques	283
D.5	Coupe lithologique déduite du sondage à la tarière	283
D.6	Application des techniques statistiques du découpage pour le sondage S4. Code couleur : Tratio , Intra-classe et Barlett . Longueur de fenêtre $W_d = 1\text{m}$	284
D.7	Limites de couches détectées avec les méthodes statistiques. Code couleur : Tratio , Intra-classe et Barlett . Avec W_d la longueur de la fenêtre testée.	284
D.8	Analyse profil X. Limites de couches détectées avec les méthodes statistiques pour une longueur de la fenêtre égale à 1.5m. Code couleur : Tratio , Intra-classe .	285
D.9	Mesures de résistivité issus du premier panneau électrique (ERT1).	286

D.10 Mesures de résistivité issus du deuxième panneau électrique (ERT2).	287
D.11 Mesures de résistivité issus du troisième panneau électrique (ERT3).	287

Liste des tableaux

6.1	Caractéristiques des échantillons prélevés à proximité de l'essai E663s2	49
7.1	Interprétation de l'abaque (Robertson (1990)) en fonction de l'indice de comportement I_c	52
7.2	Récapitulatif des méthodes existantes. 1) <i>Probabiliste - floue</i> ; 2) <i>T ratio</i> ; 3) <i>Intra-classe</i> ; 4) D^2 ; 5) <i>Bartlett</i> ; 6) <i>Analyse par cluster</i> ; 7) <i>Abaque de classification</i> ; 8) <i>Bayésienne - Robertson</i> ; 9) <i>Bayésienne</i> ; 10) <i>Ondelettes</i> . . .	57
7.3	Valeurs seuils des paramètres des tests d'évaluation proposées dans la littérature	64
8.1	Caractéristiques physiques des matériaux compactés	67
8.2	Caractéristiques granulométriques des matériaux	67
8.3	Valeur moyenne de q_d (MPa) des sondages réalisés en pointe de 4 cm^2	69
8.4	Coefficient de variation de q_d des sondages réalisés en pointe de 4 cm^2	69
8.5	Valeurs du n_p pour les matériaux utilisés	71
8.6	Paramètres d'identification et d'état des échantillons prélevés	83
8.7	Synthèse des résultats - profondeur (m) des interfaces identifiées à partir des différents tests paramétriques étudiés pour les sondages Panda	85
9.1	Récapitulatif des conclusions sur les différentes méthodes	96
12.1	Plage des paramètres physiques des matériaux en laboratoire	122
12.2	Plage des paramètres physiques des matériaux in situ	122
12.3	Classes de sortie proposées	125
12.4	Répartition des pénétrogrammes de la base de données	126
12.5	Tamisâts à 2 mm et 80 μm des matériaux dans les moules de calibration en laboratoire	133
12.6	Pénétrogrammes bien classés selon les matériaux et la taille de la pointe . . .	134
12.7	Caractéristiques des matériaux	135
12.8	Plage des paramètres physiques des nouveaux matériaux in situ	137
12.9	Identification et classification des échantillons prélevés sur le site de Castello d'Empúries	138
13.1	Composition des 5 mélanges de matériaux étudiés	142
13.2	Classes de sortie proposées pour le modèle PMC	143
15.1	Fonctions d'autocorrélation théoriques (source : Vanmarcke (1977a) , Vanmarcke (1983))	161

15.2	Principaux modèles de variogramme (source : Bacconnet (1991))	163
15.3	Rapports d'anisotropie courants suite au processus géologique d'après (Journal and Huijbregts (1978)), (Koltermann and Gorelick (1996)), (Deutsch and Journal (1998)), (Chiles and Delfiner (1999)), (Deutsch (2002)), (Bakhtiari (2011)), (Huber (2013)), (Wackernagel (2013))	163
15.4	Correspondance entre les quatre formes d'anisotropie et les formations naturelles et anthropiques (source : Zhu and Zhang (2013)).	164
15.5	Échelles de fluctuation de la résistance à la pénétration reportées dans la littérature (source : Huber (2013) et élaboration propre)	166
15.6	Valeurs typiques de distances d'auto-corrélation d'après (source : Favre (2004))	167
16.1	Indice de variabilité horizontale HVI pour les différents modèles testés	190
16.2	Indices HVI calculés pour les pénétrogrammes au sein de la couche argileuse	192
16.3	Indices HVI calculés pour $\ln(q_d)$ au sein de la couche argileuse	193
16.4	Modèles de dérives et caractéristiques statistiques des résidus de la formation sableuse	194
16.5	Échelles de fluctuation verticales estimées à l'aide des différents modèles d'autocorrélation théoriques sur chaque unité	196
16.6	Paramètres du champ aléatoire selon le modèle de découpage choisi	201
16.7	Réduction de l'erreur (%) par rapport à ne considérer qu'une seule couche homogène	205
17.1	Indices de variabilité horizontale calculés pour la première campagne d'essais. Profondeur d'étude égale à 4 m.	217
17.2	Indices de variabilité horizontale calculés pour les essais dans l'axe X. Profondeur d'étude égale à 3.4 m.	218
17.3	Indices de variabilité horizontale calculés pour les essais dans l'axe Y. Profondeur d'étude égale à 2 m.	218
17.4	Résultats de l'analyse de variabilité spatiale pour le panneau ERT1	221
17.5	Résultats de l'analyse de variabilité spatiale pour le panneau ERT2	221
17.6	Résultats de l'analyse de variabilité spatiale pour le panneau ERT3	222
17.7	Résultats de l'inférence bayésienne pour θ_H avec une loi a priori uniforme peu informative	224
17.8	Résultats de l'inférence bayésienne pour θ_H avec une loi informative	226
17.9	Synthèse des valeurs de θ_H estimées	226
18.1	Récapitulatif des conclusions sur les différentes méthodes	231
C.1	Sondages Panda, nombre de mesures, coordonnées UTM et profondeur du sondage	272
C.2	Statistiques de base du q_d par sondage	273
C.3	Statistiques du modèle M2, unité 1 (U1)	274
C.4	Statistiques du modèle M2, unité 2 (U2)	274
C.5	Statistiques du modèle M3, unité 1 (U1)	275
C.6	Statistiques du modèle M3, unité 2 (U2)	275
C.7	Statistiques du modèle M4, unité 3 (U3)	276
C.8	Statistiques du modèle M4, unité 4 (U4)	276

C.9	Réduction de l'incertitude en fonction de l'échelle de fluctuation horizontale et du nombre de sondages	279
D.1	Sondages Panda, nombre de mesures, coordonnées UTM et profondeur du sondage	282
D.2	Analyse profil X. Résultats Tratio	285
D.3	Caractéristiques des profils électriques réalisés et paramètres de l'inversion .	286

Notations

α		: seuil de significativité ou niveau de signification
B_{stat}		: statistique du test de Bartlett
C_U		: coefficient d'uniformité de Hazen
C_C		: coefficient de courbure
C		: covariance
CV		: coefficient de variation
ξ		: pas de régularisation
E_M	[MPa]	: module pressiométrique Ménard
F_r		: rapport de frottement
f_s	[kPa]	: frottement latéral unitaire local
γ_d	[kN/m ³]	: poids volumique sec
$\gamma(\tau)$: variogramme
I	[%]	: incertitude du champ
Ic		: indice de comportement
P_a	[kPa]	: pression atmosphérique
p_L	[MPa]	: pression limite pressiométrique
ρ	[Ωm]	: résistivité
ρ_I		: coefficient de corrélation intra-classe
$\rho(\tau)$: fonction d'autocorrélation
N_{SPT}		: nombre de coups mesuré au SPT
q_c	[MPa]	: résistance de pointe statique
q_d	[MPa]	: résistance de pointe dynamique
q_t	[MPa]	: résistance de pointe corrigée
Q_{tn}		: résistance au cône normalisé
u	[kPa]	: pression interstitielle
u_0	[kPa]	: pression hydrostatique
μ		: moyenne
σ		: écart-type
σ^2		: variance
σ_{KO}^2		: variance du krigeage ordinaire
θ	[m]	: échelle de fluctuation
W_d	[m]	: longueur de la fenêtre

Abréviations

COV	: covariance
CPT	: Cone Penetration Test (essai de pénétration statique au cône)
CPTU	: essai de pénétration statique au piézocône
CSX	: modèle de corrélation cosinus
DMT	: Dilatomètre Marchetti
DPT	: Dynamic Penetration Test (essai de pénétration au cône dynamique)
ERT	: Electrical Resistivity Tomography (Tomographie de la résistivité électrique)
GTR	: Guide des Terrassements Routiers
HET	: moule de calibration "hétérogène"
HVI	: Horizontal Variability Index (Indice de Variabilité Horizontal)
KO	: krigeage ordinaire
KS	: krigeage simple
LPC	: Laboratoire des Ponts et Chaussées
MCMC	: Monte Carlo Markov Chain (méthodes de Monte Carlo par chaînes de Markov)
MCO	: régression par moindres carrés ordinaire
PANDA	: essai de pénétration dynamique léger à énergie variable (Pénétromètre Autonome Numérique Dynamique Assisté par ordinateur)
PMT	: Ménard Pressumeter Test (essai pressiométrique Ménard)
RMSE	: Root Mean Square Error (erreur quadratique moyenne)
RBF	: Radial Basis Function Network (réseau à fonction de base radiale)
RNA	: Réseau de Neurones Artificiels
PMC	: Multilayer Feedforward Perceptron (Perceptron Multicouche)
PNN	: Probabilistic Neural Network (réseau de neurones probabiliste)
SNX	: modèle de corrélation exponentiel
SQX	: modèle de corrélation gaussien
SMK	: modèle de corrélation Markov d'ordre 2
SPT	: Standard Penetration Test (essai de pénétration au carottier)
SVM	: Support Vector Machine (machines à vecteurs de support)
USCS	: Unified Soil Classification System (Système de classification des sols universel)

Introduction générale

La géotechnique est devenue une science pluridisciplinaire qui concerne à la fois des études expérimentales et théoriques, la conception et l'exécution des ouvrages, les interactions des sols et des structures, la reconnaissance de sites, les études de matériaux naturels, les risques naturels, la stabilité et le suivi des ouvrages en terre, les problèmes de pollution des sols, etc. La variété des projets concernés par le domaine de la géotechnique a quand même un point de départ commun : la caractérisation du site par le biais d'une campagne de reconnaissance des sols. Elle peut comporter la visite du terrain, des levés géologiques, des sondages, des essais in situ et en laboratoire et des reconnaissances géophysiques. L'information récoltée sert de base pour alimenter un modèle géotechnique afin de concevoir et justifier les ouvrages.

Dans la pratique, les ingénieurs disposent de protocoles normalisés pour utiliser et interpréter les différents moyens de reconnaissance des sols, des règles de calcul pour dimensionner les ouvrages et de guides ou recommandations pour l'exécution des travaux. Cependant, il n'existe pas de règles ou de procédures permettant de formaliser les processus d'abstraction et de synthèse intellectuelles qu'impose la conception d'un modèle géotechnique du terrain. Elle présente alors un fort caractère subjectif puisque repose fondamentalement sur l'expertise et le jugement de l'ingénieur mais dépend aussi de la quantité et de la qualité d'information disponible et de l'objectif final du projet. Cependant, la collecte des informations sur site présente souvent un coût élevé par rapport aux budgets et moyens de financement destinés aux campagnes de reconnaissance. On se retrouve donc souvent dans la situation où les données du terrain sont insuffisantes et le modèle géotechnique peu fiable ([Magnan \(2002\)](#)).

Du fait qu'ils sont souvent moins chers, qu'ils permettent de tester directement le sol dans son état naturel et qu'ils fournissent un sondage presque en continu, les essais in situ sont plus utilisés que les essais de laboratoire dans la pratique courante de la géotechnique, et en France en particulier ([Plumelle et al. \(2013\)](#)). Parmi la grande palette d'essais in situ disponibles, les essais pénétrométriques sont les plus couramment utilisés. L'intérêt principal des essais de pénétration réside dans le fait de fournir verticalement une information mécanique en continu du sol permettant de caractériser la variabilité d'une formation donnée en fonction de la profondeur, ce qu'aucun autre outil d'investigation géotechnique n'est en mesure de proposer ([Jacquard and Boutet \(2016\)](#)).

Dans le cadre de la caractérisation des sols de surface, l'essai pénétrométrique dynamique à énergie variable Panda® ([Gourvès \(1991\)](#)) présente de nombreux avantages, tels que sa grande résolution de mesure et la possibilité de collecter une grande quantité de données du fait de sa rapidité de mise en œuvre et d'utilisation, de son coût abordable et grâce également à la possibilité d'adapter l'énergie de battage. Depuis son apparition, l'essai Panda a bénéficié de nombreux progrès théoriques et technologiques grâce aux divers travaux menés autour de l'appareil ([Zhou \(1997\)](#), [Benz-Navarrete \(2009\)](#), [Escobar \(2015\)](#)). Cet essai s'avère donc être un moyen de reconnaissance développé et moderne par rapport au pénétromètre dynamique à cône classique, longtemps considéré comme un appareil de reconnaissance rudimentaire et non satisfaisant ([Waschkowski \(1983\)](#)).

Si le pénétromètre Panda offre des perspectives intéressantes en terme de collecte de données et de caractérisation des sols, les possibilités offertes par les méthodes d'analyse et de traitement du profil des mesures enregistrées pour caractériser un site, n'ont toutefois jamais été étudiées et appliquées au profit de l'exploitation et de l'interprétation de ce genre d'essais. Dans ce contexte, le travail présent a pour objectif principal d'étudier les potentialités des différentes méthodes d'analyse de données pour l'exploitation du signal pénétrométrique Panda dans l'aide à l'obtention d'un modèle de terrain. L'ensemble des

techniques proposées s'inscrit dans le cadre d'une méthode globale qui suit la démarche traditionnelle du géotechnicien qui cherche, à partir des données du terrain, à répondre aux objectifs de base de la reconnaissance géotechnique : quelle est la nature et la disposition des formations homogènes de sol et quelles sont ses propriétés mécaniques.

Pour mener à bien ce travail, on s'appuie fondamentalement sur l'analyse et le traitement des essais in situ mais aussi effectués sur des éprouvettes confectionnées en laboratoire. La démarche envisagée à partir des données issues des sondages au pénétromètre dynamique Panda est la suivante :

- 1) identifier systématiquement les unités homogène de sol ;
- 2) pour chaque formation homogène détectée à l'étape précédente, développer des outils basés sur les techniques d'intelligence artificielle pour identifier la nature des matériaux traversés à partir de leur classification ;
- 3) proposer une répartition dans l'espace de la résistance de pointe intégrant l'incertitude spatiale et aboutir à un modèle de terrain probabiliste en 3D ;
- 4) développer des approches permettant le passage du modèle de terrain au modèle géotechnique.

Ce manuscrit est structuré en quatre parties, chacune abordant un objectif spécifique.

- i. Premièrement, on rappelle les objectifs et les principaux moyens de la reconnaissance des sols, en détaillant un peu plus l'essai de pénétration et notamment l'essai de pénétration dynamique Panda. Le modèle géotechnique de terrain, les étapes pour sa conception et son cadre normatif sont présentés ensuite. Les sources d'incertitudes sont ensuite rappelées, pour finir sur les méthodes statistiques et probabilistes envisagées par les comités techniques, qui permettent de décrire l'incertain géotechnique dans la caractérisation des propriétés du sol.
- ii. La deuxième partie porte sur l'identification automatique des unités homogènes du terrain, à partir du signal pénétrométrique Panda. Les résultats d'une enquête réalisée auprès d'une dizaine d'utilisateurs du Panda introduit la problématique. Ensuite, des méthodes par fenêtre glissante ont été sélectionnées sur la base de l'étude bibliographique. L'application de ces méthodes au Panda a été étudiée et validée grâce à une campagne d'essais en laboratoire et d'essais in situ. Une analyse comparative entre les résultats du découpage pour l'essai Panda et ceux obtenus à l'aide d'autres techniques appliquées au pénétromètre statique CPT conclut cette partie.
- iii. La troisième partie porte sur l'identification automatique des matériaux composant les unités homogènes détectées dans le signal Panda à partir des méthodes proposées en partie II. Suite à une étude bibliographique sur les techniques d'apprentissage supervisées, une approche par réseaux de neurones artificiels a été retenue. Elle a été étudiée pour deux problématiques : une première dans laquelle on cible la caractérisation de sols naturels, à partir d'une banque de données composée d'essais de laboratoire et in situ et une deuxième sur la classification d'un matériau argileux industrialisé (bentonite) formé par un mélange de billes sèches et de matériau concassé.
- iv. Enfin, la dernière partie est consacrée à la production d'un modèle du terrain basé sur la modélisation et la simulation de la résistance de pointe dynamique au moyen de fonctions aléatoires de l'espace. Cette modélisation est basée sur une approche par champs aléatoires conditionnés par les sondages Panda sur le terrain. Les différentes étapes de sa mise en œuvre ont été appliquées sur un site réel situé dans la plaine

deltaïque méditerranéenne en Espagne. Sur un deuxième site expérimental, on étudie l'apport de données ou procédures complémentaires, telles que l'analyse structurale d'une croix de sondages permettant une caractérisation plus fine de la variabilité naturelle du site. L'étude de l'autocorrélation des mesures de résistivité ou l'utilisation des procédures bayésiennes pour tenter de mieux décrire ou modéliser la variabilité horizontale ont également été explorées.

Première partie

De la reconnaissance des sols au modèle géotechnique du terrain

1. Introduction

Une étude de sol commence par une reconnaissance géologique qui peut être accompagnée par des prospections géophysiques. La reconnaissance géologique est complétée par la détermination des caractéristiques mécaniques qui peuvent être mesurées à l'aide d'essais in situ ou en laboratoire. Après une synthèse sur les objectifs d'une reconnaissance des sols et sur le cadre réglementaire cette première partie propose une présentation générale sur les différents moyens de reconnaissance des sols.

Dans la pratique courante, les essais in situ sont beaucoup plus usités que les essais de laboratoire et parmi eux les essais de pénétration ont attiré notre attention. Ils sont capables de fournir un profil de mesures presque en continu. Leur faible pas de mesure et leur excellente répétabilité permettent une caractérisation aisée de la variabilité naturelle du sol selon la profondeur. C'est pourquoi, les travaux présentés ici se focalisant sur cet essai, nous présentons par la suite une synthèse sur les essais de pénétration en insistant plus en détail sur le pénétromètre dynamique Panda qui est l'outil de reconnaissance retenu dans notre étude.

L'interprétation, l'analyse et la synthèse des données de terrain issues de la campagne de reconnaissance permettent d'élaborer le modèle géotechnique du terrain. Une réflexion sur la notion de modèle géotechnique, sur les étapes pour sa conception et son cadre normatif est présentée en fin de cette partie. On finira par un exposé sur les incertitudes en géotechnique et les différents facteurs à l'origine de ces incertitudes, en insistant plus particulièrement sur la variabilité spatiale naturelle des matériaux et sa prise en compte.

2. Reconnaissance des sols

2.1 Contexte

D'une manière générale la reconnaissance des sols a pour objectif de décrire le terrain en fournissant le maximum d'informations sur la nature physique, mécanique et hydraulique des différentes formations présentes en s'appuyant sur différentes méthodes d'investigation. La norme en vigueur, dans le contexte de la reconnaissance de sols à faible profondeur qui constitue le domaine d'application de ce travail ([NF P 94-261 \(2013\).Annexe K](#)) définit les objectifs de la reconnaissance géotechnique :

- identifier de façon fiable la disposition et la nature des formations concernées par la réalisation des fondations superficielles ;
- fournir les propriétés géotechniques, mécaniques, hydrauliques et électrochimiques des couches de terrain nécessaires à la conception et à la justification de la stabilité de la fondation superficielle ;
- fixer le niveau des nappes (libres ou souterraines) pour le calcul de la fondation superficielle.

Bien entendu, ces indications visent à la reconnaissance des couches de terrain impactées par les actions issues des fondations superficielles et celles permettant de définir les résistances géotechniques, ainsi que les conditions des nappes extérieures et souterraines, afin d'établir le modèle géotechnique de calcul. Par ailleurs, les paramètres utilisés pour décrire les propriétés physiques et mécaniques des sols, sont de natures très variées ([Magnan \(1993\)](#)) :

- paramètres d'identification et d'état (porosité, indice des vides, densité relative, limites d'Atterbeg, etc.) ;
- paramètres de déformabilité (indices de compression et de gonflement, module oedométrique, module pressiométrique, etc.) ;
- paramètres de résistance (cohésion et angle de frottement interne, pression limite pressiométrique, résistance de pointe statique ou dynamique, etc.) ;
- paramètres de perméabilité.

L'ensemble des paramètres cités permet ainsi une caractérisation du sous-sol, représentant notre connaissance sur les conditions du terrain, qui dépend de l'importance et de la qualité des reconnaissances géotechniques. Dans la pratique courante, on cherche à déterminer de façon économique avec le moins d'incertitudes possible les paramètres et informations nécessaires permettant de bâtir un modèle géotechnique du terrain afin de mener le calcul de l'ouvrage.

2.2 Cadre normatif

Les documents normatifs de référence sur la reconnaissance de sols sont l'Eurocode 7 et la norme française (NF P 94-500 (2013)) consacrée aux missions géotechniques.

Dans l'EC-7, la section 3 de la partie I (NF P 94-251-1 (2005)) « Données géotechniques » fournit la définition des reconnaissances géotechniques, l'évaluation des paramètres géotechniques, le contenu d'un rapport de reconnaissance de terrain et les notions de reconnaissances préliminaires et de projet. La partie 2 (NF P 94-252 (2007)) fournit des règles complémentaires à la partie 1 relatives, entre autres, à la planification et la consignation des résultats des reconnaissances.

La norme (NF P 94-500 (2013)) définit les différentes missions types de l'ingénierie géotechnique ayant pour objectif de déterminer les propriétés du sol et leurs incidences sur les ouvrages, constructions, aménagements de terrains existants ou à réaliser. Elle définit six missions types, classées de G1 à G5, leur contenu, leurs limites et leur enchaînement pour un projet de construction. La norme précise aussi les éléments à réaliser par le géotechnicien et les éléments à fournir par le client. La problématique de maîtrise des risques est aussi mise en avant. La partie technique concernant les missions géotechniques n'est pas traitée par cette norme.

2.3 Les principaux moyens de reconnaissance

Les praticiens de la géotechnique disposent d'une large panoplie de moyens de reconnaissance permettant d'obtenir les informations recherchées sur les terrains (disposition des couches, paramètres mécaniques de résistance et de déformation des terrains, perméabilité, etc.). Les moyens de reconnaissance peuvent être divisés en deux catégories principales :

- i. **les moyens de reconnaissance volumiques** : ils fournissent des mesures sur des surfaces ou volumes du sous-sol à caractériser. Il s'agit fondamentalement des essais de prospection géophysique ;
- ii. **les moyens de reconnaissance ponctuels** : ils fournissent une information locale sur le terrain en profondeur ou des informations continues le long d'un sondage. Les méthodes de reconnaissance géotechnique appartiennent à cette famille. On distingue :
 - les outils de forage et carottage ;
 - les essais de laboratoire ;
 - les essais in situ.

Les forages et carottages constituent des éléments ponctuels de reconnaissance géologique permettant d'affiner le modèle géologique par des sondages qui intéressent un volume de terrain limité. Mis à part des objectifs d'ordre géologique, ces outils permettent de réaliser des prélèvements destinés à des essais de laboratoire. Cette reconnaissance géologique précède ainsi la campagne d'essais géotechnique en laboratoire ou in situ.

Par la suite, on présente un bref aperçu des différentes techniques de reconnaissance du sous-sol.

2.3.1 Les essais géophysiques

Les perturbations à faible énergie, utilisées par les méthodes géophysiques, fournissent des informations très utiles sur les propriétés du sol et leur variation spatiale (Santamarina and Park (2016)). En effet, la prospection géophysique met en œuvre un ensemble de méthodes indirectes basées sur les propriétés physiques du milieu, notamment la densité, la vitesse de propagation des ondes élastiques (de compression V_P et cisaillement V_S), les propriétés électriques telles que la résistivité (la valeur inverse de la conductivité) et la permittivité diélectrique et thermique. Les mesures peuvent être réalisées à partir de la surface, de puits ou d'excavations préexistantes. Les méthodes géophysiques les plus utilisées en reconnaissance de sols sont les suivantes :

- prospection gravimétrique, microgravimétrie ;
- prospection sismique ;
- méthodes électriques.

Elles permettent une caractérisation de la structure du milieu souterrain de manière non destructive et avec un échantillonnage spatial suffisamment dense permettant de décrire les variations latérales et verticales de la propriété mesurée. Le degré de finesse avec lequel le sous-sol va être décrit est fonction de la propriété choisie et de la méthode utilisée. Le principe des techniques géophysiques appliquées au génie civil, leurs objectifs et les paramètres obtenus sont décrits dans les ouvrages géotechniques généraux (Plumelle et al. (2013)). Pour une description plus détaillée de la géophysique appliquée, on peut citer le livre de (Reynolds (2011)). Les mesures fournies par les essais géophysiques sont très utiles en reconnaissance de sols pour guider l'implantation des sondages et forages géotechniques ainsi que pour détecter les anomalies préjudiciables au projet. Cependant, dans la plupart des cas il n'est pas possible d'obtenir des paramètres géotechniques juste par le biais des essais géophysiques. De même, les résultats des essais géophysiques sur terrain doivent le plus souvent être validés par des forages ou par des sondages géotechniques.

L'objectif n'est pas ici de détailler toutes les méthodes géophysiques. Nous proposons dans la suite une présentation synthétique de la tomographie électrique car dans le cadre de ce travail, nous avons étudié l'apport de l'utilisation de la résistivité électrique ρ (en partie IV de ce document) à la méthodologie développée dans cette thèse. Il existe principalement deux types de méthodes pour mesurer la résistivité électrique du sol : les mesures directes de type sondage ou tomographie électrique (basées sur la loi d'Ohm) et les méthodes électromagnétiques (basées sur les équations de Maxwell). Dans ce travail, le dispositif utilisé a été la tomographie électrique afin d'étudier la variabilité spatiale d'un des terrains d'étude.

La tomographie électrique

Parmi les différentes méthodes géophysiques, la méthode de résistivité électrique est très utilisée dans les investigations du sol de surface et elle s'avère être un outil éprouvé pour la caractérisation des propriétés physiques du sol (Banton et al. (1997)). La mesure de résistivité du sol nécessite quatre électrodes (Fig.2.1), afin de former un quadripôle. l'essai consiste à mesurer la différence de potentiel ΔV entre deux électrodes (P_1, P_2), résultant de la circulation d'un courant électrique continu I injecté dans le sol, à partir d'une autre paire d'électrodes, dites de courant (C_1, C_2). En mesurant le potentiel en des points différents des points d'injection, on a pour un milieu homogène, une mesure proportionnelle à la résistivité.

La résistivité apparente, ρ_a , est celle qu'aurait le terrain homogène donnant la même valeur $\Delta V/I$ avec la même disposition des électrodes.

$$\rho_a = k \frac{\Delta V}{I} \quad (2.1)$$

où k est un facteur correctif qui dépend de la géométrie du quadripôle utilisé (disposition relative des électrodes de courant et de potentiel). La figure (Fig.2.1) présente plusieurs quadripôles, ainsi que leur facteur géométrique associé.

La résistivité électrique dépend de plusieurs propriétés du sol (Samouëlian et al. (2005)) :

- i. nature de la phase solide (distribution granulométrique et minéralogie) ;
- ii. l'indice de vides et son arrangement ;
- iii. le degré de saturation en eau ;
- iv. la salinité de l'eau ;
- v. la température.

Lorsqu'on réalise une série de mesures individuelles grâce à des dispositifs multiélectrodes le long d'un profil 2D, on peut "imager" la résistivité réelle du sol selon un plan vertical (2D) ou dans un volume de sol (3D). C'est le principe de la tomographie ou panneau électrique. Les mesures sont réalisées selon une séquence prédéfinie. Les valeurs de résistivité apparente ρ_a obtenues pour chacun des quadripôles de mesure sont reportées dans un plan vertical appelé pseudo-section. Afin d'obtenir la résistivité réelle du sol en chaque point de la coupe verticale, les valeurs de résistivité apparente doivent être inversées (Reynolds (2011)). Le principe de l'inversion consiste à trouver des modèles de sol (résistivité réelle) permettant d'expliquer les résistivités apparentes mesurées (pseudo-section). L'objectif de l'inversion est de minimiser l'écart entre la pseudo-section calculée pour un modèle de sol et la pseudo-section mesurée, celui-ci étant généralement quantifié par une fonction d'erreur quadratique moyenne (RMSE) (éq.2.2) comme suit :

$$RMSE(\%) = \sqrt{\frac{1}{N} \sum_{i=1}^N x_i^2} \quad (2.2)$$

où x_i représente l'écart entre la résistivité apparente mesurée et la résistivité calculée en chaque point de la pseudosection. Plusieurs algorithmes existent pour minimiser cette fonction au cours d'un processus itératif permettant d'aboutir à un modèle de sol qui explique mieux les mesures. Cependant le modèle de sol obtenu au terme de l'inversion constitue une solution qui n'est pas nécessairement unique pour représenter la résistivité réelle du sol. Dans le cadre de la reconnaissance des sols les méthodes électriques sont très utilisées pour identifier le niveau de la nappe phréatique, la recherche de matériaux et la position des horizons lithologiques ou la détection des réseaux karstiques. Le problème du contact sol-électrode reste alors la principale difficulté posée par l'emploi de la méthode électrique vis-à-vis de l'auscultation des sols construits. Dans ce cas, il s'avère nécessaire d'employer la méthode électrostatique ou des méthodes électromagnétiques basse fréquence.

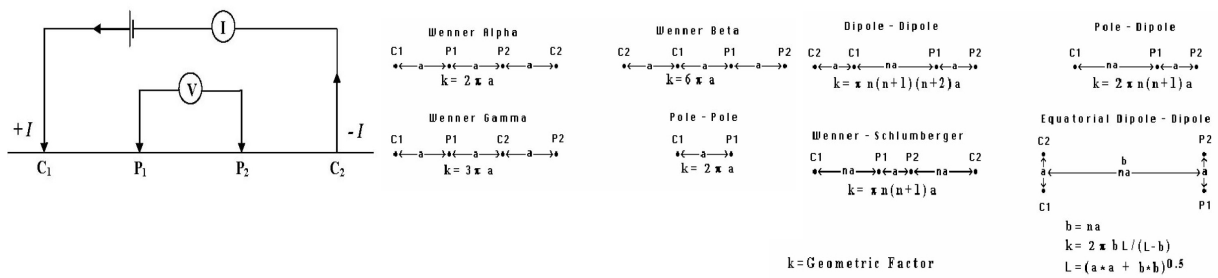


FIGURE 2.1 – Prospection électrique : à gauche principe de la mesure utilisant un quadripôle et à droite principaux quadripôles utilisés (Loke (2002))

2.3.2 Les forages et carottages

Les techniques de forage et de carottage du sol permettent de mener des investigations géologiques à caractère qualitatif voire quantitatif afin de :

- fournir une description du sol grâce à la réalisation d'une coupe lithologique du terrain ;
- obtenir des éprouvettes pour des essais en laboratoire. Le type d'essai (d'identification ou de comportement mécanique) dépend de la classe de prélèvement. D'ailleurs, le prélèvement comporte plusieurs opérations telles que la mise en caisse et transport des terrains rencontrés, le stockage et la paillasse au laboratoire. Pendant ces phases, le sol subit des déformations et des modifications de son état de contrainte (même si la teneur en eau ne varie pas) qu'il convient de minimiser autant que possible ;

Parmi les différentes techniques de forage, on distingue principalement :

- les sondages carottés qui fournissent un échantillonnage continu, permettant une bonne interprétation géologique, des analyses de sol et des essais de comportement. Le sondage carotté est le seul type de forage qui permet le prélèvement d'échantillons « intacts ». Il permet d'établir une coupe lithologique du terrain mais il reste un essai assez coûteux ;
- les forages destructifs valorisés par les digraphies fournissent des caractéristiques physiques des terrains avec des indications d'ordre géomécanique (ex. bancs durs, présence de blocs, etc.). Il est moins coûteux et plus rapide que le sondage carotté permettant de déduire une coupe lithologique indicative du terrain ;
- les tarières hélicoïdales ou manuelles (exclusivement adaptées aux sols fins peu compacts en surface : 3-4m), permettent le prélèvement d'échantillons remaniés. La compacité du sol et les éboulements éventuels limitent la mise en œuvre de ces essais, notamment dans le cas d'une tarière manuelle, où la présence d'un seul caillou peut arrêter la réalisation de l'essai. Par ailleurs, il peut être parfois difficile de préciser la côte de chacune des formations traversées à 20 ou 30cm près.
- les sondages à la pelle sont généralement limités à 6m et ne permettent que des prélèvements d'échantillons remaniés. Ils peuvent être pratiqués dans tous les sols non rocheux et les sols rocheux altérés mais la présence d'obstacles souterrains s'avère contraignante.

Pour plus de renseignements sur la pratique des forages et sondages, l'évolution des techniques et l'exploitation des différents essais, le lecteur intéressé peut s'adresser à l'ouvrage

(Reiffsteck et al. (2012)).

2.3.3 Les essais de laboratoire

Les essais de laboratoire ont pour but l'analyse et l'étude des échantillons prélevés sur terrain. Il a été défini trois catégories de méthodes de prélèvement (cf. Fig. 2.2) notées de A à C en fonction de la qualité pouvant être obtenue pour des échantillons de sols destinés aux essais en laboratoire. La catégorie A est la plus exigeante, c'est la catégorie pour laquelle on suppose que les propriétés des sols et des roches ne changent pas lors du prélèvement, de la manipulation, du transport et du stockage des échantillons.

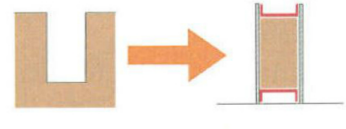
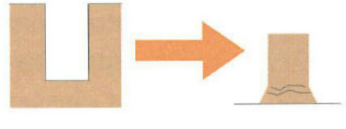
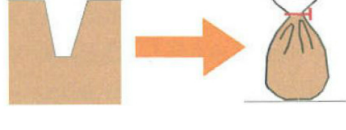
Type de prélèvement	Qualité	Catégorie	Nature des essais réalisables
	Précautions à toutes les étapes	A	Échantillon permettant de réaliser tout essai de laboratoire d'identification physique et chimique, de comportement mécanique
	Absence de précautions à une étape	B	Échantillon permettant de réaliser tous les essais sauf les essais mécaniques
	Sans précautions particulières	C	Échantillon permettant de réaliser tous les essais d'identification de nature mais pas d'état.

FIGURE 2.2 – Essais réalisables en fonction du type de prélèvement (source : Reiffsteck et al. (2012)).

Parmi les essais réalisables sur les échantillons prélevés, on peut distinguer :

- les essais d'identification destinés à déterminer des paramètres de nature afin notamment d'établir la classification du sol.
- les essais mécaniques dont les résultats permettent de construire les lois de comportement des sols en prenant en compte les phénomènes de compressibilité, de consolidation, de fluage et de cisaillement auxquels ils sont soumis.

Le principe et la description des principaux essais de laboratoire sont décrits dans (Plumelle et al. (2013)).

2.3.4 Les sondages in situ

Les sondages in situ permettent d'aboutir à une caractérisation mécanique des sols (en place) dans leur état de contrainte naturel. Ils peuvent être classés principalement en trois catégories, suivant le mode de sollicitation du sol : les essais de cisaillement, d'expansion et de pénétration (cf. Fig. 2.3). Tous ces essais sont des essais à la rupture, sauf l'essai pressiométrique qui sollicite le sol des petites aux grandes déformations. Une présentation succincte de ces trois familles d'essais sera réalisée par la suite. Une description plus en détail

est donnée dans des ouvrages géotechniques généraux (Plumelle et al. (2013), Reiffsteck et al. (2012)) et (Sanglerat (1972)) en ce qui concerne les essais de pénétration.

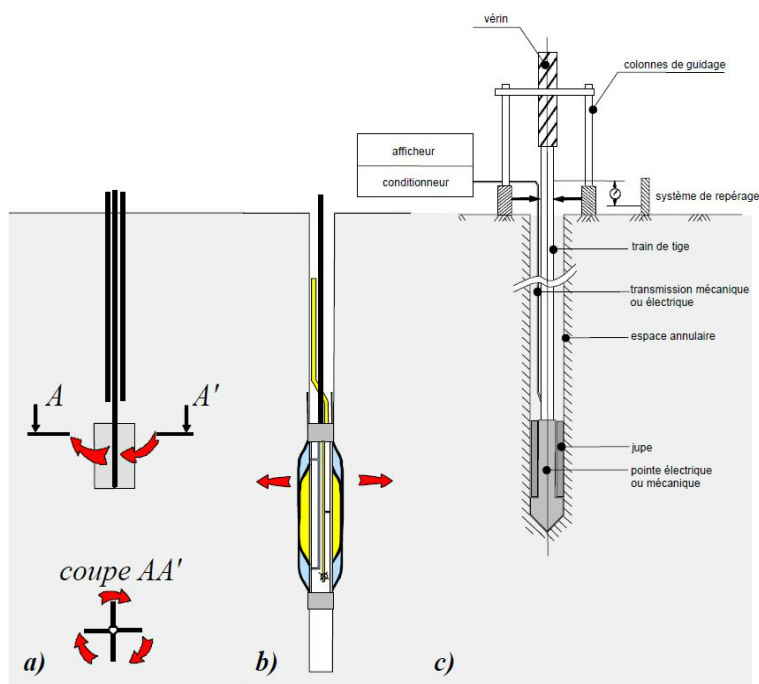


FIGURE 2.3 – Différents principes d'essais en place : a) exemple d'essai de cisaillement, b) d'expansion, c) de pénétration (source : Nasreddine (2004)).

Essais de cisaillement

Les essais de cisaillement ont pour but d'obtenir les caractéristiques de résistance des sols en imposant une contrainte de cisaillement. Parmi les essais de cisaillement, le plus connu est l'essai scissométrique (*Vane Probe Test*) (ex. Fig. 2.3.a). Étant un appareil simple d'emploi et peu encombrant, il s'avère être un moyen fiable pour connaître la résistance au cisaillement (cohésion) non drainée des argiles molles naturelles. Par ailleurs, il est possible de réaliser des essais de cisaillement en forage afin de mesurer sur site les caractéristiques de cisaillement.

Essais d'expansion

Les essais d'expansion permettent de déterminer les modules de déformation d'un terrain dans un état plus ou moins proche de son état initial suivant le mode de mise en place. Le principe est de dilater une sonde cylindrique radialement dans un forage pour déterminer la relation entre la pression appliquée par la sonde sur les parois du forage et le déplacement des parois du forage (ex. Fig. 2.3.b). Le principal essai est l'essai pressiométrique Ménard (PMT, *Ménard Pressumeter Test*) mis au point par Louis Ménard au début des années 1960. Il a pour objectif de mesurer les caractéristiques in situ du sol en petites déformations dans un domaine « pseudo-élastique » et à la rupture en grandes déformations.

Cet essai peut être réalisé dans tous les types de sols, saturés ou non, y compris le rocher (avec plus d'incertitude).

Le sondage pressiométrique comporte les deux opérations suivantes :

- i un forage destructif, opération délicate car de la qualité du forage dépend la fiabilité des résultats.
- ii. la réalisation de l'essai pressiométrique à proprement parler.

Le pressiomètre autoforeur, le pressio-pénétrromètre, l'essai de pression latérale au dilatomètre de Marchetti DMT (même s'il diffère par sa géométrie des essais d'expansion cylindrique) sont d'autres essais d'expansion reconnus en géotechnique.

Essais de pénétration

Le pénétrromètre (ex. [Fig.2.3.c](#)) permet une mesure ponctuelle de la résistance d'un terrain en fonction de la profondeur. Les essais de pénétration peuvent être réalisés par fonçage d'une pointe conique (pénétrromètre statique CPT), par battage à l'aide d'un mouton, d'une pointe conique (pénétrromètre dynamique DPT) ou par battage à l'aide d'un mouton, d'un carottier (essai de pénétration au carottier SPT).

L'essai de pénétration statique CPT s'applique à tous les sols fins et aux sols grenus dont la dimension moyenne des éléments ne dépasse pas 20 mm. Les essais de pénétration dynamique peuvent être réalisés dans tous les sols fins et grenus dont la dimension moyenne des éléments ne dépasse pas 60 mm, à l'exception de l'essai SPT, pour lequel la dimension moyenne des éléments doit rester inférieure à 20mm.

2.3.5 Couplage des données

Malgré la possibilité offerte par les méthodes géophysiques, d'ausculter de manière non intrusive un grand volume de sol, leurs principales contraintes sont d'une part la diminution de la résolution avec la profondeur et d'autre part le fait que les propriétés mécaniques du sol ne sont pas obtenues directement. Par contre, les essais in situ fournissent des paramètres mécaniques mais seulement en quelques points de l'espace. Par ailleurs, le modèle de sol obtenu au moyen des essais géophysiques nécessite une analyse inverse pour laquelle l'unicité de la solution n'existe pas. S'appuyer sur des données mécaniques mesurées à l'aide des sondages géotechniques s'avère intéressant puisqu'il peut permettre un meilleur calage de certains paramètres de l'inversion. C'est pourquoi, ces deux approches s'avèrent complémentaires. L'utilisation conjointe des mesures géophysiques et géotechniques pour améliorer et affiner les campagnes de reconnaissance devient de plus en plus répandu et constitue un sujet d'intérêt dans la communauté scientifique et des praticiens.

A titre d'exemple, dans le cadre de la reconnaissance des sols en surface, ([Cosenza et al. \(2006\)](#)) présente l'étude conjointe d'une campagne d'essais géotechniques (pénétrromètre dynamique léger, scissomètre et pressiomètre) et géophysiques (tomographie électrique, radar-sol et sismique de réfraction) dans un site alluvial. L'interface entre les formations alluviales et calcaires est bien mise en évidence par la tomographie électrique. De même, le couple de mesures de résistance du (DPT) et de la résistivité permettent de distinguer correctement l'interface entre dépôts graveleux et argileux. Par ailleurs, dans le travail de thèse de ([Sal-loum \(2015\)](#)) les résultats combinés d'une campagne d'essais géophysiques et géotechniques ont permis de mieux caractériser la lithologie de la plaine alluviale de Beyrouth (Liban). ([Guerrero \(2014\)](#)) présente une approche basée sur la tomographie électrique pour l'aide au

positionnement des sondages géotechniques et étendre les mesures de résistance aux zones présentant les mêmes propriétés physiques.

En outre, plusieurs travaux dans la littérature portent sur la relation entre le module pressiométrique E_M et la résistivité électrique (Monnet et al. (2007), Denis et al. (2011)). La corrélation linéaire entre ces deux paramètres, mise en évidence par ces travaux, a été utilisée par (Imanzadeh (2013)) pour fournir une estimation du module E_M dans le plan horizontal (X, Y) par le biais du cokrigeage¹.

2.3.6 Bilan

La reconnaissance géologique représente la première étape d'une étude de sols dont le but est la compréhension du site d'étude et la caractérisation des propriétés physico-mécaniques. Les prospections géophysiques permettent de tracer les grands traits du modèle géologique, affiné grâce aux forages géologiques, qui intéressent un volume de terrain limité. L'étude géologique est complétée de la mesure de divers paramètres de sol, notamment les caractéristiques mécaniques qui peuvent être mesurées, soit in situ, soit en laboratoire sur des échantillons prélevés.

L'intérêt de combiner à la fois des méthodes géotechniques et géophysiques dans le cadre d'une campagne de reconnaissance a été aussi évoqué. Les sondages mécaniques permettent de valider les modèles obtenus par l'inversion des mesures géophysiques et d'accéder aux paramètres de calcul des ouvrages géotechniques. Les méthodes géophysiques peuvent guider l'implantation des sondages mécaniques sur place et compléter le caractère ponctuel de ces essais.

De nos jours, les essais in situ ont reçu beaucoup d'attention grâce, d'une part, aux importants progrès réalisés dans le domaine de la métrologie, des capteurs, de l'acquisition, du traitement et du stockage de données et d'autre part, à l'évolution des approches théoriques de plus en plus rigoureuses (Schnaid (2005)). De ce fait, les essais in situ sont souvent beaucoup plus usités que les essais de laboratoire dans la pratique courante.

Parmi la grande palette d'essais in situ disponibles à l'heure actuelle, les essais pénétrométriques sont les plus couramment utilisés car leur mise en place et leur réalisation sont relativement rapides et permettent d'ausculter la plupart des sols superficiels. Leurs applications sont variées, allant de la reconnaissance classique des sols en vue de calculs de fondations, jusqu'à l'estimation des paramètres rhéologiques des sols rencontrés. Par ailleurs, ils permettent de fournir une mesure mécanique quasi-continue permettant de caractériser la variabilité des différentes formations du sous-sol en fonction de la profondeur, ce qu'aucun autre sondage mécanique n'est en mesure de proposer. La section suivante est dédiée à l'essai de pénétration.

1. Le cokrigeage est une extension du krigeage, technique géostatistique d'estimation conditionnelle, pour plusieurs variables. Le principe sera présenté dans la partie IV de ce document.

3. L'essai pénétrométrique

Utilisé depuis le début du 20^{me} siècle, l'essai pénétrométrique consiste à foncer dans le sol « un pieu à échelle réduite » muni d'une pointe conique. Il existe des pénétromètres de toutes tailles et de toutes puissances de pénétration. Toutefois, comme évoqué précédemment, la principale classification est fonction du mode de fonçage. On distingue ainsi :

- les pénétromètres statiques (ou quasi-statiques)
- les pénétromètres dynamiques

3.1 L'essai de pénétration au cône statique (CPT et CPTU)

Les pénétromètres statiques (ou quasi-statiques) consistent à foncer la pointe dans les terrains à vitesse faible et constante à l'aide d'un vérin. On distingue le pénétromètre à pointe mécanique (CPTM) et le pénétromètre à pointe électrique connu sous l'acronyme de CPT (*Cone Penetration testing*) en anglais, beaucoup plus précis que le CPTM. La nouvelle norme européenne (NF EN ISO 22476-1 (2010)) régit l'utilisation de cet essai. L'essai au piezocône (CPTU) équipé d'une couronne filtrante piezométrique, permet la mesure d'un paramètre additionnel, la pression interstitielle u . Il fait l'objet de la norme (CPTU) (NF P 94-119 (1995)). Le pénétromètre statique, dans ses deux versions (CPT et CPTU) est reconnu depuis des dizaines d'années comme un outil puissant et fiable pour la mesure sur place de résistances de sols. Il permet de déterminer rapidement le profil des sols et leurs caractéristiques stratigraphiques. Il offre aussi la possibilité d'aborder l'identification des sols en termes de nature et d'état. Cet aspect sera abordé dans la partie II de ce document, où l'on approfondira sur l'interprétation de cet essai.

3.1.1 Principe et caractéristiques mesurées

Comme décrit précédemment, cet essai consiste à enfoncer dans le sol, à vitesse faible et constante, un train de tiges terminé par une pointe munie d'un cône et à mesurer de manière continue la résistance à la pénétration q_c du cône et généralement le frottement sur le manchon.

La force totale Q_c nécessaire pour enfoncer la pointe conique divisée par la section projetée du cône correspond à la contrainte q_c . La force Q_s nécessaire à l'enfoncement du manchon divisée par sa surface latérale A_s permet de calculer le frottement latéral unitaire local f_s . Le rapport de frottement F_r (*friction ratio*), à une profondeur donnée est défini par le rapport f_s/q_c . Pour le piezocône, il mesure aussi de manière continue la pression interstitielle u .

3.1.2 Équipement et procédures

La gamme des tailles de pointes disponibles varie entre 2cm^2 pour les reconnaissances à faible surface jusqu'à 40cm^2 pour les sols graveleux. Cependant, le pénétromètre le plus souvent employé utilise une pointe de 10cm^2 dont le diamètre de 36mm est égal à celui des tiges.

La vitesse d'enfoncement standard du cône est égal à 2 cm/s. En ce qui concerne l'intervalle entre deux mesures, la plupart des pénétromètres sont équipés de systèmes permettant une acquisition des mesures entre 10 et 50mm. Cependant, un pas de mesure de 20 mm est le plus commun, ce qui permet d'avoir une information détaillée des conditions du sol.

Par rapport au CPTU, on peut trouver différentes notations pour la mesure de la pression interstitielle en selon la position de la couronne filtrante :

- position u_1 sur la pointe ;
- position u_2 à l'arrière de la pointe. Celle ci étant la position la plus commune ;
- position u_3 directement après le manchon de frottement ;



FIGURE 3.1 – CPT monté sur un camion (source : Robertson and Cabal (2015))



FIGURE 3.2 – CPT sur dispositif portable pour des zones d'accès limité (source : Robertson and Cabal (2015))

3.1.3 Avantages et inconvénients de l'essai

Cet essai fournit un profil de mesure en continu du sol, de manière rapide et sans forage préalable (environ 30 minutes pour une profondeur de 20m). Il est un essai répétable et fiable. Par ailleurs, une grande panoplie de travaux théoriques ont été menés afin d'améliorer l'interprétation de l'essai. Une grande partie des connaissances acquises grâce aux expériences réalisées, a été regroupée dans l'ouvrage (Lunne et al. (1997)). Un autre travail de référence plus orienté vers l'utilisation pratique est celui de (Robertson and Cabal (2015)).

Parmi les principales contraintes de cet essai, on peut signaler que c'est un essai "aveugle" du fait qu'il n'est pas possible de prélever des échantillons directement durant l'essai. Par ailleurs, dans la plupart des cas, ces pénétromètres sont montés sur camions ou chenillards (cf.Fig.3.1) (même si des versions portables existent (cf.Fig.3.2)) ce qui peut être contraignant

pour la réalisation d'essais sur les sites difficiles d'accès. En outre, comme évoqué auparavant, le domaine d'application de l'essai est limité aux sols dont la dimension moyenne des éléments ne dépasse pas 20 mm. L'essai n'est donc pas adapté aux terrains où ils existent des couches de sols graveleux ou de sables cimentés. Dans ce type de sol, lors du fonçage, le pénétromètre peut rester bloqué par une couche plus dure. La pointe électrique, qui comporte des capteurs, peut également être abîmée dans ce cas.

3.2 Les pénétromètres dynamiques

Le principe des pénétromètres dynamiques est d'enfoncer par battage un train de tiges terminé par une pointe débordante. Vis-à-vis de la reconnaissance de sols, la norme française distingue deux types de pénétromètres :

- i. le pénétromètre dynamique de type A (NF P 94-114 (1990)) si un système d'injection de boue de forage au-dessus de la pointe est prévu afin de diminuer le frottement latéral ;
- ii. le sondage dynamique de type B (NF P 94-115 (1990)).

Par ailleurs, on distingue également l'essai dynamique de pénétration au carottier SPT, qui fait l'objet de la norme française (NF P 94-116 (1991)).

L'essai de pénétration dynamique est sans doute le plus ancien des essais géotechniques in situ. Il avait été conçu afin de mesurer d'une manière qualitative la résistance à la pénétration du sol et particulièrement pour déterminer la compacité des sols non cohésifs dans lesquels le prélèvement des échantillons est souvent compliqué (Sanglerat (1972)). De manière générale, il permet de déterminer :

- la succession des différentes couches de terrain ;
- l'homogénéité d'une couche ;
- la position d'une couche résistante.

Ils peuvent également être utilisés pour déterminer des propriétés de résistance et de déformation des sols généralement sans cohésion, mais aussi des sols fins, et ce par des corrélations appropriés. Toutefois en France, l'essai de pénétration dynamique est déconseillé pour le calcul de la capacité portante bien que les essais de type A peuvent fournir un ordre de grandeur de cette capacité. Les résultats des essais au pénétromètre dynamique de type B peuvent permettre de guider le choix du type de fondation.

Dans ce qui suit, on présente tout d'abord l'essai de pénétration au carottier SPT et ensuite les essais au cône dynamique DPT en distinguant :

- les essais à énergie constante ;
- les essais à énergie variable.

3.2.1 Essai de pénétration au carottier (SPT)

L'essai SPT est l'un des essais les plus anciens. Il a été développé aux États-Unis en 1902. Son principe est assez simple : on enfonce un carottier normalisé battu en fond d'un forage préalable, en laissant tomber un mouton de 63.5 kg d'une hauteur de chute de 0.76m. La résistance à la pénétration dynamique est calculée à partir du nombre de coups (N_{SPT}) pour enfoncer le carottier de 30cm, après un enfoncement d'amorçage de 15cm. Toutefois, ce nombre de coups peut être influencé par plusieurs facteurs tels que la quantité d'énergie

transmise du mouton au train de tiges, les pertes par frottement, les longueurs des tiges, le diamètre du forage ou l'état de contraintes au niveau de l'essai. En effet, l'inefficacité de l'énergie délivrée représente une grande contrainte pour cet essai et il est considéré que le système utilisé en Europe ou aux États-Unis présente un taux d'énergie libéré valant 60% de l'énergie théorique. C'est pourquoi, la pratique internationale (ISO (2005)) suggère une normalisation par plusieurs facteurs de correction du nombre de coups (N_{SPT}) en calculant la valeur $(N_1)_{60}$.

Un des grands intérêts de cet essai est la grande base de données de résultats existants ayant permis d'établir une grande panoplie de corrélations empiriques avec divers paramètres du sol, tels que la densité relative, l'angle de frottement des sols sableux ou la résistance à la compression simple d'une argile. Une grande partie de ces relations ont été publiées dans le livre de Terzaghi et Peck (Terzaghi et al. (1996)). Cet essai a aussi servi d'essai de référence pour évaluer la susceptibilité à la liquéfaction des sols sous séisme. Néanmoins, certains auteurs ont critiqué cette « capacité universelle » d'évaluer tous les paramètres du sol uniquement par le biais du nombre de coups fourni par l'essai (Mayne et al. (2009)).

Il existe de nombreuses corrélations entre l'essai au carottier SPT et le pénétromètre statique CPT. Un grand nombre de ces relations ont été regroupées par (Cassan (1978)).

3.2.2 Essai de pénétration au cône dynamique (DPT) à énergie constante

Pour cet essai, on mesure l'enfoncement d'une pointe conique débordante (perdue ou non) soumise à une énergie de battage constante, fournie par une masse frappante à une cadence de 15 à 30 coups par minute, tombant en chute libre d'une hauteur normalisée. Le nombre de coups est noté en fonction de la profondeur atteinte par la pointe, mais aucun échantillon n'est prélevé.

La norme européenne (NF EN ISO 22476-2 (2005)) spécifie les caractéristiques des cinq types d'appareillage pour l'essai de pénétration dynamique suite à l'énergie spécifique de battage par coup, à savoir :

- essai au pénétromètre dynamique léger (DPL) : essai effectué avec la masse la plus faible (10 ± 0.1 kg) de la gamme des pénétromètres dynamiques ;
- essai au pénétromètre dynamique moyen (DPM) : essai effectué avec la masse moyenne (30 ± 0.3 kg) de la gamme des pénétromètres dynamiques ;
- essai au pénétromètre dynamique lourd (DPH) : essai effectué avec la masse moyenne à très lourde (50 ± 0.5 kg) de la gamme des pénétromètres dynamiques ;
- essai au pénétromètre dynamique ultra lourd (DPSH-A ou DPSH-B) : essai effectué avec la masse la plus élevée (63.5 ± 0.5 kg) de la gamme des pénétromètres dynamiques.

La hauteur de chute normalisée est égale à 500 ± 10 mm sauf pour le type DPSH-B (750 ± 20 mm).

Les résultats d'essais de pénétration dynamique sont généralement présentés sous forme de nombre de coups pour 10 cm de pénétration N_{10} en fonction de la profondeur pour les DPL, DPM, DPH et N_{10} ou N_{20} (nombre de coups pour 20 cm de pénétration) pour les DPSH-A et DPSH-B. Une autre possibilité d'exploitation des résultats d'essai est l'utilisation de la résistance dynamique de pointe q_d . Le paramètre q_d est calculé par le biais de la formule de

battage dynamique des Hollandais obtenue à partir de la théorie des chocs de Newton. Les hypothèses de cette formule sont :

- i. le choc est parfaitement inélastique
- ii. les déformations élastiques du pénétromètre et du sol sont négligeables

et la formule s'écrit comme suit :

$$Q_d = \frac{MgH}{e} \frac{1}{(1+a)} \quad (3.1)$$

avec M le poids de la masse frappante, H la hauteur de chute, g l'accélération de la pesanteur et e l'enfoncement moyen par coup. La constante a est égale au rapport de masses $a = M'/M$, où M' représente le poids des parties frappées (enclume placée en tête du train de tiges et sur laquelle s'exercent les chocs, train de tiges et pointe). Pour plus de renseignements sur la démonstration théorique de cette formule ainsi que des formules de battage dynamique selon la théorie des chocs, on consultera ([Benz-Navarrete \(2009\)](#)).

Par ailleurs on peut trouver des études détaillant les corrélations entre les pénétromètres statiques et dynamiques dans les livres ([Sanglerat \(1972\)](#), [Cassan \(1978\)](#)).

3.2.3 Essai de pénétration au cône dynamique (DPT) à énergie variable

Une variante de l'essai au pénétromètre dynamique classique a été développée en France en 1989 ([Gourvès \(1991\)](#)), il s'agit du pénétromètre dynamique léger à énergie variable Panda®. L'essai consiste à fournir une énergie variable par le choc d'un marteau sur une tête de battage solidaire d'un train de tiges, équipé d'une pointe de pénétration. L'un des grands avantages de ce pénétromètre est qu'il est possible de réaliser une auscultation fine des couches de faible à moyenne résistance en adaptant l'intensité manuelle de frappe. L'énergie délivrée au système et l'enfoncement de la pointe sont mesurés pour chaque coup de marteau. L'essai a été normalisé par la norme française ([NF XP P 94-105 \(2000\)](#)).

La petite taille de l'outil de mesure ainsi que sa portabilité, facilitent la possibilité d'effectuer l'essai dans des contextes difficiles d'accès voire des essais à l'horizontale. La taille des pointes utilisées varie entre 2 et 10cm² avec des tiges de 14 mm de diamètre. Le domaine d'application pratique du Panda concerne, outre la reconnaissance des sols, le contrôle de compactage des remblais routiers, celui des digues et barrages en terre, le diagnostic de remblais (dallages fissurés, canalisations enterrées, murs de soutènement...).

3.3 Le pénétromètre dynamique Panda

3.3.1 Introduction

C'est pour répondre à certaines problématiques de la reconnaissance des sols de surface (essais sur sols de faible portance, essais sur site d'accès difficile, budget des campagnes d'essais réduits,...) qu'a été créé le pénétromètre dynamique léger à énergie variable Panda® (Pénétromètre Autonome Numérique Dynamique Assisté par ordinateur). L'objectif était de développer un pénétromètre simple, rapide à mettre en place et à utiliser, économique et autonome. Il a été conçu et développé au C/U/S/T de l'Université Blaise Pascal ([Gourvès \(1991\)](#)) (cf.[Fig.3.3](#)), au début des années 1990.

Depuis presque une trentaine d'années depuis son apparition, les développements technologiques et les développements théoriques sur la propagation des ondes ont permis son évolution d'abord vers le Panda2 (Benz-Navarrete (2009)), puis vers le Panda 3 (Benz-Navarrete (2009), Escobar (2015)).

La partie suivante présente le principe de fonctionnement de l'appareil ainsi que ses évolutions.

LE PANDA



**Pénétrömètre dynamique léger
à énergie variable
pour la reconnaissance des sols**

Ruediger Gierke - LABORATOIRE DE GÉNIE CIVIL - CNRS - UNIVERSITÉ BREST - FACULTÉ DE GÉNIE CIVIL - FRANCE
BP 208 - 45174 ANNECYN CEDEX - Tél. 71 42 75 30 - Télécopie 71 42 75 10



FIGURE 3.4 – Exemple de réalisation d'un sondage Panda

FIGURE 3.3 – Première plaquette de diffusion de l'essai Panda (source : (Gourvès (1991)))

3.3.2 Principe de la mesure

Pour chaque coup donné par un marteau standardisé, on enregistre une mesure énergétique au niveau de la tête de battage ainsi que l'enfoncement de la pointe mesuré à l'aide d'un capteur à courroie. À partir de ces deux mesures, la résistance de pointe q_d est calculée à l'aide de la formule des Hollandais (éq.3.1) par le boîtier calculateur. Pour ce faire, on remplace le terme correspondant à l'énergie potentielle (MgH) par :

- l'énergie cinétique délivrée par le marteau (Panda 1) (cf.Fig.3.5) (Gourvès (1991))(Zhou (1997)) ;
- l'énergie de déformation élastique du corps de la tête de battage (Panda 2) (cf.Fig.3.6).

La différence entre les deux versions de l'appareil réside dans le mode de mesure de l'énergie de battage ainsi que dans le type des capteurs installés dans la tête.

La mesure de l'énergie de battage dans le pénétrömètre Panda 1 est basée sur la mesure de la vitesse du marteau au moment de l'impact v_m , ce qui permet de déterminer l'énergie cinétique de battage. La mesure consiste à enregistrer le temps de passage dt d'un aimant solidaire du piston mobile de la tête par le biais d'un capteur magnétique à effet Hall, dont les cellules ont un écartement l . L'intervalle de temps dt est calculé comme la différence entre l'instant t_2 (l'aimant passe face à la cellule B) et l'instant t_1 (l'aimant passe face à la cellule A), tel qu'illustré dans la figure (Fig.3.7) (Zhou (1997)).

Le Panda 2 a été développé afin d'améliorer la transmission de l'énergie au sein de l'appareil et s'affranchir d'une certaine méconnaissance de la quantité d'énergie cinétique délivrée manuellement par la masse frappante. Dans cet appareil, l'évaluation de la résistance de

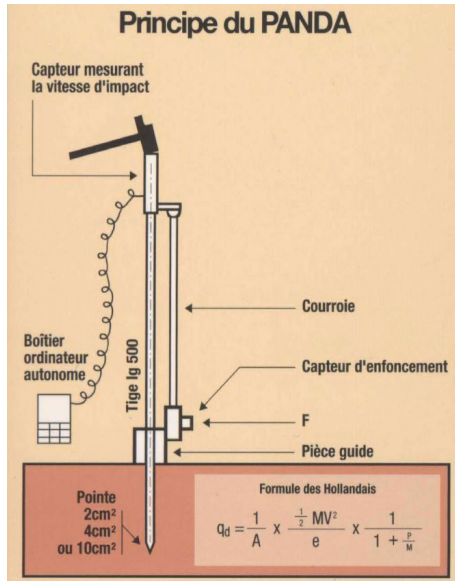


FIGURE 3.5 – Principe du Panda 1

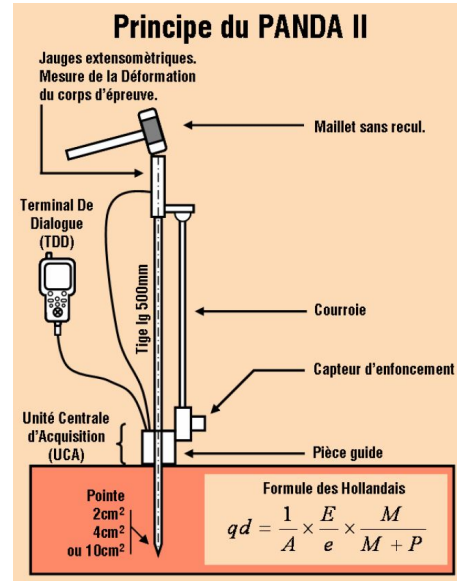


FIGURE 3.6 – Principe du Panda 2

pointe n'est plus fondée sur la formule de battage dynamique basée sur les lois de Newton, mais à partir de la théorie de propagation longitudinale d'une onde de compression dans une tige élastique. L'onde de contrainte entraînée par l'impact du marteau est ainsi enregistrée par des jauges de déformations installées au niveau de la tête de battage (cf. Fig. 3.8), permettant de mesurer l'énergie « vraiment » transmise au niveau de la tête juste après l'impact. Le Panda 2 est ainsi le premier pénétromètre dynamique qui incorpore ce type de mesures pour l'exploitation de l'information au cours d'un sondage traditionnel. Pour plus de renseignements sur la conception et la validation de cette nouvelle version de l'essai le lecteur intéressé peut lire (Benz-Navarrete (2009)).

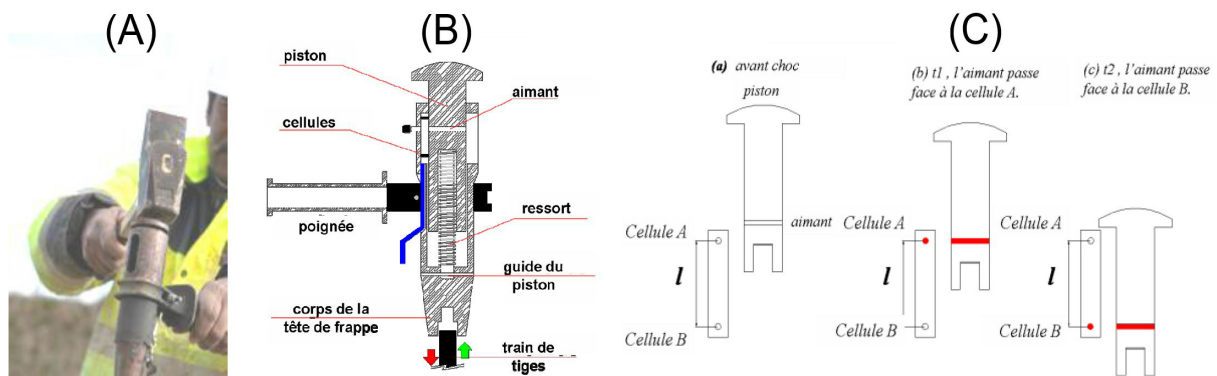


FIGURE 3.7 – (A) Photo de la tête de battage (B) schéma de la tête de battage (C) principe de mesure de la vitesse d'impact

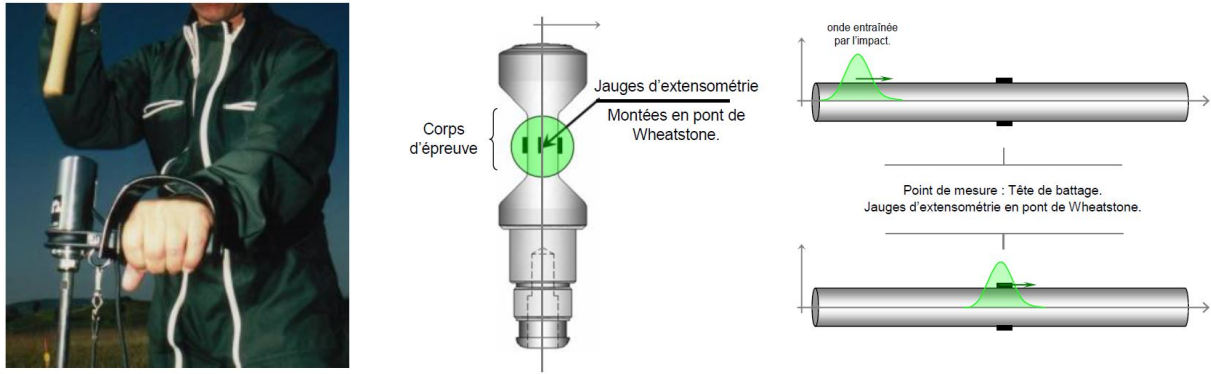


FIGURE 3.8 – Schéma du mode de mesure Panda 2 de l'énergie apportée au système juste après l'impact du marteau (source : Benz-Navarrete (2009))

3.3.3 Résultats, exploitation et domaine d'application

Les investigations avec le pénétromètre dynamique Panda peuvent être menées jusqu'à une profondeur d'environ 5-6 m en fonction de la résistance du sol (dans la limite de 50 MPa) et du frottement agissant sur les tiges. Il peut être réalisé dans tous les sols fins et grenus dont la taille maximale des particules est inférieure à 50 mm (Zhou (1997), Benz-Navarrete (2009)).

Par ailleurs, (Chaigneau (2001)) recommande des enfoncements compris entre 2 mm et 20 mm durant la réalisation de l'essai afin que les hypothèses de la formule des Hollandais soient vérifiées en réduisant ainsi la marge d'erreur sur la mesure de la résistance de pointe q_d . Le fait que cet essai soit à énergie variable permet d'adapter la puissance de battage. Les mesures obtenues au pénétromètre dynamique permettent d'établir des pénétrogrammes possédant une très grande résolution tel qu'il est illustré dans la figure (Fig.3.9). La grande collecte de données fournies par cet appareil facilite la mise en œuvre des études statistiques pour caractériser la réponse mécanique du milieu et étudier ainsi sa variabilité spatiale.

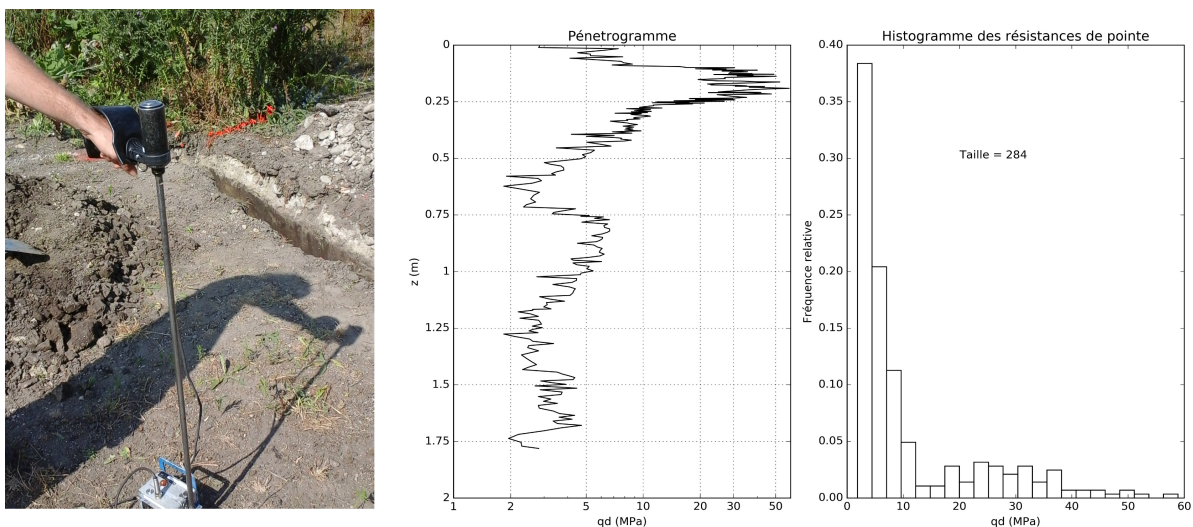


FIGURE 3.9 – Exemple de pénétrogramme obtenu avec l'essai Panda et histogramme des mesures de résistance de pointe

Compte tenu des caractéristiques évoquées, ce pénétromètre fournit un profil de la résistance du sol en fonction de la profondeur presque en continu. Ceci permet une identification aisée des épaisseurs des couches traversées ainsi que des insuffisances de compactage dans des ouvrages de terrassement. C'est pourquoi son utilisation a été préconisée principalement pour la reconnaissance des sols à faible profondeur et pour le contrôle de compactage des remblais routiers, des tranchées, des digues ou encore des barrages en terre ou de résidus miniers.

Parmi les possibilités d'exploitation des résultats pénétrométriques issus de l'essai Panda, l'estimation de la densité des sols en place a été développée notamment pour permettre le contrôle de compactage des ouvrages en terre (NF XP P 94-105 (2000)). Le principe de cette estimation de la densité des sols en place repose sur l'hypothèse d'une relation semi-logarithmique entre la résistance de pointe q_d et le poids volumique sec pour un matériau donné dans un état hydrique donné (quand il est sensible à l'eau) telle que $\gamma_d = f(q_d, w)$ (Chaigneau (2001)). Cette relation a été établie à partir d'essais pénétrométriques Panda réalisés en moule de calibration pour un ensemble de sols représentatifs de la classification des sols du Guide des Terrassements Routiers (NF P 11-300 (1992)) et pour différents états hydriques. Cette banque de sols a été élaborée au sein du laboratoire de Polytech Clermont Clermont Ferrand depuis plusieurs années (Chaigneau (2001)) et chaque matériau a été identifié et est conservé physiquement au sein du laboratoire. Cette banque contient actuellement plus d'une trentaine de matériaux et son exploitation permet de relier les paramètres de nature et d'assemblage du sol avec la réponse mécanique à une sollicitation donnée ou de corréler des caractéristiques mécaniques entre elles.

Reconnaissance des sols à faible profondeur

Le travail de (Zhou (1997)) avait déjà montré les avantages d'un pénétromètre léger à énergie variable tel que le Panda dans le cadre de la reconnaissance des sols en surface. Entre autres, la possibilité d'exécution des sondages sur des sites d'accès difficile et dans toutes les directions, d'une manière rapide en assurant la fiabilité des résultats fournis.

Suite à (Zhou (1997)) plusieurs travaux se sont appuyés sur le pénétromètre dynamique Panda dans le cadre de la prospection géotechnique à faible profondeur. Pour citer quelques exemples (Galinié et al. (2003)) ont utilisé le Panda pour évaluer et caractériser des couches archéologiques dans le milieu urbain mais aussi afin de quantifier la variabilité spatiale. Par ailleurs, (Athapaththu et al. (2015)) proposent une méthodologie de reconnaissance des talus naturels difficiles d'accès basée sur le Panda et des travaux précédents ont montré l'intérêt de l'outil vis-à-vis de l'évaluation de la variabilité spatiale du sous-sol (Athapaththu et al. (2007)).

Techniques de couplage

Malgré ses avantages, le pénétromètre Panda ne permet pas de connaître la nature des matériaux traversés, puisque c'est un sondage mécanique « aveugle », comme la plupart de pénétromètres. C'est pourquoi, d'autres techniques complémentaires ont été développées afin de fournir des mesures complémentaires à la résistance de pointe dynamique q_d .

Dans cet esprit, il a été proposé la technique dite de « géoendoscopie » qui permet de compléter le sondage mécanique. L'utilisation de l'endoscopie et de l'analyse d'images permet

ainsi de récolter et d'étudier des images des matériaux in situ. De même, il rend possible la détermination des caractéristiques physiques des matériaux auscultés, notamment leur distribution granulométrique. Le couplage de l'endoscopie à l'essai de pénétration Panda a déjà fait l'objet de deux travaux de thèse (Breul (1999b), Haddani (2004)).

Enfin, dans le cadre d'une méthode de diagnostic de digues basée sur des techniques originales économiques (Lepetit (2002)) a bâti une sonde pour la mesure de la perméabilité in situ dans la cavité réalisée suite à un essai au Panda pour l'estimation de la stabilité mécanique de l'ouvrage.

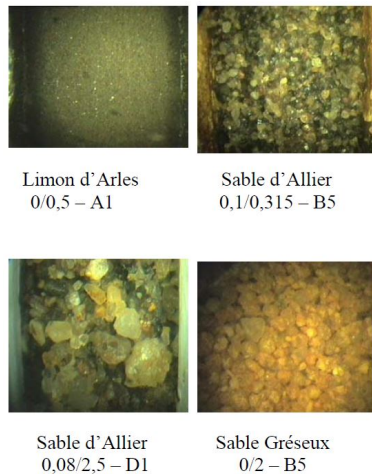


FIGURE 3.10 – Exemples d'images de matériaux obtenues en laboratoire grâce à l'endoscopie (source : Breul (1999a))



FIGURE 3.11 – Exemples de la réalisation de l'essai de géondoscopie sur terrain et d'images obtenues

3.3.4 Le pénétromètre Panda 3

Le Panda 3 correspond à la dernière évolution du pénétromètre Panda (Benz-Navarrete (2009)) (Escobar (2015)). Dans cette version, l'analyse du phénomène de battage dynamique change complètement par rapport à la version 1 ou 2. En effet, le principe de l'essai est fondé sur l'analyse de la propagation des ondes engendrées par le choc du marteau sur la tête du pénétromètre durant l'enfoncement. À partir du calcul de la force, de l'accélération et de la vitesse des ondes au niveau de la pointe, il est ainsi possible d'obtenir une courbe charge-enfoncement représentative du comportement du sol soumis à la pénétration dynamique.

Le principe de l'essai est schématisé dans la figure (Fig.3.12). Lors du battage, la variation de déformation $\varepsilon(t)$ et d'accélération $a(t)$ entraînée par le passage de l'onde de compression créée par l'impact est mesurée au niveau de la tête de mesure. Pour chaque coup de marteau, à condition de découpler les ondes descendantes et remontantes ε_d et ε_r , on peut calculer l'enfoncement $s_p(t)$ et la force $F_p(t)$ résultante en pointe. En faisant certaines hypothèses simplificatrices, il est possible de tracer la courbe de contrainte en fonction de l'enfoncement dynamique $\sigma_p = f(s_p)$ pour chaque impact du marteau. Une méthodologie analytique d'exploitation de cette courbe permet d'estimer différents paramètres du sol, à savoir :

- des paramètres de rupture (résistance de pointe statique et dynamique)
- le module de déformation dynamique E_d^{P3} au chargement et déchargement
- la vitesse des ondes de compression c_p^{P3} et des ondes de cisaillement c_s^{P3} dans le sol

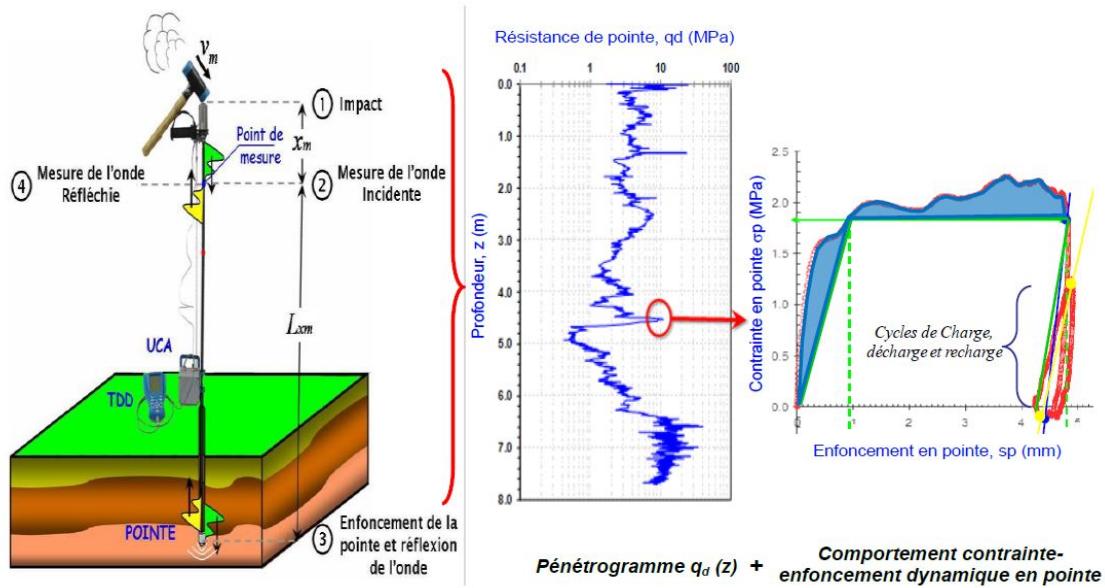


FIGURE 3.12 – Principe du Panda 3 (source : Benz-Navarrete (2009))

- un coefficient d'amortissement J_s

Un graphique typique fournissant l'évolution des paramètres mesurés à l'aide du Panda 3 en fonction de la profondeur est fourni à la figure (Fig.3.13) pour deux éprouvettes de sable d'Allier à différentes états de densité (D_1 et D_2). Les valeurs de résistance de pointe mesurées lors d'un essai Panda 2 sont aussi incluses.

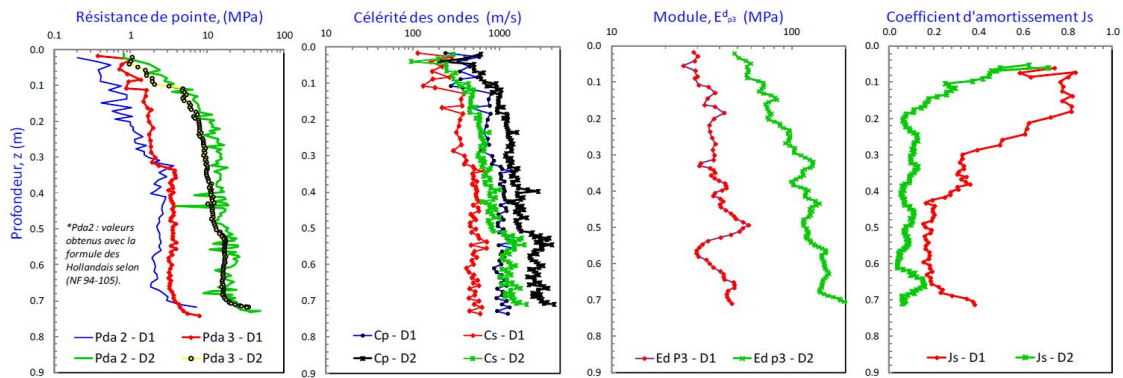


FIGURE 3.13 – Profils de mesures recueillies par l'essai Panda 3 obtenus en moule d'étalonnage pour un sable d'Allier sec à deux états de densité différents D_1 et D_2 (source : Escobar (2015))

3.4 Bilan

Les essais de pénétration dynamique ont été très utilisés pour étudier la stratigraphie et le comportement dynamique des sols ou l'évaluation du risque de liquéfaction. Cependant, la plupart sont lourds et encombrants ce qui limite leur utilisation dans les sites difficiles d'accès ou pour certaines applications géotechniques et rend le temps d'exécution des sondages relativement long. Par ailleurs, du fait que l'énergie de battage soit importante et constante,

leur résolution de mesure n'est pas très élevée notamment dans les sols de faible portance. De même, la mesure de l'énergie transmise lors du battage n'est pas bien maîtrisée, entraînant des incertitudes sur la précision de la résistance de pointe mesurée.

Le pénétromètre léger à énergie variable Panda, dont l'apparition date d'une trentaine d'années permet de palier ces limitations puisque :

- il a une résolution d'acquisition de la résistance des sols en place très fine grâce notamment à la possibilité d'adapter manuellement l'énergie de battage ;
- du fait de sa légèreté et de sa rapidité de mise en œuvre, il permet de réaliser des sondages dans tous les sites même ceux d'accès difficile et de multiplier les sondages sur un même site fournissant ainsi un grand nombre de données pouvant permettre l'étude de la variabilité du site ;
- dans les version 2 et 3 du Panda, l'énergie de battage étant complètement maîtrisée, la précision des mesures est améliorée

Cependant l'essai Panda, comme le reste des pénétromètres dynamiques, ne fournit qu'un seul paramètre : la résistance dynamique du sol. Par ailleurs cet essai reste « aveugle » puisqu'il ne permet pas la prise d'échantillon et donc la description naturaliste des sols traversés. C'est pourquoi, afin d'améliorer et d'enrichir l'information obtenue lors d'un sondage dynamique, est née l'idée du Panda 3 ([Benz-Navarrete \(2009\)](#)) qui est basée sur la mesure et le découplage des ondes créées suite au battage de la masse frappante. Cette nouvelle version permet l'obtention à une profondeur donnée et pour chaque impact, de plusieurs paramètres mécaniques du sol. Par ailleurs, le développement d'autres essais (géoendoscopie, perméamètre léger) pouvant être couplés à l'essai Panda, permet de compléter l'information des caractéristiques mécaniques du sol mesurées avec le Panda par une information sur les caractéristiques physiques des matériaux sondés.

4. Modèle géotechnique du terrain

Dans l'état actuel des connaissances, l'élaboration d'un modèle géotechnique constitue fréquemment le point crucial des études géotechniques. En effet, l'ingénieur géotechnicien a besoin d'avoir une image claire de l'organisation spatiale des matériaux constituant le sol, de leurs interactions et de leurs caractéristiques. Le modèle géotechnique constitue ainsi un recueil de données sur lesquelles on s'appuie pour justifier le dimensionnement d'un ouvrage (Reiffsteck et al. (2012)).

La qualité du modèle dépendra de différents facteurs (stratégie et organisation de la reconnaissance, moyenne des essais, interprétation...). Au cours de la réalisation des études, le modèle géotechnique passe d'une description de l'environnement géologique général qui peut être très sommaire, imprécise voire peu fiable à la description du cadre géomécanique local (Fig.4.1). Le but de la reconnaissance est de rendre ce modèle fiable, précis, et le plus complet possible afin de permettre une conception de l'ouvrage qui soit la plus sûre possible. Autrement dit, la reconnaissance du sol se traduit par l'élaboration d'un modèle géotechnique qui trouve son application dans le modèle de calcul.

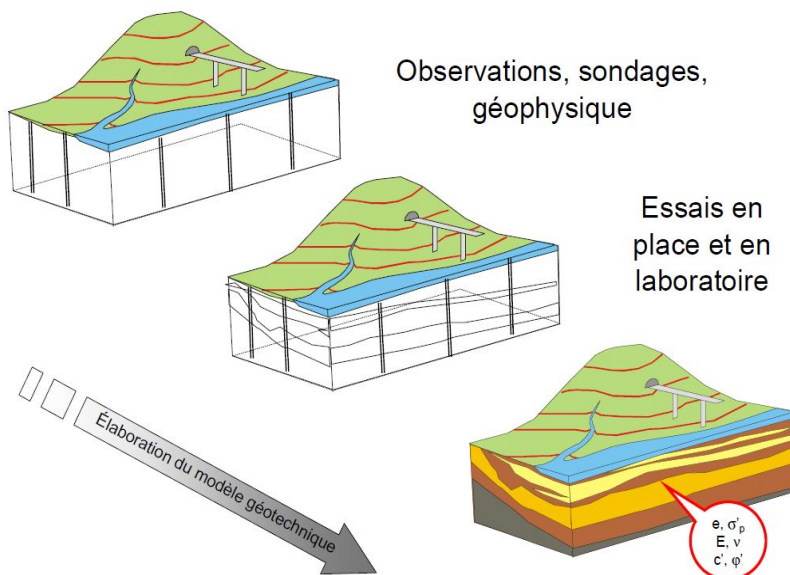


FIGURE 4.1 – Schéma sur la progression de la reconnaissance géotechnique vers l'élaboration du modèle géotechnique (Magnan (2010))

4.1 Le modèle géotechnique

4.1.1 Notion de modèle géotechnique

D'une manière simple, on peut définir le modèle géotechnique comme une représentation synthétique et simplifiée du sous-sol qui peut être plus ou moins complexe. Il regroupe les données relatives à la géométrie des terrains, à leur nature et à leurs propriétés mécaniques. Ces informations et données ont été collectées lors de la campagne de reconnaissance géotechnique. Le modèle géotechnique est alors une synthèse des données sur lesquelles on s'appuie pour alimenter un « modèle de calcul » afin de justifier le dimensionnement d'un ouvrage.

4.1.2 Étapes pour la conception du modèle géotechnique

L'élaboration du modèle géotechnique nécessite l'utilisation d'outils informatiques de collecte, d'analyse et d'interprétation des données (SIG, base de données...), puis de synthèse et de représentation. Le modèle géotechnique constitue ainsi une synthèse du modèle géologique, hydrologique et des données géotechniques sous la forme d'une maquette numérique. Sa conception et son développement comporte une démarche à plusieurs étapes, lesquelles sont synthétisées dans le schéma de la figure (Fig.4.2). Les principales étapes dépendant de la reconnaissance de sols sont les suivantes :

- i. identification des couches homogènes¹ du terrain et définition de leur géométrie ;
- ii. identification et classification de la nature des matériaux rencontrés ;
- iii. définition des caractéristiques mécaniques du sol au sein de chaque unité.

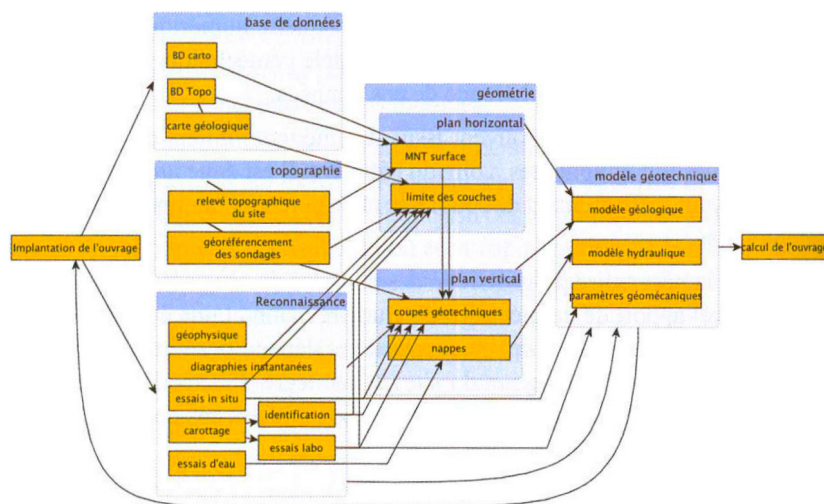


FIGURE 4.2 – Étapes pour l'élaboration du modèle géotechnique (source : Reiffsteck et al. (2012))

1. On souligne que la notion de couche homogène peut varier selon le contexte de l'étude. Notamment la notion de couche en géologie peut être différente de celle en géotechnique. Cette notion sera discutée dans la partie II de ce document.

4.1.3 Cadre normatif

L'annexe B.1 de la partie 2 de l'Eurocode 7 (NF P 94-252 (2007)) résume le processus du calcul géotechnique en plusieurs étapes successives (cf.Fig.4.3). La première porte sur la reconnaissance et les essais géotechniques. La seconde sur la détermination des valeurs caractéristiques. La dernière phase comporte les calculs de vérification du dimensionnement. Le modèle géotechnique apparaît comme une synthèse des résultats de la reconnaissance géotechnique et l'évaluation des paramètres géotechniques pondérés par leurs coefficients pondérateurs correspondants. Cette démarche simplifiée met en évidence l'ampleur et la complexité du problème de l'élaboration d'un modèle géotechnique.

Cependant il existe une vision commune sur le caractère évolutif du modèle géotechnique, qui devrait être complétée au fur et à mesure que de nouveaux apports sur le projet ou le site sont disponibles. C'est pourquoi à l'heure actuelle, la méthode observationnelle semble être une stratégie privilégiée par rapport à la conception du modèle géotechnique pour des ouvrages d'une certaine taille (Reiffsteck et al. (2012)) pour lesquels on dispose de moyens budgétaires importants. La méthode observationnelle consiste à adapter la conception ou la réalisation de l'ouvrage en fonction des modifications constatées par rapport aux hypothèses envisagées en phase de conception. L'instrumentation de l'ouvrage géotechnique permet de l'ausculter et de réaliser le suivi pendant sa construction. On peut ainsi renforcer le dimensionnement si les mesures mettent en évidence cette nécessité. L'Eurocode 7 la conseille (NF P 94-251-1 (2005)) quand il est compliqué de prévoir le comportement géotechnique d'un ouvrage.

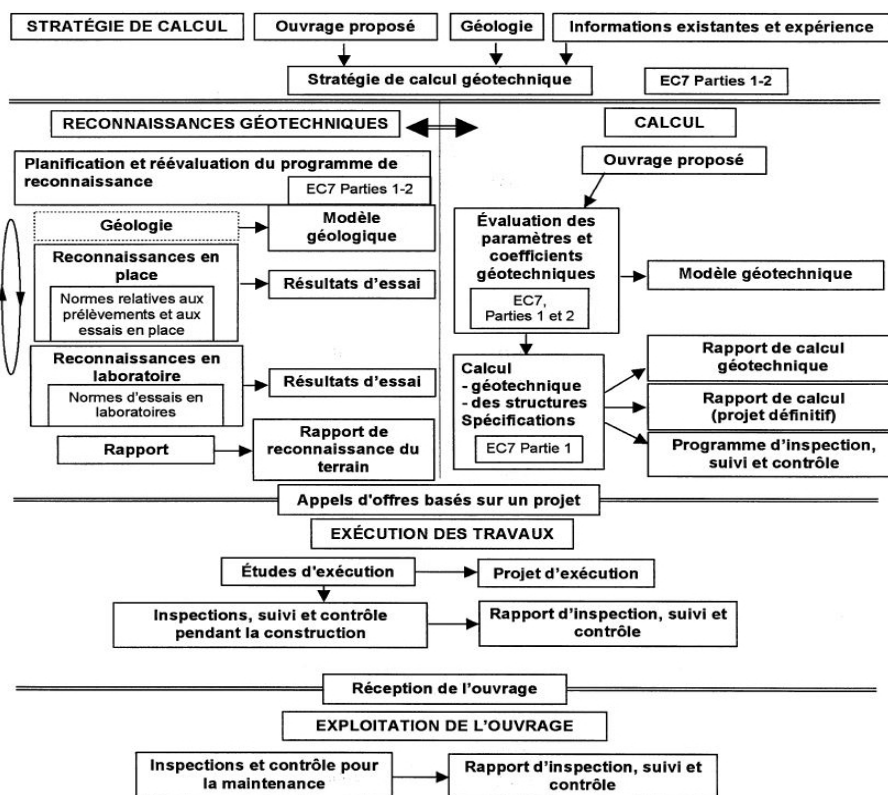


FIGURE 4.3 – Organisation des différentes phases d'une étude géotechnique (source : NF P 94-252 (2007))

4.2 Sources d'incertitudes en géotechnique

La conception du modèle géotechnique est confrontée à plusieurs sources d'incertitudes, que l'on peut résumer d'une manière simpliste en : variabilité naturelle du sol, erreurs de mesure et incertitude du modèle (Vanmarcke (1977a)). Plusieurs auteurs ont traité ce sujet parmi lesquels on peut citer (Favre (2004)) côté francophone, (Phoon and Kulhawy (1999), Christian and Baecher (2003)) côté anglo-saxon et (Auvinet (2002)) côté hispanophone. Il est reconnu que les principales sources d'incertitude de la géotechnique peuvent être regroupées en deux classes principales : incertitude *active* ou *aléatoire* et *passive* ou *épistémologique*. Une description plus approfondie des facteurs à l'origine de ces sources d'incertitude est présentée dans ce qui suit.

4.2.1 Incertitude active ou aléatoire

Cette incertitude est dite active puisqu'elle est inhérente au désordre naturel du milieu physique à étudier, le sol. Elle est souvent la principale cause de la dispersion que présentent l'ensemble des mesures recueillies lors de la campagne de reconnaissance.

Dans la littérature, on admet que la variation spatiale d'une propriété du sol est due principalement au processus géologique naturel qui produit et modifie continuellement les massifs de sol in situ (Phoon and Kulhawy (1999), Fenton (1999a)). Un autre aspect de l'incertitude naturelle des sols et des roches est qu'elle peut se manifester à diverses échelles (cf. Fig. 4.4). Ainsi, la structure de la variabilité des milieux géotechniques est également fonction de l'échelle d'observation, entravant ainsi la définition et la caractérisation des propriétés physiques, mécaniques et hydrauliques. En conséquence, il faut définir un volume élémentaire représentatif où l'échantillon peut être considéré comme homogène. Cependant, en géotechnique ce volume est souvent imposé par l'échelle de l'ouvrage à dimensionner (Auvinet (2002)), qui est différente de celle des échantillons testés en laboratoire.

Dans un contexte géotechnique plus large (Uzielli et al. (2006)) décompose l'incertitude aléatoire en trois niveaux d'hétérogénéité comme suit :

- i. *stratigraphique* comme résultat des processus géologiques et géomorphologiques à l'échelle du site. Dans ce sens, l'ingénieur est souvent confronté à des stratigraphies complexes, avec des unités de sols qui présentent une forte hétérogénéité mécanique ;
- ii. *lithologique* représentée par des alternances de couches minces à faible/forte résistance à l'intérieur d'une couche à faible/forte résistance ou par la présence de lentilles dans un massif de sol uniforme ;
- iii. *intrinsèque ou naturelle* représentée par les variations des propriétés à l'intérieur d'une unité de sol supposée homogène au sens géotechnique.

On cite le travail de (Leroueil and Hight (2003)) dans lequel les auteurs ont élaboré une liste exhaustive de l'ensemble des facteurs responsables de la variabilité naturelle des sols. Il s'agit principalement de leur composition, des paramètres micro-mécaniques et de leur histoire mécanique. Vis-à-vis de l'histoire mécanique des géomatériaux, on peut distinguer celle, plus ancienne, liée au processus de formation et celle, plus récente, qui prend en compte les effets de l'érosion, les fluctuations de la nappe phréatique, ainsi que les modifications du milieu naturel par l'action de l'homme (Jardine (1995)).

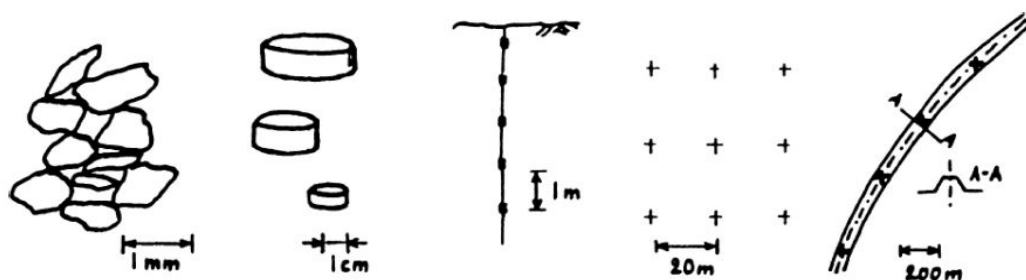


FIGURE 4.4 – Échelles de variabilité spatiale des paramètres géotechniques (source : Vanmarcke (1978) et Uzielli et al. (2006))

4.2.2 Incertitude passive ou épistémique

Cette incertitude est issue de la connaissance imparfaite du sol. À la différence de l'incertitude aléatoire qui ne peut pas être supprimée, cette incertitude est artificielle et peut être réduite en augmentant le nombre de mesures ou leur qualité par exemple. En ce qui concerne l'incertitude passive vis-à-vis de la caractérisation des sols on distingue les erreurs d'observation et les erreurs d'enquête (Favre (2004)). Par ailleurs, trois types d'erreurs d'observation peuvent être considérées :

- i. *les erreurs de mesure proprement dites* qui sont inévitables lors de la réalisation des essais in-situ ou en laboratoire et donc liées à l'appareillage et à l'opérateur. L'étude publiée par (Phoon and Kulhawy (1999)) donne une information quantitative, sous la forme de coefficient de variation et d'étendue, sur les erreurs de mesures des principaux essais de laboratoire ainsi que celles qui sont liées aux essais géotechniques in situ ;
- ii. *les erreurs de représentativité* à cause de la transformation de la mesure physique en un paramètre mécanique ;
- iii. *les erreurs de l'instant*, bien que les campagnes de reconnaissance géotechniques soient réalisées durant un intervalle de temps donné, les propriétés du sol peuvent présenter une variabilité temporelle plus ou moins importante. D'après (Rétháti (1988)) cité par (Phoon and Kulhawy (1999)), cette variabilité temporelle peut être négligée lorsque la campagne géotechnique est réalisée durant une période inférieure à 1-2 semaines.

En outre, parmi les erreurs d'enquête on peut distinguer :

- i. *les erreurs d'enquête proprement dites* qui sont les erreurs liées à une mauvaise conduite des reconnaissances. (Favre (2000)) donne deux exemples simples d'une reconnaissance géotechnique incapable de fournir une information représentative du site (Fig.4.5). Sur la figure de gauche sont représentés deux appuis sur un remblai récent supposé homogène alors qu'il existait un remblai ancien en biseau sous la fondation de gauche qui n'a pas été reconnu. Sur la figure de droite, la reconnaissance qui a été réalisée à une profondeur inférieure à 3-5 fois la largeur de la fondation superficielle, n'a pas permis d'identifier la présence d'une couche de faible portance qui pourrait piloter les tassements. (Magnan (2000)) cite un cas réel qui illustre comment une mauvaise implantation des essais géotechniques peut conduire à la conception d'un modèle géotechnique erroné. En effet, suite à la faible variabilité des résultats d'essais oedométriques sur des échantillons prélevés, un modèle géotechnique simple constitué par des couches horizontales et basé sur les valeurs moyennes des essais en labora-

toire avait été retenu. Cependant, le modèle n'était pas représentatif du site ausculté, puisque les sondages carottés avaient été implantés le long de l'axe d'une vallée fossile que des profils pénétrométriques ont pu révéler ultérieurement (Fig.4.6).

- ii. *les erreurs d'échantillonnage* : La caractérisation des propriétés d'un massif du sol est souvent basée sur un nombre réduit d'échantillons. Il peut donc exister un biais et une incertitude au sens statistique lors des analyses géotechniques (Uzielli (2008)).

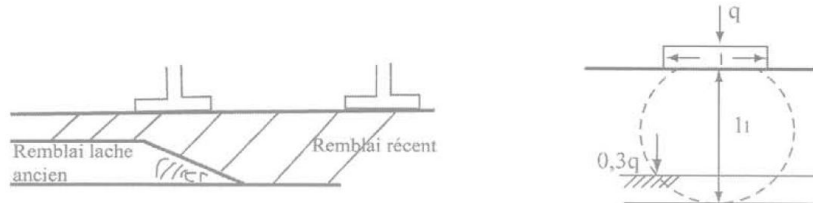


FIGURE 4.5 – Exemples de problèmes d'enquête lors d'une reconnaissance géotechnique (source : Favre (2000))

Finalement, il est reconnu que les modèles géomécaniques, qu'il soient analytiques ou numériques, sont basés sur des hypothèses simplificatrices d'une réalité complexe. La pertinence et la précision de ces modèles est donc une source d'incertitude omniprésente dans le domaine. Elle comporte aussi l'erreur des modèles de transformation mathématique utilisés pour délivrer les paramètres mécaniques de calcul à partir des mesures réalisées lors des essais géotechniques.

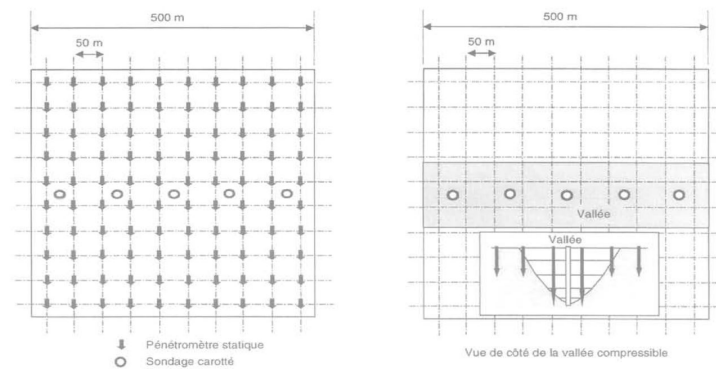


FIGURE 4.6 – Plan schématique du site qui comportait une vallée fossile (source : Magnan (2000))

4.3 Prise en compte de l'incertitude

4.3.1 Mise en évidence du problème

D'une manière très simpliste, vis-à-vis du dimensionnement des ouvrages géotechniques, le principe général est de calculer les effets S des actions et la résistance maximale R que peuvent opposer le sol ou des éléments de structure. On distingue ainsi 3 groupes de facteurs d'incertitude (Favre (2004)) :

- les incertitudes sur les valeurs de résistance du sol et des éléments de structure ;

- les incertitudes sur les valeurs des actions ;
- les incertitudes sur les sollicitations calculées.

Les paramètres géotechniques déterminés d'une manière directe ou indirecte grâce aux différents moyens de reconnaissance, sont ceux qui interviennent principalement du côté des résistances vis-à-vis de la sécurité des ouvrages géotechniques (même si en géotechnique, les actions et les résistances sont liées et interagissent entre elles). Cependant, comme on vient de le voir, ils présentent souvent une forte variabilité naturelle spatiale que l'on doit intégrer et modéliser au sein des différentes unités de terrain constituant le modèle géotechnique. L'ingénieur doit alors aborder l'étape d'analyse numérique pour décider des valeurs de calcul des propriétés d'un sol ou d'une roche. Pour ce faire, il doit souvent s'appuyer sur un faible nombre de résultats d'essais et toute autre information qu'il peut trouver sur le site, comme des données géologiques ou des expériences locales.

Compte tenu de ces constats, l'interprétation des mesures fournies par une campagne de reconnaissance des sols comporte un fort caractère subjectif. Ce fait a été mis en évidence dans la littérature. À titre d'exemple, on peut fournir deux exemples dans les figures (Fig.4.7 et Fig.4.8) montrant plusieurs interprétations « expertes » des profils de résistance mesurés à l'aide des sondages SPT. On constate une grande diversité des choix des valeurs représentatives à utiliser dans les calculs. Cette manière traditionnelle de définir les valeurs à considérer pour l'élaboration du modèle géotechnique, n'est régie par aucune règle particulière mise à part l'expertise de l'ingénieur. Dans cette optique, le but des Eurocodes était d'ouvrir une voie vers des approches plus rationnelles de ces valeurs, en proposant des méthodes de traitement statistique des données de sol.

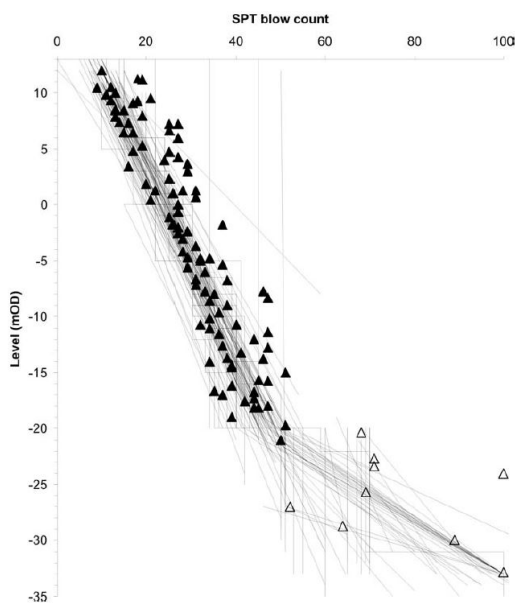


FIGURE 4.7 – Interprétation des ingénieurs de la valeur représentative à incorporer dans les calculs pour les mesures de SPT dans les argiles plastiques de Londres et Lambeth (source : [Bond and Harris \(2008\)](#))

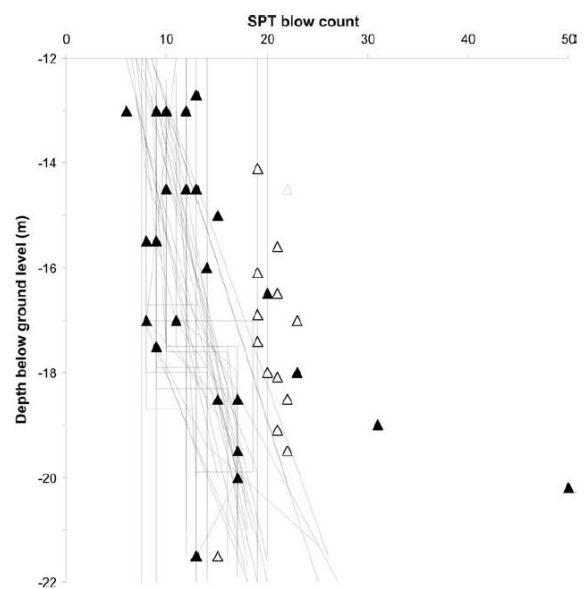


FIGURE 4.8 – Interprétation des ingénieurs de la valeur représentative à incorporer dans les calculs pour les mesures de SPT dans les sols graveleux de Thames (source : [Bond and Harris \(2008\)](#))

4.3.2 La valeur caractéristique X_k des propriétés géotechniques du sol

Dans le cadre des Eurocodes, la justification des ouvrages est fondée sur une démarche de définition des valeurs de calcul prudentes vis-à-vis de la vérification des états limites de service (ELS) et des états limites ultimes (ELU). L'aléa, fondamentalement causé par la variabilité naturelle du sous-sol est pris en compte grâce à la définition des valeurs caractéristiques (X_k). Ensuite, les valeurs de calcul sont déduites grâce à l'application des coefficients partiels (γ) dont leur rôle est de majorer les charges et de minorer les résistances. La norme en vigueur propose ainsi des méthodes de calcul semi-probabilistes pour tenir compte de différents aléas géotechniques.

En principe, la valeur caractéristique des paramètres géotechniques est choisie afin de tenir compte de :

- la variabilité inhérente au sous-sol ;
- l'incertitude vis-à-vis de la détermination de ces paramètres ;
- la dimension de la zone du sol concernée par le mécanisme de défaillance.

Néanmoins, la définition de la valeur caractéristique fournit par l'Eurocode-7 (NF P 94-251-1 (2005)) reste assez générale :

« La valeur caractéristique d'un paramètre géotechnique doit être une estimation prudente de la valeur influençant l'occurrence de l'état limite. »

Cette définition montre que l'identification d'une valeur représentative (ou caractéristique) de la variabilité spatiale des terrains n'est pas évidente. Toutefois, la directive donnée par l'Eurocode-7 - Partie 1 (NF P 94-251-1 (2005)) cible mieux cette définition. Elle explique bien que le paramètre qui contrôle le comportement de l'ouvrage est souvent une valeur moyenne sur une certaine surface ou un certain volume de sol. Il y a donc un effet "moyenne" et la défaillance du système aura plutôt lieu quand le paramètre s'écarte de sa moyenne en tout point de la zone concernée par la rupture. Pourtant l'Eurocode -7 statue : « La valeur caractéristique est une estimation prudente de cette valeur moyenne ». De même, la réduction de la variabilité (*effet de réduction de la variance*) dépend naturellement de la dimension de la masse de sol soumise au chargement et de la façon dont le sol est sollicité (une faiblesse locale peut dans certains cas entraîner la rupture, tandis que, dans d'autres cas, il faut qu'un volume de sol important soit en état de plasticité pour que la rupture se développe).

Les Eurocodes proposent l'utilisation de la statistique descriptive afin de prendre en compte la variabilité naturelle des paramètres du sol pour définir ces valeurs caractéristiques. Plusieurs approches statistiques ont été ainsi proposées dans la littérature basées sur la modélisation des propriétés du sol comme une variable aléatoire afin de calculer X_k . Les principales approches proposées, qu'on décrira par la suite, sont regroupées dans la figure (Fig.4.9). La valeur moyenne est notée par X_m . Le lecteur intéressé pourra se reporter à (Orr (2016)) pour une présentation plus en détail de la définition des valeurs caractéristiques dans le cadre des Eurocodes.

Fractile 5%

Les valeurs caractéristiques correspondent à une probabilité de mise en défaut de 5% et X_k peut être calculée comme le fractile à 5% de la distribution du paramètre géotechnique

étudié. Pour une loi normale ou gaussienne, on obtient l'équation (éq.A) avec CV le coefficient de variation.

Cependant, vis-à-vis du dimensionnement des ouvrages géotechniques, cette équation fournit souvent des valeurs de calcul trop conservatrices (Orr (2016)). En effet, l'Eurocode-7 cible bien le cadre d'application de cette équation. Elle ne doit être appliquée que lorsqu'on prévoit une rupture de terrain au niveau local. Cependant, pour les ouvrages géotechniques, les états limites concernent souvent des grands volumes ou des grandes surfaces de sol.

Estimation de la moyenne avec un niveau de confiance de 95%

Cette approche est équivalente à considérer le *paramètre étendue* défini dans (Baguelin and Kovarik (2000)) comme une moyenne spatiale du paramètre local ou ponctuel (grandeur résultant des mesures) sur une surface ou un volume de sol correspondant au mécanisme de rupture à vérifier. On calcule ainsi la valeur caractéristique comme une estimation de la valeur moyenne avec un niveau de confiance de 95% (ou une probabilité maximale de mise en défaut de 5%). Pour ce faire, on peut utiliser l'équation (éq.B), proposée par la *Société Japonaise de Géotechnique*. Avec $t_{\alpha,\nu}$ la loi de Student pour un niveau de confiance $\alpha\%$ et ν le nombre de degrés de liberté égal à $n - 1$ sous l'hypothèse de distribution normale.

Zone concernée par la rupture et variabilité spatiale

Une démarche simplifiée a été proposée par (Schneider and Schneider (2012)) (éq.C.1), (éq.C.2), sur la base du modèle additif du coefficient de variation CV_{total} proposé par (Phoon and Kulhawy (1999)), qui prend en compte les effets des incertitudes aléatoires et épistémiques. L'équation simplifiée ne considère que l'incertitude inhérente causée par la variabilité naturelle du matériau. L'auteur propose deux lois de densité de probabilité théoriques (normale ou log-normale) et le modèle exponentiel de réduction de variance, sous l'hypothèse de corrélation parfaite selon la direction horizontale. La fonction de réduction de variance permet de prendre en compte le volume de sol concerné par l'état limite et la corrélation spatiale pour calculer la variance « réduite » de la moyenne dans l'espace du paramètre géotechnique.

Une approche similaire avait été proposée en France par (Baguelin and Kovarik (2000)).

	fractile 5%	Hypothèse loi normale $X_k = X_m(1 - 1.645CV)$ (éq.A)
	nombre de mesures	Hypothèse loi normale $X_k = X_m(1 - \frac{t_{\alpha,\nu}CV}{\sqrt{n}})$ (éq.B)
zone concernée par la rupture Référence : (Schneider and Schneider (2012))		
$COV_{total} = \sqrt{\Gamma_s^2 CV_{inh}^2 + CV_{meas}^2 + CV_{trans}^2 + CV_{stat}^2}$ Référence : (Phoon and Kulhawy (1999))	Hypothèses $CV_{meas}^2 \cong 0$	Hypothèse loi normale (éq.C.1) $X_k = X_m(1 - 1.645CV_{inh}\sqrt{\Gamma_s^2})$ si $CV < 0.3$
	$CV_{stat}^2 \cong 0$	Hypothèse loi log-normale (éq.C.2) $X_k = X_m 0.2 \sqrt{\ln(1+\Gamma_s^2 CV_{inh}^2) / \sqrt{1+\Gamma_s^2 CV_{inh}^2}}$ si $CV > 0.3$

FIGURE 4.9 – Approches statistiques pour le calcul des valeurs caractéristiques. Notation : CV_{inh} : CV de la variabilité inhérente, CV_{meas} : CV des erreurs de mesure, CV_{stat} : CV des erreurs statistiques, CV_{trans} : CV des erreurs de transformation du modèle.

4.3.3 Évolution de la réglementation internationale : modélisation d'une reconnaissance pour la prise en compte de la variabilité naturelle

La définition des valeurs caractéristiques envisagées par l'Eurocode consiste à déterminer statistiquement la valeur du paramètre géotechnique au sein des couches de sols définies dans le modèle géotechnique. Le paramètre du sol est alors modélisé comme une variable aléatoire, telle qu'illustrée dans la figure (Fig.4.10 (a)).

La dispersion naturelle des propriétés des sols est causée par des phénomènes physiques complexes. Cependant, cette variabilité n'est pas totalement imprévisible. Bien que les sols présentent des caractéristiques physiques et/ou mécaniques différentes d'un point à un autre, ils exhibent une certaine structure spatiale. En effet, les propriétés mesurées sur deux échantillons ont tendance à présenter des caractéristiques similaires lorsqu'ils ont été récoltés à proximité l'un de l'autre. Or, dès que l'écartement entre les points échantillonnés augmente, le degré de similarité diminue. Cette variabilité peut alors être modélisée de manière explicite grâce au formalisme des champs aléatoires (Vanmarcke (1977a)).

D'une manière très simpliste, un champ aléatoire utilise aussi une loi de probabilité théorique pour modéliser la propriété du sol à étudier, mais il intègre son organisation dans l'espace via une fonction spatiale, dont la définition mathématique la plus usitée est celle de la fonction d'autocorrélation. À titre d'illustration, on montre dans la figure (cf. Fig.4.10 (b)) un dessin schématique sur la modélisation du sol comme un champ aléatoire. Cette approche permet de mener des études probabilistes sur les effets de la variabilité naturelle vis-à-vis du dimensionnement des ouvrages géotechniques.

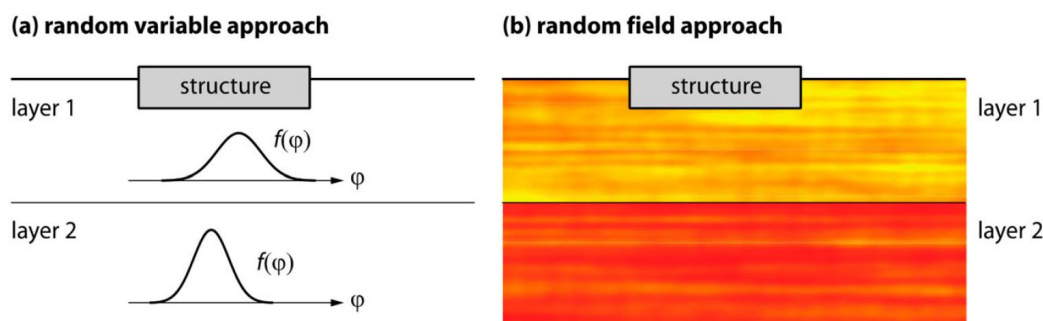


FIGURE 4.10 – Schéma sur les deux approches possibles pour modéliser la variabilité spatiale des paramètres du sol : (a) variable aléatoire et (b) champ aléatoire (source : Papaioannou and Straub (2016))

La modélisation probabiliste du sol grâce à la théorie des champs a fait l'objet de nombreuses études qui ne datent pas d'aujourd'hui. À titre d'exemple, en France, on cite les travaux du Laboratoire des Ponts et Chaussées, commencés à la fin des années 1970, regroupés dans (Magnan (1996)). L'ensemble des travaux portant sur la variabilité des sols présentés au cours des années ont permis une prise de conscience de l'importance de la variabilité des sols². Cette inquiétude est aussi partagée par les comités techniques vis-à-vis de l'élaboration de la nouvelle réglementation internationale. La preuve la plus récente à notre

2. Une synthèse bibliographique sera présentée dans la partie IV de ce rapport.

connaissance est issue du document (Li et al. (2017)), inclus dans le dernier rapport publié par le groupe de travail au niveau européen TC205/TC304 sur la discussion des méthodes statistiques et fiabilistes dans le cadre des Eurocodes. Deux approches basées sur le formalisme des champs aléatoires ont été envisagées pour modéliser et incorporer dans la phase de calcul la variabilité spatiale des sols. La démarche générale a été synthétisée dans la figure (Fig.4.11) dans laquelle on distingue trois étapes principales :

- (1) estimation des paramètres statistiques à partir de mesures fournies par la campagne de reconnaissance géotechnique ;
- (2) modélisation probabiliste dans l'espace des propriétés de sol ;
- (3) phase de calcul géotechnique.

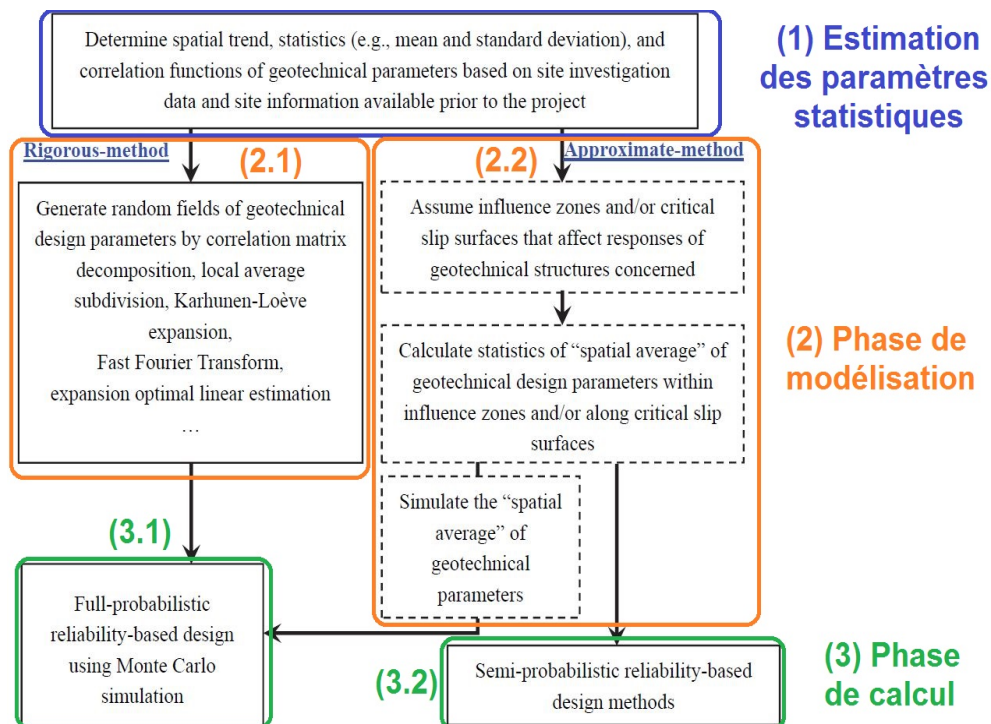


FIGURE 4.11 – Incorporation de la variabilité spatiale dans des méthodes de calcul fiabiliste sur la base de la théorie des champs aléatoires (source : Li et al. (2017))

Dans ce travail, on est concerné par les deux premières étapes (estimation des paramètres statistiques et phase de modélisation). On remarque que vis-à-vis de la phase de modélisation, deux approches ont été distinguées :

- (2.1) *Rigorous-method* : la méthode dite "rigoureuse" qui consiste à simuler directement le champ aléatoire à partir des paramètres statistiques déduits lors de la première phase (1). Un des grands intérêts des simulations est le couplage avec la méthode des éléments finis par le biais de la technique de Monte Carlo puisqu'il est possible de générer à chaque fois un champ corrélé représentant les caractéristiques des matériaux qu'on attribue au modèle par éléments finis³.

3. Il s'agit d'une des approches de la méthode des éléments finis stochastiques (MEFS). L'autre alternative possible est la méthode des perturbations. Dans l'optique de la géotechnique, on propose au lecteur de se reporter à (Auvinet et al. (2000)) pour plus de renseignements.

Le principal avantage de cette approche est qu'il n'est pas nécessaire de définir la géométrie de la zone d'influence concernée par l'état critique lors de la modélisation de la variabilité spatiale. Pour chaque simulation, c'est le modèle de calcul qui va déterminer la géométrie de cette zone ;

(2.2) *Approximate-method* : la méthode approximative nécessite de définir la géométrie de volume ou surface du sol concernée par l'état critique à étudier. Ensuite, il faut intégrer dans l'espace, les paramètres géotechniques définis comme des champs aléatoires et appliquer la réduction de variance suivant le volume où ils ont été intégrés. Les paramètres géotechniques sont traités comme des variables aléatoires à variance réduite, à ce moment deux options sont possibles :

(3.1) on utilise la technique de Monte Carlo pour calculer la probabilité de défaillance en fonction à l'état critique à étudier, après plusieurs tirages ou simulations de la variable aléatoire ;

(3.2) on estime la valeur caractéristique avec un niveau de confiance donné dans le cadre d'une méthode de calcul semi-probabiliste. Ceci représente une approche simplifiée similaire à celle proposée par (Schneider and Schneider (2012)).

4.4 Bilan

La première étape du processus de conception géotechnique consiste à réaliser la reconnaissance du site pour produire un modèle géotechnique qui sert de base pour la conception de l'ouvrage. Bien que le contenu de ce modèle soit clair en théorie : la définition de l'ensemble des propriétés (géologiques, mécaniques et hydrauliques) en tout point du massif considéré, sa définition et conception reste encore très subjective. Le fait qu'il soit un « modèle » implique la réalisation de choix de la part de l'ingénieur en fonction de l'information disponible, de sa qualité et de ce dont il a besoin.

Ce modèle géotechnique est construit à partir des données collectées durant la campagne de reconnaissance géotechnique. L'ingénieur doit alors faire un travail d'analyse et de synthèse de ces informations. Toutefois, on a vu que l'ingénieur doit travailler dans un monde incertain, compte tenu des nombreuses sources d'incertitude de la géotechnique. Il existe des incertitudes dues à la variabilité naturelle des terrains à plusieurs échelles, mais aussi celles qui proviennent de toutes les opérations nécessaires à l'acquisition de l'information issue de la campagne de reconnaissance géotechnique. La phrase qui résume le mieux cette problématique est celle proposée par Karl Terzaghi lors de la première conférence internationale de mécanique des sols en 1936 : *Unfortunately, soils are made by nature and not by man, and the products of nature are always complex.*

Pour prendre en compte la variabilité spatiale naturelle des matériaux au sein de différentes couches de terrain dans l'élaboration du modèle géotechnique et tenter de palier la subjectivité inhérente à l'approche basée uniquement sur l'expertise de l'ingénieur, l'Eurocode a introduit les notions de valeurs caractéristiques. Ces valeurs sont basées sur la description des propriétés du sol comme une variable aléatoire grâce aux outils de la statistique descriptive. Toutefois, ces approches ne permettent pas de prendre en compte dans le calcul la structure spatiale de ces propriétés. Pour ce faire, des méthodes basées sur la théorie de champs aléatoires et/ou de la géostatistique ont été étudiées au cours de ces dernières années dans la littérature.

5. Conclusion, problématique et démarche proposée

5.1 Conclusion

La définition d'un modèle géotechnique de terrain nécessite une campagne de reconnaissance de sol. A partir de la présentation des enjeux et des objectifs d'une reconnaissance géotechnique et des différents moyens de reconnaissance disponibles, l'intérêt des essais in situ par rapport aux essais de laboratoire, notamment du fait qu'ils permettent de sonder le milieu dans son état naturel et de caractériser la variabilité du terrain, a été mis en évidence.

Parmi les essais in situ, l'essai de pénétration est l'un des essais les plus usités pour la reconnaissance géotechnique des sols. Néanmoins, les essais de pénétration dynamiques "classiques" ont certaines limitations notamment en ce qui concerne leur résolution et la précision de leur mesure. Dans ce contexte, le développement de l'essai pénétrométrique dynamique léger Panda®, de par son mode de mesure de l'énergie de battage et sa finesse d'acquisition et haute résolution de mesure ainsi que du fait de sa rapidité de mise en œuvre, peut s'avérer un outil intéressant pour améliorer la reconnaissance des sols de surface et renforcer ainsi la fiabilité des modèles de terrain.

Suite à la campagne de reconnaissance des sols, une analyse et une synthèse des mesures doit être effectuée afin de créer une maquette numérique du terrain, « le modèle géotechnique ». Ce modèle géotechnique du terrain doit permettre de définir en chaque point du terrain les caractéristiques physico-mécaniques probabilisées du sol. Les étapes à suivre pour son élaboration sont connues, à savoir :

- i. identifier et préciser l'organisation des formations ou unités homogènes du terrain d'après un critère donné ;
- ii. fournir une classification de la nature des matériaux rencontrés dans ces formations ;
- iii. modéliser les propriétés mécaniques des différentes unités du terrain.

De fait, l'élaboration du modèle géotechnique de terrain est un compromis à la fois entre l'information disponible et l'objectif précis auquel il faut répondre. En ce sens, le point (iii) de son élaboration représente à ce jour l'une des étapes les plus délicates puisque l'on analyse un paramètre du sol mesuré en général seulement en quelques points de l'espace. Jusqu'à il y a peu, le modèle géotechnique de terrain reposait essentiellement sur une vérité experte puisque c'est l'ingénieur qui choisissait sur la base des résultats de la campagne de reconnaissance, comment représenter le terrain dans ce modèle. Or, aujourd'hui ce modèle pour être plus fiable doit pouvoir intégrer la variabilité du terrain et les incertitudes rencontrées dans l'ensemble de la chaîne d'acquisition et d'analyse des données. Les contraintes budgétaires et physiques ainsi que la variabilité naturelle des matériaux rendent difficile l'analyse et la

modélisation de données géotechniques. La norme en vigueur (EC-7) propose la définition de valeurs caractéristiques grâce aux outils statistiques classiques, permettant une caractérisation globale des caractéristiques physico-mécaniques du sous-sol. Les formalismes des fonctions aléatoires ou des champs aléatoires permettent une caractérisation probabiliste de l'incertitude naturelle dans l'espace. À ce jour, ces techniques mathématiques commencent à être reconnues par les comités techniques afin d'intégrer au mieux la variabilité spatiale du sol vis-à-vis des calculs géotechniques.

Cependant, dans la pratique courante de la géotechnique, la caractérisation de la variabilité spatiale du terrain pour la mise en œuvre de ces techniques s'avère compliquée faute de données en quantité et qualité suffisante. En reconnaissance de sols, dans la plus part des cas, la quantité de données disponibles est faible au regard de la complexité à représenter. Il est reconnu qu'une bonne définition de la variabilité des sols nécessite des profils de mesures des propriétés de sol quasi en continu afin de pouvoir apprécier correctement leur caractère aléatoire ainsi que la corrélation spatiale des propriétés d'un point à un autre. Le faible pas de mesure des pénétromètres statiques (2 cm) par rapport aux autres essais, 1 mètre en général pour le SPT et le presiomètre représente donc un grand intérêt. Pour cette raison l'essai au cône statique a été largement utilisé à cette fin. Toutefois, certains travaux (ex. [Deplagne and Bacconnet \(1993\)](#), [Gaouar \(1997\)](#), [Chaigneau \(2001\)](#), [Villavicencio \(2009\)](#), [Athapaththu et al. \(2015\)](#)) ont déjà mis en évidence la force d'un pénétromètre dynamique à énergie variable comme le Panda, vis-à-vis de l'analyse de la variabilité des données de résistance du sol.

C'est pourquoi, développer des méthodes d'analyse pour mieux interpréter les profils des données mesurées afin d'en extraire le maximum d'information disponible et améliorer les études sur la caractérisation des variations des propriétés du sol et sur leur modélisation probabiliste dans l'espace est précieux pour l'ingénieur. Face à ce constat, il apparaît que l'exploitation de l'essai pénétrométrique Panda et son évolution peuvent offrir des potentialités intéressantes en vue de mieux caractériser les terrains et leur variabilité et permettre ainsi d'alimenter le modèle géotechnique afin de le rendre plus fiable.

5.2 Problématique traitée et démarche proposée

Ce travail s'inscrit dans le contexte de la reconnaissance des sols à faible profondeur avec l'objectif d'améliorer la construction d'un modèle du terrain. On s'appuie fondamentalement sur des données de terrain majoritairement recueillies à l'aide de l'essai de pénétration dynamique à énergie variable Panda. La question principale posée est la suivante : *comment mieux exploiter une campagne de reconnaissance géotechnique basée sur l'utilisation du Panda pour construire et améliorer un modèle de terrain géotechnique ?* La démarche générale envisagée pour répondre à cette problématique est schématisée (cf. [Fig.5.1](#)).

Les questions secondaires inhérentes à la démarche et se rattachant à cette problématique sont :

- i. Peut-on détecter les différentes couches caractérisant le terrain à partir de l'exploitation des sondages pénétrométriques Panda ?
- ii. Est-il possible d'obtenir des indices sur l'identification des sols à partir des résultats de l'essai Panda ?
- iii. Comment intégrer ces nouveaux outils d'exploitation du signal Panda et améliorer

la connaissance de la variabilité du sol à partir de ces données en vue d'alimenter et construire un modèle de terrain ou d'optimiser une campagne de reconnaissance complémentaire ?

iv. Comment passer du modèle de terrain au modèle de terrain géotechnique ?

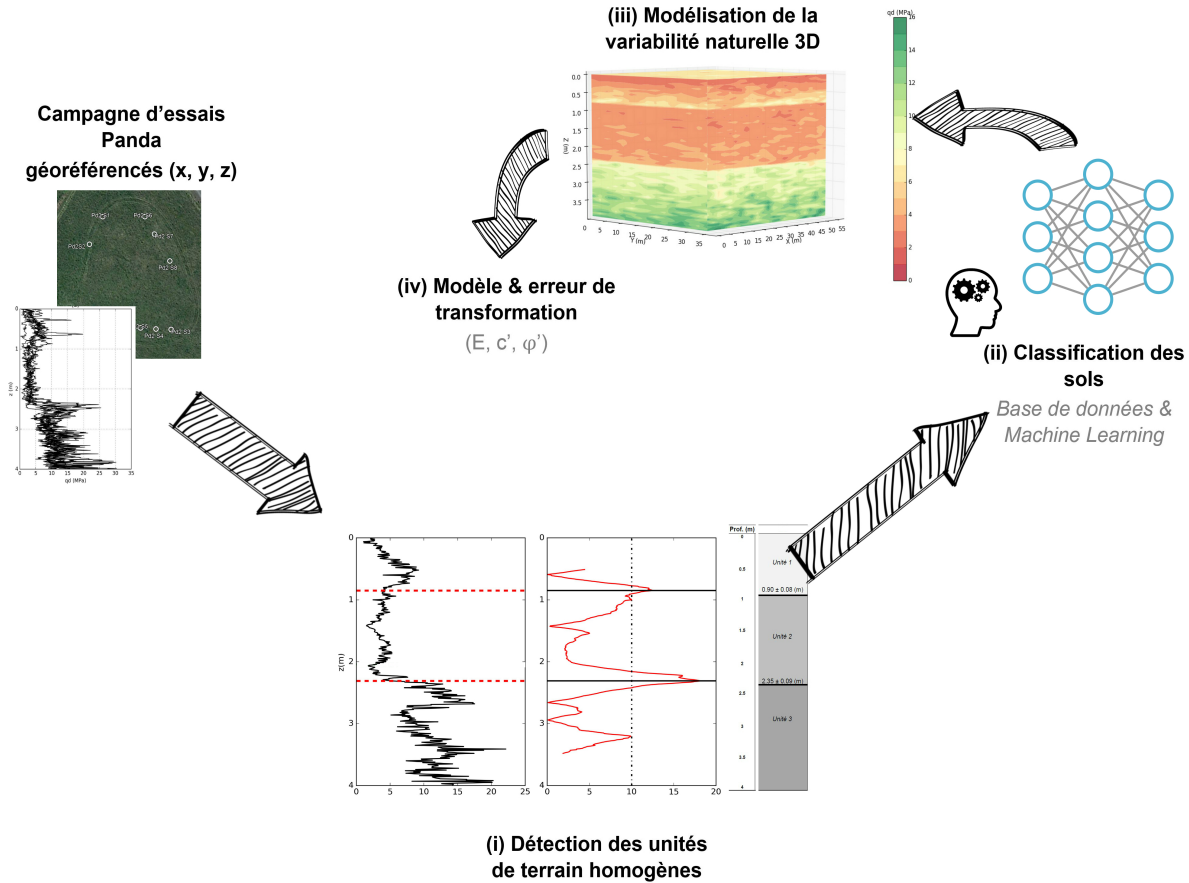


FIGURE 5.1 – Schéma de la démarche proposée

Dans ce contexte, deux domaines d'application étaient possibles :

- traiter les données de reconnaissance d'ouvrages/terrains importants (ville, parc d'ouvrages, grand linéaire, grand projet...);
- traiter les données de reconnaissance issues de campagnes géotechniques pour des projets plus circonscrits et modestes.

Suivant le cas choisi, les questions scientifiques à relever sont différentes. Dans le premier cas, on dispose en général d'un grand nombre d'essais mais relevant de techniques multiples (essais géophysiques, sondages pressiométriques, pénétrométriques, forages,...) et ayant pu être réalisés à différentes périodes ou sur une durée relativement longue. Dans ce cas, le principal verrou concerne la formalisation et la calibration des données puis leur couplage et fusion en vue d'alimenter le modèle du terrain. Dans le second cas, l'ingénieur dispose en général d'un nombre d'essais restreint et la question scientifique majeure réside dans l'optimisation de l'exploitation de ces données (du point de vue de la représentativité statistique et de l'enrichissement par des informations complémentaires) en vue de réaliser un modèle de terrain avec une fiabilité accrue. Ce travail s'est focalisé sur le second domaine d'applica-

tion tout en sachant que les méthodes et les outils développés dans ce cadre pourraient être transposés au premier domaine d'application.

Ce manuscrit présente ainsi la mise en œuvre de méthodes d'analyse appliquées aux données Panda afin de bâtir un modèle du terrain, dans le cadre d'une méthode globale. L'approche envisagée est parallèle aux procédures du géotechnicien, qui à partir des données disponibles vise à représenter le plus fidèlement possible toute en incorporant certaines simplifications le modèle de terrain qu'il étudie. On parle de modèle du terrain plutôt que de modèle géotechnique puisque le passage aux paramètres géotechniques à partir des mesures recueillies par l'essai Panda (question secondaire iv) pour aborder le dimensionnement d'un ouvrage géotechnique aux états limites n'a pas été abordé dans ce travail.

Enfin, pour répondre aux questionnements (i, ii, et iii), la démarche envisagée à partir des données issues des sondages Panda est la suivante :

- 1) développer des outils d'analyse automatique permettant de préciser l'organisation des formations homogènes du sol à partir de l'analyse du profil de résistance dynamique ;
- 2) pour chaque formation homogène détectée à l'étape précédente, développer des outils basés sur les techniques d'intelligence artificielle permettant d'identifier la nature des matériaux traversés ;
- 3) proposer une répartition dans l'espace de la résistance de pointe mesurée au pénétromètre dynamique Panda. Le but est de mettre en œuvre un modèle probabiliste 3D, qui à partir du positionnement des sondages, permettra de voir comment se propage l'incertitude à l'intérieur des différentes unités de terrain et de rendre compte in fine grâce aux simulations stochastiques de cette incertitude spatiale.

Chacune des questions envisagées est l'objet d'une partie de ce manuscrit.

Deuxième partie

Identification automatique des couches de sol homogènes

6. Problématique

6.1 Contexte

D'une manière générale la reconnaissance des sols a pour objectif de décrire le terrain du point de vue de sa stratigraphie et de ses propriétés physiques et mécaniques. Cela entraîne l'identification des couches de terrain nécessaires pour la conception et justification de la stabilité de l'ouvrage géotechnique en question.

Un des grands intérêts des pénétrètres est leur capacité à enregistrer pratiquement en continu la résistance du sol. Le pénétrogramme ainsi obtenu est très riche d'informations sur la stratigraphie (Shahrou et Gourvés (2005)) grâce au grand nombre de mesures et au faible pas d'échantillonnage permettant de caractériser la variabilité de nature et de comportement d'une formation donnée. D'ailleurs, l'Eurocode 7 (NF P 94-261 (2013)) établit que la stratigraphie d'un site peut être déduite à l'aide des essais de pénétration. Pour notre cas d'étude, la succession et l'homogénéité des couches du terrain seront donc déduites à partir de l'analyse du profil de résistance dynamique de pointe q_d du Panda.

Concernant l'essai Panda, cette stratigraphie est habituellement décelée sur la base d'une analyse experte réalisée par l'ingénieur, grâce à une analyse visuelle des variations du pénétrogramme qui est souvent complétée par un levé géologique du site. Compte tenu de la subjectivité de cette procédure, il existe un intérêt de compter sur des outils et techniques d'aide à l'ingénieur pour proposer un découpage automatique à partir de l'analyse du profil pénétrométrique Panda. D'ailleurs, la division du profil géotechnique en unités homogènes s'avère une étape préalable à l'application des méthodes statistiques qu'on utilise pour modéliser les propriétés du sol, lesquelles ne peuvent porter que sur des populations que l'on a clairement identifiées comme homogènes afin que la variabilité puisse être modélisée par une loi statistique (Uzielli et al. (2006)).

6.2 Notion de "couche"

Avant de développer des outils permettant la détermination des unités de sol physiquement homogènes, il est nécessaire de définir la notion de "couche" au sein d'un sol. C'est pourquoi, on présente par la suite une réflexion sur la notion de "couche", puis on formalisera quelle est la définition retenue dans ce travail.

Le terme "couche" représente un ensemble du terrain homogène vis-à-vis des différentes caractéristiques (pétrographiques, minéralogiques, granulométriques, mécaniques, hydrauliques, tectoniques ou paléontologiques), en fonction du contexte d'étude. Par exemple en sédimentologie, une couche est égal à une unité de sédimentation qui s'est déposée sous des

conditions physiques relativement stables, et dont l'épaisseur est normalement supérieure à 1cm. On note que la nature physique du sol n'est pas prise en compte dans ce cas là. Un autre exemple, est celui du contrôle de compactage, la couche est définie et considérée comme une épaisseur de matériau déposée et compactée par les engins.

Dans le cadre de la reconnaissance des sols en géotechnique, la couche est une simplification de l'ingénieur pour représenter un volume de sol homogène. Voici donc la première question, comment est-ce définit-on cette homogénéité, en fonction de la nature et composition des matériaux ou d'après son comportement mécanique ?

Les données de la géologie ne sont pas toujours pas celles de la géotechnique : des sols de même origine géologique peuvent avoir des propriétés géotechniques très différentes, et réciproquement. Une stratigraphie lithologique nécessite l'identification des sols en laboratoire, en fonction de leur granulométrie et plasticité. Cependant et à titre d'illustration, il est évident qu'une formation de sol sableux qui peut être considérée homogène au sens géologique, peut présenter une large fourchette de valeurs de résistance de pointe. Elle peut donc, dans le cadre d'une étude géotechnique, être divisée en plusieurs sous-couches avec différentes propriétés mécaniques et d'état. En effet, plusieurs travaux dans la littérature ont montré qu'une stratigraphie sur la base de résistances mécaniques peut être différente du levé géologique d'un site. ([Cherubini et al. \(2002\)](#)) montre comment les limites entre couches géotechniques statistiquement homogènes ne coïncident pas exactement avec la position des interfaces lithologiques. ([Hegazy and Mayne \(2002\)](#), [Zhang and Tumay \(2003\)](#)) discutent du fait qu'il n'existe pas de "corrélation" parfaite entre une homogénéité en termes de nature et la réponse mécanique mesurée à l'aide du CPT, souvent utilisé pour déterminer le profil géotechnique du site. ([Wang et al. \(2013\)](#)) montrent aussi comment une couche d'argile avec des passages limoneux, considérée comme homogène suite au forage réalisé sur site et à l'identification au laboratoire, peut être subdivisée en plusieurs unités à partir des différentes techniques stratigraphiques proposées pour l'essai CPT.

Un autre exemple de formations de sols qui peuvent être considérées homogènes en termes de composition mais pas vis-à-vis de leur réponse mécanique est donné par ([Leroueil and Hight \(2003\)](#)) pour l'argile de Bothkennar. Les essais granulométriques sur différents échantillons prélevés ont révélé des courbes granulométriques très similaires. Cependant, cette formation est caractérisée par trois faciès du fait des différents processus de dépôt, entraînant des comportements mécaniques assez différents comme le montrent les résultats des essais triaxiaux.

Voici donc pourquoi la définition de "couche" s'avère un problème complexe qui n'a pas une réponse unique. La détermination du profil géotechnique est fonction du problème posé, et différents découpages peuvent être envisagés pour un même site selon l'ouvrage à concevoir et le calcul ou l'analyse à mener. Dans ce travail, on retiendra comme définition de "couche" une unité de sol homogène à la fois d'un point de vue minéralogique, de ses propriétés physico-chimiques et de la résistance de pointe dynamique q_d mesurée à l'aide de l'essai Panda. On remarque que cette définition est cohérente avec celle des couches homogènes (géologiquement et mécaniquement), préconisée par l'Eurocode 7 pour un calcul de résistance de pointe équivalente vis-à-vis du calcul des fondations.

6.3 Problème posé par une reconnaissance de couches "experte"

Afin d'évaluer la pertinence et la fiabilité d'une reconnaissance de couches par analyse experte, on présente ici les résultats d'une étude menée auprès de neuf personnes possédant une connaissance éprouvée de l'essai dynamique Panda. On a demandé à ces 9 utilisateurs expérimentés de réaliser un découpage expert de 12 pénétrogrammes de 1 à 6 mètres de profondeur réalisés sur différents sites de la région Auvergne-Rhône-Alpes. Les sondages ont été réalisés dans le cadre de diverses missions de reconnaissance géotechnique de surface. Pour chacun de ces sondages, on dispose du levé géologique du site grâce à un sondage destructif ou à la tarière réalisé à proximité du sondage pénétrométrique. Lors de leur analyse des sondages, mis à part le profil de résistance dynamique q_d , les experts ne possèdent pas d'informations sur la géologie du site ni d'autres informations supplémentaires. Enfin, il a été demandé aux experts d'analyser et d'effectuer le découpage des sondages selon la définition de "couche" retenue dans notre étude.

La figure (Fig.6.1) fournit les résultats obtenus par cette enquête où le nombre de couches détectées par chaque expert est reporté pour chaque sondage analysé. Une deuxième figure récapitule la moyenne ainsi que les valeurs maximales et minimales des couches identifiées par les experts (Fig.6.2). Ce graphique fait également apparaître le nombre de couches lithologiques du site suite à l'analyse du géologue à partir des sondages carottés ou à la tarière. Il s'agit donc des couches de matériaux avec des propriétés minéralogiques différentes. Lorsque le forage a été mené jusqu'à la même profondeur que le sondage pénétrométrique, ce nombre de couches est indiqué par une ligne bleue et dans le cas où les profondeurs sont différentes, la ligne est tracée en rouge. L'analyse des résultats met en évidence des écarts importants entre les différents experts utilisant pourtant les mêmes règles d'analyse. On voit que pour un même sondage, on peut trouver de 2 à 10 couches suivant l'expert qui analyse le sondage. On constate sur cette enquête que deux experts (le "1" et le "5") ont tendance à détecter beaucoup plus de couches que les autres. Néanmoins, il n'est pas rare de trouver une différence de 3 à 4 couches sur un même sondage. Enfin on constate quand même que sur un bon nombre de sondages, plusieurs experts fournissent des découpages identiques. Lorsque l'on compare le découpage donné par les experts avec le nombre de couches lithologiques fourni par une analyse géologique, le nombre de couches décelées par les experts est en moyenne plus important. Toutefois cette différence paraît "normale" puisque l'analyse des experts prend en compte des variations mécaniques qui comme cela a déjà été dit ne sont pas forcément en adéquation avec les variations "géologiques" du sol.

Par la suite, nous avons calculé pour chaque sondage, les écarts issus de la comparaison des analyses réalisées par chaque expert vis-à-vis des autres. La distribution de ces écarts est représentée en figure (Fig.6.3). Malgré les écarts importants observés pour certains sondages (jusqu'à 10 couches), il existe environ 60% de chance que cet écart soit inférieur à deux couches entre les experts.

Afin de mieux illustrer notre propos et les différences d'analyse entre les experts, nous présentons par la suite (Fig.6.4), un comparatif de l'analyse des couches estimées par 4 experts sur le sondage E663s2. Il s'agit d'un profil constitué de sols fins. On dispose de la caractérisation au laboratoire de deux échantillons récoltés à 0.5 et 1.5 mètres de profondeur, dont les paramètres physiques déterminés par les essais classiques au laboratoire sont four-

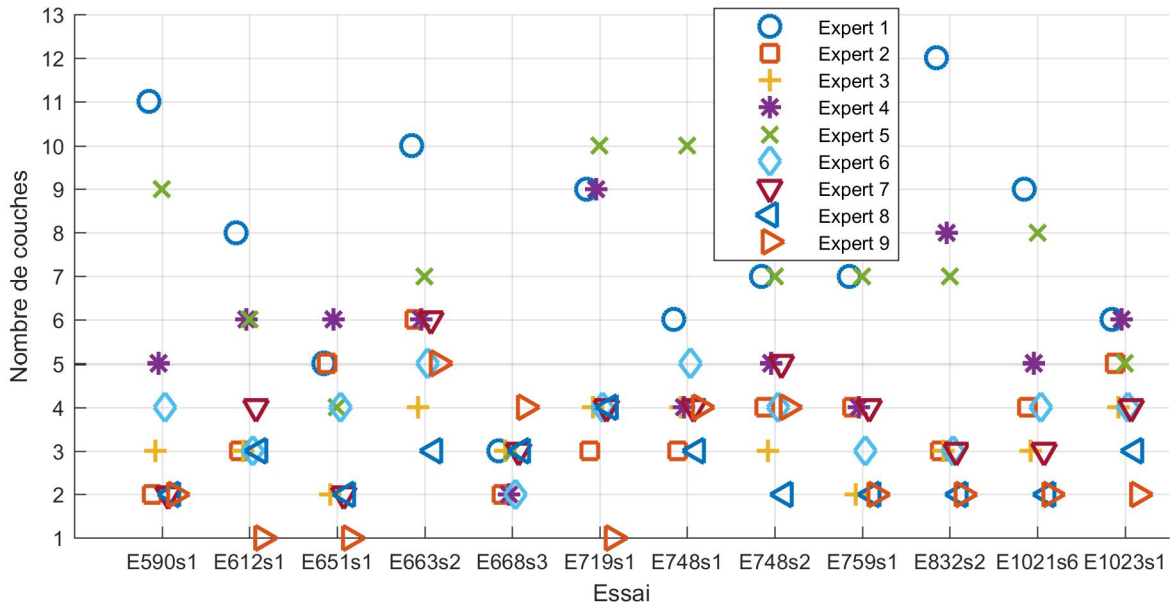


FIGURE 6.1 – Nombre de couches décelées pour chaque pénétrogramme de sondage étudié d’après l’analyse des experts

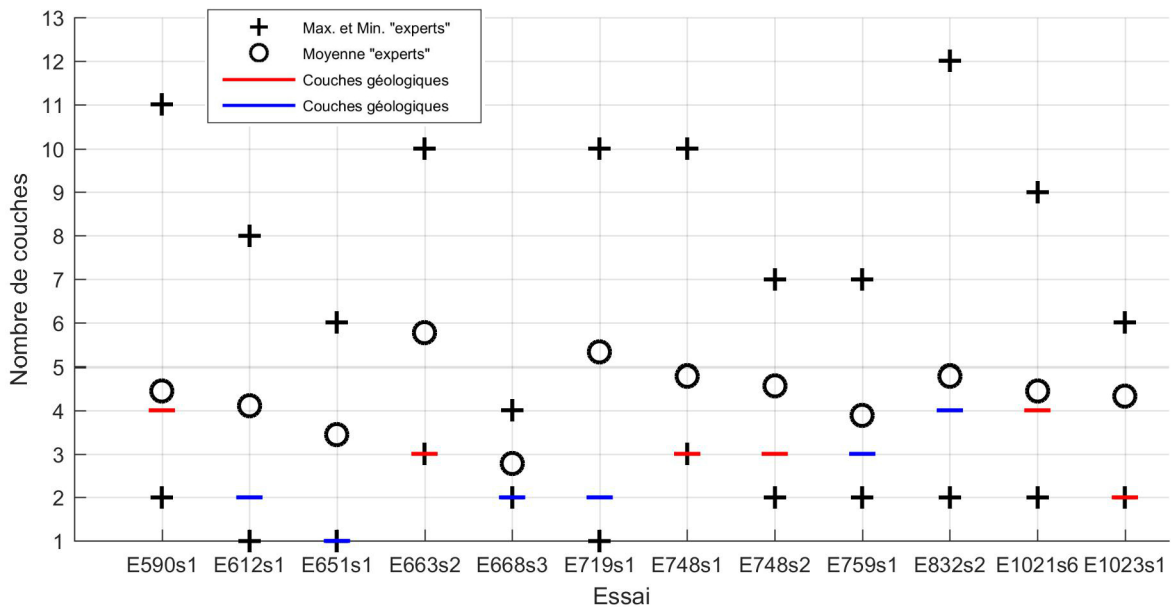


FIGURE 6.2 – Synthèse comparative du nombre de couches pour chaque essai d’après l’analyse des experts et le nombre de couches identifiées par le géologue. Code couleur : **bleu** = si le forage géologique a été mené jusqu’à la même profondeur que le sondage pénétrométrique ; **rouge** = si les profondeurs sont différentes.

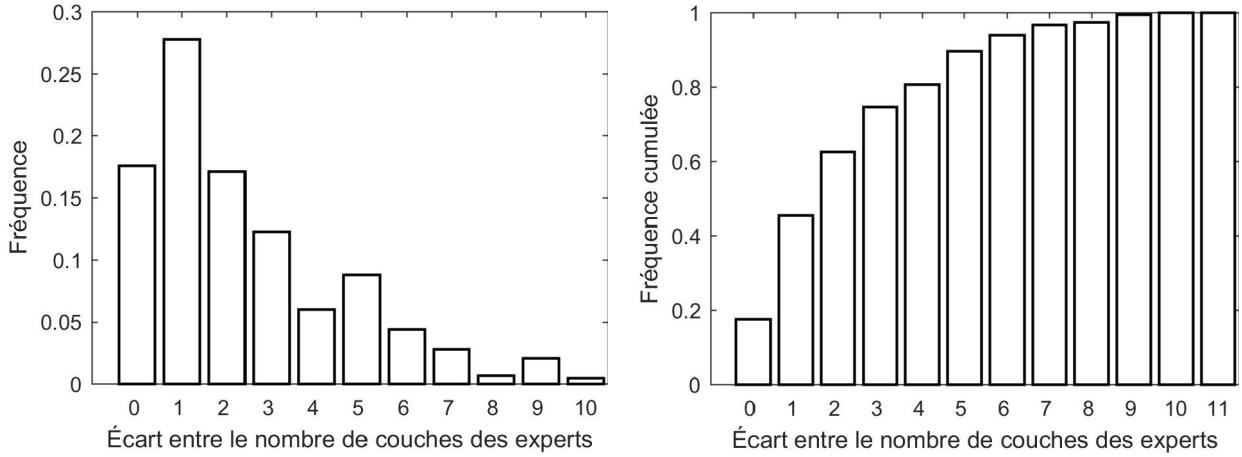


FIGURE 6.3 – Distribution des fréquences

TABLE 6.1 – Caractéristiques des échantillons prélevés à proximité de l’essai E663s2

Prof. (m)	Nature	Passant à 80 μm	Passant à 2mm	VBS	w(%)	GTR
0.5	Argiles peu plastiques	89.5	99.5	4.1	24.1	A2
1.5	Argiles ocres plastiques	81.7	95.5	7.8	35.2	A3

nis dans le tableau (Tab.6.1). Les matériaux présentent des distributions granulométriques similaires. Toutefois, les argiles peu plastiques qui composent l’horizon de terres noires existantes en surface présentent une plasticité plus faible que les sols argileux rencontrés au delà d’un mètre. Nous avons constaté qu’intuitivement, les experts cherchent à diviser le pénétrogramme en unités où la résistance de pointe q_d est constante ou présente une augmentation systématique selon la profondeur par l’effet de la consolidation. Certains sujets nous ont même raconté qu’ils s’aident d’une règle pour réaliser cette tâche, comme on peut le voir sur le second découpage (Fig.6.4), où l’ingénieur a même tracé les droites représentant la résistance de calcul au sein de la couche. D’autres par contre mènent des analyses plus détaillées, tel qu’observable sur le dernier découpage (le plus à droite de la figure) (Fig.6.4) où l’expert met l’accent sur des faibles épaisseurs où une chute importante de la résistance mécanique est constatée.

On a aussi représenté la coupe géologique issue d’un sondage destructif jusqu’à trois mètres de profondeur réalisé à côte du sondage en (Fig.6.4). Trois formations ont été reconnues par le géologue (jusqu’à 3 mètres) tandis que ces quatre experts ont signalé 3, 4, 5 et jusqu’à 7 couches. Une première formation constituée d’argiles grumeleuses a été reconnue jusqu’à environ 1 mètre de profondeur. Les caractéristiques mécaniques relevées au sein de cette formation sont faibles et elle a été distinguée par les 4 experts. Après 1 mètre de profondeur, une formation argileuse a été repérée avec deux horizons. Pour le premier découpage situé à gauche, (Fig.6.4) la formation d’argile plastique a été considérée comme un seule unité. Or, pour le reste, une interface à environ 1.7 mètres de profondeur a été repérée, où une transition lithologique a été signalée par l’un des experts entre les deux horizons argileux. En général, il s’avère que les stratigraphies proposées par les experts à partir de l’analyse de la seule courbe pénétrométrique ne coïncident pas avec la coupe lithologique du

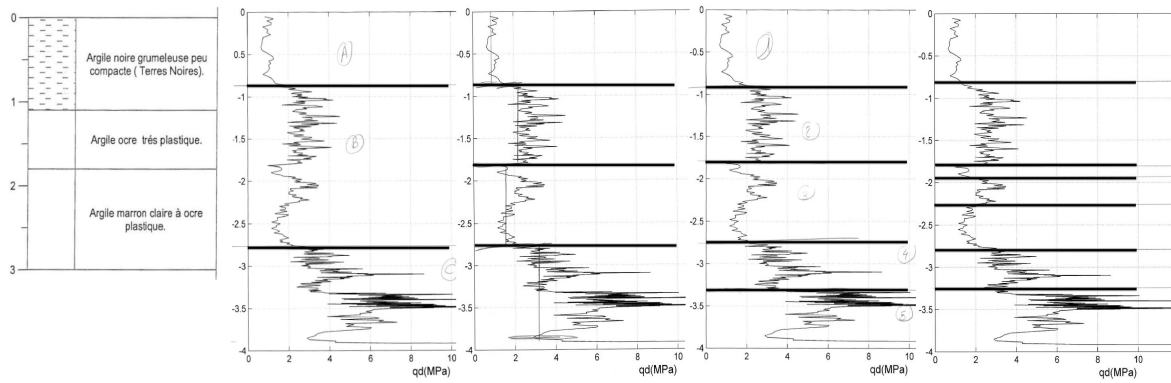


FIGURE 6.4 – Coupe schématique du sol déduite du sondage à la tarière et découpages proposés par les experts pour l'essai E663s2

site déterminée par le géologue.

Suite aux résultats de l'enquête, on conclut que le découpage "expert" sur la base d'une analyse visuelle d'un pénétrogramme, fournit comme on s'y attendait une différence avec un découpage lithologique mais est également l'emprunt de beaucoup de subjectivité entre les différents experts ce qui amène à des différences d'analyse qui peuvent être assez importantes. On propose donc d'étudier des procédures numériques afin de formaliser cette procédure, apporter une aide à l'ingénieur et permettre ensuite une aide à la décision sur l'analyse des couches ainsi détectées.

7. État de l'art

7.1 Méthodes proposées dans la littérature

Dans ce qui suit, on présente une synthèse bibliographique sur les principales approches de découpage proposées pour le pénétromètre statique CPT/CPTU, l'essai pressiométrique et le pénétromètre dynamique Panda.

7.1.1 Méthodes de découpage pour l'essai CPT et CPTU

On présente premièrement une synthèse bibliographique sur les différentes approches permettant d'établir une stratigraphie grâce à l'analyse des mesures issues du pénétromètre statique (CPT) et du piézocône (CPTU). Compte tenu du grand nombre de méthodes existantes pour cet essai, on présente un tableau récapitulatif (Tab.7.2) avec les caractéristiques principales des différentes approches. Les différentes méthodes peuvent être regroupées de la façon suivante :

1) Abaques de classification empiriques :

Plusieurs abaques de classification des sols ont été proposés sur la base des résultats des essais CPT, permettant de déterminer une stratigraphie du profil géotechnique. Suite aux abaques présentés par (Schmertmann (1978), Douglas and Olsen (1981)), le système non normalisé le plus couramment utilisé a été suggéré par (Robertson et al. (1986)).

Compte tenu des effets de la pression de confinement (la résistance à la pénétration et le frottement latéral unitaire augmentent avec la profondeur), des abaques de classification sur la base de mesures normalisées de l'essai CPT ont été aussi présentés dans la littérature. Le plus utilisé est celui de Robertson (Robertson (1990), Robertson (2009)) repris par la norme (NF P 94-261 (2013)), en proposant un raffinement de l'abaque initial, présentant une résistance au cône et un rapport de frottement (*friction ratio*) normalisés Q_{tn} et F_r :

$$Q_{tn} = \frac{q_t - \sigma_{vo}}{P_a} \cdot C_N \quad (7.1)$$

$$F_r = \frac{f_s}{q_t - \sigma_{vo}} \cdot 100\% \quad (7.2)$$

avec la résistance de pointe corrigée $q_t = q_c + (1 - a)u_2$ où a est le rapport de la surface non soumise à la pression interstitielle à la surface totale du cône, $P_a = 101.3kPa$ la pression atmosphérique, σ_{vo} et σ'_{vo} les contraintes verticales totales et effectives initiales. On note u_2 et u_0 les pressions interstitielle et hydrostatique initiales. Le

TABLE 7.1 – Interprétation de l’abaque (Robertson (1990)) en fonction de l’indice de comportement I_c .

I_c	Zone	Nature du sol
-	9	Sables cimentés ou dilatants
-	8	Sols fins intermédiaires très raides
$I_c < 1.31$	7	Sables à sables graveleux
$1.31 < I_c < 2.05$	6	Sables propres à sables silteux
$2.05 < I_c < 2.60$	5	Sables silteux à silts sableux
$2.60 < I_c < 2.95$	4	Silts argileux à argiles silteuses
$2.95 < I_c < 3.60$	3	Argiles à argiles silteuses
$I_c > 3.60$	2	Sols organiques et tourbes
-	1	Sol fin (argile) ou silt sensible

terme C_N est le facteur de correction de la pression de confinement défini comme suit :

$$C_N = \left(\frac{P_a}{\sigma'_{vo}} \right)^n \quad (7.3)$$

où l’exposant n varie entre 0.5 et 1 en fonction de la nature du sol (*Soil Behaviour Type* SBT) ($n = 1$ pour les sols argileux et $n = 0.5$ pour les sols sableux) (Robertson (2009)) :

$$n = 0.381 \cdot I_c - 0.05 \cdot \frac{\sigma'_{vo}}{P_a} - 0.15 \quad (7.4)$$

avec $n \leq 1$. On précise que dans le premier travail (Robertson (1990)) la résistance au cône normalisée est définie avec le terme Q_t :

$$Q_t = \frac{q_t - \sigma_{vo}}{\sigma'_{vo}} \quad (7.5)$$

et dans (Robertson (2009)) la notation Q_{t1} est proposée puisque ce paramètre est équivalent à la résistance de pointe corrigée q_t avec un facteur de correction sur la pression de confinement dont l’exposant n est égal à 1.

En complément (Robertson and Wride (1998)) intègrent l’expression de l’indice de comportement I_c dans l’abaque normalisé (Robertson (1990)), permettant d’approcher les frontières des zones par des arcs de cercle et d’identifier ainsi le type de sol (Tab.7.1). A titre d’exemple, la valeur $I_c = 2.6$ sépare le comportement argileux du comportement sableux. Cet indice est issu d’un calcul itératif à partir de l’équation :

$$I_c = \sqrt{[3.47 - \log(Q_{tn})]^2 + [1.22 + \log(Fr)]^2} \quad (7.6)$$

Afin de tenir compte des mesures de pressions interstitielles recueillies au piézocône. (Robertson (1990)) inclut un abaque additionnel qui tient compte du rapport de pression interstitielle, B_q avec la résistance au cône normalisée Q_t :

$$B_q = \frac{u_2 - u_0}{q_t - \sigma_{vo}} = \frac{\Delta u}{q_n} \quad (7.7)$$

où q_n est la résistance de pointe nette.

Plus récemment (Schneider et al. (2008)) propose aussi un abaque de classification basé sur la résistance au cône normalisée Q_{t1} et la pression d'eau nette normalisée ($\Delta u/\sigma'_{vo}$).

2) Techniques statistiques

2.1) Test paramétrique par fenêtre glissante

Ces méthodes appliquent un test paramétrique¹ pour comparer deux populations définies à l'aide d'une fenêtre glissante de taille fixe. La position des interfaces entre couches est associée à la profondeur où l'indicateur, ou statistique, du test atteint un maximum local et dépasse une certaine valeur seuil. On distingue plusieurs approches dans la littérature suite au type de test statistique utilisé.

Initialement, trois méthodes ont été proposées par (Webster and Beckett (1968), Webster and Wong (1969), Webster (1973)) :

- i. *T ratio* : basé sur un test d'égalité de deux moyennes ;
- ii. *Intra-classe* : basé aussi sur une comparaison des moyennes de deux échantillons. Toutefois, la statistique résultant de ce test est normalisée dans un intervalle $[0, 1]$.
- iii. D^2 : à la différence des deux tests précédents il permet de prendre en compte plusieurs variables.

(Wickremesinghe (1989)) a étudié l'application de ces trois approches aux mesures pénétrométriques du CPT afin d'établir une stratigraphie du site et (Mora (2013)) pour l'identification de couches de sols volcaniques. D'autres chercheurs ont utilisé ces techniques pour les comparer avec de nouvelles approches : dans (Phoon et al. (2003)) le coefficient intra-classe est retenu et dans (Ching et al. (2015a)) l'approche T ratio.

(Phoon et al. (2003), Phoon et al. (2004)) proposent l'utilisation d'un autre test statistique, celui de Bartlett. La méthode proposée se déroule en deux étapes. Tout d'abord, on identifie à l'œil les limites entre les différentes couches de sol en fonction des maxima locaux du test de Bartlett. Ensuite, au sein de chacune des couches détectées, on applique l'approche dite "géostatistique", puisqu'il s'agit d'une version modifiée du test original (Bartlett (1937)) capable de rejeter l'hypothèse nulle de stationnarité en moyenne d'ordre 2 en fonction du modèle théorique d'autocorrélation spatiale. En d'autres termes, pour un niveau de risque égal à 5%, cette technique permet de déterminer si la moyenne et l'écart-type (ou la variance) du profil de résistance mesurée le long de chacune des formations identifiées peuvent être considérés constants. Pour définir la valeur seuil, la méthode tient compte du modèle théorique de corrélation spatiale choisi, de la distance de corrélation et de l'intervalle de mesure.

2.2) Analyse par cluster

L'analyse par cluster, connue comme *clustering* en littérature anglaise, consiste à regrouper les données en ensembles ou groupes homogènes d'après un critère de similarité exprimé par le biais d'une mesure de distance (ou métrique) entre plusieurs variables. Cette technique a fait l'objet de plusieurs publications comme méthode de regroupement des données pénétrométriques en ensembles géomécaniquement homogènes. (Hegazy (1998), Hegazy and Mayne (2002)) ont été les premiers à étudier

1. On parle de tests paramétriques lorsque l'on fait l'hypothèse que les populations à comparer suivent une distribution théorique, généralement gaussienne.

l'application de cette technique au profil géotechnique issu de l'essai CPTU. L'analyse a été appliquée aux paramètres normalisés (Q_t, B_q). Afin d'évaluer la stratigraphie du sol à partir des résultats de l'analyse par cluster, ces auteurs ont proposé les critères d'interprétation suivants :

- i. si l'épaisseur du cluster reste inférieure à 0.5 mètres, il s'agit d'une zone de transition, d'une lentille ou d'une erreur de la méthode ;
- ii. lorsque l'épaisseur du cluster se trouve entre 0.5 et 1 m, il s'agit d'une couche secondaire ou une formation composée par un mélange des différents matériaux ;
- iii. des épaisseurs supérieures à 1 mètre sont considérées comme des couches primaires. (Facciorusso and Uzielli (2004)) réalisent l'analyse statistique sur le couple de paramètres normalisés (Q_t, F_r). Ils trouvent une bonne correspondance entre les stratigraphies déduites suite à l'analyse par cluster et les profils lithologiques du site d'après des forages et sondages carottés. (Mlynarek et al. (2007)) ont étudié différents algorithmes de clustering, ainsi que plusieurs distances métriques, pour différents couples de paramètres d'entrée (Q_t, F_r et B_q). Par ailleurs, les travaux (Hegazy and Mayne (2002) et Mlynarek et al. (2007)) ont conclu que la distance "cosinus" s'avère être le meilleur choix pour déterminer le profil géotechnique du site. Enfin, l'analyse par cluster en 3D a été aussi explorée dans le travail de (Liao and Mayne (2007)). Cette nouvelle approche facilite l'identification des couches d'argiles raides qui sont parfois incorrectement classées comme « sables » par les abaques de classification empiriques basés sur la résistance à la pénétration et le rapport de frottement.

3) Techniques probabilistes

Ces approches visent à déterminer le nombre et les limites entre couches les plus probables.

3.1) Basées sur la théorie des ensembles flous

(Zhang and Tumay (2003)) présentent un bilan de l'ensemble des études réalisées pour la mise au point d'une technique probabiliste basée sur la théorie des ensembles flous permettant l'identification des unités physiquement homogènes à travers l'analyse du profil CPT. L'approche est basée sur l'analyse d'un indice de classification noté U qui est calculé à partir d'une transformation de coordonnées des paramètres de l'abaque de classification de Douglas (Douglas and Olsen (1981)) (q_c, F_r). Ce paramètre est modélisé comme une variable aléatoire qui suit une loi normale et il présente une corrélation statistique avec la nature du sol en termes de granulométrie et de pourcentage des fines. Cette technique fournit à chaque profondeur la probabilité du type de sol ainsi que le degré d'appartenance aux ensembles flous définis (très probablement sable HPS , très probablement mélange intermédiaire HPM ou très probablement argileux HPC). L'application de la méthode statistique Intra-classe avec une fenêtre glissante de 1.5 m au profil de l'indice de classification U en fonction de la profondeur, permettait d'identifier la stratigraphie finale du profil de sol.

3.2) Basées sur des approches bayésiennes

Le raisonnement bayésien permet d'incorporer de l'information *a priori* lors des analyses à réaliser en fusionnant ainsi plusieurs sources d'information. Les généralités et notions de base sur la statistique bayésienne seront décrites dans la partie IV de ce document.

(Wang et al. (2013)) propose une méthode stratigraphique en intégrant l'abaque de classification normalisé de Robertson (Robertson (1990)) sur la base des variables

(Q_{tn}, F_r) dans une approche bayésienne. La méthode permet d'identifier le nombre de couches le plus probable et d'estimer ainsi l'épaisseur de couche la plus probable. Une classification de la nature du sol en terme de probabilité est aussi fournie grâce à l'abaque de classification.

Une méthode plus générale aussi fondée sur la statistique bayésienne a été proposée par (Cao and Wang (2013)). Cette technique a pour but l'identification des couches statistiquement homogènes vis à vis de paramètres géotechniques, tel que l'angle de frottement, déduites grâce aux relations empiriques proposées dans la littérature pour le CPT.

4) Analyse par ondelettes

Cette méthodologie est la plus récente. Elle a été développée par (Ching et al. (2015a)), et repose sur l'identification des limites entre couches à partir des variations du profil de l'indice de comportement I_c issue de la formulation de (Robertson and Wride (1998)), qui est analysé par une transformée continue en ondelettes. Très succinctement, les ondelettes sont des outils mathématiques introduits pour combler la lacune entre les deux modes de représentation du signal, temporel et fréquentiel. La transformée par ondelettes permet de représenter un signal comme une combinaison de fonctions élémentaires de moyenne nulle nommées ondelettes (Mallat (1989)). Il existe plusieurs familles d'ondelettes avec différentes propriétés, le choix d'une ondelette mère est un choix crucial conditionné par le problème à résoudre. Dans l'approche proposée par (Ching et al. (2015a)), l'ondelette dérivée première de gaussienne est retenue. Le principal intérêt de cette technique est qu'elle permet de mener une analyse multi-échelle par contraction ou dilatation dans le domaine spatial de l'ondelette analysante.

La principale contrainte de cette méthode est qu'elle ne permet que la prise en compte d'une source d'information. De même, elle ne permet pas de distinguer différentes unités géotechniques ayant la même valeur de I_c (ex. couches de sols argileux normalement consolidés et surconsolidés). Néanmoins, l'application de cette méthode à la résistance au cône corrigée s'avère une perspective intéressante.

Bilan

Les performances ainsi que les caractéristiques principales des différentes approches proposées dans la littérature sont résumées dans le tableau (Tab.7.2).

i. Nombre de paramètres pris en compte :

Parmi tous les tests paramétriques par fenêtre glissante, seule la méthode D^2 est une méthode multi-variable. Cependant la méthode Intra-classe pourrait aussi prendre en compte plusieurs paramètres (Cherubini et al. (2002)). De même, la méthode basée sur l'analyse par ondelettes n'est applicable qu'à un paramètre.

Une analyse statistique de plusieurs variables est possible au moyen de l'analyse par cluster. D'ailleurs, les méthodes bayésiennes permettent de prendre en compte plusieurs sources d'information.

ii. Applicabilité pour d'autres essais ou paramètres géotechniques :

Les méthodes basées sur les abaques de classification empiriques, ainsi que la méthode bayésienne proposée par (Wang et al. (2013)), sont inhérentes à l'essai CPT, tout comme l'approche basée sur les ensembles flous proposée par (Zhang and Tumay

(2003)). On souligne que toutes ces techniques fournissent aussi une classification de la nature du sol au sein de la couche détectée.

L'approche par ondelettes quant à elle, est appliquée à l'indice de comportement qui est un paramètre qui ne peut être déduit qu'à partir des mesures recueillies par le CPT mais elle pourrait être adaptée à d'autres variables. Les techniques restantes s'avèrent être des méthodes très générales applicables à d'autres profils de mesures géotechniques.

iii. Efficacité et épaisseur minimum détectable :

Pour les tests paramétriques par fenêtre glissante l'épaisseur minimum de la couche à identifier est fonction de la taille de la fenêtre. L'analyse par clusters s'avère être une technique performante pour la détection des lentilles ou couches de faible épaisseur. Cependant, les résultats sont fonction du nombre de clusters ou d'ensembles choisis représentant l'échelle d'observation ou le degré de résolution de l'analyse. Par contre, le temps d'analyse pour ces deux méthodes est très rapide.

Par ailleurs, les approches bayésiennes peuvent prendre en compte des épaisseurs variables puisque l'épaisseur des couches à identifier est modélisée par une variable aléatoire et elles sont les seules permettant d'incorporer explicitement des connaissances ou informations a priori. Ce sont des méthodes très rigoureuses dont la principale contrainte est qu'elles sont très gourmandes en temps de calcul informatique.

Les ondelettes sont quant à elles un outil naturel pour réaliser une analyse multi-zoom à plusieurs échelles. La méthode basée sur ce principe permet ainsi d'identifier des couches d'épaisseur variables avec un temps de calcul faible.

iv. Prise en compte de la variabilité spatiale :

Il semblerait que seules la procédure modifiée par Bartlett et les méthodes fondées sur le raisonnement bayésien permettent la prise en compte de la variabilité spatiale grâce à la modélisation du profil de mesure comme un champ aléatoire.

v. Automatisation :

L'analyse par cluster nécessite l'interprétation de l'ingénieur qui doit aussi tester l'influence du nombre de clusters. Les procédés fondés sur des tests statistiques par fenêtres glissantes sont automatisables lorsque l'on fixe une taille de fenêtre et une valeur seuil. À l'exception près de la procédure du test modifié de Bartlett pour laquelle on ne peut pas retenir une valeur seuil au préalable. Les approches fondées sur le principe bayésien nécessitent de définir les paramètres liés à l'information a priori et de répéter l'analyse pour des nombres différents de couches possibles, ce qui peut entraîner des temps de calcul longs. Enfin, l'approche par ondelettes peut ne pas être facile à interpréter pour les ingénieurs et nécessite d'un étalonnage à partir d'une base de données pour sa mise en œuvre.

7.1.2 Méthodes de découpage pour le pressiomètre

L'essai pressiométrique conduit à une courbe effort-déformation dont l'analyse permet de calculer un module de déformation E_M (module de Ménard ou pressiométrique), une pression limite p_l et de fluage p_f . Pour identifier les couches de sol lors des reconnaissances géomécaniques, l'analyse par cluster a été présentée dans (Monnet et al. (2002), Monnet and Broucke (2012)). Le choix des variables statistiques se porte sur le couple (p_l, E_M) . Parmi les deux distances métriques étudiées, euclidienne et cosinus, la distance cosinus semble

TABLE 7.2 – Récapitulatif des méthodes existantes. 1) *Probabiliste - floue*; 2) *T ratio*; 3) *Intra-classe*; 4) D^2 ; 5) *Bartlett*; 6) *Analyse par cluster*; 7) *Abaque de classification*; 8) *Bayésienne - Robertson*; 9) *Bayésienne*; 10) *Ondelettes*.

Méthodes	Classification de la nature du sol	Plusieurs variables	Uniquement CPT	Procédure automatique	Information a priori	Variabilité spatiale	Test paramétrique	Temps calcul
1	Oui	Non	Oui	Oui	Non	Non	Non	Rapide
2	Non	Non	Non	Oui	Non	Non	Oui	Rapide
3	Non	Oui	Non	Oui	Non	Non	Oui	Rapide
4	Non	Oui	Non	Oui	Non	Non	Oui	Rapide
5	Non	Non	Non	Non	Non	Oui	Oui	Rapide
6	Non	Oui	Non	Non	Non	Non	Non	Rapide
7	Oui	Oui	Oui	Non	Non	Non	Non	Rapide
8	Oui	Oui	Oui	Non	Oui	Non	Non	Lent
9	Non	Oui	Non	Non	Oui	Oui	Non	Lent
10	Oui	Non	Oui	Oui	Non	Non	Non	Rapide

être la mieux adaptée. D'après les résultats, l'application de l'analyse par cluster à l'essai pressiométrique permet d'établir une meilleure constitution du profil géotechnique du site.

7.1.3 Méthodes de découpage pour l'essai Panda

Dans le cadre du contrôle de compactage, (Chaigneau (2001)) propose une technique d'identification de la stratigraphie appliquée à l'analyse du signal pénétrométrique obtenu avec le Panda. Les hypothèses retenues pour détecter une couche sont les suivantes : la résistance de pointe dynamique suit une loi lognormale et le coefficient de variation est constant pour un matériau homogène. La méthode proposée consiste à définir un intervalle de confiance à 95% du $\ln(q_d)$ à partir des cinq premières valeurs mesurées. À chaque fois que le logarithme d'une valeur mesurée est situé en dehors de l'intervalle, une limite entre couches est repérée à cette profondeur. Ensuite, un autre intervalle basé sur les cinq mesures suivantes est créé. L'algorithme complet, propose d'appliquer cette méthode de contrôle de régularité depuis la profondeur minimale jusqu'au bas du sondage (sens "aller") et ensuite de l'appliquer en sens inverse (sens "retour"). La profondeur finale des interfaces est calculée comme la moyenne des limites issues du découpage dans les deux sens.

7.1.4 Bilan

À ce jour il existe une seule méthode de détection des couches pour l'essai Panda, développée dans le cadre du contrôle de compactage. Elle est fondée sur l'hypothèse du coefficient de variation (CV) constant pour un même matériau considéré homogène. La valeur du CV est choisie en fonction du type de matériau dans la base de données des matériaux de compactage (Chaigneau (2001)). Cependant vis-à-vis de la problématique de la reconnaissance des sols, cette information n'est pas disponible. La méthode n'est donc pas applicable pour l'analyse des profils pénétrométriques in-situ.

Par ailleurs, la détermination du profil géotechnique à partir des mesures recueillies par l'essai CPT/CPTU est un sujet de recherche actif. Il existe une grande panoplie de méthodes proposées dans la littérature, dont certaines peuvent s'avérer intéressantes pour notre problématique. Vis-à-vis de l'application à l'essai Panda :

- on est confronté à l'analyse d'un seul paramètre, la résistance de pointe dynamique q_d . On préconise donc des techniques d'analyse pour une seule variable ;
- on ne compte pas sur de l'information a priori concernant la résistance du sol des couches à identifier. C'est pourquoi, une approche fondée sur la technique bayésienne a été écartée.

Suite à la synthèse bibliographique présentée dans le tableau (Tab.7.2), il s'avère que les tests paramétriques par fenêtre glissante semblent être un choix pertinent pour notre problématique. Ce sont des méthodes basées sur des critères statistiques qui visent à diviser le profil de mesures d'étude en zones statistiquement homogènes au sein desquelles la moyenne et l'écart type (ou variance) peuvent être considérés constants. Cette homogénéité statistique dépendra du test paramétrique utilisé. Comme on le montrera par la suite, les tests T ratio et Intra-classe sont principalement concernés par la différence entre les moyennes des deux échantillons à comparer alors que pour le test de Bartlett, il s'agit de la variance des échantillons. Théoriquement, ces techniques vont ainsi permettre de diviser le profil de résistance q_d en unités ou couches de sol homogènes au sein desquelles les variations de q_d

pourront être considérées comme aléatoires autour d’une valeur moyenne. Ceci est cohérent avec les méthodes statistiques classiques utilisées en géotechnique qui visent à modéliser la dispersion des données relative à la formation étudiée par une loi statistique. De même, les approches par champs aléatoires étudiées dans la partie IV de ce manuscrit, nécessitent aussi la définition d’une loi théorique d’une population de données du sol identifiée comme statistiquement homogène.

Enfin, une méthode basée sur l’analyse du profil pénétrométrique par ondelettes pourrait être une autre approche intéressante pour notre problématique. En effet, elle permet d’identifier les positions (profondeur dans notre cas) où le pénétrogramme change ou présente des variations, pour différentes échelles d’observation. Toutefois, elle nécessite d’un étalonnage sur une base de données pour permettre de choisir, entre les discontinuités détectées, lesquelles représentent une interface entre couches. Une description plus détaillée de sa mise en œuvre pour l’essai Panda ainsi que les premiers résultats obtenus sont fournis dans l’(Annexe A.3).

7.2 Méthode de la fenêtre glissante

7.2.1 Principe de la méthode

Les méthodes retenues pour notre étude sont basées sur l’utilisation d’une fenêtre glissante de taille fixe W_d (Fig. 7.1). La fenêtre est divisée en son centre d_o en deux parties qui contiennent les deux populations Ω_1 et Ω_2 à comparer, de tailles n_1 et n_2 respectivement. Soient μ_1 , et μ_2 les moyennes empiriques et σ_1^2 , σ_2^2 les variances empiriques estimées pour Ω_1 et Ω_2 et calculées à partir des formules suivantes :

$$\mu_i = \frac{1}{n} \sum_{i=1}^{n_i} x_i \quad (7.8)$$

$$\sigma_i^2 = \frac{1}{n-1} \sum_{i=1}^{n_i} (x_i - \mu_i)^2 \quad (7.9)$$

avec x_i les valeurs de la population analysée.

Une fois la fenêtre définie et les paramètres calculés, on fait glisser la fenêtre le long du profil de mesures, avec un pas de décalage égal à l’intervalle d’échantillonnage des mesures à analyser. Pour chaque point d_o , on évalue l’homogénéité statistique des deux échantillons à l’aide des tests statistiques décrits par la suite (*T ratio*, *Coefficient Intra-classe* et *test de Bartlett*). Pour l’analyse des mesures issues des sondages géotechniques, on obtient ainsi une courbe fournissant l’évolution de la statistique du test utilisé en fonction de la profondeur, où les maxima locaux (pics du profil) indiquent des zones où, pour l’échelle d’observation retenue, l’homogénéité statistique n’est pas vérifiée. Ces maxima sont donc censés représenter la position des limites entre couches que l’on cherche à identifier. Comme on le montrera par la suite, plus la valeur de la statistique du test analysé est grande et moins les populations comparées sont homogènes.

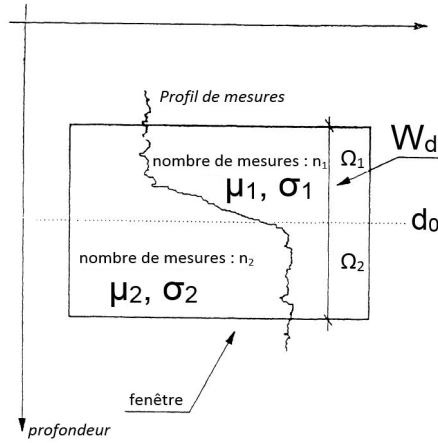


FIGURE 7.1 – Schéma de la procédure par fenêtre glissante

7.2.2 Test paramétriques pour la comparaison de deux échantillons

On présente par la suite les trois tests paramétriques applicables à la procédure par fenêtre glissante et permettant de définir les interfaces entre couches.

T ratio

L'approche *T Ratio* (Webster and Wong (1969)), consiste à utiliser la valeur de la statistique *t de Student* à travers une procédure par fenêtre glissante. Le test *t de Student* est couramment utilisé pour tester l'hypothèse d'égalité de l'espérance de deux variables aléatoires suivant une loi normale et de variance inconnue. Pour comparer deux échantillons, la statistique du test *T ratio*, est définie comme suit :

$$T_{ratio} = \sqrt{\frac{n_1 n_2}{n_1 + n_2}} \frac{\mu_1 - \mu_2}{\Upsilon_w} \quad (7.10)$$

avec Υ_w^2 la variance pondérée des échantillons à analyser. Son calcul est donné par la formule suivante :

$$\Upsilon_w^2 = \frac{1}{n_1 + n_2 - 1} (n_1 \sigma_1^2 + n_2 \sigma_2^2) \quad (7.11)$$

D'après la formule (Eq.7.10), T augmente lorsque l'écart entre les deux moyennes devient plus grand et la somme des deux variances est plus petite. Plus les valeurs de la statistique T sont élevées, plus les deux ensembles de données seront différents l'un de l'autre.

Coefficient de corrélation Intra-classe

Le coefficient de corrélation intra-classe (*intraclass correlation coefficient*) qu'on notera par ρ_I , est calculé selon l'expression suivante (Webster and Beckett (1968), Wickremesinghe (1989)) :

$$\rho_I = \frac{\Upsilon_b^2}{\Upsilon_b^2 + \Upsilon_w^2} \quad (7.12)$$

où Υ_w^2 est la variance pondérée (Eq.7.11) et Υ_b^2 est la variance totale :

$$\Upsilon_b^2 = \frac{1}{n_1 + n_2 - 1} \sum_{i=1}^{n_1+n_2} (x_i - \mu)^2 \quad (7.13)$$

où μ est l'espérance de l'ensemble des données à l'intérieur de la fenêtre. Lorsque les deux échantillons à analyser ont la même taille ($m = n_1 = n_2$), (Zhang and Tumay (2003)) fournissent une autre expression de calcul pour ρ_I :

$$\rho_I = \frac{1}{1 + \frac{1}{\frac{m-1}{m} + \frac{(\mu_1 - \mu_2)^2}{2(\sigma_1^2 + \sigma_2^2)}}} \quad (7.14)$$

D'après l'équation (Eq.7.14) on peut remarquer que la valeur de ρ_I est contrôlée principalement par la différence au carré entre les moyennes des deux échantillons testés et la somme de leurs variances. En conséquence, la valeur de la statistique ρ_I augmente dès que la différence $(\mu_1 - \mu_2)^2$ augmente ou bien que la somme $\sigma_1^2 + \sigma_2^2$ diminue. À la différence de la statistique T , ρ_I est bornée supérieurement par 1. Lorsque le rapport entre ces deux termes $(\mu_1 - \mu_2)^2 / (\sigma_1^2 + \sigma_2^2) \gg 1$, la valeur de ρ_I s'approche de l'unité. Par contre, dès que ce rapport présente des valeurs proches de 0, il n'existe pas de différences significatives entre les deux populations comparées. Le coefficient ρ_I s'approche de sa limite inférieure donnée par $(m - 1) / (2m - 1)$, égale à 0.5 lorsque la population m des échantillons devient grande (Phoon et al. (2003)).

Test de Bartlett

Le test d'égalité des variances de Bartlett est utilisé pour évaluer l'homogénéité des variances de plusieurs échantillons indépendants (condition d'homoscédasticité). Une des contraintes que présente ce test, est qu'il repose fortement sur la normalité des données. Notons par $\sigma_1^2, \dots, \sigma_k^2$ les variances des échantillons de tailles n_1, \dots, n_k . La statistique du test noté B_{stat} (Bartlett (1937)) est donnée par :

$$B_{stat} = \left(n \ln \sigma^2 - \sum_{i=1}^k n_i \ln \sigma_i^2 \right) / C \quad (7.15)$$

σ^2 étant la variance pondérée des k variances considérées et n l'effectif total des k groupes. Le dénominateur C est un facteur de correction fonction de la taille et de la quantité k des échantillons testés.

Lorsque on utilise ce test avec la procédure par fenêtre glissante décrite auparavant, le nombre d'échantillons à comparer devient égal à 2, avec des variances (σ_1^2, σ_2^2) . En utilisant les logarithmes décimaux le test peut être exprimé d'après (Phoon et al. (2003)) :

$$B_{stat} = \frac{2.30259(m-1)}{C} [2 \log_{10} \sigma_w^2 - (\log_{10} \sigma_1^2 + \log_{10} \sigma_2^2)] \quad (7.16)$$

où m est le nombre de mesures utilisées pour calculer les variances σ_1^2 et σ_2^2 , et la variance pondérée σ_w^2 est définie comme suit :

$$\sigma_w^2 = \frac{\sigma_1^2 + \sigma_2^2}{2} \quad (7.17)$$

La valeur de la statistique B_{stat} de Bartlett tend vers 0 dès que les deux variances σ_1^2 et σ_2^2 présentent des valeurs similaires. La constante C du dénominateur prend la valeur suivante :

$$C = 1 + \frac{1}{2(m-1)} \quad (7.18)$$

Dans la procédure décrite dans (Phoon et al. (2003), Phoon et al. (2004)) des valeurs critiques de la statistique B_{stat} (Eq. 7.16) pour un niveau de signification (risque d'erreur accepté) égal à 5% ont été reportées. Ces valeurs critiques tiennent compte de la structure de corrélation spatiale de la variable à étudier, notamment du modèle théorique de corrélation et de la valeur de l'échelle de fluctuation. Le but est de proposer une méthode rigoureuse capable d'accepter l'hypothèse de stationnarité d'ordre 2. Pour plus de détails, le lecteur intéressé peut se reporter aux travaux de (Phoon et al. (2003), Phoon et al. (2004)).

7.2.3 Taille de la fenêtre et valeurs seuils

Toutes les méthodes décrites auparavant utilisent une fenêtre qui glisse sur le profil des mesures à analyser. La taille de cette fenêtre s'avère donc être un choix préalable à faire mais elle n'est pas un paramètre intrinsèque du profil de sol. Cette taille est directement liée à l'épaisseur minimale des couches à détecter. L'épaisseur minimale des couches que nous souhaitons pouvoir détecter est donc le premier critère permettant de fixer la taille de la fenêtre. (Wickremesinghe (1989)) a étudié la sensibilité à la longueur de la fenêtre vis-à-vis de l'analyse des paramètres mesurés à l'aide du CPT. Il a conclu que le choix d'une taille faible entraîne plus de bruit vis-à-vis de la statistique à estimer. Toutefois, le fait de choisir des tailles plus importantes pourrait ne pas permettre la détection de couches d'épaisseurs faibles.

Concernant l'essai de pénétration statique à pointe électrique, plusieurs recommandations et règles par rapport à la taille de la fenêtre sont prescrites. D'une part, ces recommandations sont liées aux règles de la statistique, la taille doit être assez large pour recueillir un nombre de mesures suffisant afin que les tests statistiques à réaliser soient représentatifs. Mais ils existent aussi des raisons physiques, fondées essentiellement sur la zone d'influence de l'enfoncement de la pointe lors de l'essai de pénétration et sur la réponse du sol au moment de la pénétration :

i. Zone d'influence

Sur la base de la théorie d'expansion des cavités et suite aux différentes expériences menées au laboratoire, (Treadwell (1976)) montre comment la réponse à la pénétration du sol est influencée par les matériaux autour de la pointe (avant et après). Cet auteur a donc suggéré une épaisseur minimale de sol influencée par la pénétration en fonction du diamètre de la pointe conique, de 10 à 20 fois le diamètre pour des transitions des sols meubles à raides, et moins de 10 fois pour le cas inverse (passages des sols raides à des sols mous). Cette problématique sur les zones de transition a été aussi abordée par le biais des analyses numériques.

(Ahmadi and Robertson (2005)), montrent comment, en fonction de la résistance et de la pression de confinement du sol, l'influence d'une limite entre couches se fait sentir à une distance jusqu'à 15 fois le diamètre de la pointe.

Plus récemment, (Salgado (2012)) a présenté une synthèse des travaux expérimentaux réalisés en laboratoire dans des chambres de calibration et en centrifugeuses estimant le champ de déplacement autour de la pointe grâce à des techniques de corrélation digitale des images. Les résultats expérimentaux utilisant un sable siliceux montrent des zones d'influence comprises entre 2 et 4 fois le diamètre de la pointe.

ii. Règles statistiques

En ce qui concerne la taille de la fenêtre permettant de respecter les lois de la statistique, (Vivitrat (1978)) suggère que la couche de sol doit contenir au moins 20 points pour être statistiquement significative.

(Lacasse and Nadim (1997)) préconisent une longueur minimale en termes de nombre d'effectifs, d'au moins 10 afin que la variance estimée soit représentative. Cette dimension minimale, a été aussi retenue dans le travail de (Phoon et al. (2003)).

D'une manière plus concise, on expose par la suite, différentes valeurs des longueurs de fenêtre préconisées dans la littérature :

- (Webster (1973)) suggère de choisir une longueur de fenêtre égale à 2/3 de la distance attendue entre les limites de couches ;
- (Wickremesinghe (1989)) réalise une étude paramétrique pour différentes tailles de fenêtre vis-à-vis de l'essai CPT : 0.5, 1.5 et 2 mètres. Suite aux analyses réalisées le choix d'une longueur inférieure à 1 mètre ne semblait pas pertinent pour les profils analysés ;
- (Zhang and Tumay (1996)) a considéré une taille de fenêtre égale à 1.50 m.
- (Phoon et al. (2003)) utilisent une fenêtre de 5 mètres de hauteur totale pour l'analyse des essais CPT menés à 30 mètres de profondeur. Cependant, pour tester l'hypothèse de stationnarité au sein d'une même couche, la hauteur de la fenêtre de calcul est la même que la valeur de l'échelle de fluctuation estimée au sein de la formation ;
- (Uzielli (2008)) garde une longueur de fenêtre égale à 0.25 m de hauteur pour calculer les moments statistiques des différents paramètres le long des différents sondages.
- (Ching et al. (2015a)) testent trois hauteurs de fenêtre vis-à-vis de la méthode T ratio : 1, 2 et 5 mètres. Une fenêtre égale à 1 m permet l'identification des couches fines du sol mais elle détecte aussi de « fausses couches ». Ce problème est résolu pour les fenêtres de plus grandes tailles (2m et 5m) alors que la capacité à détecter des couches de sols de faible épaisseur est fortement réduite.

En ce qui concerne les valeurs seuils des paramètres des tests d'évaluation de l'homogénéité statistique, dans le tableau (Tab.7.3), on a synthétisé quelques valeurs seuils empiriques proposées dans la littérature pour les méthodes T ratio et Intra-classe. Par rapport aux valeurs limites proposées par (Wickremesinghe (1989)) pour la statistique ρ_I , on remarque que :

- lorsque $\rho_I \geq 0.8$, il s'agit d'une limite entre couches primaires ;
- lorsque $0.8 > \rho_I \geq 0.65$, il s'agit d'une limite entre sous-couches.

TABLE 7.3 – Valeurs seuils des paramètres des tests d'évaluation proposées dans la littérature

Méthode	Seuil	Références
<i>T ratio</i>	10	Wickremesinghe (1989) et (Ching et al. (2015a))
<i>Intra-classe</i>	0.65	Wickremesinghe (1989)
<i>Intra-classe</i>	0.8	Wickremesinghe (1989)
<i>Intra-classe</i>	0.7	Zhang and Tumay (1996) et Phoon et al. (2003)

Bilan

Dans le contexte de notre application, celle de la reconnaissance de sols en surface à l'aide de l'essai Panda, on retiendra les choix suivants :

i. Zone d'influence

La taille de la pointe utilisée pour la reconnaissance in situ fait 2.2 cm de diamètre (4cm^2 de section). La zone de sol influencée par la pénétration de la pointe aurait une épaisseur maximale d'environ 30cm d'après (Ahmadi and Robertson (2005)) ou serait de l'ordre de 5 à 10 cm si l'on suit les prescriptions de (Salgado (2012)).

ii. Taille de la fenêtre

Le pénétromètre Panda présente un pas d'échantillonnage variable, normalement entre 2 et 10 mm. Pour le pas de régularisation proposé dans cette étude, 10 mm, afin de tester des échantillons contenant au moins 10 points de mesures (Lacasse and Nadim (1997)), une taille minimale de fenêtre égale à 20 cm est donc retenue.

Vis-à-vis de l'analyse des sondages de pénétration in situ, des fenêtres d'une taille inférieure à 1 m ont été déconseillées (Wickremesinghe (1989)). Une fenêtre avec une longueur entre 1 et 2m semble être un choix pertinent pour la profondeur d'étude maximale fixée dans ce travail, 5 mètres.

iii. Valeurs seuils

Les valeurs seuil retenues dans notre étude seront celles reportées dans la littérature, à savoir :

- *T ratio* : seuil égal à 10
- *Intraclasse* : seuil égal à 0.7
- *Bartlett* : l'identification des interfaces entre couches doit être réalisée manuellement à partir de l'inspection du profil de la statistique du test réalisé.

8. Démarche proposée

8.1 Démarche

Pour mener cette étude sur la reconnaissance des couches et le découpage automatique du pénétrogramme, on s'est basé sur trois campagnes d'essais Panda. La première est constituée par des essais réalisés dans des moules de calibration en laboratoire avec des pointes de pénétromètre de différentes tailles. La deuxième est constituée de 12 sondages in-situ réalisés dans le cadre des missions de reconnaissance géotechnique de la société Sol Solution. Ils s'agit des essais analysés par les neuf experts. Puis finalement, on a testé la technique sur des essais Panda réalisés sur le site expérimental de Castelló d'Empúries (Gérome, Espagne). Pour ce terrain on dispose d'une caractérisation géotechnique préalable du site au travers de deux essais CPTU mais aussi grâce aux échantillons prélevés lors d'un sondage carotté caractérisé en laboratoire. Sur ces essais, la démarche proposée est la suivante :

- i. Premièrement, les trois tests statistiques par fenêtre glissante seront évalués et critiqués sur les essais de laboratoire en moules de calibration réalisés avec des sols « modèles » et caractérisés selon un protocole de mise en place connu. On s'intéressera notamment à évaluer l'influence de la taille de la fenêtre, l'effet des différents pas de régularisation du signal, la sensibilité des résultats obtenus selon que la méthode est appliquée sur les mesures brutes de q_d ou sur sa transformée logarithmique et l'effet de la taille de la pointe ;
- ii. Deuxièmement, on étudiera si les méthodes mises en place en laboratoire sont opérationnelles sur des essais Panda in situ. Pour ce faire :
 - on testera les méthodes sur les essais réalisés sur le site expérimental de Castelló d'Empúries, afin d'évaluer leurs efficacités vis-à-vis de l'identification des matériaux fins, des couches de différentes épaisseurs et de l'influence de la longueur de la fenêtre ;
 - ces méthodes seront aussi testées sur les sondages in situ analysés par les neuf experts. Les couches identifiées à l'aide des tests paramétriques utilisant différentes tailles de fenêtre seront comparées aux coupes lithologiques du terrain déduites par les géologues. Enfin, la répétabilité des résultats fournis par les méthodes statistiques sera comparée à celle des analyses expertes.
- iii. Finalement, on comparera les stratigraphies issues des différentes méthodes utilisées pour l'essai CPT à celles déterminées grâce aux approches développées pour l'essai Panda. Le but est d'étudier quelle est la complémentarité et quelles sont les différences entre les approches développées pour l'essai statique et dynamique.

8.2 Étude sur des essais au laboratoire

8.2.1 Protocole expérimental

Moules de calibration

Afin de tester les différentes méthodes proposées, nous nous sommes basés sur la confection d'échantillons en moule de calibration. Ces moules ont un diamètre de 400 mm et une hauteur maximale de 1600 mm. Un total de quatre matériaux granulaires ont été employés pour élaborer quatre éprouvettes :

- sable grossier (Sg)
- ballast (B)
- microballast (M)
- sable de Dunkerque (Sd)

L'ordre des matériaux a été modifié pour chaque module comme le montre la figure (Fig.8.1). Les couches de matériaux compactés ont une épaisseur constante égale à 0.27m. Chaque matériau a été mis en place par compactage (compactage dynamique de type Marshall réalisé avec une masse de 20 kg tombant d'une hauteur 60 cm sur une plaque de 396 mm de diamètre). Pour chaque moule, trois essais pénétrométriques Panda ont été effectués : un premier sondage S1, avec une pointe de 2 cm² fixe, le deuxième et le troisième avec des pointes perdues de 2 cm² et 4 cm² de section. Lors des essais, les tiges des essais précédents ont été laissées en place afin de ne pas perturber le milieu et l'agencement granulaire qui s'est formé autour de la tige. On précise qu'on a privilégié dans notre étude les essais réalisés en pointe de 4 cm², puisqu'il s'agit du type de pointe utilisée pour la reconnaissance de sol. Les essais en pointe de 2 cm² fixe sont utilisés dans le cadre du contrôle de compactage des remblais (NF XP P 94-105 (2000)) de moins de 1.20 m de profondeur, mais aussi pour l'auscultation des voies ferrées.

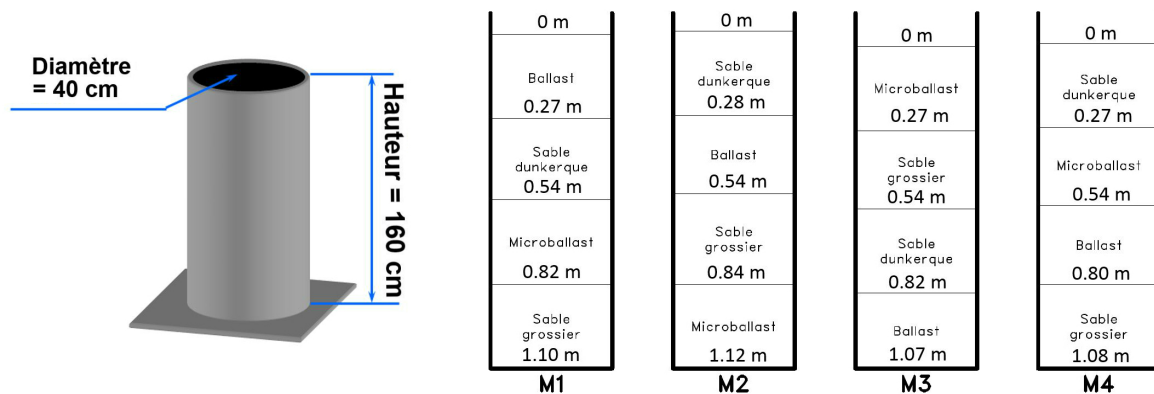


FIGURE 8.1 – Répartition des matériaux pour les éprouvettes confectionnées

Les caractéristiques physiques et d'état des couches mises en place sont données dans la tableau (Tab.8.1), notamment le poids volumique γ , le poids volumique sec γ_d et la teneur en eau $w(\%)$. L'acronyme *HET* signifie moule hétérogène. Les courbes granulométriques des matériaux sont données (cf. Annexe A).

TABLE 8.1 – Caractéristiques physiques des matériaux compactés

HET 1				HET 2			
Matériau	$\gamma(KN/m^3)$	$\gamma_d(KN/m^3)$	w (%)	Matériau	$\gamma(KN/m^3)$	$\gamma_d(KN/m^3)$	w (%)
B	14.6	14.6	0.5	Sd	15.1	14.5	4
Sd	15.7	15.1	4	B	15.1	15.1	0.5
M	14.8	14.8	0.5	Sg	14.2	14.1	1
Sg	14.8	14.7	1	M	14.6	14.6	0.5
HET 3				HET 4			
Matériau	$\gamma(KN/m^3)$	$\gamma_d(KN/m^3)$	w (%)	Matériau	$\gamma(KN/m^3)$	$\gamma_d(KN/m^3)$	w (%)
M	14.6	14.6	0.5	Sd	15.3	14.7	4
Sg	16.1	16	1	M	15.4	15.4	0.5
Sd	15	14.3	4	B	15.4	15.4	0.5
B	15.6	15.6	0.5	Sg	15	14.9	1

TABLE 8.2 – Caractéristiques granulométriques des matériaux

	Sable de Dunkerque	Sable grossier	Microballast	Ballast
D_{50} (mm)	0.21	1.4	6	38.9
C_C	0.94	1.2	0.84	1.06
C_U	2.81	2.13	1.87	1.64

On regroupe dans le tableau (Tab.8.2) les caractéristiques granulométriques des matériaux testés. On note par C_U le facteur d'uniformité de Hazen (d_{60}/d_{10}) et par C_C le facteur de courbure ($d_{30}^2/d_{10} \cdot d_{60}$).

Analyse visuelle des pénétrogrammes obtenus

L'ensemble des pénétrogrammes bruts obtenus, est présenté dans la figure (Fig.8.2) ci-après. On a représenté par des lignes en noir les interfaces physiques entre les différentes couches.

En général, on observe des résistances de pointe plus élevées pour les sondages réalisés en pointe de 2 cm² perdue (pénétrogrammes en bleu) pour les matériaux plus grossiers, ballast et microballast. Cependant les sondages en pointe de 4 cm² (pénétrogrammes en rouge) semblent présenter des résistances plus fortes pour la couche de sable grossier dans les moules HET1 et HET4.

Visuellement on constate des contrastes sur les allures des pénétrogrammes des diverses couches mises en place. Cependant la transition entre les deux sables dans le moule n°3 ne semble pas être identifiable à l'œil nu.

8.2.2 Analyse statistique des pénétrogrammes

Les deux premiers moments statistiques des mesures de la résistance de pointe q_d sont analysés ici. Tout d'abord on présentera les résultats en fonction du type de matériau. Ensuite on étudiera l'influence du diamètre de la pointe et de la granulométrie du matériau.

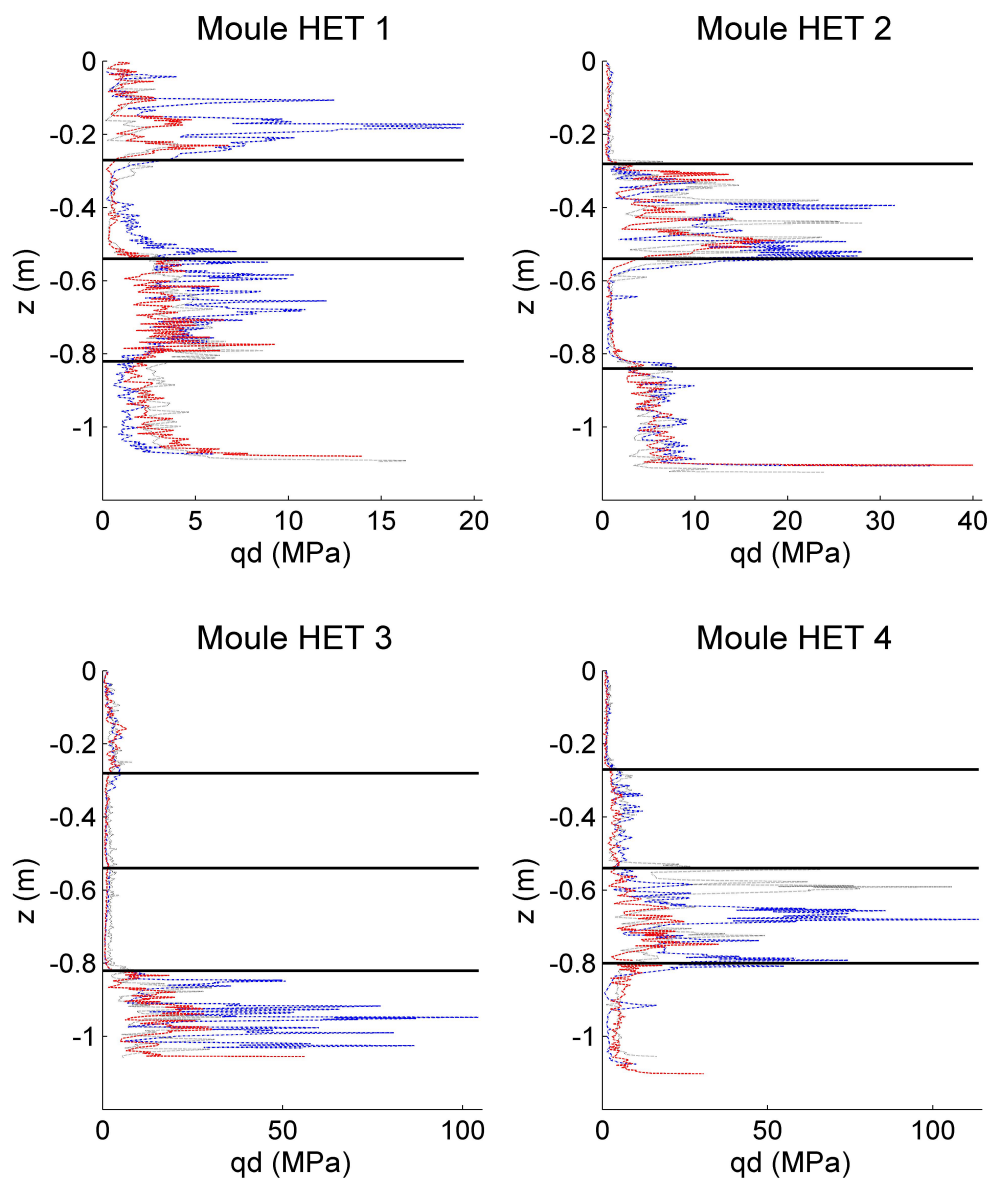


FIGURE 8.2 – Pénétrogrammes bruts obtenus dans les moules en laboratoire. *Noir*- pointe 2 cm² de section fixe, *bleu*- pointe 2 cm² de section perdue et *rouge*- pointe 4 cm² de section perdue.

TABLE 8.3 – Valeur moyenne de q_d (MPa) des sondages réalisés en pointe de 4 cm^2

Matériaux	Moules			
	HET 1	HET 2	HET 3	HET 4
Ballast	2.15	7.29	13.63	11.91
Microballast	3.45	4.81	2.31	3.54
Sable grossier	2.89	1.07	0.97	5.69
Sable de Dunkerque	0.7	0.58	1.05	1.18

TABLE 8.4 – Coefficient de variation de q_d des sondages réalisés en pointe de 4 cm^2

Matériaux	Moules			
	HET 1	HET 2	HET 3	HET 4
Ballast	0.63	0.62	0.67	0.58
Microballast	0.41	0.34	0.61	0.3
Sable grossier	0.54	0.46	0.36	0.46
Sable de Dunkerque	0.63	0.27	0.25	0.34

On précise que sur la base des tests réalisés, il semblerait ne pas être possible d'estimer la zone d'influence de la pointe lors de la réalisation de l'essai de pénétration. L'influence des zones de transition n'a donc pas été prise en compte lors du calcul des statistiques présentées ici et ceci a pu introduire un biais.

Moments statistiques en fonction du matériau

Les analyses présentées ici sont celles relatives aux sondages effectués en pointe de 4 cm^2 . Les valeurs moyennes de q_d sont synthétisées dans le tableau (Tab.8.3). À l'exception du premier moule (HET1), le ballast est le matériau qui présente les valeurs moyennes de résistance les plus fortes. Au contraire, les valeurs moyennes les plus faibles ont été constatées pour le sable de Dunkerque, à l'exception du moule HET3. En effet, dans le moule n°3, les deux sables montrent des résistances moyennes très similaires. En outre, on peut noter que les différentes couches de matériaux présentent des valeurs moyennes de q_d différentes, à l'exception de deux sables dans le moule HET3.

En outre, la dispersion relative des pénétrogrammes est analysée sous la forme de coefficients de variation CV en fonction des différents matériaux (Tab.8.4). Les courbes de pénétration du matériau le plus grossier, le ballast, sont celles qui présentent les valeurs de coefficient de variation les plus fortes, avec une valeur assez constante de 0.62. Le sable de Dunkerque, dont la fraction grossière est moins importante, présente toutefois les valeurs de CV les plus faibles avec une valeur autour de 0.28, à l'exception du moule (HET1) dont la valeur du CV est très forte. Ceci pourrait être expliqué par l'effet de la zone d'influence. On peut observer dans la figure (Fig.8.2) pour le moule HET1, comment la résistance commence à augmenter au-dessus de l'interface entre le sable et le microballast en entraînant une perturbation sur la valeur du CV calculée. Sinon on note une grande dispersion des valeurs du CV du microballast entre 0.3 et 0.61.

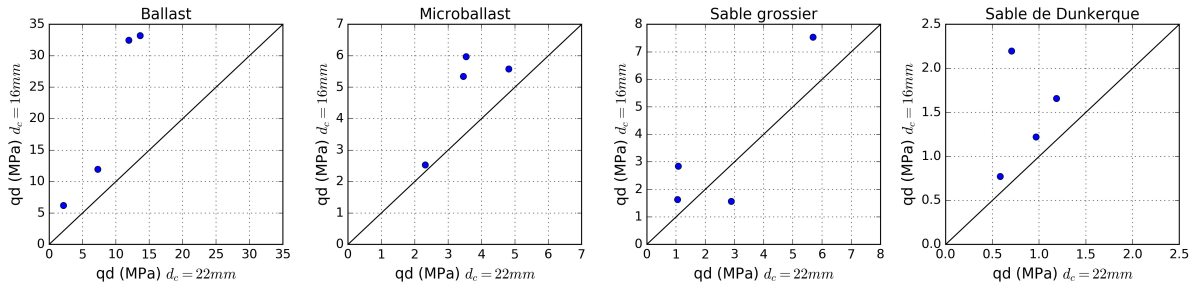


FIGURE 8.3 – Comparaison du q_d moyen en fonction de la taille de la pointe (d_c : diamètre de la pointe)

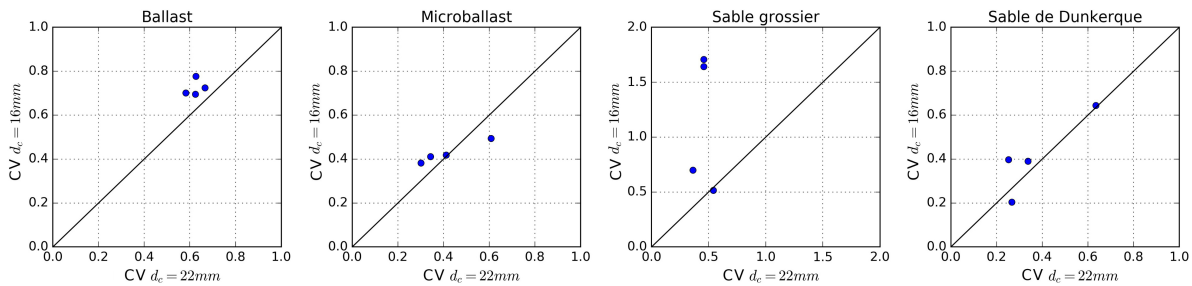


FIGURE 8.4 – Comparaison du CV de q_d en fonction de la taille de la pointe (d_c : diamètre de la pointe)

Effet de l'influence de la taille de la pointe

Pour le protocole au laboratoire, on dispose d'essais réalisés avec deux tailles de pointe différentes et des moules de calibration confectionnés avec des matériaux possédant des distributions granulométriques différentes. On étudie ici l'influence de ces deux facteurs vis-à-vis des moments statistiques des pénétrogrammes obtenus.

Tout d'abord, on évalue l'effet de la taille de la pointe sur la valeur moyenne de la résistance q_d . Les résultats, en fonction des matériaux, sont donnés dans la figure (Fig.8.3). Sur l'axe des abscisses on a représenté les q_d moyens des sondages réalisés en pointe de 4cm^2 , dont le diamètre de la pointe (d_c) est égal à 22 mm et sur l'axe des ordonnées les valeurs moyennes de q_d des sondages réalisés en pointe de 2cm^2 , dont le diamètre de la pointe (d_c) est égal à 16 mm. Il est alors clair que les sondages effectués avec la pointe plus petite exhibent une résistance moyenne légèrement plus forte indépendamment de la granulométrie du matériau mis en place. Les plus grands écarts ont été observés pour le ballast.

Ensuite on examine les valeurs du coefficient de variation en fonction de la taille de la pointe utilisée. Une comparaison pour les quatre matériaux testés est présentée dans la figure (Fig.8.4). On constate que les valeurs du CV sont similaires indépendamment de la taille de la pointe, à l'exception du sable grossier. En effet, les CV pour les essais réalisés en pointe de 2cm^2 au sein de la couche de sable grossier dans les moules HET2 et HET4 sont beaucoup plus forts.

Mat.	d_c (mm)	d_{50} (mm)	n_p
B	16	38.9	0.4
	22	38.9	0.6
M	16	6	2.67
	22	6	3.67
Sg	16	1.4	11.4
	22	1.4	15.7
Sd	16	0.21	76.2
	22	0.21	104.8

TABLE 8.5 – Valeurs du n_p pour les matériaux utilisés

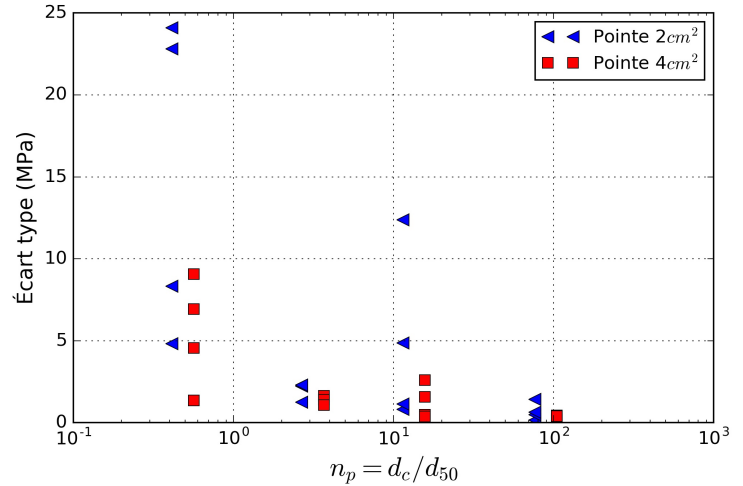


FIGURE 8.5 – Relation entre le rapport non dimensionnel n_p et la dispersion du pénétrogramme

Effet de l'influence de la granulométrie

Pour évaluer l'influence de la granulométrie du matériau de manière conjointe avec le diamètre de la pointe utilisée on calcule le rapport adimensionnel (n_p) :

$$n_p = \frac{d_c}{D_{50}} \quad (8.1)$$

avec d_c le diamètre de la pointe du pénétromètre et D_{50} le diamètre moyen des particules. Suite aux essais de pénétration menés sur des matériaux granulaires dans une chambre de calibration virtuelle (VVC) utilisant des modèles 3D basés sur la méthode des éléments discrets (DEM), (Arroyo et al. (2011)) suggèrent que la dispersion des signaux pénétrométriques peut être expliquée par cet indice. Quand la valeur du rapport (n_p) augmente, l'effet d'échelle du diamètre de la pointe devient plus important et les pénétrogrammes ont une allure plus lissée.

Les valeurs du rapport n_p calculé pour chaque matériau en fonction de la taille de la pointe utilisée sont données dans le tableau (Tab.8.5). Dans la figure (Fig.8.5) on a représenté l'écart-type de q_d pour chacune des couches des matériaux testés en fonction de la valeur du rapport n_p . On voit que :

- les pénétrogrammes réalisés avec une pointe de 2cm^2 présentent une dispersion plus élevée que ceux réalisés avec une pointe de 4cm^2 de section. Ceci confirme l'effet de moyenne qu'entraîne la taille de la pointe. Le volume du sol sollicité lors de l'essai augmente avec la taille de la pointe conique et les mesures deviennent plus lissées ;
- la dispersion du pénétrogramme augmente avec la taille des éléments du matériau. On observe ainsi que l'écart-type des pénétrogrammes pour les matériaux grossiers est plus élevé par rapport aux matériaux fins. Cependant, les pénétrogrammes issus des couches constituées par le microballast ne suivent pas cette tendance.

Bilan

Pour chacune des quatre éprouvettes réalisées, on a observé que les deux premiers moments statistiques de q_d au sein des couches de différents matériaux granulaires ont des valeurs différentes. Par conséquent, l'application des méthodes statistiques par fenêtre glissante devrait fournir de bons résultats. En outre, la taille de la pointe influence la valeur de ces deux moments. En effet, les résistances de pointe q_d mesurées sont plus importantes pour les essais effectués avec une pointe de 2cm^2 de section. L'écart type augmente aussi et les matériaux plus grossiers entraînent en général plus de dispersion sur le signal pénétrométrique. La valeur du coefficient de variation est par contre très peu affectée par le diamètre de la pointe.

8.3 Étude sur les tests statistiques par fenêtre glissante

8.3.1 Influence du pas de régularisation

Régularisation

Par définition, le Panda est un pénétromètre dynamique à énergie variable. C'est pourquoi lors d'un essai, les enfoncements enregistrés à chaque coup de marteau varient. En conséquence, les mesures de résistance ne sont pas uniformément réparties sur la profondeur du sondage. Pour réaliser une étude statistique sur les valeurs de q_d , il est nécessaire de disposer d'une variable équirépartie sur toute la longueur du sondage. Pour ce faire, on peut calculer une valeur de q_d régularisée q_{dreg} sur une fenêtre glissante de hauteur constante. On pondère les mesures de résistances de pointe contenues dans la fenêtre par leur enfoncement respectif afin d'éviter de biaiser le calcul de q_d moyen dans la fenêtre mobile :

$$q_{dreg} = \frac{\sum q_{di} \cdot e_i}{\sum e_i} \quad (8.2)$$

avec q_{di} les mesures de résistance se trouvant dans la fenêtre et e_i les enfoncements mesurés.

Ce traitement est appelé *régularisation* (Deplagne (1991)) quand la hauteur de la fenêtre est égale au pas d'avancement de la fenêtre, appelé pas de régularisation.

Influence du pas de régularisation sur les deux premiers moments statistiques de q_d

Tout d'abord on étudie l'influence du pas de régularisation des trois tests statistiques préconisés, à savoir T ratio, Intra-classe et Bartlett. Dans la figure (Fig.8.6), on représente la distribution des enfoncements mesurés lors du battage pendant la réalisation des essais. On peut constater que leur étendue varie principalement entre 1 et 10 mm. De même, les valeurs de pénétration moyenne sont très faibles (2, 3 ou 4 mm). L'influence du pas de régularisation vis-à-vis des deux premiers moments statistiques de q_d est examiné par la suite.

Premièrement, on montre la valeur de la moyenne de q_d en fonction du pas de régularisation retenu pour les différents matériaux dans chaque moule (Fig.8.7). En général la valeur moyenne ne varie pas en fonction du pas de régularisation utilisé (noté ξ) à l'exception des couches réalisées avec du ballast. Toutefois, l'analyse de l'écart-type de q_d en fonction du pas

de régularisation utilisé (Fig.8.8), montre que la dispersion des données pénétrométriques est plus influencée. D'après cette figure, il devient clair que cet effet de lissage augmente avec l'importance du pas de régularisation par rapport à l'enfoncement moyen. De même, on constate de manière générale que la chute de la valeur de ce paramètre de dispersion est plus forte pour le ballast, à l'exception des couches de matériaux mises en oeuvre dans le moule HET1.

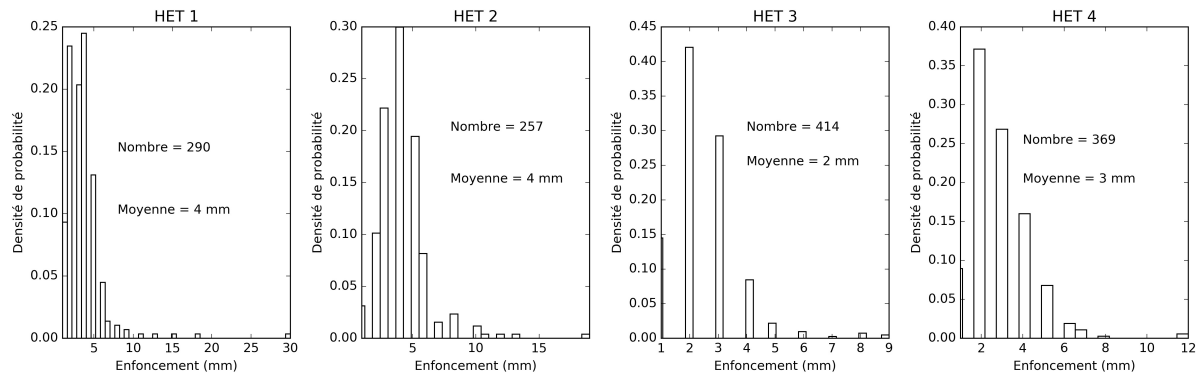


FIGURE 8.6 – Histogramme des enfoncements mesurés (pointe de 4 cm^2).

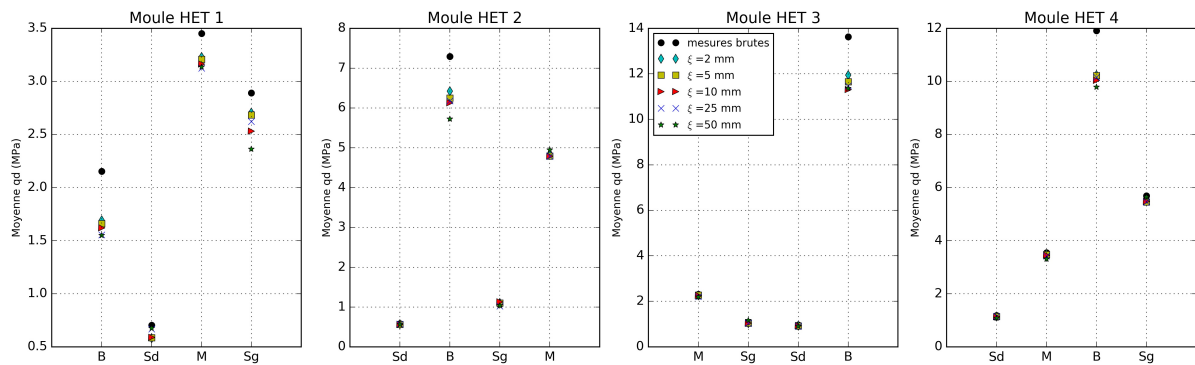


FIGURE 8.7 – Influence du pas de régularisation ξ (pointe de 4 cm^2). Notation : B - ballast, M - microballast, Sg - sable grossier et Sd - sable de Dunkerque.

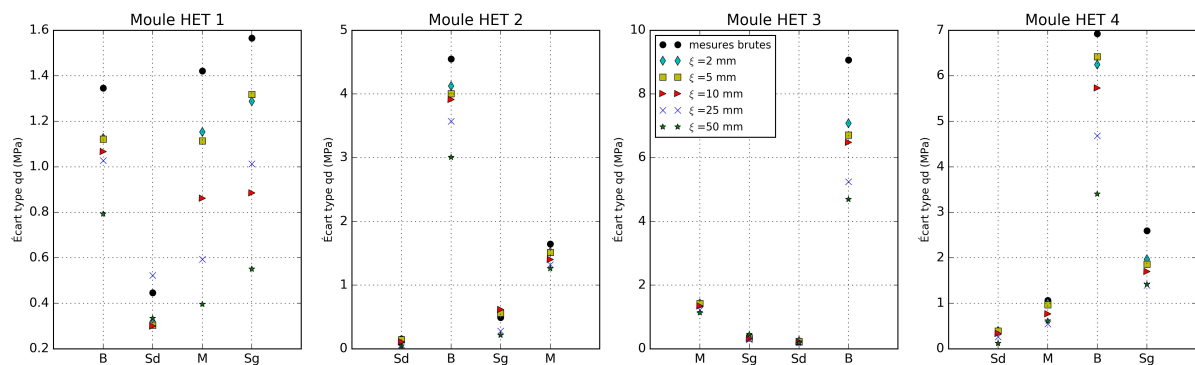


FIGURE 8.8 – Représentation des pénétrogrammes pour différentes valeurs de régularisation ξ (pointe de 4 cm^2). Notation : B - ballast, M - microballast, Sg - sable grossier et Sd - sable de Dunkerque.

Pour les analyses concernant l'influence du pas de régularisation, présentées par la suite, on retiendra une taille de fenêtre égale à 0.40 m en référence à (Webster (1973)), qui recommande une longueur de la demi-fenêtre égale à 2/3 de l'épaisseur des couches mise en œuvre (0.27 m). Toutefois l'influence de la hauteur de la fenêtre sera étudiée dans la section suivante.

Méthode T ratio

La figure (Fig.8.9) montre l'influence des différents pas de régularisation vis-à-vis du test T ratio. La valeur seuil pour le test de T ratio est aussi affichée dans les graphiques. On peut remarquer que les valeurs du profil T ratio calculé deviennent plus faibles dès que le pas de régularisation augmente. En effet, les maxima locaux du profil statistique censés signaler la position des interfaces entre couches, sont moins « proéminents » dès que le pas de régularisation utilisé s'écarte de la valeur de l'enfoncement moyen. L'interprétation du profil et l'identification des maxima locaux semblent être plus claires pour les mesures brutes ou lorsque le pas de régularisation a une valeur plus faible ($\xi = 2mm$ ou $\xi = 5mm$) et donc proche de la valeur de l'enfoncement moyen.

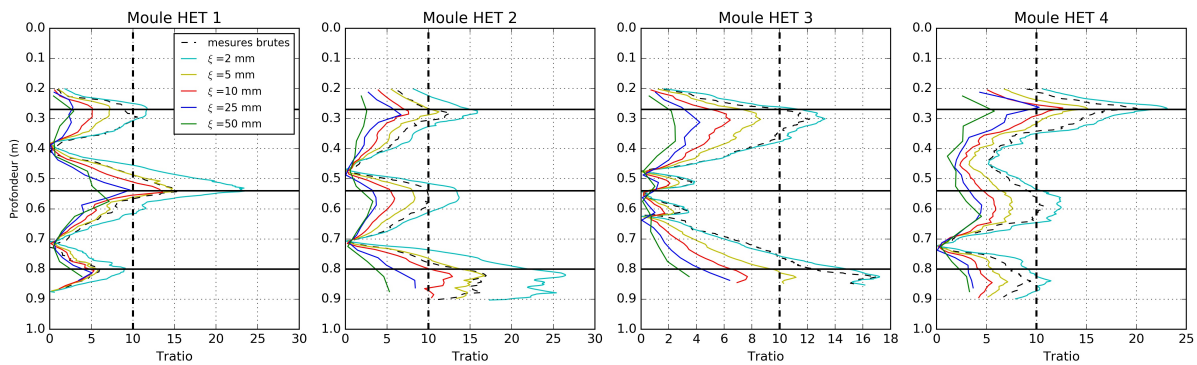


FIGURE 8.9 – Test T ratio : influence du pas de régularisation (ξ) (pointe de 4 cm^2)

Méthode Intra-classe

Les résultats des tests réalisés pour évaluer l'influence du pas de régularisation sont présentés dans la figure (Fig.8.10). La ligne pointillée en "tirets" représente la valeur seuil égale à 0.7 suggérée par (Zhang and Tumay (1996)). Les valeurs seuils suggérées par (Wickremesinghe (1989)) et égales à 0.65 et 0.80 sont représentées par des lignes pointillées en "points et tirets".

A contrario du test T ratio, très sensible au pas de régularisation, on ne constate pas d'écarts aussi importants entre les courbes ρ_I pour des pas de régularisation égaux ou inférieurs à 25mm. Les maxima locaux du coefficient intra-classe atteignent parfois des valeurs plus élevées pour des pas de régularisation plus importants. Ceci est observé par exemple dans les moules HET2 et HET4 pour la première interface entre couches situées plus en surface, où la valeur maximale de ρ_I est atteinte pour un pas de régularisation égal à 50mm. Cependant, le comportement du test T ratio qui présentait des valeurs plus faibles lorsque le pas de régularisation augmentait par rapport à l'enfoncement moyen n'a pas été constaté ici.

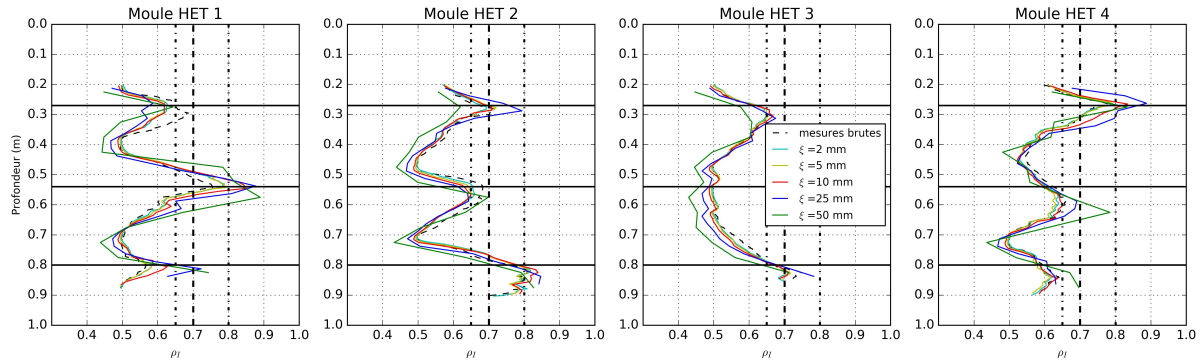


FIGURE 8.10 – Test intra-classe (ρ_I) : influence du pas de régularisation (ξ) (pointe de 4 cm^2)

Méthode Bartlett

Pour appliquer ce test statistique il est nécessaire de compter sur un pas d'échantillonnage constant. La régularisation est donc une étape nécessaire. Les profils statistiques issus du test de Bartlett (B_{stat}) pour différents pas de régularisation sont regroupés dans la figure (Fig.8.11). On signale que les pics ou maxima locaux, censés indiquer la position de la limite entre couches, présentent des écarts très importants en fonction du pas de régularisation choisi. L'identification visuelle des interfaces devient plus claire pour les pas de régularisation les plus faibles 2 ou 5mm, ceux les plus proches de l'enfoncement moyen. Quand le pas utilisé est supérieur à 25mm, la valeur des maxima locaux devient très faible et il ne parait pas possible de décélérer visuellement les interfaces entre les couches.

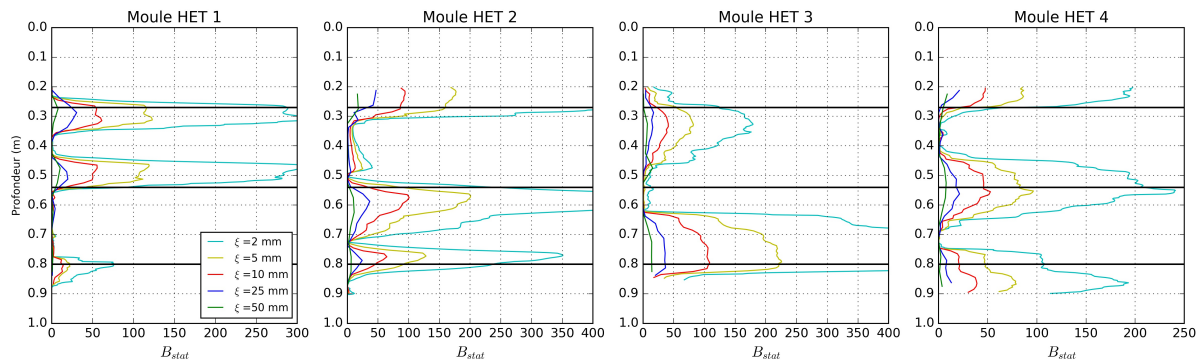


FIGURE 8.11 – Test de Bartlett (B_{stat}) : influence du pas de régularisation (ξ) (pointe de 4 cm^2)

Conclusion

Dans le contexte d'une détection de couches de faible épaisseur constituées par des matériaux granulaires homogènes, un effet de lissage et d'affaiblissement des valeurs a été constaté vis-à-vis du test T ratio et Bartlett dès que le pas de régularisation augmente par rapport à l'enfoncement moyen. Ce comportement n'a pas été observé pour le test Intra-classe et aucune règle générale concernant le pas de régularisation n'a pu être déduite pour cette méthode.

Par la suite, les pénétrogrammes régularisés avec un pas égal à la valeur de l'enfoncement moyen seront analysés.

8.3.2 Influence de la taille de la fenêtre

On s'intéresse à évaluer l'influence de la taille de la fenêtre vis-à-vis des tests statistiques étudiés. Une des contraintes des approches par fenêtre glissante est la détection des interfaces entre couches aux extrémités du profil des mesures à analyser. En effet, la première valeur calculée par le biais du test statistique est située à une distance du début du profil égale à la demie-hauteur de la fenêtre utilisée. Le dernier point calculé présentera le même écartement par rapport au dernier point du profil. Compte tenu que la première interface à identifier est située à environ 0.27m de profondeur, une taille de fenêtre maximale égale à 0.5m a été retenue.

Méthode T ratio

Les profils statistiques calculés à l'aide de la méthode T ratio pour différentes longueurs de fenêtre (0.2, 0.3, 0.4 et 0.5) sont présentés dans la figure (Fig.8.12). On constate un effet de lissage en fonction de la taille de la fenêtre utilisée. Les maxima locaux du profil calculé avec une fenêtre de 0.2 m sont inférieurs à la valeur seuil égale à 10 que l'on s'est fixée. Par contre, pour les autres tailles testées les pics du profil statistique sont bien situés au delà de cette valeur limite et les interfaces sont donc bien identifiées. Toutefois, on observe que les statistiques calculées ont des valeurs supérieures quand on augmente la taille de la fenêtre. Les tests réalisés avec une fenêtre de 0.4 ou 0.5m fournissent des profils statistiques avec des allures similaires et des courbes peu bruitées, permettant une identification simple des maxima locaux dont les positions sont les mêmes ou sont très proches de celles des interfaces physiques à détecter. Par contre, les pics correspondant aux premières et dernières limites sont clairs avec une fenêtre de 0.4 m, qui semble donc être optimale.

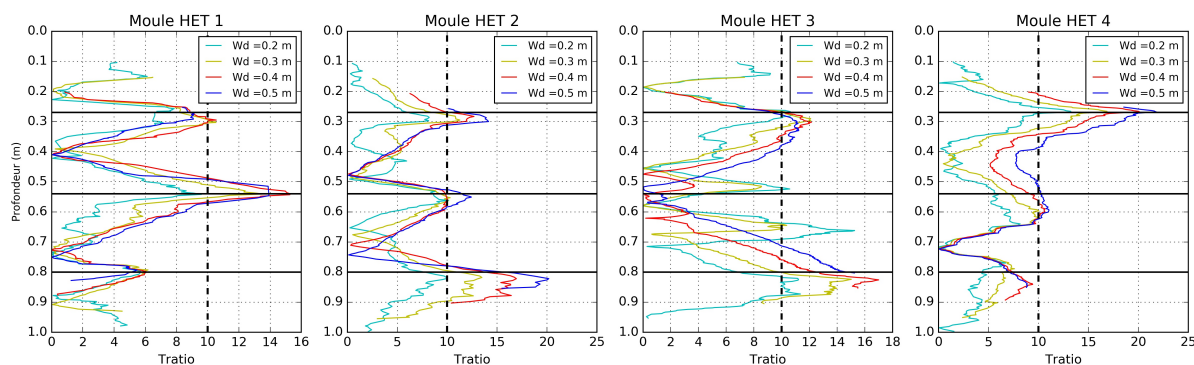


FIGURE 8.12 – T ratio : influence de la longueur (W_d) de la fenêtre glissante (pointe de 4 cm^2)

Méthode Intra-classe

Dans la figure (Fig.8.13) on regroupe les différents profils statistiques calculés pour les différentes tailles de fenêtres testées. On constate des résultats très similaires à ceux obtenus

par la méthode T ratio, avec des allures de courbes très semblables entre les deux tests. Cependant les courbes statistiques issues de la méthode Intra-classe semblent être plus proches quelle que soit la taille de la fenêtre utilisée. Cette méthode semble donc moins influencée par la taille de la fenêtre. Elles sont aussi plus lissées pour la taille de fenêtre choisie. Pour des raisons similaires au cas précédent, on retiendra une fenêtre de 0.4m pour les analyses futures.

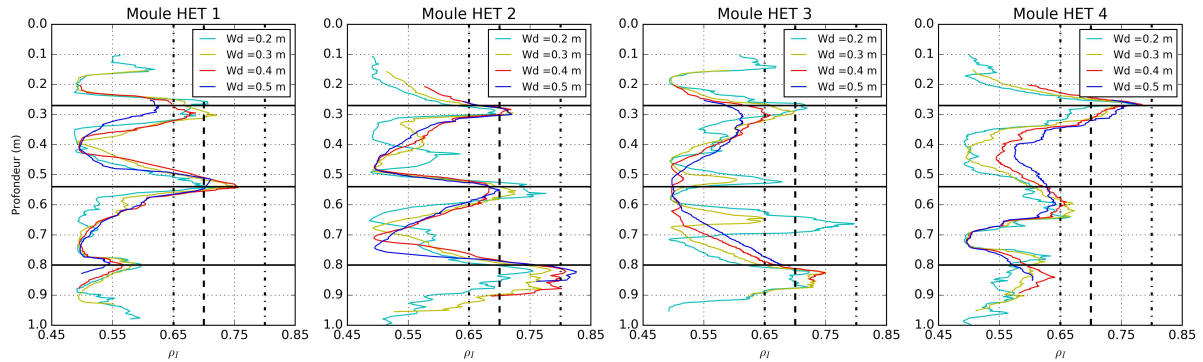


FIGURE 8.13 – Intra-classe : influence de la longueur de la fenêtre (W_d) glissante (pointe de 4 cm^2)

Méthode Bartlett

Les courbes de la statistique de Bartlett pour les 4 tailles de fenêtre étudiées sont données dans la figure (Fig.8.14). Des résultats très similaires aux deux tests précédents sont observés. En outre, on repère une augmentation systématique de la valeur du profil statistique avec la taille de la fenêtre utilisée uniquement pour les essais réalisés dans les moules HET3 et HET4. Enfin, pour une hauteur de fenêtre de 0.4 m, les pics de la courbe qui indiquent la profondeur des limites entre couches deviennent plus prononcés et le repérage des interfaces situées en extrémité du sondage (surface et fond) est facilité.

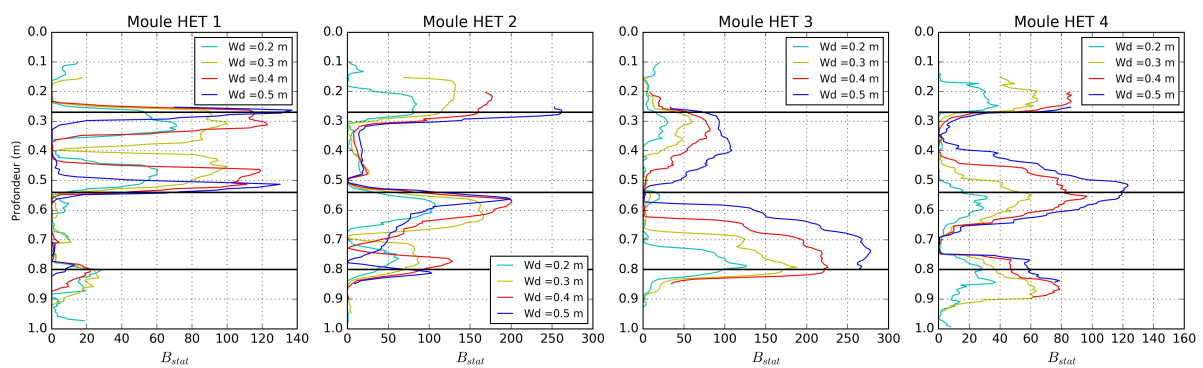


FIGURE 8.14 – Bartlett : influence de la longueur (W_d) de la fenêtre glissante (pointe de 4 cm^2)

Conclusion

Pour tous les tests, une longueur de fenêtre comprise entre 0.4 et 0.5m permet d'obtenir des profils statistiques moins bruités. Cependant, une hauteur égale à 0.4m a été retenue

comme taille optimale afin de pouvoir mieux analyser les limites des couches aux extrémités du moule.

8.3.3 Influence de la transformation logarithmique

Finalement, on évalue l'effet de travailler en échelle logarithmique $\ln(q_d)$ en vertu de l'hypothèse de normalité. La comparaison est faite sur les mesures brutes (aucune régularisation) avec une fenêtre de 0.40 m.

Méthode T ratio

Les profils statistiques issus des méthodes T ratio sont présentés dans la figure (Fig.8.15). On observe que le profil du paramètre T ratio présente des valeurs supérieures pour le logarithme de la résistance de pointe q_d , avec des maximaux locaux plus marqués. L'application d'une transformation logarithmique au profil de mesures q_d est donc favorable vis-à-vis de l'application de ce test paramétrique.

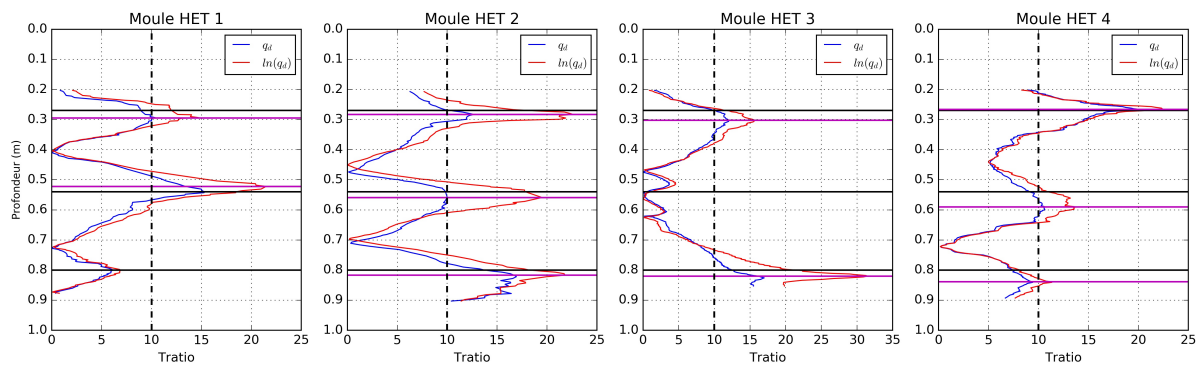


FIGURE 8.15 – T ratio : influence de la transformation logarithmique des valeurs de q_d (pointe de 4 cm^2)

Dans la figure (Fig.8.15) on a aussi représenté les limites entre couches (lignes en magenta) déterminées par le biais de la méthode T ratio pour la variable $\ln(q_d)$ et pour la valeur seuil retenue égale à 10. On constate que toutes les limites ont été détectées correctement, à l'exception de :

- la transition entre le microballast et le sable grossier dans le moule HET1. Le profil du paramètre T ratio présente un maximum local à cette profondeur mais sa valeur reste inférieure à 10 ;
- la transition entre le sable grossier et le sable de Dunkerque dans le moule HET3. L'application de cette technique au profil de résistance mesuré n'est pas sensible à cette interface puisque les moyennes des résistances mesurées au sein de ces deux couches confectionnées avec des matériaux sableux ont des valeurs très similaires.

Méthode Intra-classe

De la même manière que pour le test T ratio, la transformation logarithmique entraîne des valeurs plus élevées du paramètre ρ_I et permet de mieux identifier les limites entre couches

comme on l'observe dans la figure (Fig.8.16). Les valeurs seuils proposées dans la littérature sont aussi représentées dans les graphiques. Suite aux résultats obtenus pour $\ln(q_d)$ et avec une valeur seuil de 0.7, toutes les interfaces entre couches ont été identifiées sauf la transition entre le ballast et le sable grossier dans le moule HET4. Toutefois, un seuil plus faible égal à 0.65 permettrait d'identifier correctement cette transition. De manière similaire que pour le test T ratio, l'interface entre le sable grossier et de Dunkerque (moule HET3) ne peut toujours pas être détectée au moyen de ce test. D'ailleurs, le pic qui indique la limite entre le microballast et le sable grossier dans le moule HET1 présente une valeur beaucoup plus faible que les valeurs seuils retenues. On constate pourtant des résultats très similaires par rapport à ceux du test T ratio.

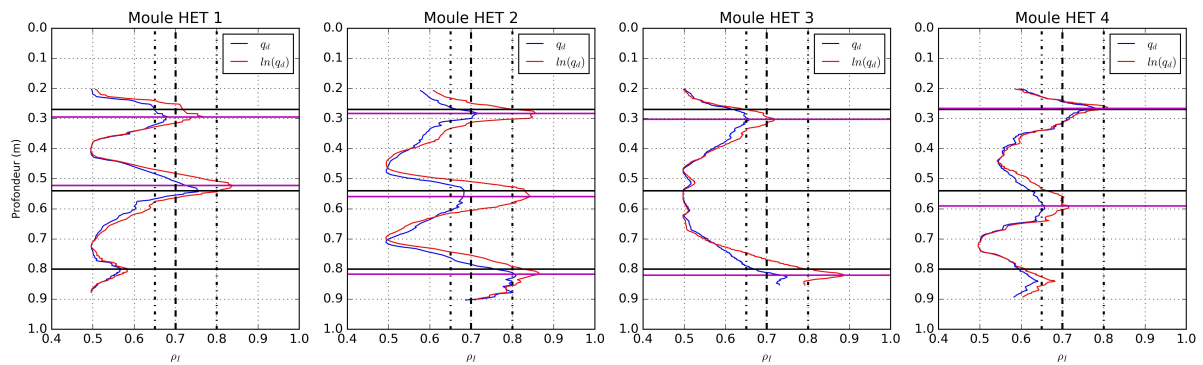


FIGURE 8.16 – Intra-classe : influence de la transformation logarithmique des valeurs de q_d (pointe de 4 cm^2)

Méthode Bartlett

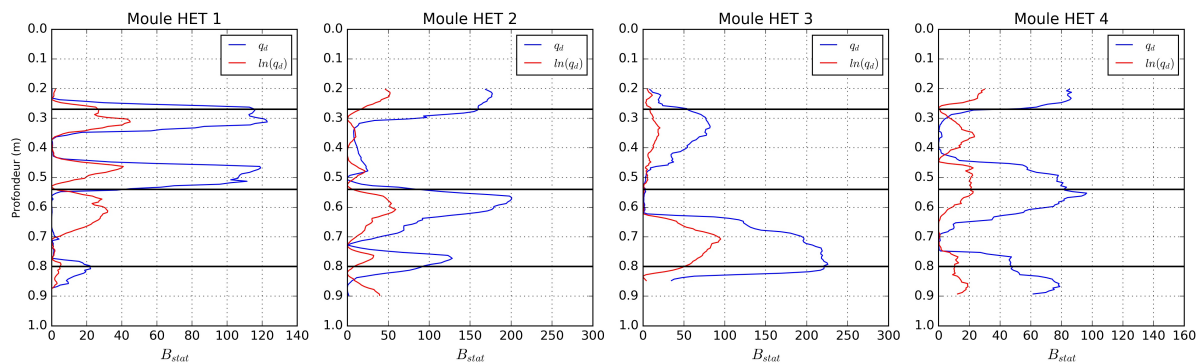


FIGURE 8.17 – Bartlett : influence de la transformation logarithmique des valeurs de q_d (pointe de 4 cm^2)

À la différence du test T ratio et Intra-classe, la transformée logarithmique des mesures de résistance q_d est déconseillée pour le test de Bartlett, comme le mettent en évidence les résultats présentés dans la figure (Fig.8.17). Les pénétrogrammes analysés ont été régularisés pour un pas de 5 mm. En effet, on observe que les pics du profil statistique de Bartlett pour $\ln(q_d)$ ne coïncident pas avec les interfaces à identifier. Enfin, on remarque que les mêmes

limites entre couches que celle repérées grâce à la méthode T ratio pourront être identifiées en appliquant la méthode de Bartlett au profil de résistance dynamique q_d .

Conclusion

La transformée logarithmique de la résistance de pointe q_d s'avère favorable vis-à-vis des méthodes T ratio et Intra-classe, mais pas pour le test de Bartlett.

8.3.4 Influence de la taille de la pointe

Les conclusions déduites des analyses réalisées auparavant seront ici retenues pour évaluer finalement la sensibilité de ces méthodes à la taille de la pointe (d_c) utilisée. On comparera donc, pour chacun des tests, les courbes statistiques obtenues pour les sondages réalisés avec la pointe de 2 cm^2 et ceux déjà analysés effectués avec une pointe de 4 cm^2 .

Méthode T ratio

Le test statistique T ratio a été appliqué au $\ln(q_d)$. Les résultats sont fournis dans la figure (Fig.8.18). En bleu on a représenté les courbes obtenues pour les sondages en pointe de 2 cm^2 ($d_c = 16\text{mm}$) et en rouge ceux avec une pointe de 4 cm^2 ($d_c = 22\text{mm}$). On voit que les deux courbes statistiques ont des allures très similaires. Il devient aussi clair qu'en appliquant la méthode T ratio aux sondages en pointe de 2 cm^2 on pourrait identifier correctement toutes les interfaces à l'exception de celle située entre les deux matériaux sableux dans le moule HET3. En outre, on observe un maximum local près de 0.8 m de profondeur dans le moule HET1, lequel indique correctement la limite entre la couche de microballast et de sable grossier. Par contre, cette interface ne pouvait pas être détectée par l'analyse du pénétrogramme effectué en pointe de 4 cm^2 pour une valeur seuil égal à 10. L'interface entre la couche de ballast et le sable grossier dans le moule HET4 devient aussi plus nette puisque la valeur du paramètre calculée pour le sondage en pointe de 2 cm^2 est plus élevée.

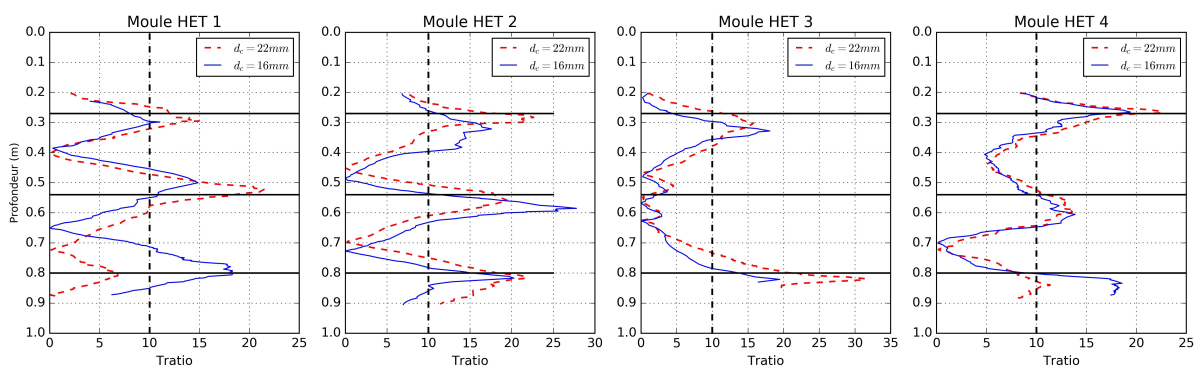


FIGURE 8.18 – T ratio ($\ln(q_d)$) : influence de la taille de la pointe (d_c)

Méthode Intra-classe

La méthode Intra-classe a été aussi appliquée à la transformée logarithmique de la résistance de pointe dynamique. Les courbes statistiques calculées sont données dans la figure

(Fig.8.19). On peut déduire des conclusions très similaires à celles issues des tests pour la méthode T ratio. On voit quand même, que le test est plus sensible (valeur du pic plus forte) à l'interface entre la couche de micro-ballast et de sable grossier dans le moule HET3 pour le sondage effectué avec une pointe de 2cm^2 .

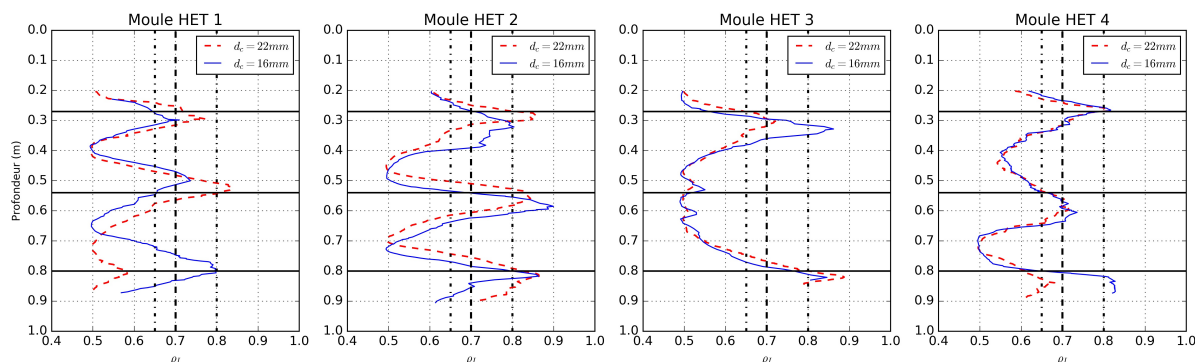


FIGURE 8.19 – Intra-classe ($\ln(q_d)$) : influence de la taille de la pointe (d_c)

Méthode Bartlett

Les profils issus du test statistique de Bartlett appliqué aux valeurs de q_d mesurées avec les deux tailles de pointe différentes sont donnés dans la figure (Fig.8.20). On observe des maxima locaux plus prononcés pour les sondages réalisés en pointe de 4cm^2 permettant donc une meilleure identification des limites entre couches. Par contre, comme pour les deux tests précédents, l'interface située entre la couche constituée par le microballast et le sable grossier dans le moule HET1 est mieux identifiée par l'analyse du test réalisé en pointe de 2cm^2 .

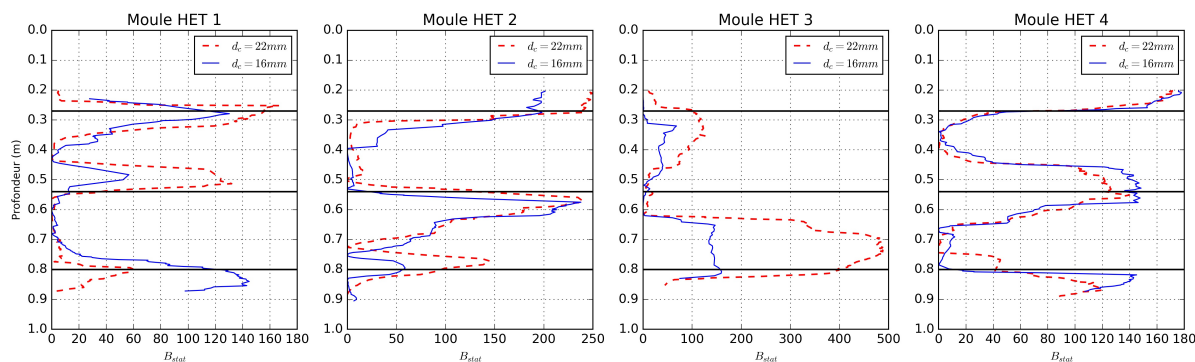


FIGURE 8.20 – Bartlett (q_d) : influence de la taille de la pointe (d_c)

Conclusion

Pour le cas d'étude présenté on n'a pas observé une influence systématique de la taille de la pointe pour la détection des couches à l'aide des méthodes étudiées.

8.3.5 Bilan

L'étude réalisée sur des essais en laboratoire porte sur la détection des couches de faible épaisseur confectionnées avec des matériaux granulaires. Suite aux analyses réalisées, pour chacune des méthodes testées on peut déduire les conclusions suivantes :

- T ratio

On a constaté une meilleure efficacité dans l'identification des interfaces entre les différentes couches pour des mesures régularisées avec pas de régularisation égal à l'enfoncement moyen. De même, cette méthode est plus sensible à la détection des interfaces quand on analyse le logarithme de la résistance dynamique. La taille de fenêtre optimale est égale à 0.4 m et une valeur seuil égale à 10 s'avère correcte.

- Intra-classe

Les courbes statistiques obtenues pour les différents pas de régularisation montrent des allures très similaires, ce test semble donc peu sensible au pas de régularisation et aucune règle générale n'a pu être déduite. En outre une fenêtre de 0.4 m de hauteur a été retenue et une valeur seuil égale à 0.7 a permis une identification correcte des différentes interfaces, favorisée par la transformée logarithmique de q_d .

- Bartlett

Des conclusions similaires à celles déduites pour la méthode T ratio ont été constatées vis-à-vis de l'effet de la régularisation. Plus la valeur du pas de régularisation choisie est différente de la valeur de l'enfoncement moyen, moins l'identification des interfaces par le biais de ce test est évidente. Comme pour les autres tests, on a retenu une hauteur de fenêtre égale à 0.4m. Par contre, l'identification des interfaces entre couches n'est possible que quand on analyse le profil de résistance dynamique q_d . L'analyse du logarithme de la résistance est déconseillée pour ce test.

Malgré le fait que les trois méthodes aient fourni des résultats similaires, la limite entre deux couches de matériaux ayant des valeurs de moyenne et d'écart-type de résistance similaires, ne paraît pas pouvoir être correctement identifiée grâce à ces méthodes. Ceci a été mis en évidence par l'analyse du profil de résistance obtenu dans le moule HET3, où la limite entre les deux couches de matériaux sableux n'a pas pu être repérée par les méthodes étudiées.

Enfin, l'efficacité de ces techniques vis-à-vis de la détection des couches de différentes épaisseurs et constituées par des matériaux fins n'a pas pu être testée. Ces deux volets devront être étudiés lors de l'analyse des essais in situ.

8.4 Analyses de l'efficacité des méthodes de découpage statistiques sur des essais in situ en conditions réelles

8.4.1 Essais sur le site expérimental de Castello d'Empúries

Sur le site expérimental de Castelló d'Empúries, nous avons réalisé (2016) une campagne de huit sondages au pénétromètre Panda. On dispose aussi de deux essais au piezocône (cf. [Annexe C.1.4](#)) réalisés par l'entreprise IGEOtest lors d'une campagne de reconnaissance en collaboration avec l'Université Polytechnique de Catalogne. La nappe phréatique a été repérée à 3.87 m de profondeur pour le premier sondage (CPTU1), et à 3.56 m pour le

deuxième (CPTU2). Une présentation de l'ensemble des essais réalisés sur site est résumé dans (Pérez et al. (2013)).

Les essais Panda ont été réalisés à proximité des essais CPTU. Concrètement, on a choisi les sondages S3 et S5, situés respectivement à 4 et 4.5 m de distance des tests CPTU2 et CPTU1 (cf. Fig. 8.21). La présence de la nappe phréatique a été confirmée par la réalisation d'essais géoendoscopiques, à environ 2.50 m de profondeur.

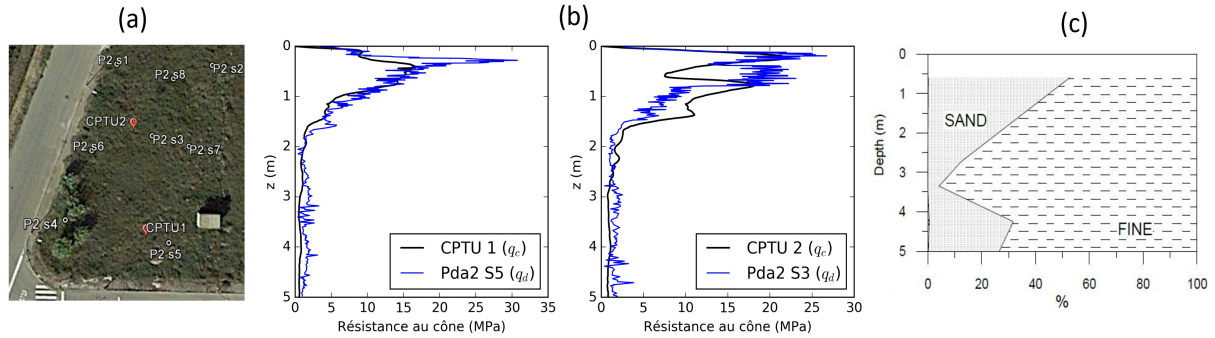


FIGURE 8.21 – (a) Implantation des sondages sur le site de Castello d'Empúries. (b) Essais pénétrométriques statiques (CPTU) et dynamiques (Panda). (c) Distribution granulométrique en fonction de la profondeur jusqu'à 5 m (source : Pérez et al. (2013))

On dispose également de l'identification des sols en surface par des essais d'identification en laboratoire de plusieurs échantillons prélevés par carottage, avec un échantillonneur Shelby de 80 mm de diamètre (Arroyo et al. (2015)). Les essais de granulométrie par tamisage et sédimentation ainsi que la détermination des limites d'Atterberg pour la fraction fine ont été réalisés par le Laboratoire GIA S.L. (*Grupo de Ingeniería y Arquitectura S.L.*) de Valence (Espagne). Dans le tableau (Tab. 8.6) on présente les caractéristiques des échantillons récoltés en surface.

TABLE 8.6 – Paramètres d'identification et d'état des échantillons prélevés

Profondeur (m)	Matériau	passant à		Limites d'Atterberg	
		75 μm	2 μm	WL %	IP %
0.5-0.7	Sable et limons argileux	47.3	15.6	-	-
2.65-2.8	Limons argileux	87.5	33.23	34.9	12.1
3.3-3.4	Argiles limoneuses	95.8	55.74	49.4	25.3
4.2-4.3	Limons argileux	68.2	31.62	25.3	9.7
5.8-5.9	Limons argileux	71.5	25.31	-	-

Les échantillons prélevés entre 2 et 6 mètres seront classés comme des argiles peu plastiques (CL) d'après le système de classification LPC et l'échantillon prélevé plus en surface est un matériaux sableux (S).

Application des techniques stratigraphiques au pénétrogramme Panda

On applique les trois tests statistiques étudiés auparavant aux sondages Panda situés à proximité des essais CPTU. Pour ce faire, on étudiera l'influence du pas de régularisation et

ensuite la sensibilité à la hauteur de la fenêtre utilisée. Les test T ratio et Intraclasse seront appliqués au logarithme de la résistance de pointe q_d tandis qu'avec le test de Bartlett, on analysera le profil des mesures q_d .

Influence de la hauteur de la fenêtre

Pour évaluer l'influence de la taille de la fenêtre, on testera quatre hauteurs possibles pour chacun des tests, à savoir 0.4 m (comme celle définie suite aux analyses en laboratoire), 1, 1.5 et 2m (suite aux recommandations de la littérature). Les pénétrogrammes ont été régularisés par un pas d'échantillonnage égal à 10mm, proche de la valeur de l'enfoncement moyen. Les profils statistiques calculés selon la profondeur pour les 4 tailles de fenêtres testées sont présentés dans (Fig.8.22) pour le sondage S5 et dans (Fig.8.23) pour le sondage S3.

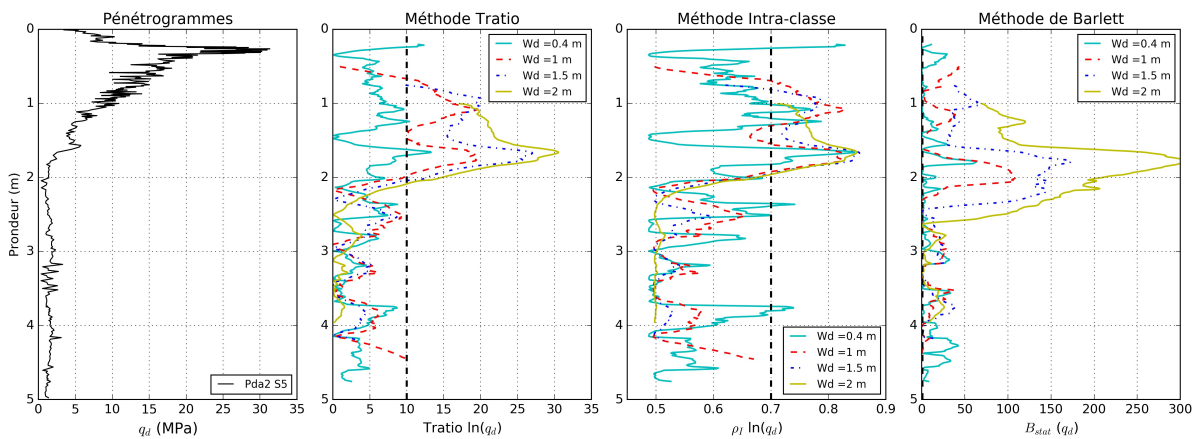


FIGURE 8.22 – Sondage Panda S5 : influence de la longueur (Wd) de la fenêtre glissante

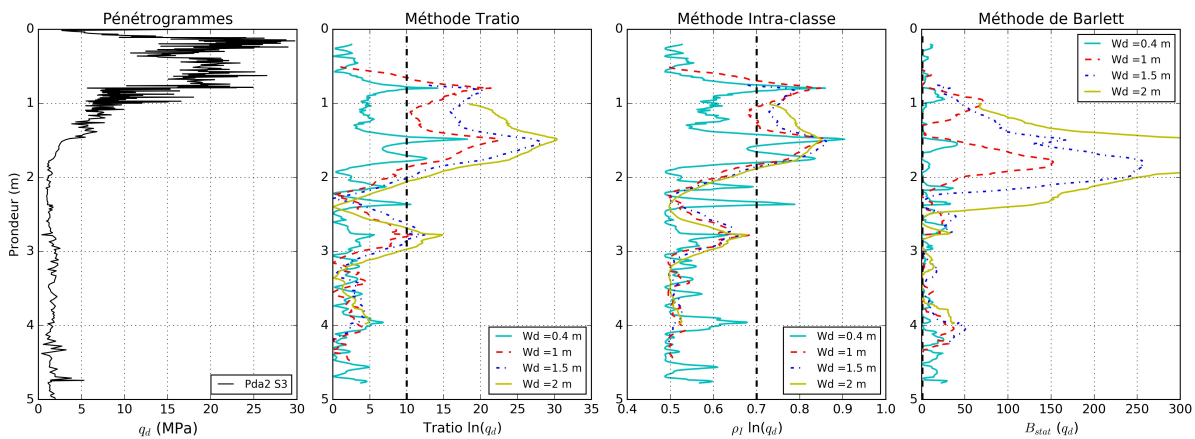


FIGURE 8.23 – Sondage Panda S3 : influence de la longueur (Wd) de la fenêtre glissante

De manière générale il est possible de remarquer :

- indépendamment de la méthode utilisée, quand on utilise une hauteur de fenêtre égale à 0.4m les profils statistiques calculés sont beaucoup plus bruités. Leurs interprétations sont plus complexes puisque on ne peut pas identifier clairement les maxima locaux indiquant la position des interfaces recherchées.

- pour les autres hauteurs de fenêtre testées, supérieures ou égales à 1 m, on peut constater des résultats similaires. En effet les profils statistiques calculés montrent des allures similaires mais il y a toujours un effet de lissage quand la hauteur de la fenêtre augmente.

Dans (Tab.8.7), on a récapitulé les profondeurs des interfaces détectées. Il se dégage que :

- un grand nombre d'interfaces sont repérées pour la fenêtre de petite taille faisant 0.4 m de hauteur. En outre, la méthode de Bartlett nécessite l'identification visuelle des maxima locaux. Cependant le profil statistique calculé comporte beaucoup de pics et cette identification n'est pas claire. C'est pourquoi on a marqué avec un X les résultats concernant ce test.
- quand on utilise des fenêtres faisant au moins 1 mètre de hauteur, on identifie uniquement deux interfaces ou 3 unités voire 2 unités pour la fenêtre de 2m de hauteur. De même, les trois méthodes signalent la présence d'une limite de couche entre 1.5 et 2 mètres de profondeur. Cette interface pourrait être associée à la transition entre les horizons sableux et argileux. De plus, une interface plus en surface a aussi été identifiée. Elle est cohérente avec l'affaiblissement des mesures de résistance que l'on observe à partir d'un mètre de profondeur.
- enfin on a constaté que les allures des courbes issues des tests T ratio et Intra-classe sont très similaires. En effet, ces deux méthodes fournissent des résultats équivalents quand à la profondeur des interfaces détectées. Par contre la position des interfaces identifiées au moyen de la méthode de Bartlett est décalée par rapport à celle signalée par les deux autres méthodes.

TABLE 8.7 – Synthèse des résultats - profondeur (m) des interfaces identifiées à partir des différents tests paramétriques étudiés pour les sondages Panda

	Pda2 S5			Pda2 S3		
	<i>T ratio</i>	<i>Intra-classe</i>	<i>Bartlett</i>	<i>T ratio</i>	<i>Intra-classe</i>	<i>Bartlett</i>
Wd = 0.4m	0.24	0.24		0.8	0.8	
	1.08	1.08		1.48	1.48	
	1.24	1.24	X	1.75	1.75	X
	1.66	1.66		2.36	2.36	
	-	2.36		-	-	
	-	3.76		-	-	
Wd = 1m	1.08	1.08	-	0.8	0.8	0.98
	1.77	1.77	2	1.49	1.49	1.76
	-	-	-	2.78	-	-
Wd = 1.5 m	0.94	0.94	-	0.82	0.82	-
	1.67	1.67	1.78	1.54	1.54	1.82
	-	-	-	2.78	-	-
Wd = 2m	1.66	1.66	1.78	1.48	1.48	1.74
	-	-	-	2.78	-	-

Bilan

Les trois méthodes statistiques ont permis une identification correcte de l'interface entre la formation des sols sableux et argileux. On a aussi constaté qu'avec une fenêtre de 0.4 m de hauteur (taille retenue à partir de l'analyse des sondages en laboratoire), les profils statistiques calculés au travers des différentes méthodes présentent beaucoup de dispersion et aboutissent à la détection d'un grand nombre d'interfaces. Avec une fenêtre de l'ordre de 1m, les résultats obtenus sont plus répétables et semblent être plus cohérents avec la lithologie du site.

8.4.2 Essais in situ analysés par les experts

Application des techniques statistiques

On applique ici les méthodes statistiques développées auparavant sur les 12 sondages Panda in situ analysés par les 9 experts. Pour chacun des essais la valeur de l'enfoncement moyen a été calculée, son étendue varie entre 9 et 19 mm. Les pénétrogrammes ont été régularisés, avec un pas égal à la valeur moyenne de l'enfoncement. Une démarche similaire a été retenue, les trois méthodes ont été testées (T ratio, Intra-classe et Bartlett). Les méthodes T ratio et Intra-classe ont été appliquées au logarithme de la résistance de pointe dynamique $\ln(q_d)$ et à q_d pour la méthode de Bartlett. D'ailleurs, quatre hauteurs de fenêtres ont été étudiées (0.4, 1, 1.5 et 2m). Les valeurs seuils retenues sont les mêmes qu'auparavant, 10 pour le test T ratio et 0.7 pour le test Intra-classe. Les courbes des statistiques issues de chacune des méthodes en fonction de la hauteur de fenêtre utilisée sont données en annexe ([Annexe A.3](#)).

Les interfaces détectées, sont fournies pour les différents sondages analysés dans la figure ([Fig.8.25](#)), représentées par des symboles afin de mieux visualiser l'ensemble des résultats. Dans cette figure on a aussi représenté les interfaces déduites à partir des levés géologiques du site (lignes magenta sur les graphiques) et l'identification visuelle de la nature des terrains reconnus. On peut remarquer que les profils analysés sont constitués essentiellement par des matériaux fins, notamment des argiles.

En ce qui concerne la comparaison entre les stratigraphies basées sur l'analyse statistique du profil de résistance dynamique et les coupes géologiques on constate que :

- il existe une bonne correspondance mais qui n'est pas univoque. En effet, les coupes géologiques sont basées sur une identification purement visuelle du forage prospecté, et certaines transitions sont basées sur un changement de la couleur des horizons traversés, qui ne sont pas toujours accompagnées par des variations significatives du profil pénétrométrique. Par ailleurs, on constate parfois un léger décalage de profondeur entre les interfaces détectées par les méthodes statistiques et les interfaces géologiques. Ce décalage peut être expliqué entre autres choses, par la zone d'influence de la pointe du pénétromètre et par l'imprécision des limites géologiques lorsque celles-ci sont identifiées à partir de sondage à la tarière.
- le sondage E1023s1 est un bon exemple pour illustrer le constat précédent. La coupe lithologique du site jusqu'à 4m comporte deux couches. Un premier remblai sableux jusqu'à environ 1 mètre de profondeur, avec des résistances mécaniques autour de 8 MPa et ensuite une formation d'argile grise plastique avec une faible portance. L'interface entre la couche de remblais et la formation argileuse a été correctement identifiée

par les trois méthodes statistiques. Par contre plusieurs sous-couches « mécaniques » ont été repérées grâce aux techniques statistiques alors que géologiquement il s'agit d'une formation argileuse homogène. D'après les approches T ratio et Intraclasse, une interface supplémentaire est mise en évidence vers 2.75 m, profondeur à partir de laquelle on observe un affaiblissement des caractéristiques mécaniques qu'on peut expliquer par la présence de la nappe phréatique vers 3 m. Par ailleurs, pour ce sondage on observe des résultats très similaires indépendamment de la taille de fenêtre utilisée ;

- un autre exemple est donné par le sondage E719s1, les trois méthodes ont détecté correctement l'interface entre le remblai anthropique et la formation argileuse quand on utilise une fenêtre égale à 1 m de hauteur. Cependant, plusieurs unités ont été repérées au sein de la formation argileuse par l'analyse statistique du profil de résistances mécaniques alors que géologiquement il s'agit d'une unité de terrain homogène.

Par rapport à la longueur de la fenêtre utilisée, on peut remarquer que :

- en général, plus la longueur de la fenêtre est importante et plus le nombre de couches identifiées est faible. Cependant il existe une répétabilité quant à la profondeur des interfaces détectées ;
- si on analyse certains sondages avec une fenêtre de taille de 0.4m, on peut obtenir des résultats très similaires à ceux issus d'une analyse avec fenêtres de plus grande hauteur, comme pour les sondages E759s1, E832s2 et E1023s1. Par contre pour d'autres sondages (E590s1 ou E748s1) réaliser l'analyse avec cette longueur entrainerait la détection d'un nombre d'unités très élevé par rapport aux autres tailles utilisées. En effet, les courbes statistiques calculées sont très bruitées et elles comportent un grand nombre de maxima locaux ;
- cependant il existe des cas où la profondeur du sondage impose la longueur de la fenêtre. Par exemple l'essai E668s3, dont le refus a été atteint à moins de 1m de profondeur. Dans ce cas, on est obligé d'analyser le profil de résistance avec une fenêtre de taille plus faible.

En ce qui concerne les méthodes étudiées on peut signaler que :

- tout comme pour les analyses précédentes, les tests T ratio et Intra-classe fournissent des résultats très similaires ;
- dans les cas observés, les maxima locaux issus de la méthode de Bartlett donnent de manière évidente le découpage entre les grands ensembles. Cette méthode semble être beaucoup moins sensible aux faibles variations de la résistance moyenne. Ceci peut être constaté pour le sondage E719s1 dans lequel les tests T ratio et Intraclasse présentent un maximum à 3 m correspondant à un changement de la résistance moyenne. Toutefois ce maximum n'est pas bien mis en évidence pour le test de Bartlett ;
- au contraire la méthode basée sur le test de Bartlett, étant un test de variance, elle est beaucoup plus sensible aux variations en termes de dispersion du profil de résistance. Un exemple clair est le sondage E1023s1 où l'on observe une grande dispersion du pénétrogramme à la fin du sondage qui est mieux repérée par ce test ;
- enfin, la position des maxima locaux est moins influencée par la hauteur de la fenêtre testée pour les méthodes T ratio et Intraclasse que pour celle de Bartlett.

Finalement, si on compare les résultats issus des méthodes paramétriques avec les interprétations fournies par les experts dans le graphique (Fig.8.24) où l'on a conservé les mêmes codes couleurs et symboles que ceux utilisés dans (Fig.8.25) pour représenter les résultats

des tests statistiques. On peut constater qu'il y a moins de variabilité entre les résultats des tests statistiques. Ce qui permet d'apporter plus d'objectivité et de « standardisation » à l'analyse des pénétrogrammes en utilisant ces procédures.

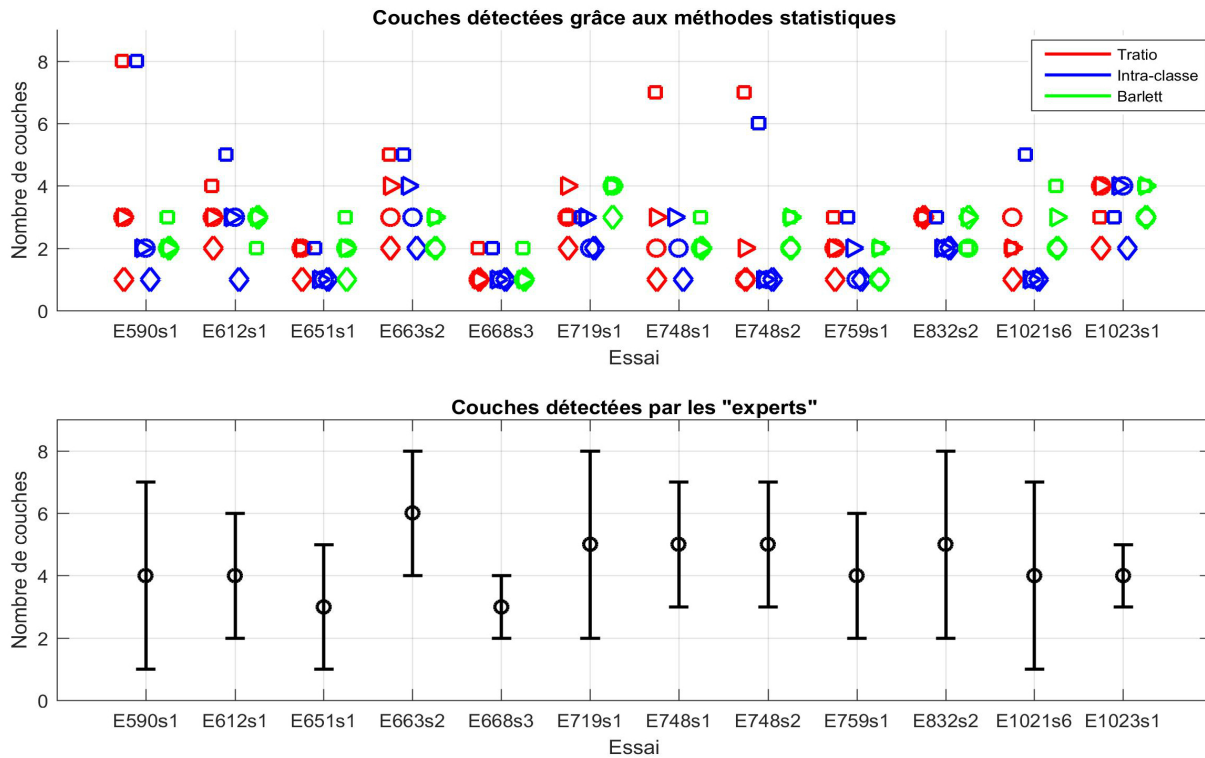


FIGURE 8.24 – Bilan sur le nombre de couches détectées grâce aux méthodes statistiques et d'après les experts. La taille de la fenêtre est représentée par un code symbole : $\square = 0.4m$, $\triangleright = 1m$, $\circ = 1.5m$, $\diamond = 2m$

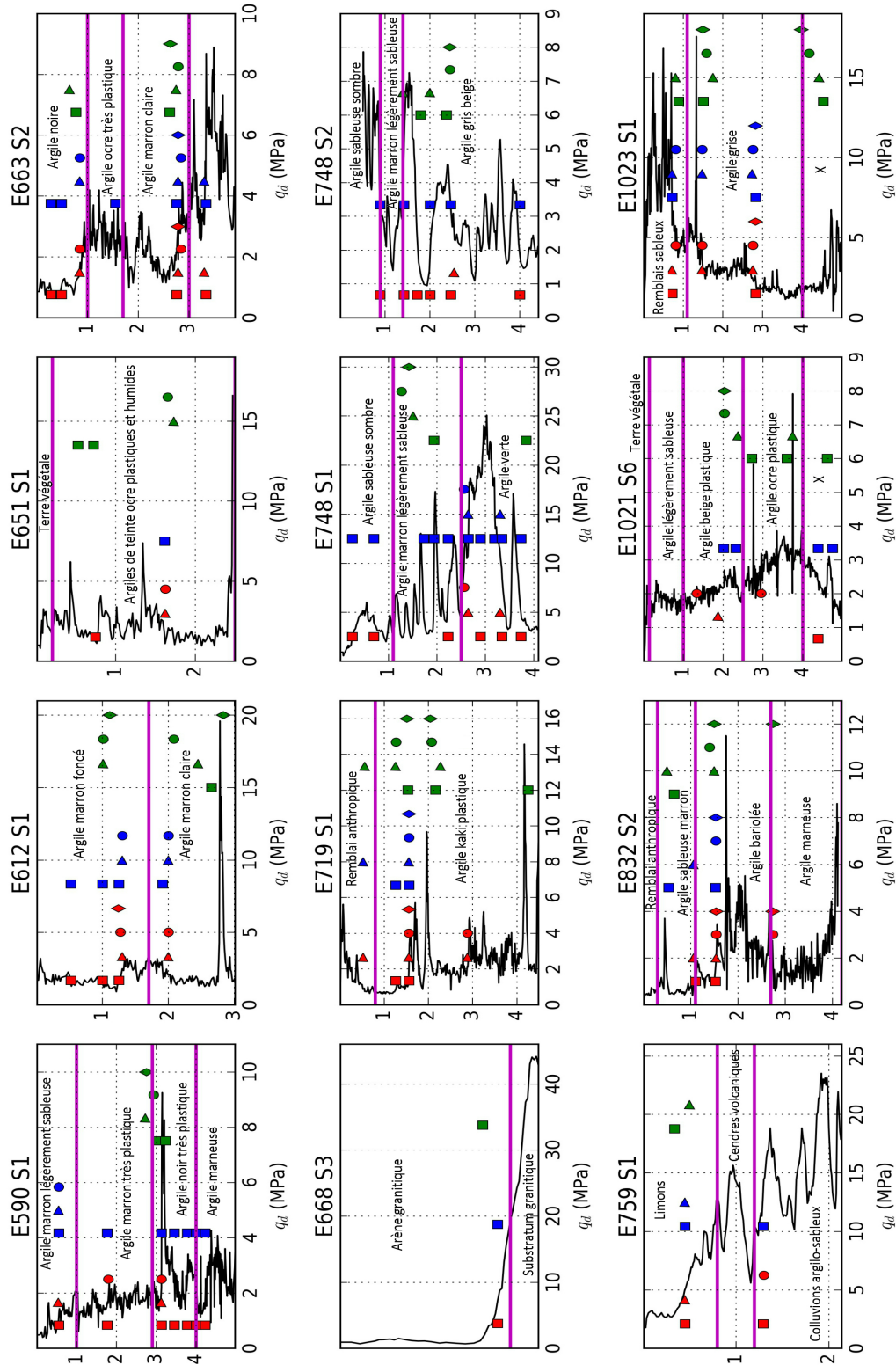


FIGURE 8.25 – Limites de couches détectées avec les méthodes statistiques. La méthode utilisée est représenté par un code couleur : rouge = **T ratio**, bleu = **Intra-classe** et vert = **Bartlett**. La taille de la fenêtre est représentée par un code symbole : $\square = 0.4m$, $\triangle = 1m$, $\circ = 1.5m$, $\diamond = 2m$.

Bilan

L'application des méthodes de découpage statistique à l'ensemble de 12 sondages réalisés in situ a permis de déduire les conclusions suivantes :

- même si il y a une correspondance entre certaines interfaces géologiques avec les interfaces identifiées à partir des profils pénétrométriques au moyen des méthodes statistiques, la coupe géologique n'est pas exactement la même que la stratigraphie décelée par le biais de ces méthodes. Les sous-couches mécaniques repérées grâce aux méthodes statistiques sont donc complémentaires à l'information issues des données géologiques. Malheureusement, on ne dispose pas d'une caractérisation complète avec des essais d'identification au laboratoire le long du profil afin de pouvoir mener une étude comparative plus rigoureuse ;
- les profils statistiques calculés avec une fenêtre d'une hauteur de 0.4 m peuvent présenter une grande dispersion, ce qui complique leur interprétation et entraîne la détection d'un grand nombre de couches pas forcément cohérent avec les données géologiques. Pour l'analyse des sondages in situ une fenêtre de l'ordre de 1 m de longueur semble être un choix plus pertinent et moins sensible à la variabilité naturelle des matériaux ;
- même si les trois approches étudiées apportent des résultats similaires, on a pu constater que les méthodes T ratio et Intra-classe sont plus sensibles aux variations de la résistance moyenne des unités à étudier tandis que la méthode de Bartlett repère mieux les variations en termes de dispersion. Par contre, le fait de ne pas pouvoir fixer au préalable une valeur seuil pour ce test rend cette technique plus subjective que les deux autres.
- enfin, globalement, on constate que les méthodes d'analyse statistiques sont moins sensibles à la subjectivité et permettent donc de "rationaliser" l'analyse du pénétrogramme Panda.

8.5 Comparatif des stratigraphies issues des essais CPTU et Panda sur le site de Castello d'Empúries

8.5.1 Techniques stratigraphiques utilisées pour l'essai CPTU

Les essais au pénétromètre statique et au piezocône sont particulièrement intéressants pour déterminer la stratigraphie des sols traversés lors d'un sondage. On compare ici la stratigraphie déduite grâce aux procédures statistiques étudiées pour les sondages Panda réalisés sur le terrain expérimental de Castello avec celles déterminées par l'analyse des mesures réalisées avec l'essai CPT. Les profils géotechniques sont comparés jusqu'à 5 m de profondeur. Parmi les différentes méthodes existantes, on utilisera les suivantes :

i. Abaque de classification normalisé

Parmi les différents abaques empiriques proposés dans la littérature, ([Robertson \(2009\)](#)) indique que les abaques normalisés fournissent des résultats plus répétables que ceux non normalisés. L'abaque normalisé proposé par Robertson ([Robertson \(1990\)](#), [Robertson \(2009\)](#)), où est donnée l'évolution de la résistance au cône normalisée Q_{tn} en fonction du rapport de frottement normalisé F_7 .

ii. Analyse par cluster

La mise en œuvre de l'analyse par cluster qu'on a réalisé, est très similaire à celle proposée par (Hegazy and Mayne (2002)). On analysera le même jeu de paramètres pris en compte par l'abaque de classification employé (Q_{tn}, F_r). Les étapes principales de la méthode par clusters utilisée sont :

- l'analyse par cluster sur la base d'une Classification Ascendante Hiérarchique (CAH). Il s'agit d'une procédure itérative où à chaque étape on cherche à regrouper deux objets en minimisant un critère d'agrégation donné ;
- le choix des variables statistiques se porte sur l'espace de mesures (Q_{tn}, F_r) ;
- afin de s'affranchir des unités, pour une profondeur i , les mesures x pour chaque variable j sont standardisées z_{ij} grâce à la méthode (z -score) :

$$z_{ij} = \frac{x_{ij} - \mu_{x_j}}{\sigma_{x_j}} \quad (8.3)$$

avec μ_{x_j} et σ_{x_j} la moyenne et l'écart-type de la variable x_j .

- plusieurs métriques pour mesurer le degré de similarité exprimé en termes d'une fonction de distance sont disponibles dans la littérature, la distance du cosinus est celle utilisée ici pour mesurer la similarité entre données (Hegazy and Mayne (2002), Mlynarek et al. (2007)) :

$$d_{ij} = \frac{\sum_{k=1}^p z_{ik} \cdot z_{jk}}{\sqrt{\sum_{k=1}^p z_{ik}^2 \sum_{k=1}^p z_{jk}^2}} \quad (8.4)$$

avec p le nombre de variables égal à 2 dans notre cas.

- la stratégie d'agrégation retenue est celle du saut minimum (*single linkage*), c'est-à-dire que la distance entre groupes est la plus petite distance entre éléments des deux groupes.
 - concernant le choix du nombre de classes ou (*clusters*), on a testé plusieurs possibilités en faisant varier le nombre de clusters entre 2 et 25. Quand le nombre de clusters est supérieur à 5, l'interprétation des résultats n'est pas claire. On a trouvé que les résultats les plus représentatifs sont ceux obtenus pour un nombre de clusters égal à 2 ou 5.
- iii. T ratio sur la courbe de l'indice de comportement I_c

La dernière technique utilisée est basée sur l'indice de comportement I_c qui dépend aussi des mesures de résistance et du frottement. Il s'agit d'un paramètre pertinent pour établir une classification vis-à-vis du comportement mécanique. On analysera les variations du profil I_c par le biais de la méthode T ratio. Deux hauteurs de fenêtre égales à 1 et 2m seront testées et une valeur seuil égale à 10 sera retenue comme suggéré dans (Ching et al. (2015a)).

8.5.2 Résultat de la comparaison avec des techniques stratigraphiques pour l'essai CPTU

Les unités géotechniques identifiées à l'aide de l'abaque de classification de Robertson, l'analyse par cluster et la méthode T ratio appliquée à l'indice de comportement sont récapitulées graphiquement dans (Fig.8.26), pour le sondage CPTU1 et (Fig.8.27), pour le sondage

CPTU2. Dans le premier graphique, on a représenté les profils de la résistance au cône normalisée Q_{tn} et le rapport de frottement F_r , dans le deuxième, le classement des mesures en fonction de l'abaque normalisé de Robertson. Le troisième et quatrième graphique correspondent aux résultats des analyses par cluster et les trois suivants aux analyses concernant le profil de l'indice de comportement. Concernant l'essai Panda, ayant constaté des résultats similaires pour les trois méthodes testées, on présente dans cette étude comparative, uniquement les résultats issus de la méthode T ratio ($\ln(q_d)$), avec une longueur de fenêtre égale à 1 mètre, donnant le nombre de couches le plus élevé.

Suite aux résultats présentés, on constate de manière générale que :

- toutes les techniques testées détectent une limite de couche à environ 1.5 m qui peut être attribuée à la transition entre les matériaux sableux et les matériaux fins. L'abaque de Robertson classe aussi les mesures recueillies jusqu'à cette profondeur comme sols sableux. Cet interface a aussi été repérée grâce à l'analyse statistique des profil de résistance dynamique mesuré à l'aide du Panda ;
- on observe quelques passages classés en zone 8 de l'abaque de Robertson en surface, indiquant la présence de sols très raides. Ils sont observés jusqu'à environ 1 m de profondeur pour l'essai CPTU1. L'analyse par cluster détermine aussi une première unité géotechnique jusqu'à 1 m de profondeur lorsque l'on choisit un nombre de clusters égal à 5. D'une manière similaire, une interface à 1 mètre de profondeur avait été repérée au moyen de l'analyse statistique du pénétrogramme Panda.

En ce qui concerne les stratigraphies déterminées par analyse de l'essai au piezocône CPTU1, il se dégage que :

- la stratigraphie mécanique déduite au moyen d'une analyse avec cinq clusters est très similaire à celle issue de l'analyse du sondage Panda avec la méthode T ratio ;
- en ce qui concerne l'analyse du profil I_c par le biais de la méthode T ratio, le nombre d'unités détectées dépend aussi de la hauteur de fenêtre utilisée. Avec une fenêtre de 1m de hauteur, on observe une interface à 2m qui a une correspondance avec le passage des limons aux argiles d'après l'abaque de Robertson et une limite à 3.5 m qui n'a aucune correspondance avec les autres méthodes. Cette limite semble avoir une correspondance avec la diminution du rapport de frottement que l'on observe à partir de 3.5m, alors que la classe de nature du comportement ne varie pas (zone 3 de l'abaque). Sinon, les interfaces entre les couches à 1.5 et 3 m ont été identifiées avec les deux fenêtres utilisées. On observe que le rapport de frottement augmente à partir de 3m, cette limite signale le passage des limons aux argiles d'après I_c .

Pour l'essai CPTU2, on peut constater que :

- une zone de transition classée comme limons argileux selon l'abaque de Robertson a été identifiée entre 1.5 et 2 m à l'aide de l'analyse par cluster. Une autre formation de limons argileux selon l'abaque de Robertson apparaît à partir de 4.5 m, du fait de la diminution des valeurs du profil F_r . Cette limite est identifiée aussi au moyen de l'analyse par cluster et à partir de l'analyse de l'indice de comportement par la méthode T ratio avec une taille de fenêtre égale à 1m ;
- une limite entre couches situées entre 2.5 et 3 m de profondeur est identifiée quand on analyse avec la méthode T ratio le profil I_c , indépendamment de la hauteur de fenêtre utilisée. Cette interface apparaît aussi quand on observe la classification issue de l'abaque de Robertson (passage de la zone 4 " limons argileux " à 3 "argiles") et sur la stratigraphie déduite de l'essai Panda.

8.5.3 Bilan

Différentes interprétations de la stratigraphie du site sont donc possibles en fonction de la technique ou de l'approche utilisée, mais aussi en fonction des variables analysées. On a pu toutefois constater que les différents profils géotechniques déterminés sont cohérents entre eux. D'ailleurs, toutes les méthodes ont identifié correctement l'interface entre les deux formations de sols sableux et fins. Sinon, un découpage plus détaillé sur la formation de sols fins est possible grâce à l'analyse de l'indice de comportement ou les résultats issus de l'abaque de classification empirique de Robertson. Il faut quand même signaler, que la pression interstitielle n'a pas été prise en compte vis-à-vis des analyses des données recueillies par l'essai CPTU.

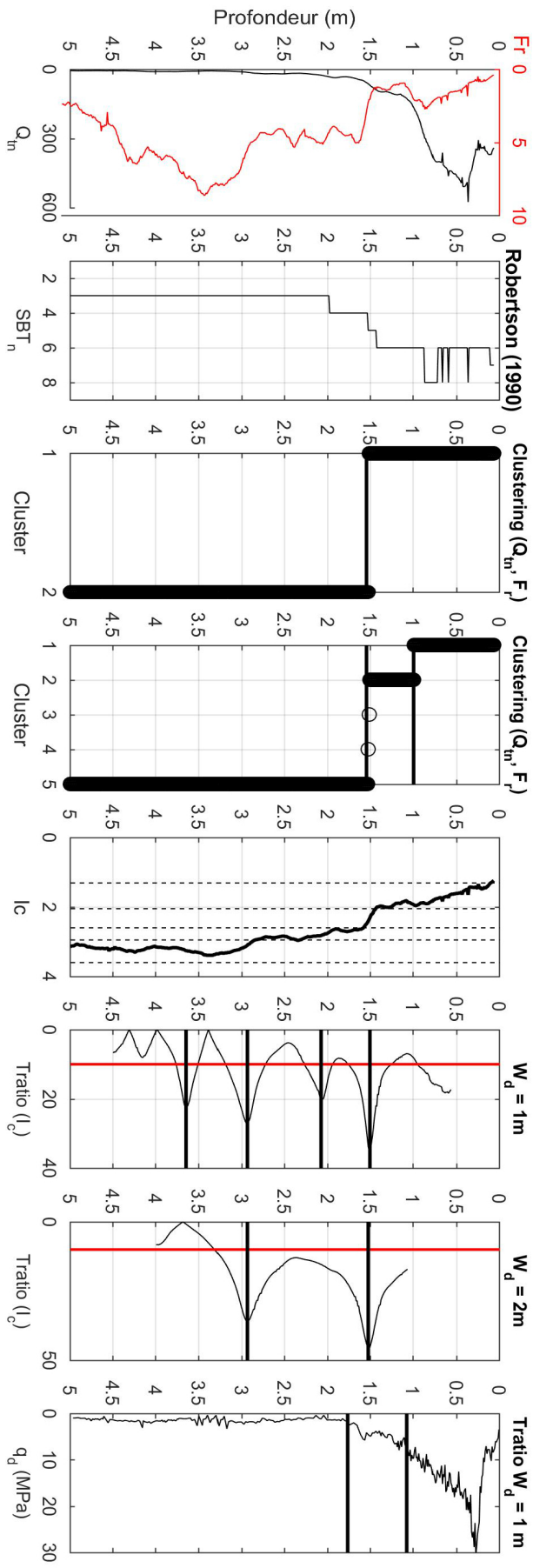


FIGURE 8.26 – Comparaison entre les stratigraphies déterminées par le sondage CPTUI et le sondage Panda S5

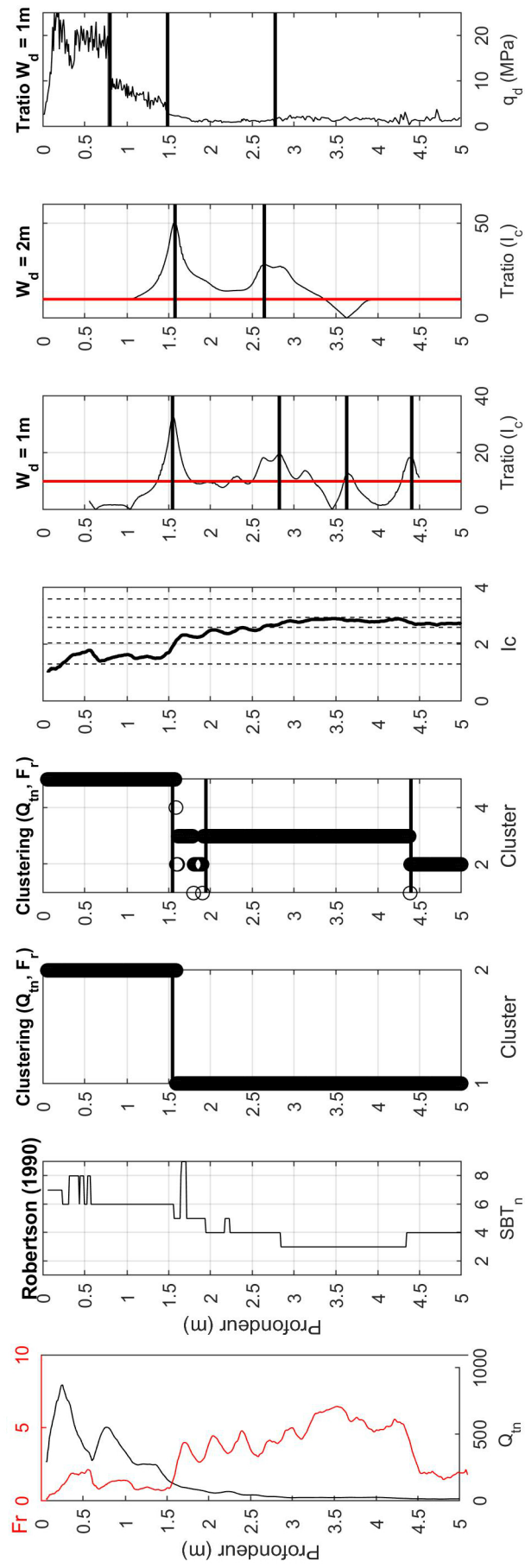


FIGURE 8.27 – Comparaison entre les stratigraphies déterminées par le sondage CPTU2 et le sondage Panda S3

9. Conclusions

Le travail mené dans cette partie a porté sur le développement d'une méthode automatisée de reconnaissance des unités homogènes de terrain à partir de l'analyse de l'essai de pénétration dynamique Panda. Suite à une réflexion sur une question qui n'a pas de réponse unique : *qu'est-ce qu'une couche ?* on a montré que la détermination d'un profil stratigraphique est un sujet qui admet plusieurs solutions en fonction du problème posé.

On a également montré à l'aide d'une enquête menée auprès d'une dizaine d'experts auxquels on a demandé de déterminer la coupe du terrain pour 12 sondages pénétrométriques réalisés en conditions réelles, sur la base de règles de définition de couches bien établies, que l'analyse "experte" d'un pénétrogramme mène à des réponses et interprétations très variables, mettant ainsi en évidence la subjectivité et le manque de fiabilité d'une telle analyse basée uniquement sur l'appréciation qualitative du profil de résistance dynamique q_d . Trois méthodes (*T ratio*, *Intra-classe* et *Test de Bartlett*) basées sur des tests paramétriques de comparaison sur la moyenne ou la variance des échantillons ont été étudiées pour l'analyse du pénétrogramme Panda.

Étude réalisée sur des essais au laboratoire

Pour valider et tester l'efficacité et la robustesse de ces méthodes, on a tout d'abord analysé une campagne de sondages réalisés en moules de calibration au laboratoire. On a pu constater qu'il est possible d'identifier correctement les limites entre les couches de faible épaisseur (0.27 m) grâce à ces trois méthodes statistiques. L'influence du pas de régularisation, de la hauteur de la fenêtre W_d et du mode de représentation de la variable (q_d ou $\ln(q_d)$) sur les méthodes proposées a été étudié. Le tableau (Tab.9.1) regroupe les conclusions principales déduites pour chacune des techniques.

TABLE 9.1 – Récapitulatif des conclusions sur les différentes méthodes

Méthodes	Pas de régularisation	Taille de la fenêtre	Variable d'étude
<i>T ratio</i>	enfouissement moyen	0.4 m	$\ln(q_d)$
<i>Intraclasse</i>	pas d'influence	0.4 m	$\ln(q_d)$
<i>Bartlett</i>	enfouissement moyen	0.4 m	q_d

Étude réalisée sur des essais in situ

Par la suite, ces méthodes ont été testées sur des pénétrogrammes réalisés in situ en conditions réelles sur des matériaux naturels. Au travers de ces analyses, nous avons mis en évidence que :

- les méthodes peuvent être appliquées également à des matériaux fins et naturels ;

- il existe une cohérence des résultats issus des différents tests statistiques. Par ailleurs, la dispersion du nombre de couches détectées par ces méthodes est plus faible que celle obtenue par des experts. Par conséquent, ces méthodes permettent une analyse plus objective et rationnelle des pénétrogrammes et une certaine "standardisation" de l'analyse ;
- la hauteur de la fenêtre d'analyse doit être adaptée à la dimension du sondage. En effet, si une hauteur de 0,4 m estimée à partir des essais en laboratoire semble bien adaptée pour les sondages courts (jusqu'à 2 ou 3 m), au delà de cette longueur de sondage et donc pour des sondages in situ, l'utilisation de cette taille de fenêtre est trop sensible aux fluctuations naturelles des caractéristiques du sol et une hauteur plus importante (de l'ordre du mètre) est à privilégier.

Néanmoins, la taille de la fenêtre qui s'avère être un paramètre fondamental pour la mise en oeuvre de ces méthodes, n'est pas un paramètre intrinsèque ni de la méthode ni du sol. C'est pourquoi la taille optimale de la fenêtre pourrait être adaptée en fonction de l'objectif de la reconnaissance géotechnique. Sur le domaine d'application du Panda, en reconnaissance des sols de surface avec une profondeur d'étude de l'ordre de 5m, pour les problèmes de fondation de bâtiment, une fenêtre de l'ordre de 1 m semble être un choix pertinent. Dans le cadre du contrôle de compactage, contexte qui n'a pas été traité dans ce travail mais est un domaine d'application courant de l'essai Panda, la taille de la fenêtre devra être adaptée à l'épaisseur des couches mises en oeuvre en choisissant une hauteur égale au double de cette épaisseur.

Conclusion sur les méthodes étudiées

Les méthodes T ratio et Intra-classe fournissent des courbes statistiques avec des allures très similaires et de ce fait sont équivalentes. Si les méthodes T ratio et Intra-classe sont contrôlées fondamentalement par les écarts de valeurs moyennes, le test de Bartlett est lui régi par les différences de variances. Cette différence induit parfois des résultats de découpage différents lors de l'analyse de certains sondages.

Les maxima locaux issus du test de Bartlett mettent en évidence les interfaces entre unités de terrain ayant des caractéristiques de résistance très distinctes que l'on peut associer à la structuration géologique. De ce point de vue, ce test est moins sensible que les deux autres méthodes et présente donc une plus faible valeur ajoutée.

Un autre avantage des méthodes T ratio et Intra-classe par rapport au test de Bartlett est la possibilité de définir une valeur seuil pour identifier les maxima locaux qui sont censés représenter une limite entre couches. La définition de ce seuil peut permettre d'envisager simplement l'automatisation de la procédure de découpage. C'est pourquoi, pour l'application à l'essai Panda, ces deux méthodes nous semblent être les mieux adaptées. Concrètement la technique T ratio appliquée sur $\ln(q_d)$ en appliquant une régularisation par l'enfoncement moyen et en fixant une valeur seuil égale à 10, permet d'identifier correctement et clairement les unités de sol homogènes.

Il reste à évaluer la robustesse de ces méthodes pour identifier sur un même terrain où l'on dispose de plusieurs essais, les interfaces des couches malgré la variabilité naturelle du terrain. Cette question sera abordée dans la dernière partie de ce manuscrit. Finalement, si la méthode de découpage proposée dans ce chapitre paraît bien adaptée, efficace et peut être automatisée, elle n'en reste pas moins un outil d'aide pour l'ingénieur, complémentaire des autres analyses constituant une campagne de reconnaissance des sols telles que par exemple les analyses stratigraphiques géologiques et qui ne peut se substituer à celles-ci.

Troisième partie

Identification des sols à partir du signal Panda

10. Introduction et contexte de l'étude

10.1 Classification des sols

10.1.1 Principe

Les systèmes de classification des sols ont pour but de ranger les sols en familles présentant des caractéristiques géotechniques et donc des comportements similaires. Ils permettent à partir des échantillons recueillis au cours d'une campagne géotechnique d'établir des coupes lithologiques du terrain. Il existe plusieurs systèmes de classification des sols fonction de l'application géotechnique pour laquelle elles sont destinées : géotechnique routière, fondations, génie parasismique... Parmi celles-ci, la classification physico-chimique reste la plus utilisée. Les deux paramètres principaux de cette classification sont la granulométrie et la plasticité. Par la suite, on décrit succinctement les deux principales systèmes de classification utilisées en France. Tout d'abord la version française du système international de classification de sols suivi de la classification géotechnique routière française. Les systèmes de classification développés à partir des essais in situ sont abordés à la fin.

10.1.2 Classification des sols LPC/USCS

La classification LPC¹ (Magnan (1997)) est une adaptation de la classification USCS² (ASTM (2010)) aux seuils granulométriques utilisés en France. Elle n'est pas dédiée pour une application géotechnique particulière mais elle est universelle. Cette classification est basée sur les paramètres classiques d'identification des sols : l'analyse granulométrique des sols et les limites d'Atterberg :

- Critères granulométriques :
 - pourcentages de gravier, sable et particules fines (tamisats à 2mm et $80\mu m$) ;
 - forme de la courbe granulométrique, basée sur la valeur du coefficient d'uniformité ou de Hazen $C_U = d_{60}/d_{10}$ et le coefficient de courbure $C_c = (d_{30})^2/d_{10}d_{60}$
- pour les sols fins on se réfère aux critères de plasticité, sur la base de la limite de liquidité w_L , l'indice de plasticité I_p et la ligne A $I_p = 0.73(w_L - 20)$ du diagramme de Casagrande.

Cette classification débouche ainsi sur la distinction de 15 sols types décrits par un symbole à deux lettres. On distingue principalement les sols grenus (passant à $80\mu m \geq 50\%$) notés par S (sables) ou G (graves) (cf.Fig.10.1), et les sols fins (passant à $80\mu m < 50\%$) notés par la lettre A (argiles) ou L (limons) (cf.Fig.10.2). Enfin, il faut noter l'existence

1. Laboratoire des Ponts et Chaussées
2. Unified Soil Classification System

d'une classification LPC modifiée qui est une extension de la classification LPC en vue de mieux décrire les sols organiques (teneur en matières organiques supérieure à 3%) (Magnan (1997)).

Définition			Symbole LCPC	Conditions	Appellation
Graves	Plus de 50 % des éléments de taille supérieure à 80 µm ont un diamètre supérieur à 2 mm	Moins de 5 % d'éléments < 80 µm	Gb	$C_U > 4$ et $1 < C_C < 3$	Grave propre bien graduée
			Gm	Une des conditions de Gb non satisfaite	Grave propre mal graduée
	Plus de 12 % d'éléments < 80 µm	GL	Limites d'Atterberg au-dessous de la ligne A	Grave limoneuse	
		GA	Limites d'Atterberg au-dessus de la ligne A	Grave argileuse	
Sables	Plus de 50 % des éléments de taille supérieure à 80 µm ont un diamètre inférieur à 2 mm	Moins de 5 % d'éléments < 80 µm	Sb	$C_U > 6$ et $1 < C_C < 3$	Sable propre bien gradué
			Sm	Une des conditions de Sb non satisfaite	Sable propre mal gradué
	Plus de 12 % d'éléments < 80 µm	SL	Limites d'Atterberg au-dessous de la ligne A	Sable limoneux	
		SA	Limites d'Atterberg au-dessus de la ligne A	Sable argileux	

FIGURE 10.1 – Classification des sols grenus (plus de 50% des éléments > 80µm)

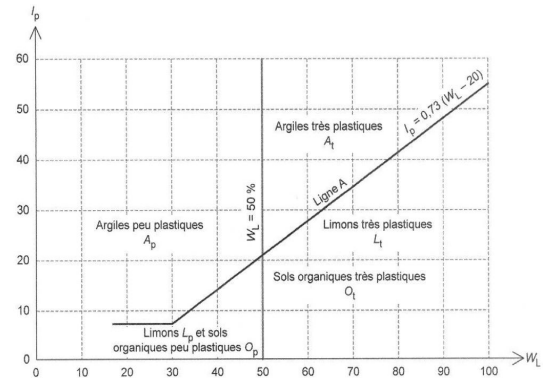


FIGURE 10.2 – Classification des sols fins. Abaque de Casagrande

10.1.3 Classification des sols pour les terrassements routiers

Cette classification dite GTR³ (NF P 11-300 (1992)) a été développée pour la géotechnique routière française, mais elle est aussi couramment utilisée pour les études géotechniques "classiques". Dans un premier temps elle fait la distinction entre :

- les sols ;
- les matériaux rocheux ;
- les matériaux particuliers.

Elle définit des classes de sols corrélées avec l'aptitude au compactage des matériaux en fonction des conditions de chantiers et de leur comportement mécanique ultérieur. Elle tient compte des mêmes caractéristiques que la classification LPC/USCS mais également de la valeur au bleu de méthylène *VBS*. Un premier classement de sols structuré en 4 classes (A, B, C et D), est défini à partir des paramètres de nature (d_{max} , tamisât à 80µm, tamisât à 2 mm et *VBS*) :

- classe A : sols fins ($d_{max} \leq 50mm$ et passant à 80µm > 35%) ;
- classe B : sols sableux et graveleux avec fines ($d_{max} \leq 50mm$) et passant à 80µm ≤ 35% ;
- classe C : sols comportant des fines et de gros éléments ($d_{max} > 50mm$) et passant à 80µm > 12% ou si 80µm ≤ 12% *VBS* est > 0.1 ;
- classe D : sols insensibles à l'eau (passant à 80µm ≤ 12% et *VBS* ≤ 0.1).

Les sous-classes (ex. A1, A2, B1...) affinent la nature du sol en fonction de son argilosité (indice de plasticité I_p et/ou *VBS*).

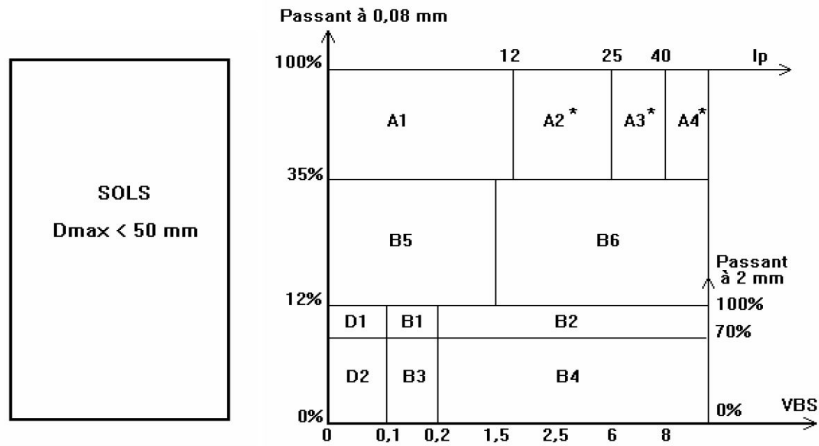


FIGURE 10.3 – Extrait du tableau synoptique (matériaux « Sols » de $D_{max} < 50$ mm) de la classification GTR des matériaux selon leur nature

10.2 Systèmes de classification développés à partir des essais in situ

Les classifications géotechniques présentées dans les paragraphes précédents fournissent une description globale de chaque sol très utile en géotechnique. Cependant, elles nécessitent d'un sondage carotté avec des prises d'échantillons et la réalisation des essais d'identification en laboratoire. Les essais in situ testent directement le sol en place et fournissent une évolution des paramètres mesurés en fonction de la nature et des propriétés mécaniques des sols traversés. Il existe donc un intérêt à corréliser les résultats des essais in situ avec ceux issus des essais d'identification des éprouvettes prélevées sur site. Dans cet esprit, plusieurs systèmes empiriques permettant une estimation de la nature des sols traversés ont été développés sur la base de plusieurs essais géotechniques in situ.

Pour l'essai pressiométrique Ménard, les normes sur les fondations superficielles et profondes (NF P 94-261 (2013)) et (NF P 94-262 (2012)) proposent d'utiliser un abaque de classification basé sur les résultats de cet essai. Il s'agit du diagramme bi-logarithmique mis au point par (Baud (2005)) sur la base de deux paramètres : la pression limite (p_L) et le coefficient rhéologique α égal au rapport du module pressiométrique sur la pression limite (E_M/p_L).

Le sondage de pénétration statique est un essai aveugle mais l'expérience acquise par son utilisation depuis de nombreuses années a permis d'établir des corrélations précises entre les paramètres mesurés pendant le sondages et le type de sol. Plusieurs systèmes de classification ont été ainsi proposés à partir des mesures recueillies par l'essai CPT et CPTU. L'abaque de Robertson est actuellement la référence (Robertson (1990)) repris aussi dans les normes de conception des fondations superficielles et profondes (NF P 94-261 (2013)) et (NF P 94-262 (2012)), il fournit une classification des terrains en fonction de la résistance à la pénétration q_c . L'abaque de Robertson basé sur les paramètres du cône normalisés est alors capable de fournir une classification fiable sans besoin d'échantillonnage. Cependant, (Robertson (2009)) souligne qu'il peut exister des divergences entre les types de sol déduits à partir des abaques, et ceux déterminés au moyen du système de classification classique USCS. Notamment, les

plus grandes différences peuvent avoir lieu pour les mélanges de sols tels que les sables mélangés ou les limons argileux. Par ailleurs, les marnes, les craies ou les roches altérées ne figurent pas dans l’abaque. Malgré tout l’intérêt de cet abaque, dans la pratique courante il est conseillé la réalisation d’un sondage carotté permettant une corrélation locale précise avec un sondage pénétrométrique proche.

Enfin, des classifications sous la forme d’abaques plus récents sont aussi en cours d’élaboration, à partir des informations obtenues lors de la pénétration d’un outil de forage (Reiffsteck et al. (2016)). Toutefois, à l’heure actuelle ces abaques ne sont pas encore exploitables à cause de plusieurs facteurs, entre autres l’importante plage de variation de la base de données prise en compte.

10.3 Problématique

Bien que le pénétromètre dynamique Panda s’avère un essai très performant pour la reconnaissance préliminaire des sites grâce à sa rapidité d’exécution et à son faible coût, il s’agit d’un essai aveugle. À la différence de l’essai de pénétration au carottier (SPT) il n’est pas possible de prélever des échantillons lors de l’essai. On ne peut ainsi pas déterminer la lithologie des terrains traversés puisqu’on ne dispose d’aucune information sur la nature du sol. Dans l’état actuel des connaissances, le seul paramètre mesuré q_d est insuffisant pour situer qualitativement la nature du sol (Waschkowski (1983)). Néanmoins, il est clair que la signature d’un pénétrogramme est liée au type et à l’état des matériaux traversés. Aussi, il apparaît intéressant d’étudier la possibilité de mieux exploiter le signal pénétrométrique Panda en vue d’essayer d’identifier partiellement les sols traversés.

Les systèmes de classification relatifs aux mesures recueillies par des essais in situ s’avèrent être des abaques empiriques simples, développés sur un important jeu de données. Toutefois, il apparaît que l’utilisation conjointe de quantités massives d’informations et d’algorithmes d’apprentissage automatique ou techniques de *machine learning* (ML) permet d’espérer pouvoir améliorer l’interprétation des essais et apporter une solution à un problème considéré jusqu’à pas très longtemps comme inaccessible. La classification s’avère une des principales applications de ces techniques et elle constitue un domaine de recherche en plein essor tandis que son application au domaine de la géotechnique s’avère encore restreinte par rapport aux autres sciences. Dans cette optique, l’essai Panda s’avère un essai particulièrement adapté à ce type de technique puisque du fait de sa résolution d’acquisition, il permet de collecter une grande quantité de données et que par ailleurs, du fait de sa simplicité d’utilisation et de mise en place, le nombre d’essais est facilement multipliable afin de créer une importante base de données. C’est pourquoi une démarche basée sur l’application des techniques de classification automatique permettant de caractériser la nature du sol ausculté semble prometteuse.

Des tentatives de classification empirique basées sur l’essai Panda ont déjà été proposées par (Athapaththu et al. (2015)) dans le cadre du développement d’une méthodologie de reconnaissance des talus naturels difficiles d’accès au Japon. Ces auteurs ont groupé les pénétrogrammes fournis par les essais Panda réalisés dans différents terrains en six familles en fonction de la tendance, de la morphologie et de la profondeur atteinte par les signaux pénétrométriques. Cependant cette classification basée sur l’analyse global du pénétrogramme ne tient pas compte des caractéristiques physiques du matériau et n’est valable que pour le site d’étude analysé.

Avant d'étudier l'adaptation des techniques de machine learning au cas de l'essai Panda, on présente quelques notions de base sur la classification automatique par intelligence artificielle. Un aperçu général des algorithmes les plus répandus, suivi d'une synthèse bibliographique sur leurs applications au domaine de la géotechnique est également présenté.

11. Cadre théorique et démarche proposée

11.1 Classification automatique

11.1.1 Introduction

De manière générale, la classification consiste à étiqueter chaque objet ou individu (typiquement représenté par un vecteur de caractéristiques) en l'associant à une classe (normalement symbolisée par une valeur discrète). Les premières approches proposées étaient des techniques heuristiques basées sur la similarité entre les individus à classer (arbres de décision, analyse discriminante...), détaillés par la suite. Au cours des deux dernières décennies, les algorithmes de *machine learning* (ML), (apprentissage automatique des machines en français) ont été appliqués à divers domaines et sont devenus très populaires.

D'une façon informelle, on pourrait définir l'apprentissage automatique comme la construction d'un modèle à partir des observations d'un phénomène afin que, sans intervention humaine, ce modèle soit capable d'analyser et de réaliser des prévisions du phénomène en question. L'apprentissage automatique se déroule en deux étapes :

- phase d'entraînement : consiste à élaborer le modèle de classification grâce à l'apprentissage d'un jeu de données, connue comme l'ensemble d'apprentissage (*training set*);
- phase de validation ou test : on teste et on valide le modèle sur un nouveau jeu de données appelé ensemble de validation (*validation set*).

Par ailleurs, on peut distinguer quatre types d'apprentissage automatique, parmi les plus répandus on peut citer (Bishop (1995), Jain et al. (1996)) :

- apprentissage supervisé : le modèle est entraîné avec une base de données d'apprentissage (couples entrée-sortie) contenant des exemples permettant au modèle de déterminer lui-même ses propres paramètres ;
- apprentissage non supervisé : il consiste à rechercher des structures naturelles dans les données sans cibler de classe de sortie au préalable. L'analyse par clusters est une technique statistique associée à ce type d'apprentissage ;
- apprentissage par renforcement : il est basé sur l'apprentissage d'une fonction d'utilité ou récompense qui représente une espérance de gain quand une action donnée est choisie. Il s'applique plus à des problèmes d'optimisation.
- apprentissage hybride : c'est un combinaison de l'apprentissage supervisé et non supervisé.

La classification automatique fait appel à l'apprentissage automatique, principale appli-

cation des techniques de ML et domaine de recherche en pleine expansion. Son utilisation dans la discipline connue comme *data mining* ou fouille de données¹ est aussi très répandue.

11.1.2 Algorithmes de classification automatique

Un exposé très général des méthodes de classification automatique est présenté par la suite. Le but est de fournir une liste des techniques couramment employées et de résumer leurs caractéristiques principales. D'après (Han et al. (2011)), on distingue des approches classiques ou traditionnelles et des algorithmes plus récents. Une liste des algorithmes traditionnels de classification automatique est donnée ci-dessous :

- **Arbres de décision inductives** : ce sont des méthodes hiérarchiques représentées par des arbres dont chaque nœud (appelés nœuds de décision) représente une caractéristique ou attribut de l'objet à classer et chaque feuille représente une valeur à prendre en compte par le nœud suivant. La classification commence ainsi à la racine de l'arbre (qui doit être la caractéristique qui divise au mieux les données de l'ensemble d'apprentissage). La classification continue en descendant en fonction des réponses des nœuds de décision. De même, un arbre de décision peut être traduit en un ensemble de règles logiques de type *IF-THEN* en créant une règle distincte pour chaque chemin de la racine à une feuille dans l'arbre. Les lecteurs le désirant pourront avoir un aperçu plus détaillé sur les principes et les travaux menés sur les arbres de décision dans (Murthy (1998)).
- **Classificateur Bayésien Naïf** : est basé sur l'estimation de la probabilité a posteriori grâce au théorème de Bayes. Des connaissances *a priori* peuvent être combinées avec des données d'observation permettant ainsi de calculer explicitement les probabilités *a posteriori* $P(C|X)$ ou la probabilité que l'objet X appartienne à la classe C . L'objet est ainsi classé dans la catégorie qui maximise $P(C|X)$. L'hypothèse de base (dite naïve) est l'indépendance des caractéristiques ou des attributs connaissant la classe cible, ce qui rend le calcul faisable.
- **Analyse discriminante** : elle consiste à approcher une fonction discriminante qui minimise le coût prévu des erreurs de classification. Il s'agit d'une méthode paramétrique puisqu'elle assume que toutes les caractéristiques suivent une distribution normale (hypothèse gaussienne). On distingue l'analyse discriminante linéaire (ADL) quand on suppose que les matrices de covariance des variables explicatives sont égales et l'analyse discriminante quadratique (ADQ) lorsque on ne retient pas cette hypothèse.

On présente maintenant les approches les plus avancées de classification automatique :

- **K-plus proches voisins (k-NN)** : l'une des techniques les plus simples de machine learning est l'algorithme surnommé *nearest neighbour* dont l'apprentissage est basé sur des groupes de données. Il s'agit d'une approche non paramétrique (aucune hypothèse a priori sur la distribution des données). Pour un échantillon inconnu, la méthode k-plus proches voisins (en abrégé k-NN) cherche dans l'espace des caractéristiques des k échantillons (stockés du jeu d'apprentissage) les plus proches de l'échantillon à classer. Cette proximité ou ressemblance est mesurée en terme de distance euclidienne généralement, mais d'autres métriques peuvent être envisagées.

1. La fouille de données est une science pluridisciplinaire portant sur l'ensemble des méthodes utilisées pour l'extraction d'un savoir ou d'une connaissance à partir des bases de données (Fayyad et al. (1996))

- **Réseaux de neurones bayésiens (RB)** : il s'agit de modèles graphiques probabilistes qui décrivent d'une manière qualitative et quantitative les liens de probabilité entre un ensemble de variables données. Les nœuds du graphe représentent les caractéristiques des objets à classer. Au contraire de la classification naïve bayésienne (modèle à caractéristiques statistiquement indépendantes), les réseaux bayésiens permettent une indépendance conditionnelle entre un sous-ensemble de variables. Dans le cadre des RB, il existe deux scénarios possibles, soit la structure du graphe est connue, soit elle est inconnue. Quand on est capable de définir les relations de causalité entre les variables explicatives et les sorties du problème, le graphe peut être construit par un expert. Néanmoins la contrainte inhérente des RB apparaît pour le deuxième scénario compte tenu du temps de calcul que peut nécessiter l'exploration de toutes les structures possibles d'un réseau d'architecture inconnue. En effet, pour un problème décrit par n caractéristiques, le nombre de structures possibles est supérieur à e^n . C'est pourquoi ces systèmes de classification sont déconseillés dès que le nombre de caractéristiques du problème devient considérable.
- **Classification par rétropropagation** : l'algorithme de rétro-propagation du gradient (*backpropagation algorithm*) (Rumelhart et al. (1988)) est la méthode d'apprentissage la plus utilisée pour l'implémentation des réseaux de neurones artificiels (RNA). Les deux familles de RNA les plus couramment utilisés en problèmes de classification (Jain et al. (1996)) sont les réseaux de rétropropagation du gradient, qui comprennent le Perceptron MultiCouche (PMC) et le réseau à fonction de base radiale (RBF). On fournira plus tard une description plus détaillée de ces deux techniques puisqu'elles seront utilisées dans la démarche proposée.
- **Machines à vecteurs de support ou séparateurs à vastes marges** en anglais *Support Vector Machines* (SVMs) : il s'agit de méthodes alternatives récentes pour la classification. Le principe de base de ces algorithmes est de déterminer l'équation d'un classificateur linéaire appelé hyperplan qui permet une séparation optimale des données. Pour deux classes d'exemples donnés, si on les schématise par des « points » dans un plan, on cherche l'hyperplan dont la distance (cette distance est dite « marge ») minimale aux exemples d'apprentissage est maximisée (cf. Fig.11.1). L'hyperplan optimal est celui qui maximise donc la marge et les points les plus proches, qui sont utilisés pour la détermination de l'hyperplan, sont appelés vecteurs de support. Cependant le classificateur de marge maximale ne peut être utilisé que si les classes de données d'apprentissage sont linéairement séparables. Dans le cas non linéairement séparable, l'astuce des SVMs est de changer l'espace des données grâce à une fonction noyau (*kernel*). On applique ainsi une transformation non linéaire aux données permettant une séparation linéaire des exemples dans un nouvel espace de plus grande dimension connue comme espace intermédiaire (*feature space*). La fonction noyau représente un produit scalaire dans l'espace intermédiaire. En pratique, il existe quelques familles de fonctions noyau paramétrables (ex. polynomiale, gaussiennes ou sigmoïde) et il faut choisir le modèle qui convient le mieux. (Hsu et al. (2003)) offre un tutoriel pratique et pragmatique sur le choix de paramètres du modèle SVM vis-à-vis des problèmes de classification. Enfin, il faut signaler que les méthodes SVM sont binaires. Les problèmes de classification à plusieurs classes exigent donc une décomposition en multiple problèmes à deux classes.

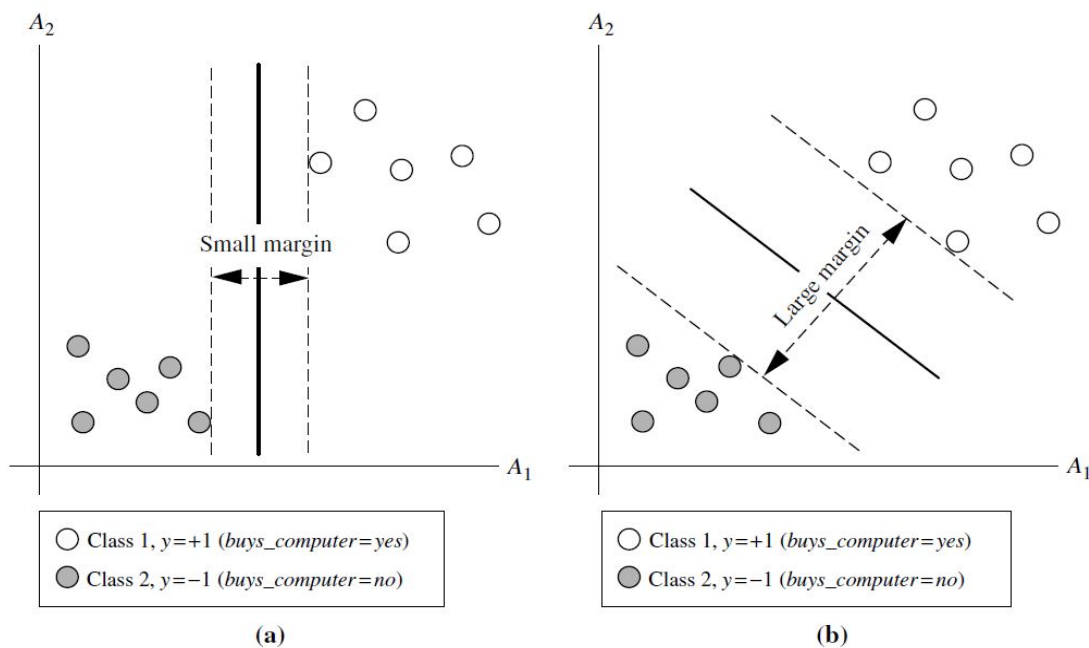


FIGURE 11.1 – Principe de l’hyperplan : pour un ensemble de données linéairement séparables, deux hyperplans séparateurs (a) et (b) possibles. L’hyperplan optimal est celui avec la marge maximale (b) (source : [Han et al. \(2011\)](#))

Bilan

Les techniques statistiques de classification traditionnelles sont souvent des modèles paramétriques qui s’avèrent performants quand les hypothèses de bases sont vérifiées. Leur efficacité dépend des hypothèses du modèle en question, qui doivent être vérifiées par le jeu des données à classer. Par contre les techniques de ML plus récentes sont basées sur un apprentissage par induction de couples d’entrée/sortie permettant ensuite de classer automatiquement des nouvelles données, sans avoir besoin d’établir des hypothèses contraignantes.

Compte tenu que chaque méthode de classification présente ses avantages et limitations, la démarche courante est de tester toutes les techniques appropriées et d’ensuite choisir celle qui fournit les meilleurs résultats. Même si on constate souvent des erreurs de classification pour les mêmes échantillons indépendamment de l’algorithme choisi, la littérature montre que la plupart des observations mal classées par une méthode peuvent être bien classées par une autre (ex. [Tam and Kiang \(1992\)](#) cité par [Kiang \(2003\)](#)). En conséquence, le choix d’un algorithme de classification dépend du problème spécifique à résoudre.

Néanmoins, on trouve quelques études comparatives dans la bibliographie qui fournissent des lignes directrices sur la performance et la validité des différentes techniques. ([Kiang \(2003\)](#)) fournit une évaluation comparative sur plusieurs techniques de classification telles que les méthodes classiques (analyse discriminante linéaire et arbres de décision), la méthode des k plus proches voisins mais aussi des algorithmes basés sur les RNA dont l’architecture implémentée est le PMC. Les expériences menées montrent que les réseaux de neurones offrent des performances supérieures au reste des algorithmes testés. Toutefois, cette méthode est la plus influencée par la taille de l’échantillon d’entraînement, suivi par l’algorithme k-NN. ([Kotsiantis et al. \(2007\)](#)) a aussi étudié la précision des différentes techniques de classifica-

tion automatique sur la base de 45 jeux de données connues. Dans l'ensemble, les méthodes RNA et SVMs fournissent les meilleurs résultats quand les caractéristiques sont des variables continues tandis que les techniques basées sur la logique inductive telles que les arbres de décision sont plus performantes pour la classification des variables discrètes. De même les méthodes SVMs et RNA s'avèrent supérieures vis-à-vis des problèmes de classification dont le nombre de variables explicatives est élevé et quand des relations non linéaires sont présentes entre les variables d'entrée et de sortie. Enfin (Paliwal and Kumar (2009)) offrent une revue bibliographique qui porte sur des études comparatives entre des approches basées sur les RNA (concrètement le PMC) et quelques techniques statistiques classiques vis-à-vis de plusieurs problèmes de classification et régression. En ce qui concerne la problématique de classification, 14 des 23 articles analysés (63%) ont montré que les PMC présentent une performance supérieure par rapport à l'analyse discriminante.

En résumé, plusieurs auteurs signalent les méthodes SVMs et RNA comme les deux algorithmes de classification supervisée les plus avancés et performants dans un contexte général, notamment grâce à leur flexibilité et leurs capacités d'apprentissage non linéaires. Toutefois la plus grande critique souvent adressée à ces deux modèles est qu'il s'agit de boîtes noires. En effet, il s'avère compliqué de comprendre leur mécanisme de fonctionnement et ensuite de tirer des règles pratiques pour améliorer les connaissances sur le phénomène modélisé.

11.1.3 Les techniques de *machine learning* en géotechnique

On présente ici une synthèse bibliographique sur l'application aux problèmes géotechniques des techniques d'apprentissage supervisé aussi connus comme techniques basées sur l'intelligence artificielle (*artificial intelligence (AI) techniques*) dans la littérature. Parmi les algorithmes de *machine learning* présentés auparavant, les RNA sont de loin les plus couramment employés en géotechnique.

Les RNA ont été largement utilisés depuis les années 1990 pour modéliser les relations non linéaires entre plusieurs paramètres des systèmes géotechniques. On commence par les réseaux de neurones bayésiens (RB) dont des exemples d'application sont donnés dans l'article hommage au professeur Wilson Tang (Lacasse et al. (2013)) qui a été un des précurseurs du raisonnement bayésien en géotechnique. Concrètement ce travail illustre leur utilisation pour deux problématiques :

- i. prédiction du risque de glissement terrain par le déclenchement d'un séisme ;
- ii. calculs de vérification de stabilité de talus lors des travaux routiers.

Un bilan détaillé sur l'applications des RNA aux différentes problématiques géotechniques est fourni dans plusieurs travaux (Shahin et al. (2001), Shahin et al. (2005), Shahin et al. (2009) et Sulewska (2011)). Les problématiques abordées par ces modèles sont :

- i. capacité des pieux battus
- ii. tassement des fondations
- iii. propriétés du sol et lois de comportement
- iv. liquéfaction
- v. caractérisation de site
- vi. murs de soutènement
- vii. stabilité de talus
- viii. tunnels

L'évaluation du potentiel de liquéfaction des sols est un sujet de recherche d'actualité dans le domaine et il s'avère un cas d'intérêt pour l'application des RNA. La plupart des méthodes conventionnelles utilisent les diagrammes empiriques basés sur l'indice SPT, la résistance CPT ou la vitesse d'ondes de cisaillement V_s . Une autre alternative de plus en plus étudiée est basée sur l'application des modèles RNA à partir des mesures in situ provenant de l'essai de pénétration au cône (CPT). La méthode consiste essentiellement à entraîner des modèles de RNA au moyen d'une base de données de cas historiques de liquéfaction pour approximer les relations non linéaires entre les divers facteurs qui déterminent le potentiel de liquéfaction et l'occurrence ou pas de ce phénomène. Premièrement (Hsein Juang et al. (1999)) avait étudié des RNA de type PMC pour la prédiction du rapport de résistance cyclique (CRR) selon la méthode basée sur l'indice de comportement (I_c) de Robertson (Robertson and Wride (1998)) et l'approche basée sur des abaques de classification empirique de Olsen (Olsen and Mitchell (1995)). (Juang et al. (2003)) utilise aussi un modèle PMC pour l'identification des sols liquéfiables et propose une méthodologie permettant de déduire une équation empirique à partir du RNA développé. Le fort taux de réussite (96%) atteint avec le PMC développé confirme l'intérêt et l'efficacité de cet outil. Plus récemment, (Goh and Goh (2007)) utilise le SVM comme un algorithme alternatif pour l'évaluation du potentiel de liquéfaction au moyen des informations mesurées par l'essai CPT. Un taux de réussite très élevé similaire à celui reporté par (Juang et al. (2003)) confirme l'intérêt d'une méthodologie basée sur cette technique.

La prédiction des tassements sur dépôts de sols grenus est aussi un des enjeux majeurs vis-à-vis de la conception des fondations superficielles. Il s'agit d'une autre problématique où les RNA apparaissent comme une alternative aux méthodes classiques basées sur les corrélations empiriques développées à partir des mesures des essais in situ. Un exemple notable est issu du travail de (Shahin et al. (2003)), qui a développé un PMC à partir d'une base de données constituée de 189 cas historiques. Le modèle basé sur le RNA s'avère plus performant, en fournissant des prédictions de valeurs de tassements plus précises que celles obtenues au moyen des modèles empiriques classiques. Dans le même esprit, (Samui (2008)) étudie aussi l'application des SVMs pour la prédiction des tassements. L'approche basée sur un SVM avec un noyau gaussien s'avère plus performante que les approches empiriques classiques. De même, les analyses de sensibilité réalisées sur les modèles SVM et RNA ont montré toutes les deux que le nombre de coups de l'essai SPT est le paramètre qui a le plus d'influence dans la prédiction des tassements.

Les RNA ont été aussi appliqués pour prédire la nature des sols à partir des mesures fournies par des essais in situ. (Kurup and Griffin (2006)) proposent l'utilisation d'un réseau de fonction à base radiale (*General Regression Neural Network* GRNN) pour caractériser la répartition granulométrique du sol à partir des mesures recueillies par le CPT. Les quatre variables d'entrée sont : la résistance de pointe statique q_c , le rapport de frottement R_f , la contrainte verticale totale σ_{v0} et la pression hydrostatique initiale u_0 . Les 3 paramètres de sortie sont le pourcentage de sols à matrice fine et à matrice grossière (argile, limon et sable) en se basant sur les tamisats de la classification USCS ($4.75mm$ et $74\mu m$). Le taux de réussite totale obtenu de 86% pour une base de données constituée de 142 effectifs confirme la validité de l'approche. Plus de difficultés ont été constatées pour les sols composés par un mélange de plusieurs fractions ou lorsque les paramètres q_c ou R_f montrent des valeurs atypiques. Un autre exemple d'application des algorithmes de classification automatique aux essais de pénétration est proposé par (Bhattacharya and Solomatine (2006)). Ces auteurs proposent

une méthodologie automatique qui peut être résumée en deux étapes : une première qui comporte un algorithme de segmentation et une deuxième d'application des techniques des algorithmes de classification (arbre à décision, PMC et SVM). Suite à l'étude comparative des trois algorithmes, le PMC est celui qui a classé correctement le plus d'échantillons mais les auteurs précisent qu'une optimisation exhaustive des paramètres du SVM n'a pas été réalisée.

Les algorithmes de *machine learning* ont été aussi appliqués dans le cadre de la caractérisation de site en 3D. (Juang et al. (2001)) proposent un approche basée sur les RNA, notamment un réseau de fonction à base radiale (GRNN) afin de fournir une cartographie de la résistance de pointe à partir des sondages CPT. De même, des approches basées sur les SVMs appliqués aux sondages SPT ont aussi été explorées dans la littérature (Samui (2012)). Les méthodes de ML peuvent donc être utilisées pour imager les propriétés des sols sur la base des mesures réalisées vis-à-vis de la caractérisation du site. Elles s'avèrent toutefois des méthodes déterministes.

Enfin, (Correia et al. (2013)) mettent aussi en évidence les avantages des méthodes de ML plus avancées telles que les RNA et les SVMs par rapport aux approches classiques vis-à-vis de plusieurs problématiques dans le contexte de l'ingénierie des transports (contrôle de remplissage et compactage, évaluation des chaussées ou le comportement mécanique des sols améliorés au moyen de la technique *jet grouting*...).

11.1.4 Bilan

En génie géotechnique, il existe une grande panoplie de problèmes complexes qui ne sont pas complètement maîtrisés. Les modèles analytiques traditionnellement développés reposent souvent sur des hypothèses simplificatrices afin de compléter le manque de compréhension physique du problème. Par ailleurs, les approches traditionnelles nécessitent souvent de fixer la structure du modèle mathématique à l'avance. À cet égard, les approches de ML nécessitent juste l'apprentissage d'un jeu de données constitué par des couples de variables d'entrée et de sortie afin de développer des modèles prédictifs. C'est pourquoi ce sont des outils de modélisation très puissants et de plus en plus populaires.

Parmi les différentes techniques existantes, les RNA apparaissent comme les outils les plus utilisés en géotechnique. Dans un contexte plus large, ils ont eu aussi un grand succès pour leur application à une large gamme de problématiques issues des différents domaines, notamment quand on dispose d'une connaissance limitée du lien entre les données d'entrée et de sortie. Ils s'avèrent être des approches robustes quand le jeu d'apprentissage contient des erreurs et être capables de fournir des précisions souvent élevées.

Les RNA basés sur l'algorithme de rétro-propagation, réseaux dits (*feed-forward*) et notamment les PMC (concrètement l'architecture de type perceptron multicouche) sont les modèles les plus employés vis-à-vis de la classification au sens large (ex. Zhang (2000), Kiang (2003)). Ceci s'explique du fait d'une des propriétés les plus remarquables de ces modèles : ils sont des « approximateurs universels », c'est-à-dire qu'une fois l'apprentissage complété, ils sont capables d'approcher une fonction quelconque avec une précision arbitraire. La démonstration rigoureuse de ce théorème est donnée dans (Hornik et al. (1989)). Les approches basées sur les RNA ont donc été privilégiées dans ce travail. Leurs principes et caractéristiques principales seront présentés dans la section suivante.

11.2 Les réseaux de neurones artificiels (RNA)

11.2.1 Généralités

Historique

La recherche sur les RNA peut être divisée en 3 périodes d'activité ([Anderson et al. \(1993\)](#)) :

- la naissance du domaine remonte aux années 1940 grâce aux travaux pionniers de Warren McCulloch et Walter Pitts sur les principes théoriques des modèles mathématiques de neurones ;
- la première application concrète des RNA est survenue vers 1960 avec le modèle de *Perceptron* créé par Rosenblatt et ses limitations ont été mises en évidence par Marvin Minsky et Seymour Papert. Concrètement ces auteurs ont démontré que ce type de réseau ne pouvait résoudre que des problèmes linéairement séparables ;
- presque 20 ans après, la communauté de chercheurs a attiré à nouveau l'attention sur les RNA. Depuis le début des années 1980, l'invention de l'algorithme de rétropropagation de l'erreur pour les perceptrons multicouches a permis une résurrection du domaine de RNA. Il a été tout d'abord proposé par Paul Werbos et par Yann LeCun et ensuite modifié et popularisé par David Rumelhart et James McClelland.

À l'heure actuelle les modèles de RNA et leurs applications à de nombreux domaines sont de nouveau des sujets de recherche d'actualité.

Fonctionnement d'un neurone artificiel

Le modèle mathématique d'un neurone artificiel est illustré à la figure ([Fig. 11.2](#)). Il s'agit d'un neurone avec R entrées représentées par la lettre (p) et une valeur scalaire comme sortie (a). Chaque entrée individuelle $\{p_1, p_2, \dots, p_R\}$ est pondérée par son poids correspondant $\{w_1, w_2, \dots, w_R\}$. Cette somme pondérée est ensuite affectée par un biais (b) et on applique une transformation via une fonction de transfert (f) :

$$a = f(n) \quad n = b + \sum_{i=1}^R w_{1,i} p_i \quad (11.1)$$

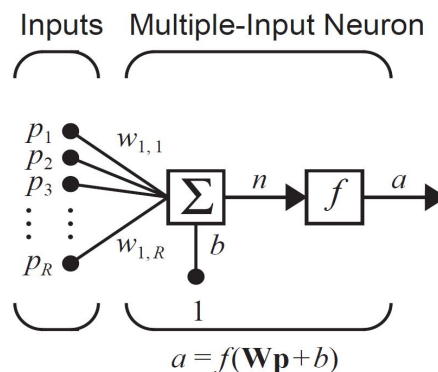


FIGURE 11.2 – Schéma de neurone formel (source : [Demuth et al. \(2014\)](#))

La fonction de transfert, aussi connue comme fonction d'activation, peut être une fonction mathématique linéaire ou non linéaire. Les quatre modèles les plus utilisés (Demuth et al. (2014)) sont :

$$\text{Fonction seuil : } \begin{cases} a = 0 & \text{si } n < 0 \\ a = 1 & \text{si } n \geq 0 \end{cases} \quad (11.2)$$

$$\text{Fonction linéaire : } a = n \quad (11.3)$$

$$\text{Fonction sigmoïde : } a = \frac{1}{1 + e^{-n}} \quad (11.4)$$

$$\text{Fonction tangente hyperbolique : } a = \frac{e^n - e^{-n}}{e^n + e^{-n}} \quad (11.5)$$

Architectures des RNA

On peut concevoir les RNA comme des graphes orientés dans lesquels les neurones artificiels sont les nœuds et les liens ou arêtes (avec un poids) sont les connexions entre les sorties des neurones et les entrées des neurones. Le schéma de connexion entre neurones détermine l'architecture du réseau. En fonction de l'architecture du RNA, on peut distinguer deux grandes familles (cf. Fig.11.3) :

- i. « feed-forward » : ou réseaux de « propagation vers l'avant » en français. Il s'agit des graphes acycliques ou statiques puisque l'information se propage toujours dans le sens de l'entrée vers la sortie du réseau ;
- ii. « feed-back » : aussi connus comme réseaux de neurones récurrents ou bouclés. Ce sont des graphes cycliques, dans lesquels l'information peut se propager dans les deux sens. Ils sont très utilisés pour le traitement des séquences temporelles de taille variable avec des données qui ne sont pas indépendantes les unes des autres, telles que les cours de la bourse, signaux de la voix humaine, écritures, etc.

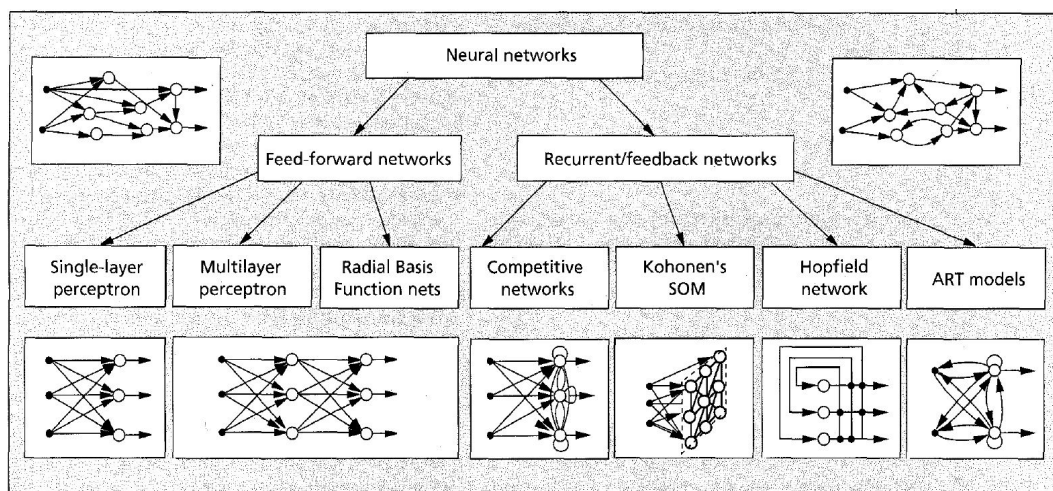


FIGURE 11.3 – Taxonomie des architectures des réseaux de neurones « feed-forward » et « feed-back » (source : Jain et al. (1996))

11.2.2 Réseaux de neurones multicouches « feed-forward »

Un réseau multicouche « feed-forward » est organisé en couches successives. Une couche représente un regroupement de neurones sans connexion les uns avec les autres mais tout neurone d'une couche est connecté à tous les neurones de la couche suivante. La structure de ces réseaux comporte ainsi au moins 3 couches, à savoir (cf. Fig.11.4) :

- une première dite "couche d'entrée" (*input layer*) où chaque neurone stocke un paramètre d'entrée ;
- ensuite une ou plusieurs couches cachées (*hidden layer*) aussi dites "couches intermédiaires" ou "de calcul". Le nombre de couches cachées ainsi que le nombre de neurones par couche sont des paramètres du réseau à déterminer ;
- la dernière unité est nommée "couche de sortie" (*output layer*). Pour un problème de classification, le nombre de neurones de cette couche est égal au nombre de classes du problème en question.

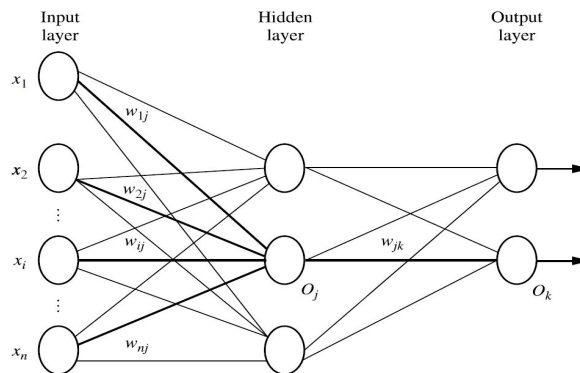


FIGURE 11.4 – Structure d'un réseau multicouche *feed-forward* (source : Han et al. (2011))

On note que la couche d'entrée n'exécute aucun calcul. C'est pourquoi, pour compter le nombre de couches totales du réseau M , par convention il est courant de ne pas tenir compte de cette première couche. Selon cette notation un RNA avec un nombre de couches égal à M aura $M - 1$ couches cachées.

Le perceptron multicouche (PMC)

Il s'agit d'un réseau de type « feed-forward » très populaire grâce à l'apprentissage par rétropropagation de l'erreur. Cet apprentissage s'avère un processus dynamique et itératif permettant de modifier les poids synaptiques w_{ij} du réseau en fonction des Q couples des variables d'entrée et sortie :

$$\{p_1, t_1\}, \{p_2, t_2\}, \dots, \{p_Q, t_Q\} \quad (11.6)$$

avec p_Q une entrée du réseau et t_Q la sortie cible correspondante. Pour chacune des entrées fournies, on calcule l'erreur entre les valeurs de sorties attendues et les sorties prédites par le réseau ($t - a$). Le but final de l'algorithme par rétro-propagation est la minimisation de cette fonction d'erreur. Dans la pratique l'erreur quadratique moyenne est couramment utilisée (Zhang (2000), Demuth et al. (2014)) :

$$F(\mathbf{x}) = E [e^2] = E [(t - a)^2] \quad (11.7)$$

avec \mathbf{x} le vecteur de poids et biais du réseau.

Ci-dessous on résume les principales étapes de l'algorithme de rétropropagation pour un PMC :

- (1) Initialiser tous les poids w_{ij}^m à des petites valeurs (normalement dans une plage comprise entre -1 et 1 ou -0.5 et 0.5)
- (2) Pour chaque couple $\{p_Q, t_Q\}$ de l'ensemble d'entraînement ou d'apprentissage²
 - (2.1) Propager les valeurs d'entrées à travers les différentes couches du réseau en calculant la sortie de chaque neurone (éq.11.1)
 - (2.2) Pour chaque neurone j de la couche de sortie calculer l'erreur δ_j :

$$\delta_j^M = f'(n_j^M)[t_j^Q - a_j^M] \quad (11.8)$$

avec f' la dérivée de la fonction de transfert ou d'activation f .

- (2.3) Propager les erreurs vers l'arrière, pour chaque neurone j dans les couches de calcul :

$$\delta_j^m = f'(n_j^m) \sum_j w_{ij}^{m+1} \delta_j^{m+1} \quad \text{pour } m = (M-1), \dots, 1 \quad (11.9)$$

- (2.4) Mise à jour des poids dans toutes les couches :

$$w_{ij}^m = \alpha \delta_j^m a_j^{m-1} \quad (11.10)$$

avec α une constante qu'on appelle taux d'apprentissage, avec une valeur comprise entre 0 et 1.

- (3) L'algorithme est fini quand on atteint le critère d'arrêt. Sinon il retourne à l'étape (2)

La performance du PMC est contrôlée principalement par trois facteurs : l'architecture en termes de nombre de couches et de neurones de calcul par couche, les fonctions d'activation choisies et les valeurs des poids synaptiques. C'est pourquoi, plusieurs enjeux interviennent vis-à-vis de la conception d'un PMC :

- combien de couches cachées sont nécessaires pour une tâche donnée ?
- combien de neurones de calcul sont nécessaires par couche ?
- quelle est la performance du réseau pour un ensemble de données indépendant de celui d'apprentissage (capacité de généralisation) ?
- quelle est la taille nécessaire de l'ensemble d'apprentissage ?

La contrainte majeure de ces réseaux est qu'il n'existe pas de règles explicites permettant de décider du nombre de couches cachées et de leurs nombres de neurones. Pour un PMC avec une seule couche cachée, il existe dans la littérature plusieurs règles empiriques pour estimer le nombre de neurones de calcul, détaillées dans (Shahin et al. (2008)). Toutefois, une méthode "essai et erreur" est souvent employée. Tout d'abord, on fixe le nombre de couches cachées et ensuite on choisit le nombre de neurones de chaque couche. Une interprétation géométrique du rôle de la couche cachée vis-à-vis de la discrimination des objets est fournie dans la figure (Fig. 11.5).

2. Le choix des associations des entrées et sorties est fait de manière aléatoire pour s'assurer que le réseau apprend vraiment les relations entre les variables d'entrée et sortie plutôt qu'une séquence des couples de variables.


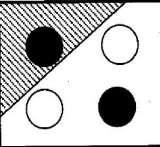
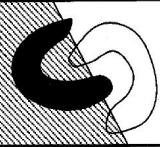
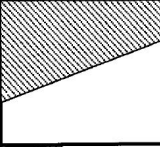
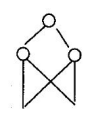
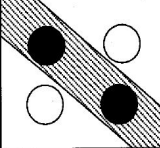
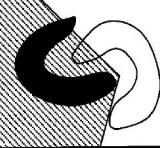
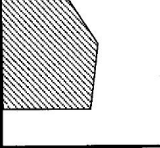
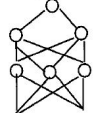
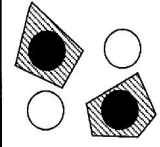
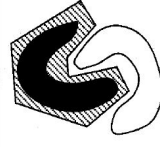
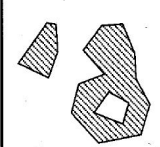
Structure	Description of decision regions	Exclusive-OR problem	Classes with meshed regions	General region shapes
 Single layer	Half plane bounded by hyperplane			
 Two layer	Arbitrary (complexity limited by number of hidden units)			
 Three layer	Arbitrary (complexity limited by number of hidden units)			

FIGURE 11.5 – Interprétation géométrique du rôle du nombre de couches cachées pour un problème de classification dans un espace d’entrées 2D (source : Jain et al. (1996))

De manière générale, la complexité des frontières de discrimination entre les classes des objets augmente avec le nombre de couches du réseau. En revanche, la capacité de généralisation du réseau peut être influencée de manière négative et des problèmes de sur-apprentissage (*overfitting* en anglais) peuvent avoir lieu. Le sur-apprentissage survient quand un RNA avec une bonne performance lors de son entraînement ne fournit pas des prédictions correctes sur des jeux de données qui ne font pas partie de la base d’apprentissage. En d’autres termes, si le nombre de degrés de liberté du modèle est grand par rapport au nombre de variables explicatives du problème et la taille de l’échantillon d’entraînement, le modèle peut apprendre les particularités ou les singularités des données au lieu de « trouver » une règle générale qui permettrait de résoudre le problème. En conséquence, l’architecture optimale contiendra juste le nombre de couches et de neurones nécessaires.

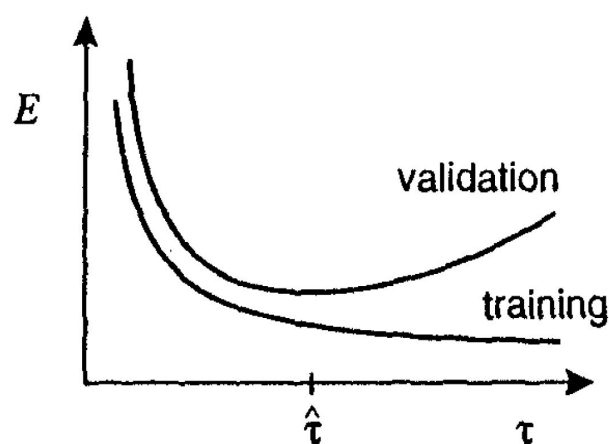


FIGURE 11.6 – Illustration schématique de l’évolution des erreurs pour les ensembles d’apprentissage et validation en fonction de la période d’entraînement τ . L’objectif d’aboutir à la meilleure capacité de généralisation suggère que l’entraînement doit être arrêté à l’itération $\hat{\tau}$ équivalente à la valeur minimale de l’ensemble de validation (source : Bishop (1995)).

Plusieurs stratégies peuvent être mises en œuvre afin d'éviter les problèmes de sur-apprentissage :

- Division du jeu de données : le modèle du RNA doit être validé sur des ensembles de données indépendants de la base d'apprentissage. La démarche courante dans la pratique est dite *hold-out* (Bishop (1995)). Elle consiste à partitionner les données disponibles en trois sous ensembles, un d'entraînement pour réaliser l'apprentissage, et deux ensembles pour valider la performance du modèle, à savoir : un ensemble "validation" et un ensemble "test". Quand on divise le jeu de données, environ 70% est normalement utilisé pour l'apprentissage, 15% pour la validation et 15% pour le test (Demuth et al. (2014)).

Compte tenu que les RNA sont conditionnés par les données d'entraînement, des techniques basées sur des test statistiques de comparaison des échantillons peuvent être envisagées pour une division optimale des données. (Shahin et al. (2004)) ont proposé une méthodologie basée sur l'analyse par clusters permettant de diviser les données d'une manière rigoureuse et systématique. D'ailleurs (Shahin et al. (2008)) fournissent une révision complète sur le sujet de la division de données dans le cadre de l'application des RNA en géotechnique.

- Critère d'arrêt : la rétro-propagation des erreurs n'offre aucune garantie quant à la convergence de l'algorithme vers le minimum global de la fonction d'erreur. Des critères d'arrêt de l'algorithme s'avèrent alors nécessaires. L'approche la plus simple consiste à limiter la durée de la phase d'apprentissage avec un nombre maximum de périodes d'entraînement. Une autre démarche consiste à fixer une valeur seuil sur la fonction d'erreur retenue. Une approche qui permet de surmonter le problème de sur-apprentissage est le critère d'arrêt basé sur la validation croisée (*cross-validation*), connue comme *early stopping* (Demuth et al. (2014)). Le principe est simple, l'apprentissage est interrompu lorsque la fonction d'erreur issue de l'ensemble de validation n'exhibe pas une tendance décroissante et commence à augmenter (cf. Fig.11.6). Ce critère est recommandé quand le nombre de données est faible et c'est une méthode très efficace contre le sur-apprentissage (Smith (1993) cité par Shahin et al. (2008)).
- Régularisation : cela consiste à modifier la fonction d'erreur en rajoutant un terme d'inertie qui pénalise la complexité du réseau. Ceci entraîne un lissage de la fonction d'approximation modélisée par le réseau, ce qui est équivalent à réduire le nombre de neurones cachés. La régularisation bayésienne s'avère être une version plus avancée de cette approche qui permet de calculer la valeur optimale de ce paramètre. Pour une description plus détaillée de ces techniques et de leur implémentation, le lecteur intéressé peut se référer à l'ouvrage de (Demuth et al. (2014)).

Fonctions de base radiale : réseaux de neurones probabilistes

Les réseaux dits à fonction de base radiale ou « Radial Basis Function network » (RBF) en anglais sont un cas particulier des réseaux PMC (Jain et al. (1996)). L'algorithme de rétro-propagation d'erreur dans le dimensionnement d'un PMC peut être regardé comme un problème d'approximation stochastique puisque, en fin de compte, il s'agit d'une technique itérative d'optimisation. La conception du RBF représente une perspective différente puisque l'apprentissage de ces réseaux est équivalent à trouver la surface dans un espace multidimensionnel qui s'ajuste au mieux aux données selon un critère statistique donné. Ces

réseaux peuvent être développés pour résoudre des problèmes de classification avec les réseaux probabilistes ou (« probabilistic neural network » (PNN)) (Specht (1990)) mais aussi de régression avec les réseaux de régression ou « general regression neural network » (GRNN) (Specht (1991)). L'architecture de ces réseaux est constituée par trois couches (cf. Fig. 11.8).

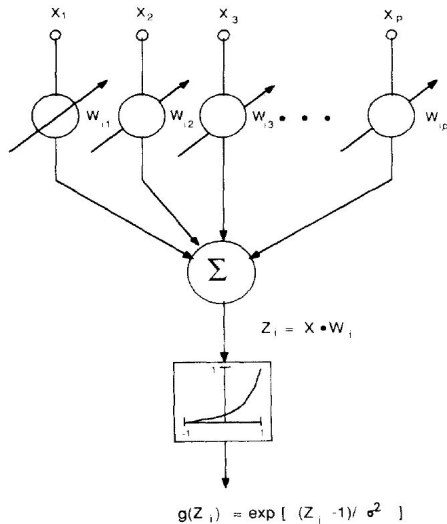


FIGURE 11.7 – Détail du neurone de la couche cachée (*pattern unit*) (source : Specht (1990))

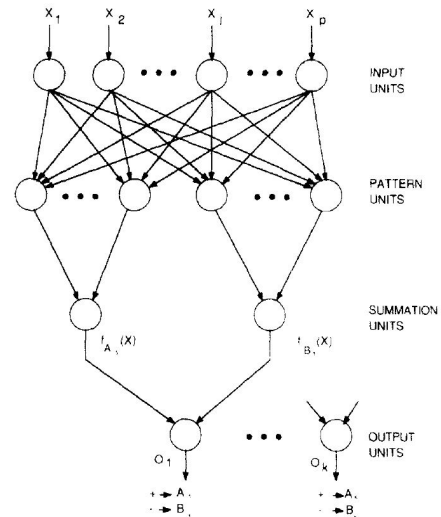


FIGURE 11.8 – Architecture du réseau de neurones probabiliste (source : Specht (1990))

Une première couche d'entrée dont le rôle est le stockage des paramètres d'entrée. La deuxième couche est la seule couche cachée du réseau dont les neurones ont une fonction à base radiale comme fonction de transfert (*pattern units*). Finalement une troisième couche appelée "couche d'agrégation" (*summation units*) généralement linéaire est responsable de fournir les sorties du réseau.

Il faut noter que la principale particularité des RBF est que les neurones de la couche cachée utilisent une fonction radiale comme fonction d'activation. Notamment le modèle courant est une fonction de type gaussien :

$$\exp\left(-\frac{(\mathbf{X} - \mathbf{W}_i)^T(\mathbf{X} - \mathbf{W}_i)}{2\sigma^2}\right) = \exp\left(-\frac{\|\mathbf{X} - \mathbf{W}_i\|^2}{2\sigma^2}\right) \quad (11.11)$$

avec \mathbf{W} les poids et \mathbf{X} les entrées du réseau et $\|\cdot\|$ désigne la norme euclidienne. La valeur de sortie des neurones dépend alors d'un voisinage autour du centre. Le centre est représenté par le vecteur de poids de dimension p , avec p le nombre de paramètres d'entrée tel que montré dans la figure (Fig. 11.7). Les neurones travaillent ainsi localement dans l'espace d'entrées et la sortie du neurone tend vers 0 dès qu'on s'éloigne du centre. L'écart-type σ ou variance σ^2 de la fonction d'activation gaussienne est appelé dans la littérature « spread constant » ou paramètre de lissage « smoothing parameter ». Il s'agit d'un facteur d'échelle de l'influence de voisinage des neurones. Quand $\sigma \rightarrow 0$ le comportement du PNN est assimilable à celui d'un classifieur de type *plus proches voisins* alors que la frontière discriminante entre classes tend vers un hyperplan souple et dérivable lorsque $\sigma \rightarrow \infty$.

3. Les paramètres d'entrée sont notés par \mathbf{X} afin d'être cohérents avec les figures présentées par la suite.

Pour un problème de classification, la couche d'agrégation est constituée d'autant de neurones que de classes cibles présentes dans le problème. Le neurone de la couche d'agrégation avec la valeur la plus élevée détermine la classe d'appartenance. Ceci peut être interprété en termes de probabilité tel qu'expliqué par (Specht (1990)), et on peut appliquer le théorème de Bayes pour estimer les probabilités a posteriori d'appartenance à chaque classe.

Pour entraîner les réseaux RBF plusieurs stratégies peuvent être envisagées. Compte tenu que les poids associés à la couche linéaire de sortie ont tendance à évoluer beaucoup par rapport aux fonctions d'activation non linéaires de la couche cachée, leur apprentissage est abordé en deux étapes distinctes :

- une première basée sur les techniques d'optimisation non linéaires consistant à estimer la position des centres des neurones de la couche cachée, leur écart-type ou paramètre de lissage σ ;
- la deuxième étape a pour but l'estimation des poids de la couche linéaire. Pour plus de détails sur les différents algorithmes d'entraînement de ces réseaux, le lecteur intéressé peut consulter (Demuth et al. (2014)).

Bilan entre PMC et RBF

RBF et PMC sont deux réseaux « feed-forward » considérés comme des approximateurs universels. Cependant ils diffèrent l'un de l'autre par les aspects suivants :

- i. Un réseau RBF a juste une couche cachée alors qu'un PMC peut en avoir plusieurs. D'ailleurs, on remarque qu'à contrario des réseaux PMC, le seul paramètre à déterminer des RBF est le paramètre de lissage σ ;
- ii. la couche cachée d'un réseau RBF est non linéaire avec une fonction d'activation normalement gaussienne, tandis que la couche de sortie est linéaire. Cependant, les couches cachées et de sortie d'un PMC utilisé pour un problème de classification sont généralement non linéaires. Quand un PMC est destiné à résoudre un problème de régression non linéaire, une couche de sortie avec une fonction de transfert linéaire s'avère le choix habituel ;
- iii. la fonction de transfert de chaque neurone caché d'un RBF calcule la norme euclidienne (distance) entre le vecteur d'entrées et le centre de chaque unité. Par contre, la fonction d'activation de chaque neurone caché d'un PMC calcule le produit scalaire entre le vecteur d'entrées et les poids synaptiques du vecteur de chaque unité. Deuxièmement, au lieu d'ajouter un biais b comme pour le PMC, on multiplie par le biais $b=1/(2\sigma^2)$;
- iv. les RBF sont ainsi caractérisés par un apprentissage local, chaque paramètre d'entrée influence certaines unités cachées alors que pour un PMC, chaque paramètre d'entrée peut être pris en compte simultanément par tous les neurones de calcul. C'est pourquoi, au contraire des RBF, les PMC apprennent de façon globale. Grâce à cette qualité, certains auteurs ont jugé l'utilisation des réseaux RBF comme un outil d'interpolation mieux adapté que les PMC vis-à-vis de la caractérisation géotechnique d'un site (Juang et al. (2001)), et ils ont mis en relation l'écart-type de la fonction d'activation, avec la longueur de corrélation d'une fonction de corrélation exponentielle.

11.3 Démarche proposée

L'objectif de cette partie est d'étudier et de développer une méthodologie capable, à partir de l'analyse des signaux pénétrométriques mesurés à l'aide de l'essai Panda, de fournir une estimation de la nature du sol traversé. Pour ce faire, il pourrait être avantageux, tel qu'il a été présenté, de mettre en œuvre des méthodes de classification automatique basées sur les RNA. Deux typologies ont notamment été ciblées : le PMC et le PNN. En effet, ces deux architectures s'avèrent des algorithmes très performants et usités pour des problèmes de classification supervisée. Or il s'agit bien du cas que l'on cherche à traiter, puisque finalement dans notre problème, on maîtrise les classes de sortie (les classes de sol), que l'on cherche à déterminer. De plus, ces deux algorithmes sont robustes puisqu'ils sont capables d'effectuer une généralisation à partir d'un nombre relativement faible de données. Aujourd'hui, des architectures faisant partie de l'apprentissage profond (*deep learning*) proposent des algorithmes encore plus développés et performants. Mais la mise en œuvre de ces approches exige une quantité des données beaucoup plus importante que celle dont on dispose actuellement dans les banques de données de sondages Panda. C'est pourquoi ces techniques n'ont pas été testées dans ce travail. Pour plus de renseignements sur les différents travaux menés avec des RNA reposant sur le *deep learning* on renvoie le lecteur à ([Schmidhuber \(2015\)](#)).

Dans la figure (Fig.11.9) on présente un schéma descriptif de la stratégie de la méthodologie proposée.

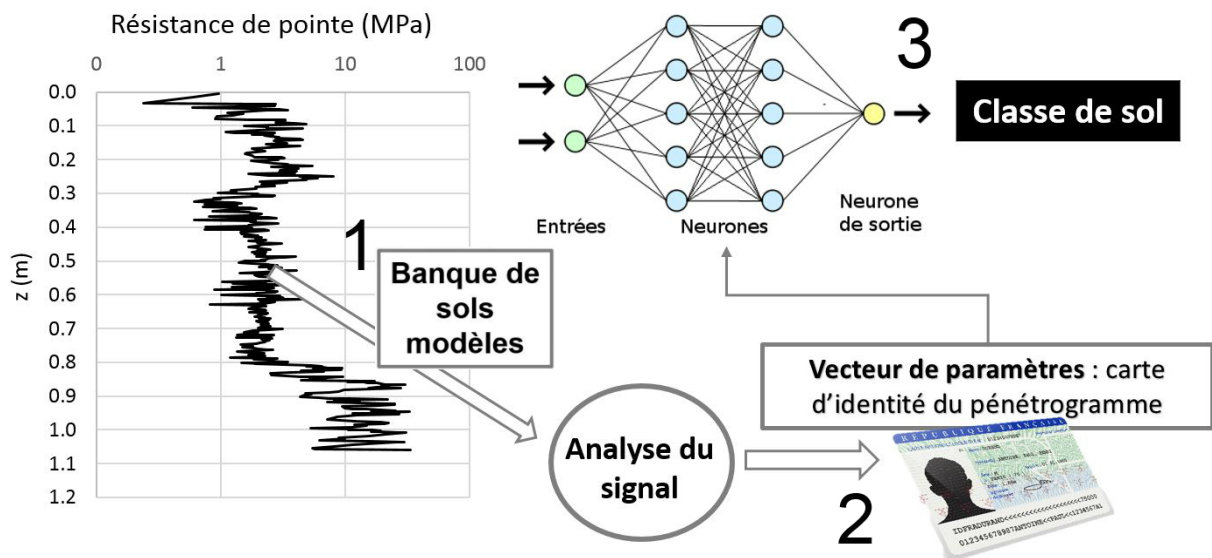


FIGURE 11.9 – Schéma de principe de la méthodologie proposée.

Sa mise en œuvre peut être décomposée en plusieurs étapes. Chacune d'elle est discutée par la suite vis-à-vis de la problématique abordée dans ce travail :

- i. Acquisition de données : les algorithmes de ML nécessitent un entraînement sur une base de données la plus représentative possible pour la problématique à aborder. La première phase consiste donc à recompiler les données afin de créer une banque de données. Dans notre cas, nous nous sommes appuyés sur des pénétrogrammes issus des essais de pénétration réalisés au pénétromètre Panda avec les paramètres

d'identification mesurés en laboratoire sur des matériaux réels "modèles", mais aussi sur des pénétrogrammes issus d'essais réalisés in situ ;

- ii. Définition des entrées du modèle : lors de cette étape on cherche à déterminer un vecteur de variables d'entrée représentatif des objets à classer vis-à-vis du problème de classification à aborder.

Il est bien connu que deux sols de natures différentes peuvent présenter des valeurs de résistance à la pénétration similaires. Il n'y a pas une correspondance biunivoque entre la résistance de pointe et la nature du sol et donc la valeur de q_d sans aucune autre information ne permet pas de déceler le type de matériau testé. Néanmoins plusieurs systèmes de classification empiriques ont été développés autour de la résistance de pointe et elle s'avère être un paramètre clé des abaques de classification développés pour l'essai de pénétration statique.

En fin de compte, un pénétrogramme représente la réponse d'un sol dans un état donné en fonction de la profondeur à une sollicitation mécanique à la rupture. Plusieurs études ont démontré la reproductibilité et la répétabilité du signal pénétrométrique fourni par l'essai Panda in situ et en laboratoire pour des formations homogènes (Chaigneau (2001)). Les pénétrogrammes mesurés dans un matériau homogène sont les mêmes. C'est pourquoi, au lieu de tenter d'établir une corrélation traditionnelle entre les mesures individuelles de résistance de pointe q_d ou une valeur moyenne de q_d avec le type de sol, on propose d'analyser la totalité du pénétrogramme. L'idée est alors d'appliquer différentes analyses du signal au pénétrogramme afin de créer un paramétrage qui va être l'identifiant de l'essai de pénétration vis-à-vis du système de classification. On cherche à créer ainsi une mémoire associative intelligente grâce aux RNA entre les pénétrogrammes et les classes de sol à définir selon les objectifs visés.

- iii. Définition des sorties du modèle : l'étape suivante consiste à fixer les critères pour définir le nombre de classes du modèle à développer. Les classifications géotechniques utilisées dans la pratique seront la référence à suivre.
- iv. Apprentissage du système : cette phase consiste à entraîner le réseau grâce à la base de données élaborée, en prenant en compte les différents volets évoqués dans l'état de l'art (division du jeu de données, problèmes de l'apprentissage...) et les particularités de la typologie du réseau à développer.
- v. Validation et capacité de généralisation du modèle : finalement on réalisera une analyse des résultats fournis par le RNA et on testera sa capacité de généralisation vis-à-vis de nouveaux sols non présents dans la base de données initialement créée (base de sols "modèles").

Par la suite, l'application de la méthode proposée sera étudiée pour deux problématiques :

1. une première dans laquelle on cible la caractérisation de sols naturels à partir d'une banque de données composée d'essais maîtrisés en laboratoire et d'essais in situ ;
2. une deuxième dans laquelle on présente un cadre d'application plus restreint, où on cherche à développer une classification pour un matériau industrialisé (Expangel). Il s'agit d'un matériau granulaire argileux (bentonite) formé par un mélange de billes sèches et de matériau concassé.

12. Application à la classification de la nature du sol

12.1 Développement du système de classification

12.1.1 Élaboration de la base de données

Un problème de classification automatique commence avec l'élaboration d'un jeu de données permettant de créer une base d'apprentissage sur laquelle caler le modèle de classification. Pour cela, on a élaboré une base contenant les sondages Panda ainsi que les caractéristiques granulométriques des matériaux testés. La base de données (BDD) créée comprend des sondages réalisés en laboratoire en conditions contrôlées mais aussi des sondages in situ. Les essais en laboratoire étant réalisés en conditions maîtrisées (de teneur en eau et de densité), ils sont plus homogènes que les essais in situ et devraient permettre de mieux calibrer les paramètres représentatifs des classes de sols. Les essais in situ quant à eux, sont plus représentatifs de l'état réel des matériaux et devraient permettre de rendre la méthode de classification automatique plus robuste en intégrant la variabilité naturelle des caractéristiques des sols testés.

En ce qui concerne les essais de laboratoire, il s'agit de sondages Panda réalisés dans des moules de calibration de 80 cm de hauteur, pour plus d'une vingtaine de sols « modèles »¹ pour lesquels on dispose de l'ensemble des informations de caractérisation géotechnique (Chaigneau (2001)). L'étendue de l'ensemble des paramètres d'état des matériaux testés inclus dans la base des essais de laboratoire sont fournis dans le tableau (Tab.12.1) : la teneur en eau w , la valeur au bleu de méthylène VBS , le poids volumique sec γ_d , les tamisats à 2mm et 80 μm . La répartition des sols testés en laboratoire et in situ en fonction de leurs tamisats à 2 mm et à 80 μm est représentée dans les graphiques de la figure (Fig.12.1). Concernant les données pénétrométriques enregistrées sur ces matériaux, ne seront prises en compte que les données du pénétrogramme situées au dessous de la profondeur critique z_c où le signal pénétrométrique est stable en moyenne et en variance. On précise que ces sondages ont été réalisés avec une pointe de 2cm².

Par ailleurs, la banque de données in situ a été élaborée à partir des essais réalisés sur le terrain lors de campanes géotechniques menées principalement en région Auvergne-Rhône-Alpes. Pour chacun des essais, on dispose de l'identification des matériaux traversés grâce aux essais de caractérisation de laboratoire sur les échantillons prélevés au moyen d'un carottage

1. On appelle sols modèles des matériaux faisant partie d'une banque de sols stockée physiquement au laboratoire IP à Polytech Clermont-Ferrand et pour lesquels on a déterminé les caractéristiques physiques et calibré un certain nombre de méthodologies développées au sein du laboratoire

ou lors d'une ouverture à la pelle réalisée au voisinage du sondage Panda (à moins de 5 m). Une profondeur limite égale à 5 m a été fixée afin de négliger les effets éventuels de la pression de confinement. Les zones des sondages pénétrométriques qui vont servir à l'apprentissage des réseaux de neurones ont été choisies sur la base d'une analyse experte des sondages en recoupement avec les couches identifiées visuellement et au sein desquelles des échantillons ont été prélevés.

Caractéristiques	min	max
Passant à $80\mu\text{m}$ (%)	0	96.73
Passant à 2mm (%)	0	100
C_U	1.4	62.5
C_C	0.5	2.57
w (%)	3.01	12.9
VBS (ml/g)	0.06	2.73
IP (%)	2.61	18.9
$\gamma_d(kN/m^3)$	14.17	21.02

TABLE 12.1 – Plage des paramètres physiques des matériaux en laboratoire

Caractéristiques	min	max
Passant à $80\mu\text{m}$ (%)	29.3	98.9
Passant à 2mm (%)	66.6	100
w (%)	12.6	71
VBS (ml/g)	0.46	12.61

TABLE 12.2 – Plage des paramètres physiques des matériaux in situ

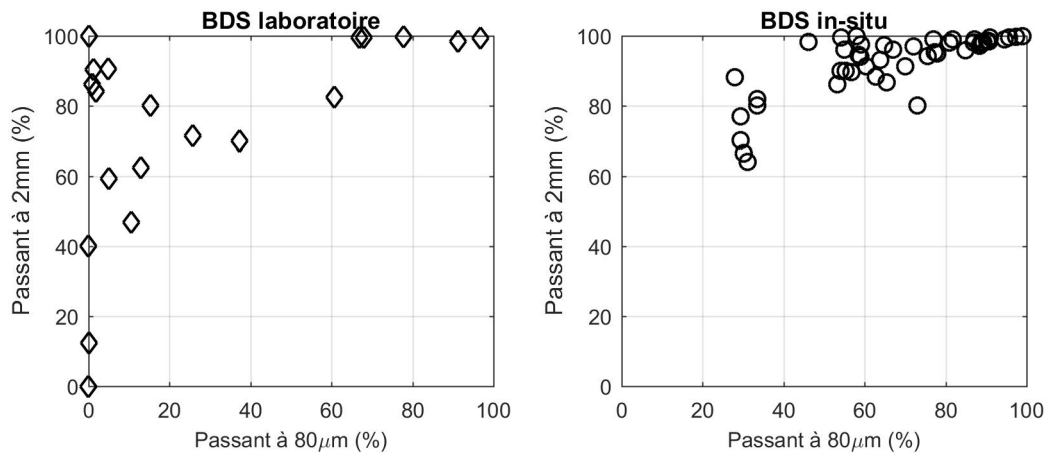


FIGURE 12.1 – Tamisats à 2 mm et $80\mu\text{m}$ des matériaux de la banque de sondages (BDS) de laboratoire (gauche) et in situ (droite)

En outre, les sondages et les échantillons pris en compte dans la base de données sont situés au-dessus de la nappe phréatique. Les paramètres d'état des sols testés ont été regroupés dans le tableau (Tab.12.2). On peut remarquer que les valeurs de teneur en eau pour les sols naturels sont plus importantes que celles des échantillons de laboratoire, dont les états hydriques pris en compte sont "sec" et "moyen". La base de sols de laboratoire dispose de sondages effectués sur des sols dans un état hydrique humide. Cependant, le nombre de mesures des pénétrogrammes réalisés dans ces éprouvettes est beaucoup trop faible pour que les analyses réalisées puissent être représentatives. Les résultats de l'essai au bleu de méthylène *VBS* et de la répartition des passants au diamètre de $80\mu\text{m}$ et 2mm (Fig.12.1) montrent

que les sols de la base de données des sols naturels contiennent une fraction argileuse plus élevée. Enfin, il faut noter que l'ensemble des sondages analysés ont été effectués avec une pointe de 4cm^2 de section correspondant à la taille de pointe préconisée pour effectuer des sondages de reconnaissance de sols.

12.1.2 Définition des entrées du modèle

Le but ici est de créer un paramétrage permettant d'avoir une « signature » du pénétrogramme à classer. On génère ainsi un vecteur de paramètres en tant qu'identifiant du pénétrogramme, qui représentent les données d'entrée du modèle de classification. Un jeu initial de 26 paramètres a été proposé. Il est issu d'une analyse statistique, non linéaire, morphologique et fréquentielle du profil de résistance dynamique q_d mais aussi d'une analyse statistique des paramètres d'enfoncement par coup e . L'analyse des pénétrogrammes et le jeu de paramètres choisi est détaillée par la suite.

Analyse statistique

L'analyse statistique de la résistance de pointe dynamique q_d comprend 9 paramètres qui sont :

1. la valeur moyenne de q_d ;
2. la valeur médiane de q_d ;
3. l'écart-type de q_d ;
4. le coefficient de variation de q_d ;
5. la variance de q_d ;
6. l'étendue de q_d ;
7. l'écart interquartile de q_d ;
8. l'asymétrie de q_d ;
9. le coefficient d'aplatissement de q_d .

Les deux premiers paramètres sont des mesures de tendance centrées. Les paramètres 3 à 7 représentent des mesures de dispersion. Le paramètre n°8 définit la symétrie de la distribution des mesures du pénétrogramme analysé. Le paramètre n°9 est un moment statistique centré d'ordre 4. Il est également appelé *kurtosis* et permet de calculer le degré d'aplatissement de la distribution.

Analyse non linéaire

Deux paramètres sont définis à partir de l'analyse non linéaire :

10. l'entropie de Shannon ;
11. l'entropie logarithmique.

La notion d'entropie est un concept introduit dans la théorie de l'information qui peut être interprété comme une mesure de la complexité du signal. La définition formelle d'entropie E_{SH} pour une variable aléatoire discrète X avec n réalisations possibles ayant une probabilité P_i d'apparaître est la suivante :

$$E_{SH} = - \sum_i^n P_i \cdot \ln P_i \quad \text{avec} \quad P_i = P(X = X_i), i = \{1, \dots, n\} \quad (12.1)$$

Il s'agit d'un paramètre souvent employé pour la classification des signaux, notamment dans le domaine de la médecine (ex. [Sabeti et al. \(2009\)](#)). En pratique, les probabilités recherchées sont estimées à l'aide de l'histogramme empirique des mesures de q_d et on applique ensuite l'équation (éq.12.1) pour calculer le paramètre n°10 dit entropie de Shannon. De la même manière, on peut calculer le paramètre n°11 qui représente l'entropie logarithmique par le biais de l'équation :

$$E_{log} = \sum_i^n \log(P_i^2) \quad (12.2)$$

Analyse statistique d'enfoncement

Des paramètres déduits de l'analyse statistique de l'enfoncement par coup lors de l'essai pénétrométrique ont été également définis. Au total 6 variables ont été prises en compte :

12. la valeur moyenne d'enfoncement e ;
13. la variance d'enfoncement e ;
14. le coefficient de variation d'enfoncement e ;
15. l'étendue d'enfoncement e ;
16. l'asymétrie d'enfoncement q_d ;
17. le coefficient d'aplatissement d'enfoncement q_d ;
18. l'espérance de q_d multiplié par l'enfoncement e .

Analyse morphologique

L'analyse morphologique a pour but de fournir des paramètres descriptifs de la forme du signal. Pour effectuer cette analyse, on divise d'abord le pénétrogramme en tranches de 1 cm. Cette valeur a été retenue afin d'avoir un minimum de 10 points de mesures pour réaliser l'analyse et du fait qu'on estime une valeur de l'enfoncement moyen pour les essais in situ de l'ordre de 1 cm. La valeur finale du paramètre est calculée comme étant l'espérance des analyses réalisées pour chaque morceau du pénétrogramme. Les paramètres retenus sont les suivants :

19. changement du signe de la pente du pénétrogramme ;
20. nombre de croisements du signal pénétrométrique avec la ligne de tendance ;
21. longitude de la forme de l'onde ;
22. pente de la droite de régression du pénétrogramme ;
23. terme indépendant de la droite de régression.

Le paramètre n°19 mesure le nombre de fois où le pénétrogramme passe d'une pente positive à une pente négative. Le paramètre n°20 représente le nombre de fois où le signal pénétrométrique croise la ligne de tendance. La longitude de la forme de l'onde est le paramètre n°21 et peut être assimilé à une forme d'entropie caractérisant l'agitation du signal ([Reiffsteck et al. \(2012\)](#)), selon l'équation suivante :

$$L(z) = \sum_{z_o}^z |q_d(z + \Delta z) - q_d(z)| \quad (12.3)$$

Les paramètres n°22 et 23 sont les variables de la tendance linéaire du signal pénétrométrique calculées à l'aide d'une régression par moindres carrés.

Analyse spectrale

On utilise une transformée de Fourier rapide pour effectuer une analyse spectrale des résidus une fois la tendance linéaire du signal enlevée. Trois paramètres sont issus de l'analyse spectrale du signal :

24. la puissance maximale ;
25. la puissance moyenne ;
26. la fréquence de la puissance maximale.

On calcule la densité spectrale de puissance qui correspond au carré du module de la transformée de Fourier divisé par la taille du domaine d'intégration, dans notre cas la profondeur du sondage. La valeur maximale et moyenne de la densité spectrale de puissance correspondent aux paramètres n°24 et 25. Le paramètre n°26 correspond à la fréquence pour laquelle la densité spectrale présente sa valeur maximale.

12.1.3 Définition des sorties du modèle

Cette étape consiste à définir les classes de sortie de l'outil de classification automatique. L'ensemble des matériaux qui constituent la base de données élaborée (laboratoire et in situ) ont été classés selon la nomenclature GTR (NF P 11-300 (1992)). Une classification basée sur les seuils granulométriques français (2mm et $80\mu m$) a été alors proposée. On a défini un total de quatre classes de sols récapitulées dans le tableau (Tab.12.3).

On rappelle que dans le contexte d'un problème de classification à k classes, chaque neurone représente une couche de sortie. En conséquence, pour notre problème, un vecteur à 4 valeurs représente les 4 neurones de la couche de sortie, tel qu'indiqué dans la colonne « codification » du tableau (Tab.12.3). Une valeur égale à "1" du neurone de sortie définit le cas d'appartenance à cette classe, tandis qu'une valeur égale à "0" indique le cas contraire.

TABLE 12.3 – Classes de sortie proposées

Classe cible	Nature	Tamiset	Codification
Classe1	<i>Sols fins</i>	$80\mu m > 35\%$	1 0 0 0
Classe2	<i>Sables/graves avec fines</i>	$12\% < 80\mu m < 35\%$	0 1 0 0
Classe3	<i>Sables fins</i>	$80\mu m < 12\% \& 2mm > 70\%$	0 0 1 0
Classe4	<i>Graves</i>	$80\mu m < 12\% \& 2mm < 70\%$	0 0 0 1

La distribution des sondages qui constituent la base de données élaborée en fonction des classes de sortie proposées est donnée dans le tableau (Tab.12.4). Il n'existe pas une équi-répartition des effectifs selon les 4 classes de sortie visées. En effet, les sondages in situ appartiennent majoritairement à la classe n°1 des sols fins, à l'exception de 5 sondages réalisés sur des sables ou graves avec des fines (classe n° 2). Ceci peut être expliqué par le fait que in situ, les analyses d'échantillons de sols sont en général réalisées sur les sols susceptibles d'être « problématiques » au sens géotechnique et qui sont le plus souvent des sols argileux ou fins. La classe n°2 est celle qui présente le moins d'effectifs (27 sondages).

TABLE 12.4 – Répartition des pénétrogrammes de la base de données

Classe cible	Nature	Laboratoire (n°)	In situ (n°)	Répartition
Classe1	<i>Sols fins</i>	47	64	51%
Classe2	<i>Sables/graves avec fines</i>	22	5	12%
Classe3	<i>Sables fins</i>	38	0	18%
Classe4	<i>Graves</i>	42	0	19%
Total		149	69	100% (218)

12.1.4 Dimensionnement et optimisation du réseau PMC

Choix de l'architecture et des techniques d'entraînement

La performance d'un RNA dépend fondamentalement de trois facteurs :

- les fonctions d'activation ou de transfert ;
- l'architecture du réseau ;
- le poids des connexions entre neurones.

Comme expliqué lors de la revue bibliographique, la détermination de l'architecture d'un réseau PMC exige de trouver le nombre optimal de couches cachées ainsi que le nombre de neurones de calcul. Compte tenu qu'il n'existe pas d'approche universelle pour accomplir cette tâche, une procédure d'essai-erreur a été retenue. La recherche exhaustive des paramètres du réseau a été réalisée sur la base des critères suivants :

- des modèles avec 1 ou 2 couches cachées ont été testés ;
- le nombre maximal de neurones pour une seule couche cachée a été fixé à 52^2 ;
- les fonctions de transfert testées sont celles couramment utilisées pour des problèmes de classification avec un PMC. Ce sont la fonction sigmoïde ou logistique et la tangente hyperbolique ;
- on a comparé les deux techniques de normalisation les plus courantes pour transformer les différents paramètres d'entrée v dans le même intervalle :

$$\text{Min-max} \quad v' = \frac{v - \min_v}{\max_v - \min_v} \cdot (\max_{v'} - \min_{v'}) + \min_{v'} \quad (12.4)$$

$$\text{Z-score} \quad v' = \frac{v - \mu_v}{\sigma_v} \quad (12.5)$$

avec μ_v et σ_v étant l'espérance et l'écart-type empiriques du paramètre v à transformer.

- à chaque fois la base de données créée a été divisée en deux ensembles de données de manière aléatoire. L'ensemble d'apprentissage est constitué par 80% de l'effectif et les 20% restant sont utilisés pour la validation du modèle ;
- pour éviter des problèmes de surapprentissage, la technique *early stopping* a été retenue.

Afin de choisir la meilleure combinaison possible des paramètres du PMC, les performances des RNA entraînés ont été comparées en fonction de la fraction des pénétrogrammes

2. (Hecht-Nielsen (1989)) suggère que la limite supérieure du nombre de neurones de calcul pour un PMC avec une seule couche cachée peut être approchée par l'expression $(2I + 1)$ avec I le nombre des entrées.

correctement classés pour l'ensemble de validation. La performance des différentes architectures du PMC en fonction du nombre de neurones est présentée ici. Dans la figure (Fig.12.2) on fournit les résultats de plusieurs architectures du PMC avec deux couches de calcul. Ensuite on présente les résultats des PMC avec une seule couche cachée, avec un pré-traitement des données d'entrée *min-max* (Fig.12.3) et *Z-score* (Fig.12.4).

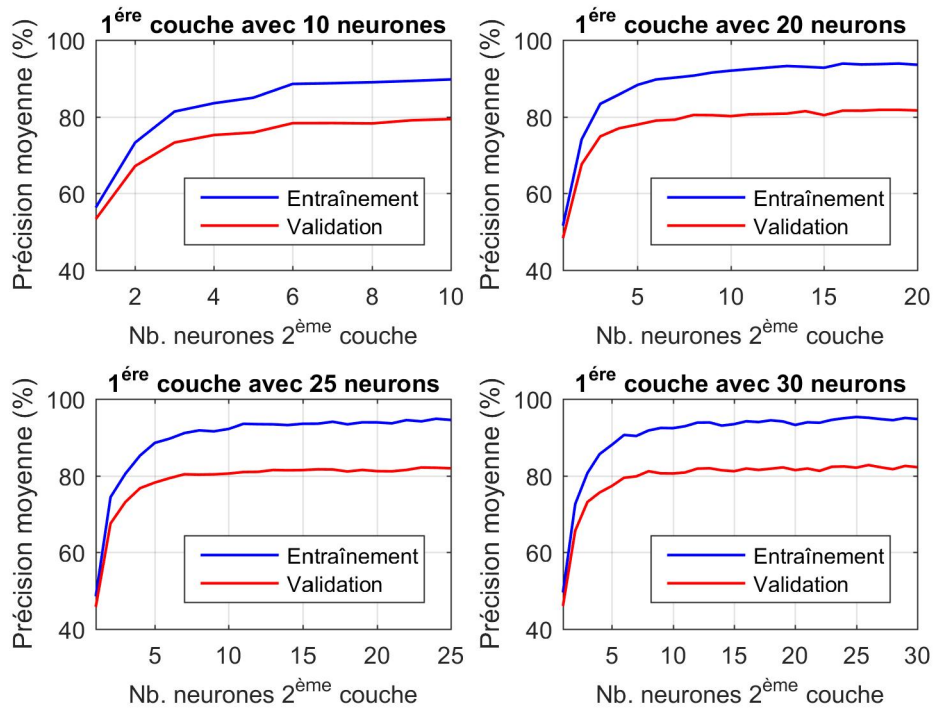


FIGURE 12.2 – Performance du PMC vs nombre de neurones de la deuxième couche de calcul. Pré-traitement des données (min-max)

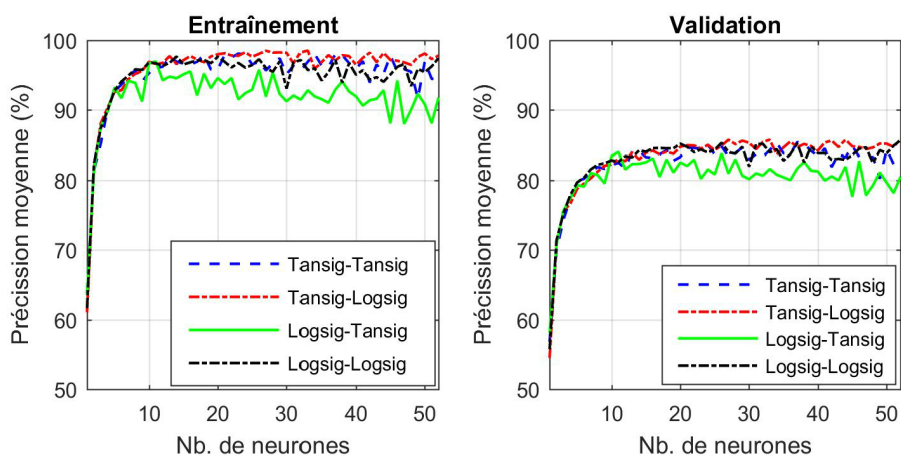


FIGURE 12.3 – Performance du PMC vs nombre de neurones de calcul. Pré-traitement des données (min-max)

Suite à la recherche menée, aucune différence significative n'a été constatée entre le fait de

prendre en compte une couche cachée ou deux. En conséquence, on a privilégié l'architecture la plus simple possible, avec une seule couche cachée. Par ailleurs, d'après les graphiques du pourcentage des échantillons bien classés en fonction du nombre de neurones de calcul, on a constaté qu'à partir de 20 neurones, l'amélioration des résultats devient négligeable.

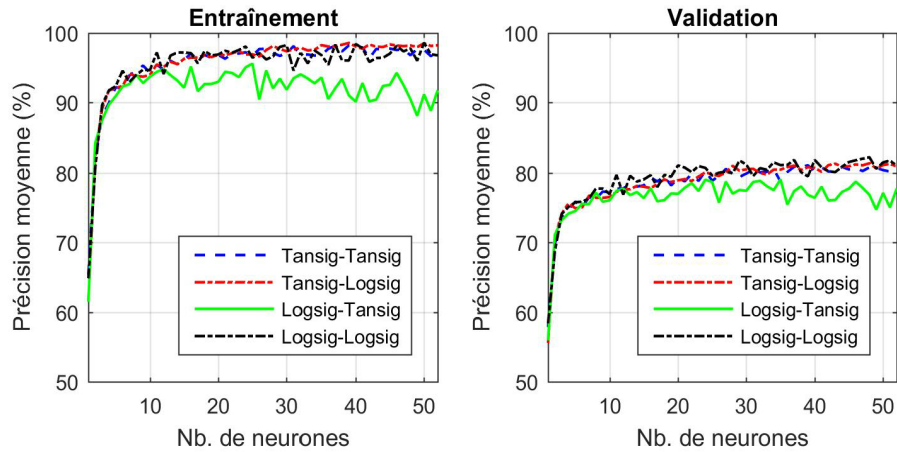


FIGURE 12.4 – Performance du PMC vs nombre de neurones de calcul. Pré-traitement des données (z-score)

Une normalisation des données d'entrée dans un intervalle $[-1, 1]$ avec le choix de la fonction tangente hyperbolique comme fonction de transfert a fourni les meilleurs résultats. Le RNA avec le meilleur taux de pénétrogrammes correctement classés a 16 neurones de calcul dans la couche cachée. Les résultats sont présentés sous forme de matrice de confusion dans la figure (Fig.12.5).

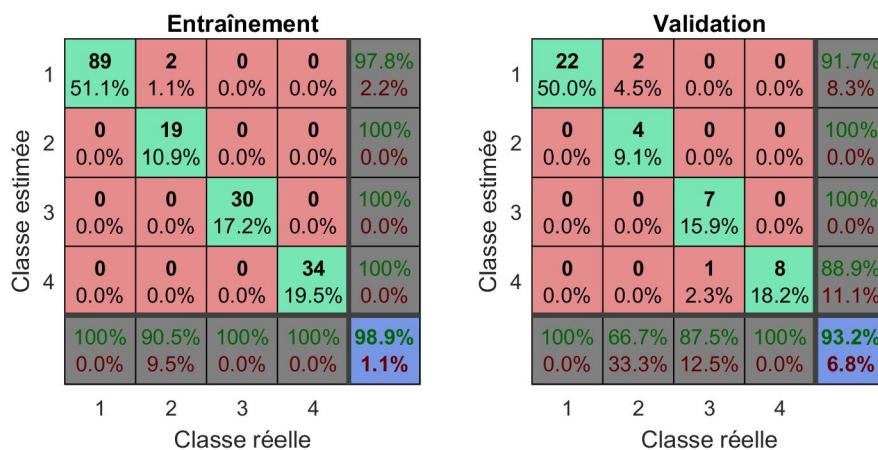


FIGURE 12.5 – Matrice de confusion du PMC choisie après la phase d'optimisation

Ces matrices s'interprètent de la façon suivante : chaque colonne de la matrice représente le nombre d'occurrences d'une classe de référence. Le numéro de la colonne représente le numéro de la classe de sortie, par exemple la colonne n°2 représente la classe n°2. En revanche, chaque ligne représente le nombre d'occurrences d'une classe estimée par le modèle.

Un des intérêts de la matrice de confusion est qu'elle montre rapidement si le système parvient à classer correctement. En ce sens, plus cette matrice de confusion approche d'une matrice diagonale, plus le système de classification est performant. On a également inclut les pourcentages d'échantillons correctement classés par ligne, colonne, et le total de chaque ensemble.

Enfin, le modèle retenu ici sera celui utilisé dans la phase d'optimisation des entrées du modèle.

Choix des entrées du modèle

Afin d'étudier l'importance relative des différents paramètres d'entrée, une analyse de sensibilité a été réalisée. L'analyse de sensibilité d'un modèle mathématique étudie comment ses variables de sortie sont influencées par des perturbations de ses entrées. La technique mise en œuvre ici est dite « one-at-a-time » (OAT) puisque on ne fait varier qu'une seule variable d'entrée alors que les autres sont fixées à leurs valeurs nominales. On a considéré quatre valeurs de perturbation possibles :

- i. 10% de l'écart-type empirique de chaque paramètre ;
- ii. 50% de l'écart-type empirique de chaque paramètre ;
- iii. l'écart-type empirique de chaque paramètre ;
- iv. la variance empirique de chaque paramètre.

Du point de vue physique, le fait de choisir la variance comme une valeur de perturbation n'a aucun sens. On cherche quand même à évaluer l'influence de la substitution de la valeur d'un paramètre quelconque par une valeur aberrante. Néanmoins, cette analyse ne peut pas être considérée comme une véritable analyse de sensibilité, puisqu'on ne mesure pas l'importance relative de chaque variable. Il s'agit plutôt d'évaluer l'effet de chaque modification en terme de performance des prédictions réalisées afin de déterminer des paramètres d'entrée qui n'apportent pas ou peu d'information et puissent ainsi être éliminés. La performance est évaluée en terme de variation totale du pourcentage d'échantillons mal classés :

$$\Delta E_i(\%) = | E(\%) - E_i(\%) | \quad (12.6)$$

avec $E(\%)$ le pourcentage total des effectifs mal classés par le RNA et $E_i(\%)$ le nouveau pourcentage des effectifs mal classés suite à la modification du paramètre d'entrée n^oi. Les résultats de l'analyse sont présentés dans la figure (Fig.12.6), où on peut noter que les perturbations introduites aux différents paramètres ne semblent pas avoir d'influence sur les prédictions du système de classification. Notamment, la variation de l'erreur est toujours inférieure à 5% pour les modifications des paramètres issues de l'analyse statistique des enfoncements (paramètre n^o12 à 18). Même observation pour les paramètres n^o25 et n^o26 calculés par le biais de l'analyse spectrale. Ces paramètres sont alors éliminés du vecteur d'entrées, dont la taille est ainsi réduite de 26 à 17 paramètres.

Choix du modèle final

Une fois que le nouveau vecteur d'entrées a été calculé, on réalise à nouveau une recherche exhaustive afin d'optimiser l'architecture du PMC. On retient la configuration du RNA déduite auparavant : normalisation des données *min-max*, fonction tangente hyperbolique comme fonction d'activation et une seule couche cachée. Maintenant, il reste à déterminer le nombre de neurones cachés ou de calcul. 70% des données ont été retenues pour

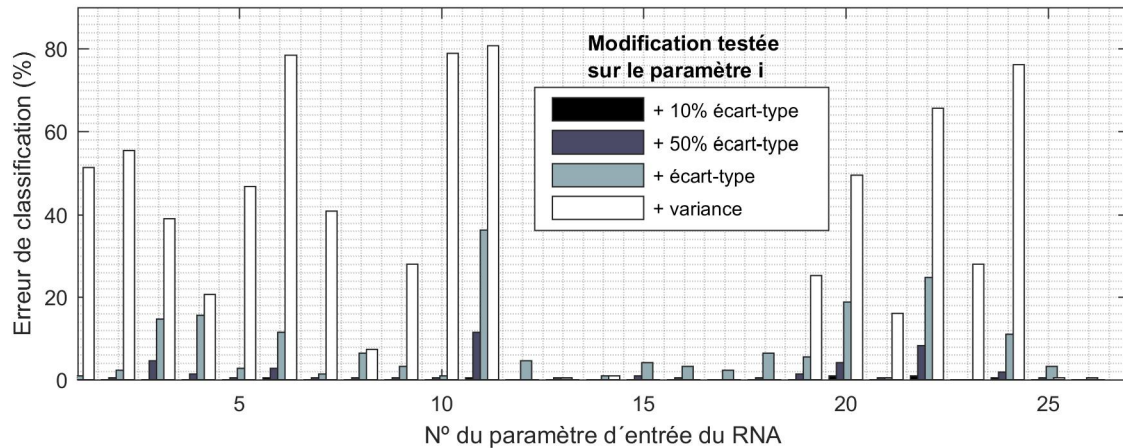


FIGURE 12.6 – Analyse de sensibilité basée sur l'étude de l'effet d'une modification de chacun des paramètres d'entrée du PMC

l'apprentissage. Une fraction de 15% a été choisie pour l'ensemble de validation et de test du réseau. On a trouvé qu'un RNA avec 12 neurones de calcul dans la couche cachée avait la meilleure capacité de généralisation et fournissait le % le plus élevé des données correctement classées pour l'ensemble de validation et de test. Les résultats sont présentés sous forme de matrice de confusion dans la figure (Fig.12.7). La capacité de généralisation du réseau est confirmée par les excellents résultats obtenus pour le jeu de données de validation et de test, avec 31 pénétrogrammes sur 33 correctement classés. Une discrimination parfaite entre les sols fins et graves a été obtenue. Plus de difficultés sont rencontrées pour les données de la deuxième classe (sables/graves avec fines), qui présente la fraction des effectifs bien classés la plus faible (70%).

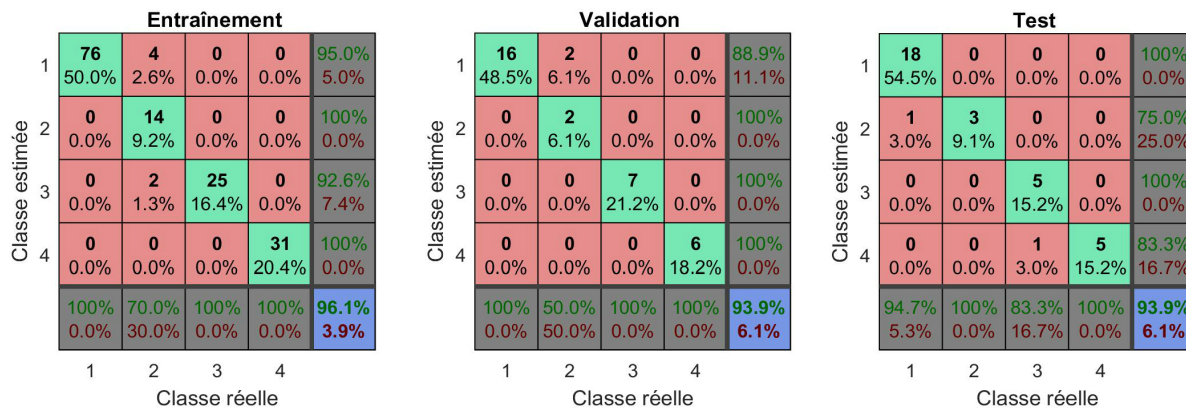


FIGURE 12.7 – Matrices de confusion du PMC avec 12 neurones cachés

12.1.5 PNN

On rappelle que les réseaux de neurones probabilistes (PNN) présentent une topologie beaucoup plus simple et qu'il ne faut optimiser qu'un paramètre, l'écart-type σ de la fonction de transfert gaussienne, qu'on appellera paramètre de lissage. C'est pourquoi, on a souhaité étudier l'application de ce réseau pour le problème de classification posé.

Afin de pouvoir comparer les résultats de la classification fournis par les deux RNA testés dans ce travail, on a gardé la même division du jeu de données que celle utilisée dans le calcul du PMC optimisé auparavant. On utilise le vecteur d'entrées constitué par les 17 paramètres déduits antérieurement. Une normalisation *min-max* a aussi été appliquée aux paramètres d'entrée. La valeur optimale de σ a été déterminée en fonction du % des pénétrogrammes de l'ensemble de validation et de test qui ont été classés correctement. Les meilleurs résultats ont été obtenus pour une valeur de σ égal à 0.2, tel que montré dans la figure (Fig.12.8).

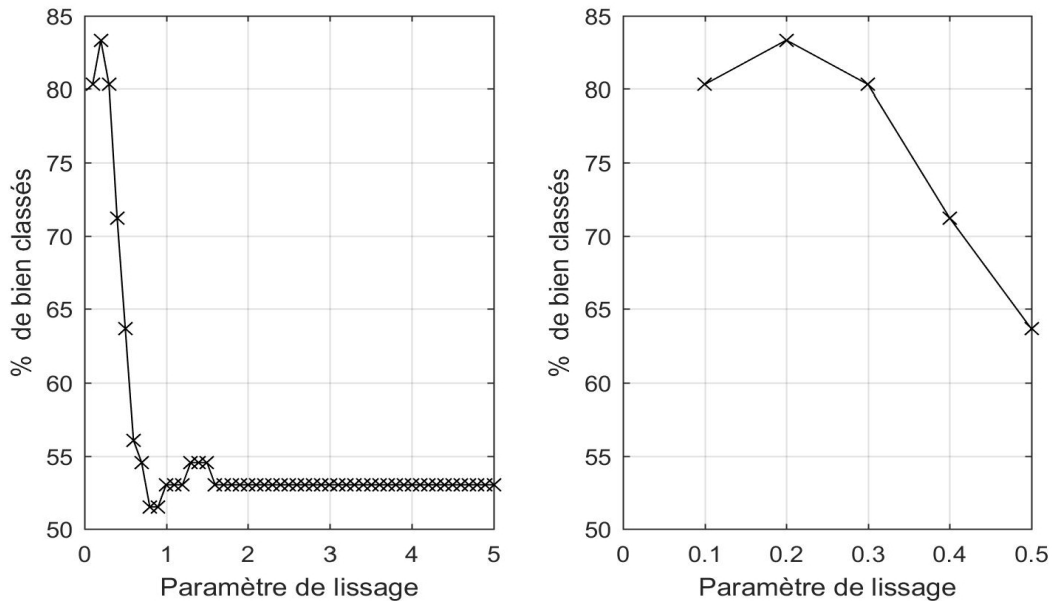


FIGURE 12.8 – Détermination de la valeur optimale du paramètre de lissage σ

Ce résultat s'avère intéressant puisqu'il permet de mieux comprendre le fonctionnement du système de classification. En effet, d'après (Specht (1991)), pour une valeur de σ aussi faible, le réseau PNN est similaire à un système de classification régi par la règle des plus proches voisins. En d'autres termes, le modèle d'apprentissage supervisé n'a pas trouvé une règle de classification générale permettant de classer les pénétrogrammes sur la base du paramétrage proposé. Il s'agit plutôt d'une classification où chaque pénétrogramme représente un point dans un espace à n dimensions avec n la taille du vecteur d'entrées du modèle (17 dans notre cas). La classe de sol assignée à un nouveau pénétrogramme est déduite à partir des pénétrogrammes « les plus similaires ou proches » dans l'espace des caractéristiques d'entrée.

Les matrices de confusion pour les différents ensembles de données étudiées sont présentées dans la figure (Fig.12.9). On observe une très bonne précision, avec 201 pénétrogrammes sur 218 correctement classés, un pourcentage similaire à celui issu du modèle PMC avec 210 sur 218. De même, les résultats de la phase d'apprentissage sont presque identiques à ceux du PMC. C'est pour les ensembles de validation et test que l'on constate le plus de différences et où le système PMC fournit une classification plus précise. Par ailleurs, ici encore, les données appartenant à la classe 2 sont celles qui sont le moins bien classées. Ce qui été également le cas avec l'utilisation du PMC.

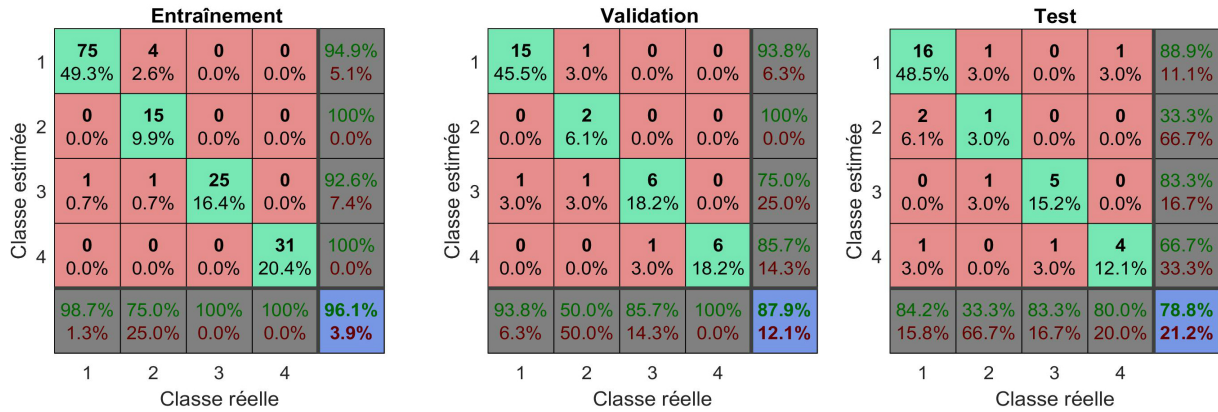


FIGURE 12.9 – Matrices de confusion du PNN avec $\sigma = 0.2$

12.2 Validation : tests sur de nouveaux sondages

12.2.1 Démarche

Le but ici est de tester et de valider la robustesse des deux modèles de classification élaborés (PMC et PNN) vis-à-vis des essais de pénétration (en laboratoire et in-situ) réalisés sur de nouveaux matériaux ne faisant pas partie de la base de données créée initialement. Pour ce faire, on a testé les échantillons suivants :

- essais de laboratoire en moules de calibration constitués par 4 couches de différents matériaux granulaires ;
- essais de laboratoire sur deux matériaux fins (argile et sable) dans un état hydrique humide (h) ;
- essais in situ issus de campagnes de reconnaissance géotechnique réalisées en région Auvergne-Rhône-Alpes et sur le site expérimental de Castelló d'Empúries (Espagne).

Les résultats pour chaque jeu de données testé sont présentés par la suite.

12.2.2 Application à des matériaux granulaires en laboratoire

Les éprouvettes testées ici sont celles utilisées dans le protocole expérimental de laboratoire mis en œuvre pour l'étude des méthodes d'identification automatique des couches à partir du pénétrogramme Panda (cf. [Partie II](#)). Les quatre matériaux granulaires utilisés ont été mis en place par couches compactées avec une épaisseur de 0.27m. Les pénétrogrammes analysés ici ont une quantité de mesures inférieure à celles des essais de la base de données en laboratoire utilisée pour l'entraînement.

Pour les quatre matériaux utilisés, les tamisats à 80 μm et 2 mm ainsi que la classe granulométrique sont présentés dans le tableau ([Tab.12.5](#)). Ces matériaux ne font pas partie de la base originale. Le ballast et microballast appartiennent à la classe n°4 de la famille des graves. Le sable dénommé grossier est classé en tant que sable fin (classe n°3) et le sable de Dunkerque en tant que sable avec fines (classe n°2).

TABLE 12.5 – Tamisats à 2 mm et 80 μ m des matériaux dans les moules de calibration en laboratoire

Matériaux	Passant 80 μ m (%)	Passant 2 mm (%)	Classe de sortie
Ballast	0	0	4
Microballast	0	0	4
Sable grossier	0	94.5	3
Sable de Dunkerque	16	99.4	2

Afin d'appliquer les modèles de classification développés, on a sélectionné la portion du pénétrogramme correspondant à chacune des couches de matériau. Pour chaque sondage réalisé dans un moule, le pénétrogramme est divisé en quatre unités. Comme on dispose de quatre moules de calibration, on compte sur un total de seize effectifs à classer. On présente d'abord les résultats de la classification pour les sondages réalisés avec une pointe de 4 cm^2 et ensuite avec une pointe de 2 cm^2 . Les matrices de confusion pour les sondages avec une pointe de 4 cm^2 sont présentées dans la figure (Fig.12.10).

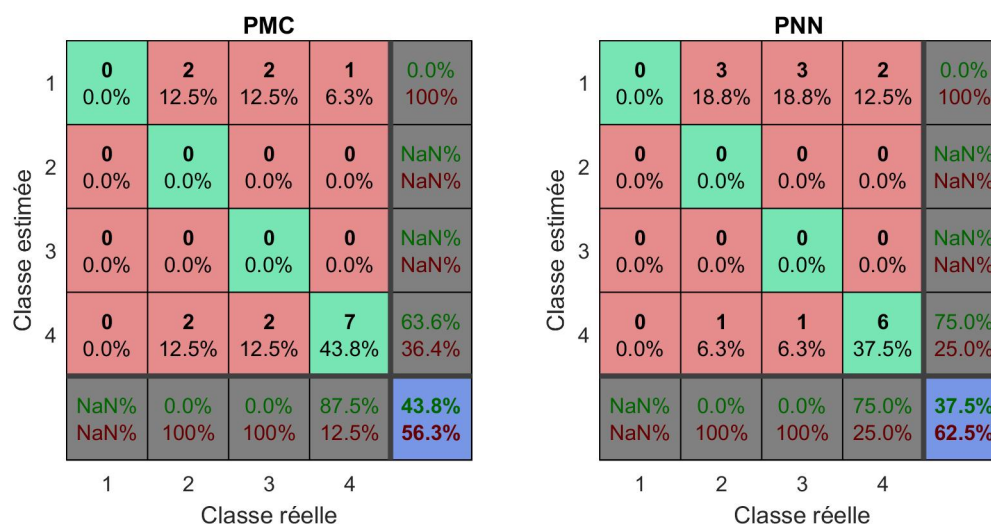


FIGURE 12.10 – Matrices de confusion pour les sondages réalisés en pointe de 4 cm^2 de section

On constate une chute importante du taux de pénétrogrammes correctement classés par rapport aux résultats antérieurs. En effet, ce pourcentage reste inférieur à 50% pour les deux modèles testés, PMC et PNN. Les résultats sont similaires indépendamment du réseau utilisé. Le PMC a bien classé un seul pénétrogramme de plus que le PNN. Cette tendance est inversée lorsqu'on examine les résultats issus de la classification des essais avec une pointe de 2 cm^2 . Les matrices de confusion sont fournies dans la figure (Fig.12.11).

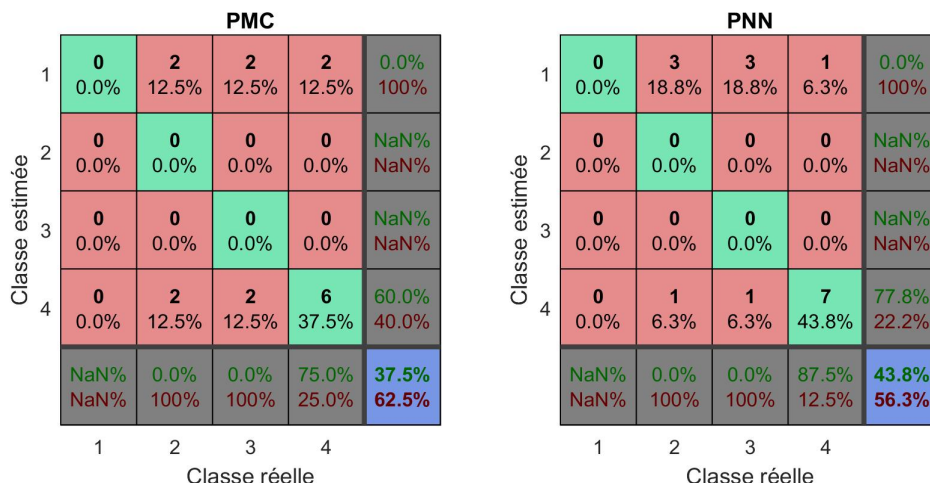


FIGURE 12.11 – Matrices de confusion pour les sondages en pointe de 2 cm^2

Les résultats de la classification en fonction des matériaux et de la taille de la pointe utilisée sont récapitulés dans le tableau (Tab.12.6). Tous les pénétrogrammes au sein du ballast sont classés correctement comme un sol à matrice grossière de type grave (classe n° 4) pour les deux modèles, PMC et PNN, indépendamment de la pointe utilisée. Au moins 50% des pénétrogrammes au sein du micro-ballast ont aussi été correctement classés. Les essais sur le micro-ballast mal classés sont ceux des moules M1 et M4. Par ailleurs, quand une pointe de 4 cm^2 est utilisée, les essais sur la couche de micro-ballast du moule M1 et M4 sont tous mal classés avec le modèle PNN, et uniquement celui du moule M1 avec le modèle PMC. La situation est inversée pour les essais avec une pointe de 2 cm^2 . Néanmoins, les pénétrogrammes au sein des couches de matériaux sableux, n'ont pas été correctement classés. Le PNN classe 3/4 essais comme des sols fins et 1/4 comme des graves. Le PMC classe la moitié des pénétrogrammes comme des graves et l'autre moitié comme des sols fins.

TABLE 12.6 – Pénétrogrammes bien classés selon les matériaux et la taille de la pointe

Matériaux	2 cm^2		4 cm^2	
	PMC	PNN	PMC	PNN
<i>Ballast</i>	4/4	4/4	4/4	4/4
<i>Microballast</i>	2/4	3/4	3/4	2/4
<i>Sable grossier</i>	0/4	0/4	0/4	0/4
<i>Sable de Dunkerque</i>	0/4	0/4	0/4	0/4

Bilan

Indépendamment de la section de la pointe utilisée lors des essais, les deux matériaux sableux n'ont pas été correctement classés et ce quel que soit le réseau utilisé. Cette difficulté avait été déjà constatée auparavant, durant la phase de développement du système de classification et peut s'expliquer par la faible représentativité des matériaux sableux (faible nombre de sondages de la classe 2) dans la base de données ayant servi à l'apprentissage des réseaux de neurones. De meilleures performances ont été constatées vis-à-vis de la classification des matériaux grossiers, notamment pour le ballast.

12.2.3 Application à des matériaux à l'état hydrique humide

L'influence de l'état hydrique humide sur deux matériaux sensibles à l'eau, un sable fin (classe de sortie 2) et une argile sableuse (classe de sortie 1), a été testée pour les deux réseaux développés. Les caractéristiques des matériaux sont données dans le tableau ci-dessous (Tab.12.7).

TABLE 12.7 – Caractéristiques des matériaux

Caractéristiques	Argile sableuse	Sable de Brive
Passant à 80 μm (%)	60.66	15.32
Passant à 2mm (%)	82.45	80.18
$\gamma_d(kN/m^3)$	18.57	18.52 // 19.45
w(%)	9.87	11.73 // 11.57
Ip (%)	18.9	n.p.
VBS (mg/l)	0.75	0.6
Classe de sortie	1	2

Trois sondages avec une pointe de 2cm^2 ont été réalisés dans chaque moule de calibration. Finalement, on dispose de trois pénétrogrammes pour l'argile sableuse et de six (pour le sable de Brive), trois pour chaque densité testée, au total deux à $18.52kN/m^3$ et à $19.45kN/m^3$. Les matrices de confusion regroupant les résultats pour les trois pénétrogrammes réalisés dans l'argile sableuse sont présentées dans la figure (Fig.12.12). Avec le PMC, 2 sondages sur 3 ont été classés correctement alors qu'avec le PNN seulement 1 sur 3.

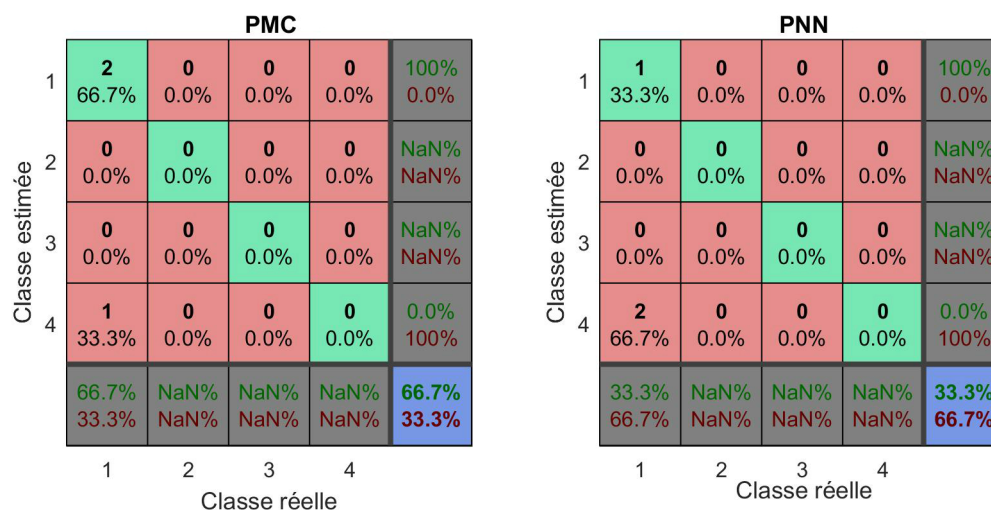


FIGURE 12.12 – Matrices de confusion pour les sondages effectués dans l'argile sableuse.

Dans la figure (Fig.12.13) on regroupe les matrices de confusion concernant les essais effectués sur le sable de Brive. Dans ce cas, les réponses fournies par les deux réseaux sont identiques. Les trois essais effectués avec un poids volumique sec $\gamma_d=18.52kN/m^3$ sont mal classés et identifiés comme des matériaux fins. Les autres trois essais réalisés sur le matériau avec un poids volumique sec $\gamma_d=19.45kN/m^3$ ont été identifiés comme des graves.

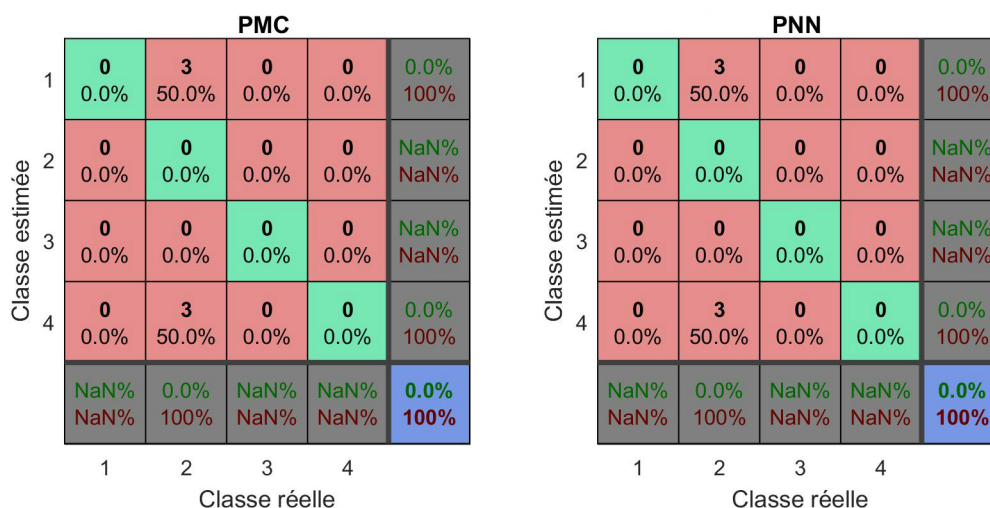


FIGURE 12.13 – Matrices de confusion pour les sondages effectués dans le sable de Brive.

Bilan

Suite aux essais réalisés, on constate tout d'abord comme pour les essais précédents, une difficulté pour les deux RNA développés pour classer correctement les essais réalisés sur des matériaux sableux. Il semble donc nécessaire d'étoffer la banque de données avec ce type de matériaux afin d'améliorer l'apprentissage des 2 RNA sur ces matériaux. On constate également que l'état hydrique ne semble pas influencer l'analyse de manière trop importante puisque pour les échantillons de sols fins, les réponses fournies par les deux RNA sont plutôt satisfaisantes.

12.2.4 Validation sur de nouveaux essais in situ

Nouveaux essais issus de la même zone géographique

La base de sondages in situ a été élargie avec de nouveaux essais in situ issus de campagnes de reconnaissance géotechnique réalisées dans la région Auvergne-Rhône-Alpes. On dispose d'un total de 64 pénétrogrammes. La majorité (52) a été réalisée sur des sols fins (classe n°1) et le reste (12 essais) sur des sols sableux avec des fines (classe n°2). Les paramètres physiques des échantillons testés en laboratoire sont récapitulés dans le tableau (Tab. 12.8) et les pourcentages des passants aux tamisats 80 μm et 2 mm ont été représentés dans la figure (Fig. 12.14).

Caractéristiques	min	max
Passant à 80 μm (%)	10.3	99.3
Passant à 2mm (%)	40.1	100
w (%)	5.9	62.9
VBS (mg/l)	0.4	13

TABLE 12.8 – Plage des paramètres physiques des nouveaux matériaux in situ

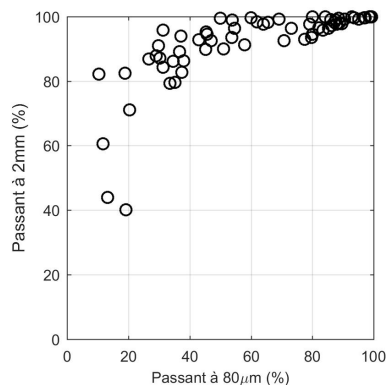


FIGURE 12.14 – Tamisats à 2mm et 80 μm des nouveaux matériaux in situ

Les résultats de la classification fournis par le PMC et le PNN sont présentés sous la forme de matrice de confusion à la figure (Fig.12.15). On constate une performance plus faible que celle obtenue précédemment lors de la phase de validation : 60% des pénétrogrammes correctement classés avec le réseau PMC et 50% avec le réseau PNN. On souligne que ni le PMC ni le PNN n'ont été capables de classer correctement les pénétrogrammes sur les matériaux sableux avec fines. Les deux modèles présentent des résultats très similaires, mais la précision de classification du PMC reste légèrement supérieure à celle du PNN. En effet, 38 essais sur 52 pour les cas des sols fins ont été correctement classés contre 32 pour le PNN. Par ailleurs, les pénétrogrammes qui n'ont pas été correctement classés pour les matériaux fins ont été identifiés comme des sols de la classe n°4, c'est-à-dire des graves, ce qui n'est pas satisfaisant.

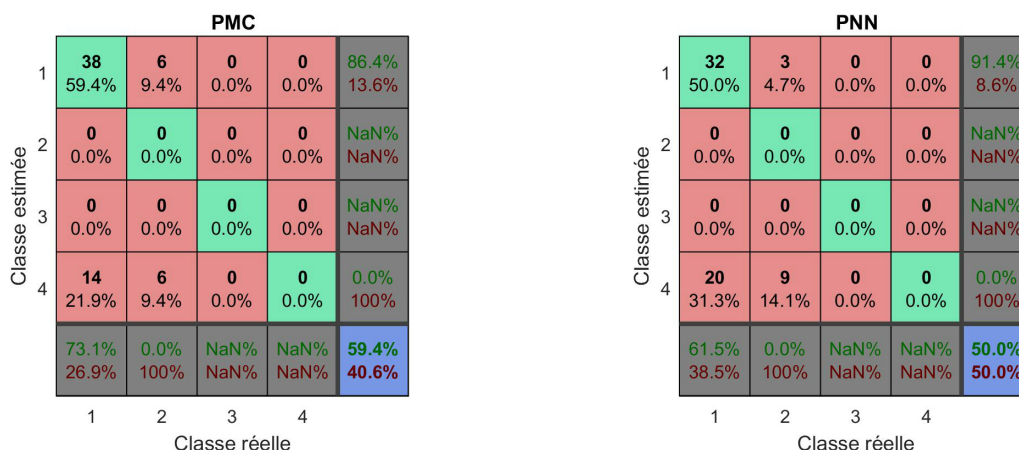


FIGURE 12.15 – Matrices de confusion

Nouveaux essais sur le site expérimental de Castelló d'Empúries

Une dernière application a été testée sur des matériaux ne faisant pas partie de la base de données d'apprentissage et issu d'une zone géographique différente des matériaux de la base d'apprentissage. Il s'agit d'essais réalisés sur le site expérimental de Castelló d'Empúries. La classification des échantillons disponibles est donnée dans le tableau (Tab.12.9). On note

TABLE 12.9 – Identification et classification des échantillons prélevés sur le site de Castello d’Empúries

Prof. (m)	Passant à	Limites d’Atterberg		Symbole	LPC	GTR
	75 μ m	WL %	IP %		Appellation	Symbole
0.5 -0.7	47.3	-	-	S	Sable	A
2.65 - 2.8	87.5	34.9	12.1	Ap	Argile peu plastique	A2
3.3 - 3.4	95.8	49.4	25.3	Ap	Argile peu plastique	A3
4.2 -4.3	68.2	25.3	9.7	Ap	Argile peu plastique	A1

que la formation située proche de la surface, à environ 0.5 m, est classée comme un sol sableux d’après le système LPC tandis que selon la classification GTR, il s’agit d’un sol fin (A). L’abaque de Robertson classe la première épaisseur de terrain, jusqu’à environ 1.5 m, dans la zone 6 comme un sable propre, en concordance avec le système de classification LPC/USCS. Selon les classes de sorties du système de classification proposé dans ce travail, tous les échantillons appartiennent à la classe n°1 des sols fins.

De même, les algorithmes de classification développés ont été testés sur l’ensemble des sondages du site de Castello, mais on ne présente ici que les résultats concernant les sondages Panda S3 et S5 réalisés au voisinage des essais au pénétromètre statique et à proximité également des endroits où les échantillons ont été prélevés. Les résultats obtenus sur les autres sondages sont très similaires à ceux obtenus pour les sondages présentés. Afin d’appliquer les procédures de classification des sols, les pénétrogrammes ont tout d’abord été traités avec la méthode de découpage mise au point dans la partie II de ce manuscrit. Cela afin d’analyser des portions de pénétrogramme correspondant à des unités de sol homogènes. Une première analyse a été effectuée avec un découpage réalisé avec une fenêtre de 1 m de hauteur. Les résultats de l’algorithme de découpage couplé avec la classification fournie par les RNA sont donnés dans la figure (Fig. 12.16) ($W_d = 1m$).

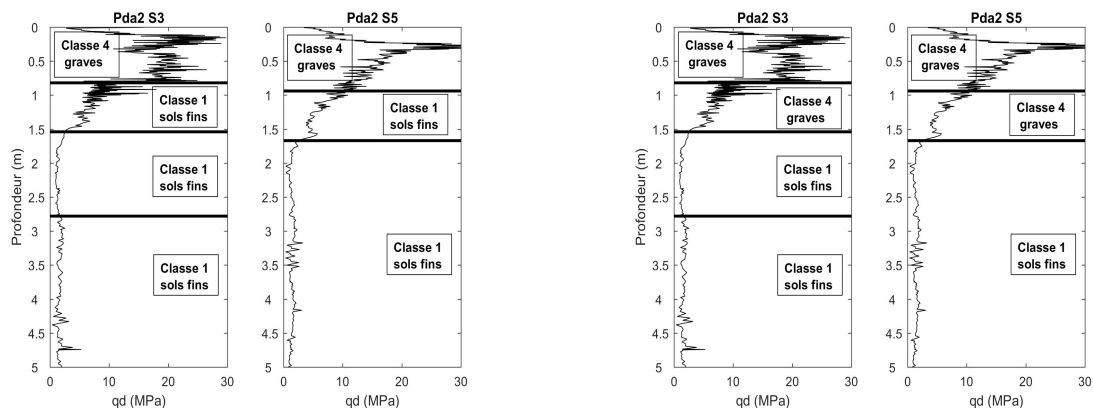


FIGURE 12.16 – Classification du RNA couplée avec le découpage par fenêtre glissante ($W_d = 1m$). À gauche classification fournie par le modèle PMC. À droite classification fournie par le PNN.

On constate que les deux RNA (PMC et PNN) ont fourni des résultats identiques, à l’exception de la deuxième unité qui a été correctement identifiée comme un sol fin (classe n°1) par le PMC alors que le PNN a déterminé qu’il s’agissait d’une grave (classe n°4). On observe

également que la première formation du terrain plus compacte est classée systématiquement comme une grave, classification incorrecte indépendamment du système de classification retenu. Les autres formations situées sous cette première couche ont été toutes correctement classées par le réseau PMC.

Afin d'examiner l'influence de l'épaisseur des couches identifiées en fonction de la taille W_d de la fenêtre de découpage choisie, nous avons comparé les premiers résultats obtenus avec ceux issus d'un découpage réalisé avec une fenêtre de 2 m de hauteur. Les résultats de l'algorithme de découpage couplé avec la classification fournie par les RNA pour cette dernière fenêtre sont présentés dans la figure (Fig.12.17) ($W_d = 2m$). Dans ce cas, le découpage est moins précis et les deux réseaux donnent exactement les mêmes résultats. Du fait de la moins bonne précision du découpage, la couche de surface allant jusqu'à 1,5 m est mal classée et identifiée comme une grave de classe n°4.

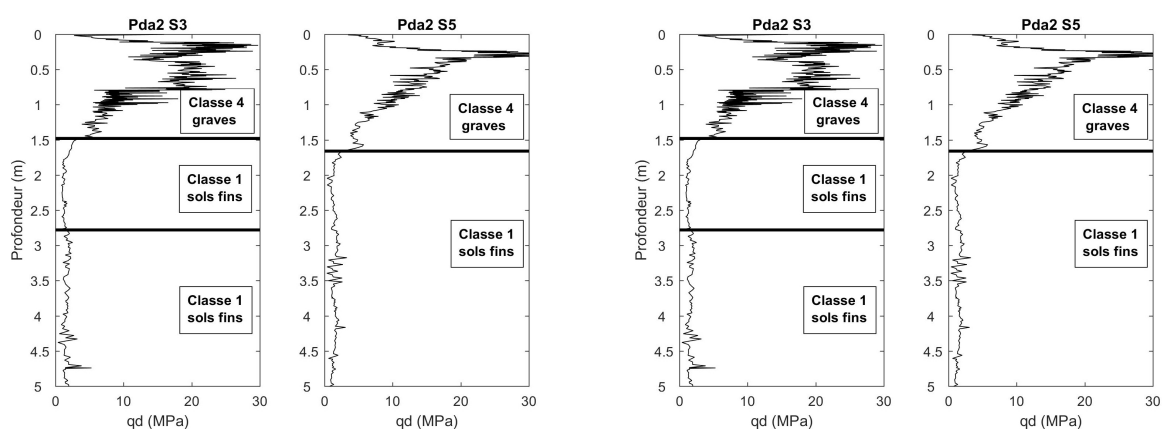


FIGURE 12.17 – Classification du RNA couplée avec le découpage par fenêtre glissante ($W_d = 2m$). À gauche classification fournie par le modèle PMC. À droite classification fournie par le PNN.

Bilan

De manière générale, des essais in situ étudiés, il se dégage que :

- la performance de classification est plus faible que celle obtenue lors de la phase de validation et ce quelque soit le réseau utilisé. Ce résultat n'est pas surprenant dans la mesure où ces nouveaux sondages ne faisaient pas partie de la banque d'apprentissage et qu'ils sont issus de zones géographiques différentes de celles des essais constituant la base de sondages utilisée. Par ailleurs, les essais in situ ont une variabilité plus importante et donc on trouve des signaux pénétrométriques moins homogènes que ceux des essais réalisés dans des moules avec des matériaux contrôlés.
- le réseau PMC est toujours plus performant que le PNN.
- ni le PMC, ni le PNN n'ont été capables de classer correctement les pénétrographes sur les matériaux sableux avec fines (classe 2). Ceci étant probablement dû au déficit de représentativité de cette classe dans la base d'apprentissage.
- on a un pourcentage très élevé de bon classement des matériaux de classe 1 (sol fins). Par contre, lorsque ces sols sont mal classés, il sont souvent identifiés comme des sols de la classe n°4, c'est-à-dire des graves.
- lors de la classification des matériaux à partir des sondages in situ, le choix de la hauteur de la fenêtre de découpage du pénétrogramme en unités joue sur la précision

de la classification. Comme il a été proposé en partie II, il semblerait qu'une hauteur de 1m soit la taille de fenêtre la mieux adaptée dans notre cas.

12.3 Bilan global

Deux architectures de RNA (PMC et PNN) ont été mises en œuvre afin de fournir une classification de la nature du sol à partir de l'analyse du signal pénétrométrique fourni par l'essai Panda. Quatre classes de sols ont été définies sur la base de critères granulométriques. Les résultats obtenus pour la première collecte de données montrent que :

- les deux architectures de RNA testées fournissent des résultats similaires avec une excellente performance sur l'ensemble des classes ;
- le PMC fournit une classification légèrement plus précise que le PNN.

En outre, une phase de validation a été conçue afin de réaliser une analyse critique des modèles de classification développés. Les deux RNA ont donc été testés à partir des essais effectués sur de nouveaux matériaux et sur des essais in situ ne faisant pas partie de la base de données originale. Les conclusions suivantes peuvent être déduites de ces analyses :

- les méthodes de classification proposées restent performantes pour classer les matériaux fins (classe 1) et les matériaux granulaires (classe 4) ;
- pour les matériaux intermédiaires (sableux - classe 2), les résultats sont plus décevants car ils ne sont pas correctement identifiés. On peut penser que ce problème est lié au déficit de représentativité de cette classe dans la base d'apprentissage. Il faudra donc veiller à renforcer la représentation de cette classe dans la base d'apprentissage afin de bien exercer et renforcer les réseaux de neurones sur l'analyse de ces matériaux ;
- l'état hydrique des matériaux et la taille de la pointe utilisée pour la réalisation des essais Panda ne semblent pas avoir une forte influence sur l'efficacité des analyses de classification ;
- le réseau PMC est toujours plus performant que le PNN.

Finalement, si les réseaux développés semblent performants sur la base d'apprentissage mise en place, on constate une diminution de leur performance pour la classification de nouveaux essais en "aveugle". La capacité de généralisation des modèles développés n'est donc pas complètement vérifiée notamment en ce qui concerne les matériaux intermédiaires des classes 2 et 3.

13. Application à la classification d'un matériau granulaire

13.1 Contexte

Les méthodes développées dans les sections précédentes ont pu être testées pour une autre application présentée ici. Ces travaux ont été réalisés dans le cadre de l'expérimentation technologique FSS (*Full Scale Seal*) (Lebon et al. (2013)) qui fait partie du projet de recherche sur le stockage des déchets nucléaires dans des formations profondes mené par l'ANDRA (Agence Nationale pour la gestion des Déchets RAdioactifs). Ce projet se déroule au sein du Centre Industriel de stockage géologique (Cigéo) sous les communes de Bure et Saudron. Un des principaux enjeux techniques du stockage profond des déchets nucléaires réside dans la fermeture du centre industriel et le maintien de son étanchéité. La bentonite est une argile gonflante connue par ses propriétés hydromécaniques permettant d'assurer l'imperméabilité des galeries de stockage et offrir une bonne capacité de cicatrisation. C'est pourquoi un essai de scellement « pleine échelle » a été réalisé à base de ce matériau argileux. C'est un matériau industrialisé et exploité sous le nom de "Expangel". Or lors de sa mise en œuvre, la bentonite peut se présenter sous la forme de granulés (concassés) et de billes sèches préfabriquées (pellets) ayant des granulométries différentes (cf. Fig.13.1). Au total, cinq formulations utilisant divers mélanges de ces deux composés de bentonite ont été testés, et il a été demandé d'évaluer le potentiel du pénétromètre dynamique Panda pour pouvoir contrôler la qualité du remplissage des galeries avec chacune de ces formulations.

13.2 Méthodologie expérimentale

13.2.1 Plan général

Dans le cadre de la mission de contrôle de remplissage, un protocole expérimental d'essais en laboratoire a été mis en œuvre afin d'obtenir les équations de la relation semi-logarithmique entre le poids volumique sec γ_d et la résistance de pointe dynamique moyenne¹ (Gourves and Barjot (1995)) :

$$\gamma_d = A \cdot \ln(q_{d1}) + B \quad (13.1)$$

Cette phase de calibration a été réalisée pour chacun des cinq mélanges de pellets et de concassés dont la composition est décrite dans le tableau (Tab.13.1).

1. Cette équation est valable pour un même matériau et pour une teneur en eau donnée (Chaigneau (2001))

Matériau	Pellets (%)	Concassé (%)
100/0	100	0
85/15	85	15
70/30	70	30
55/45	55	45
0/100	0	100



TABLE 13.1 – Composition des 5 mélanges de matériaux étudiés

FIGURE 13.1 – Photo des deux composés de bentonite : concassé à gauche et pellets à droite

13.2.2 Caractéristiques granulométriques des matériaux

La granulométrie des pellets est supposée connue par le processus de fabrication. Ils ont une forme quasi sphérique avec un diamètre de 32 mm. Les courbes granulométriques résultantes de l'analyse de quatre échantillons du matériau concassé sont présentés dans (Fig.13.2). Après les analyses réalisées, deux échantillons sont classés 0/4 et 2 autres échantillons 0/5.

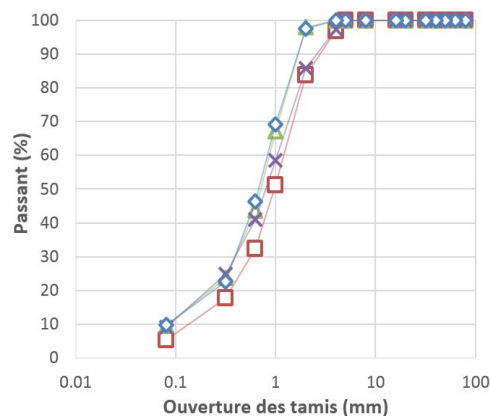


FIGURE 13.2 – Courbes granulométriques des quatre échantillons de bentonite concassée

13.2.3 Phase de calibration : essais de pénétration Panda

Pour chaque mélange calibré et présenté auparavant dans (Tab.13.1), 5 éprouvettes à 5 états de densité différents ont été confectionnées. Pour chaque éprouvette, on réalise deux essais pénétrométriques Panda avec une pointe perdue de 4cm^2 . Pour l'ensemble des éprouvettes de calibration, la profondeur critique z_c a été évaluée entre 0.10 et 0.35 m.

À partir du protocole des essais mis en œuvre, le modèle de relation poids volumique sec/résistance de pointe dynamique q_d pour chacun des matériaux a été déduit au moyen d'une régression par moindres carrés. Les courbes de calibration calculées ainsi que leurs équations sont présentées dans (Fig.13.3). Les valeurs élevées du coefficient R^2 permettent de vérifier le bon ajustement du modèle théorique aux points expérimentaux. Toutefois l'erreur du modèle n'est pas fournie. Compte tenu des résultats obtenus, il se dégage que le mélange optimal est celui constitué par une fraction de 70% de pellets et 30% de concassé (70/30) avec des densités centrées autour de 16 kN/m^3 .

Matériaux	Équations (kN/m ³)	R ²
100/0	$\gamma_S = 1.56 \cdot \ln(q_d) + 10.20$	0.99
85/15	$\gamma_S = 1.88 \cdot \ln(q_d) + 12.11$	0.88
70/30	$\gamma_S = 1.07 \cdot \ln(q_d) + 14.16$	0.98
55/45	$\gamma_S = 0.50 \cdot \ln(q_d) + 14.87$	0.93
0/100	$\gamma_S = 0.78 \cdot \ln(q_d) + 12.50$	0.88

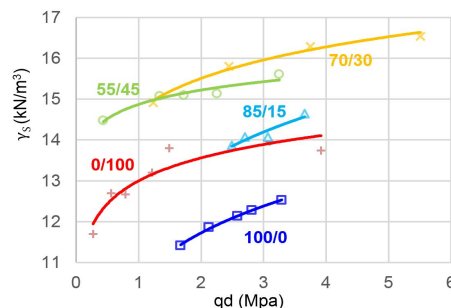


FIGURE 13.3 – Équations et courbes de calibration des cinq mélanges de matériaux

13.3 Classification supervisée

13.3.1 Démarche générale

Afin de pouvoir contrôler la densité du mélange en place et utiliser les courbes de calibration définies précédemment pour chaque formulation, il est nécessaire de reconnaître le mélange en place sans possibilité de prélever d'échantillons. C'est pourquoi, on a appliqué une démarche similaire à celle présentée auparavant afin de développer un système de reconnaissance et de classification du mélange à partir des essais Panda. La méthodologie a été appliquée sur la base des sondages issus du protocole de calibration effectué en laboratoire, pour les cinq mélanges de matériaux. L'objectif est donc d'arriver à classer un pénétrogramme parmi les 5 mélanges de pellets et concassés définis. Pour cela, nous avons utilisé un algorithme de classification RNA avec une architecture de type PMC. Les analyses réalisées ici ne concernent que la partie du pénétrogramme au dessous de la profondeur critique z_c .

13.3.2 Approche par Perceptron multicouche (PMC)

On disposait d'un total de 71 essais de pénétration. Leur répartition selon les 5 classes de sortie correspondant à chacun des 5 mélanges définis auparavant, ainsi que l'étendue des valeurs du poids volumique sec (γ_d) des matériaux testés sont fournies dans (Tab.13.2). Le nombre d'essais pour les classes n°1, 3 et 5 est supérieur à 10. En effet, des moules supplémentaires ont été réalisés pour les classes n°1 et n°5 correspondant aux formulations de pellets seuls (100/0) et de concassé seul (0/100). Par ailleurs, pour le matériau optimal, celui de la classe 3, deux moules supplémentaires avec des pressions de confinement égales à 10.5 kPa et 30 kPa ont également été effectués.

TABLE 13.2 – Classes de sortie proposées pour le modèle PMC

Classe de sortie	Composition	Nombre d'essais	Fréq. (%)	γ_d (kN/m ³)
1	100/0	19	26.7	11.3 - 12.5
2	85/15	10	14	13.5 - 14.1
3	70/30	14	20	14.5 - 16.3
4	55/45	10	14	14.4 - 15.6
5	0/100	18	25.3	11.6 - 13.9

La phase suivante consiste à définir les entrées du modèle. On définit ainsi un paramétrage grâce à l'analyse du pénétrogramme qui constitue le vecteur d'entrées vis-à-vis du problème de classification. L'ensemble des variables pris en compte pour la définition du jeu de paramètres est basé sur le paramétrage mis en oeuvre dans le cas d'application étudié auparavant. Cependant, cette fois les variables considérées ne présentent pas de corrélation entre elles. De même, les paramètres d'entropie n'ont pas été inclus puisqu'ils sont influencés indirectement par le nombre de mesures du pénétrogramme analysé, compte tenu qu'il s'agit d'un cumul des probabilités déduits de l'histogramme expérimental. Un total de sept variables d'entrées a ainsi été pris en compte, à savoir :

1. moyenne $\ln(q_d)$,
2. coefficient de variation de q_d ,
3. l'asymétrie de q_d ,
4. le coefficient d'aplatissement de q_d ,
5. pente de la tendance linéaire du pénétrogramme,
6. longueur de la forme de l'onde,
7. puissance spectrale maximale.

Une fois que les entrées et les sorties du modèles sont définies, la phase de conception du PMC est abordée. Pour ce faire, on a retenu une division du jeu de données courante : 70% pour l'entraînement et 15% pour l'ensemble de test et de validation. L'architecture optimale du RNA a été définie à l'aide d'une procédure essai/erreur et la validation croisée par *early stopping* a été adoptée comme critère d'arrêt afin d'éviter le sur-apprentissage. L'évolution selon le nombre de neurones de la moyenne du % des pénétrogrammes correctement classés pour les ensembles de validation et de test est représentée dans la figure (Fig.13.4).

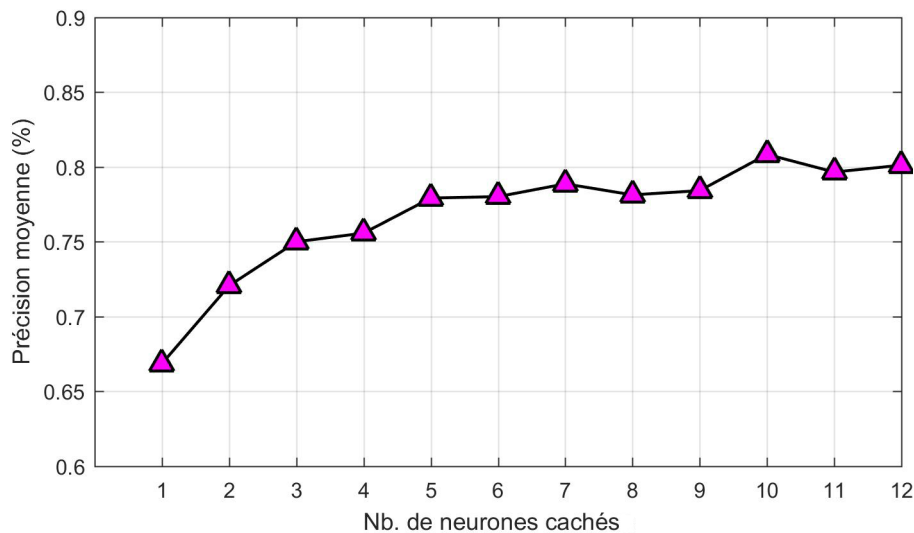


FIGURE 13.4 – Évolution du % moyen de pénétrogrammes bien classés du jeu de données de validation et de test en fonction du nombre de neurones de calcul

Le modèle final de PMC choisi est celui qui fournit la classification la plus précise des données pour l'ensemble de validation et de test. Il présente 8 neurones dans la couche cachée. Les résultats sont présentés sous forme de matrice de confusion dans (Tab.13.5). Le fort taux global de pénétrogrammes correctement classés (91.5%) met en évidence la bonne performance du PMC, avec aussi une bonne précision pour l'ensemble de validation (82%)

et de test (91%). On remarque qu'aucun essai réalisé sur un mélange constitué uniquement de pellets (100/0) ou uniquement de concassé (0/100) n'est classé de manière incorrecte. Par contre, la classification des essais effectués sur des mélanges appartenant aux classes n°2, 3 et 4 présente plus d'incertitude.

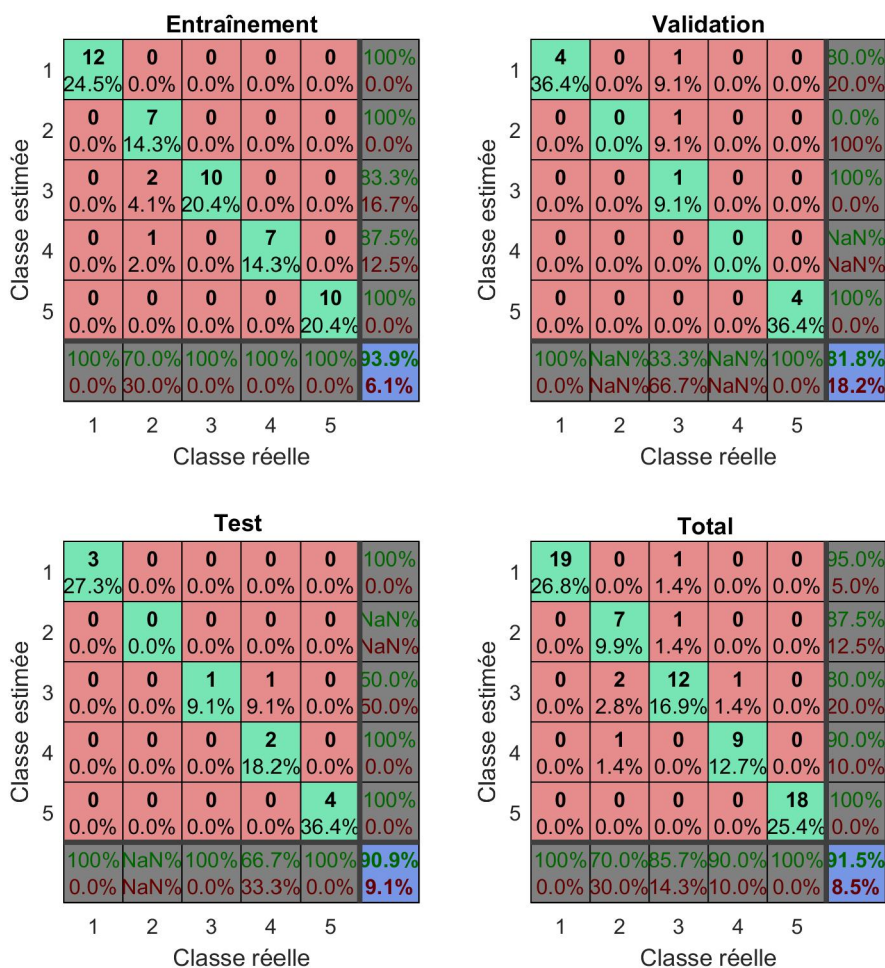


FIGURE 13.5 – Matrices de confusion du PMC

13.4 Bilan

Un cadre d'application plus simple de la méthodologie de classification proposée dans ce travail et basé sur l'application des techniques de Machine Learning a été présenté ici. Le matériau étudié est une argile bentonitique, formée par différents mélanges de billes sèches avec un diamètre de 32 mm et de concassé. Un RNA de type PMC a été étudié, afin d'identifier le cinq types de mélanges définis dans le protocole expérimental en laboratoire. Le modèle de classification développé a montré une bonne performance. Cependant, plus d'essais seraient nécessaires afin de confirmer la capacité de généralisation pour les mélanges des fractions granulaires, des classes 2, 3 et 4.

14. Conclusions

Une méthodologie de classification automatique par réseaux de neurones basée sur l'analyse du signal pénétrométrique a été proposée et deux cas d'application ont été étudiés.

Le premier cadre d'application vise la classification des sols dans un contexte géotechnique. Pour ce faire, deux types d'architectures ont été envisagées : le PMC et le PNN. L'apprentissage de ces deux outils a été réalisé grâce à une base de plus de 200 essais Panda, la plupart effectués en laboratoire mais la base comptait aussi des essais in situ. Pour ce premier jeu de données, les deux réseaux ont fourni des résultats très prometteurs, avec une performance supérieure pour le PMC. Ensuite, les deux modèles de classification ont été testés sur des essais réalisés sur de nouveaux matériaux et de nouveaux essais in situ. Les résultats obtenus ont montré des performances moindres que sur la base d'apprentissage et des difficultés à classer les sols intermédiaires (classe 2 et 3) ce qui met en doute la capacité de généralisation des modèles développés. De même, plusieurs remarques peuvent être formulées par rapport à la méthode déployée :

i. Base de données

La base de données d'apprentissage et de calibration des modèles de RNA est constituée en majorité (70%) d'essais en laboratoire et d'essais in situ. La base de données en laboratoire contient plus de 20 matériaux avec des matrices granulométriques assez différentes (argile, limons, sables et graves). Par contre, la diversité des sols des essais in situ est très restreinte (une écrasante majorité de sols fins - classe 1). De ce fait, la représentativité des sols intermédiaires de classe 2 et 3, sols qui posent le plus de difficulté de classification, s'avère trop faible et devra être étoffée à l'avenir afin de mieux exercer les méthodes RNA et les rendre plus robustes.

De manière général, les méthodes d'apprentissage sont très tributaires du nombre de données disponible pour l'entraînement et il est préférable de disposer d'un grand nombre de données plutôt que de tenter de développer des algorithmes de classification « plus intelligents » ([Domingos \(2012\)](#)). Ceci est spécialement important vis-à-vis de la problématique étudiée ici, où l'analyse est basée sur un seul paramètre du sol, la résistance de pointe dynamique et que la variabilité naturelle entraîne une difficulté supplémentaire dans l'application de ce type de méthode.

ii. Entrées du modèle

La répartition granulométrique du sol a une influence sur le comportement mécanique, alors qu'il n'existe pas un lien biunivoque entre ces deux propriétés. C'est pourquoi, on a essayé de fournir une représentation du pénétrogramme par le biais d'un jeu de paramètres comme entrées du modèle. Or, même si les RNA sont des outils qui présentent une bonne tolérance face au problème de multicolinéarité¹, il pourrait être

1. La multicolinéarité survient lorsque certaines variables d'entrée du modèle sont corrélées avec d'autres.

plus intéressant de réduire le jeu de paramètres proposé (en éliminant les variables corrélées) afin de rendre la méthode plus robuste. De même, la normalisation des paramètres d'entropie issus de l'analyse non linéaire aurait pu permettre d'améliorer la robustesse des méthodes développées.

iii. **Classes de sortie**

Les sols testés ont été classés selon le système de classification GTR. C'est pourquoi on a proposé des classes granulométriques basées sur ce système. Même si cette classification est très utilisée en géotechnique générale, une définition des classes basée sur les seuils granulométriques de la classification universelle LPC/USCS aurait été plus cohérente dans un contexte de reconnaissance géotechnique des sols.

iv. **Architectures des RNA utilisés**

Les deux RNA étudiés dans ce travail ont fourni des résultats similaires, mais on a abouti à des résultats légèrement plus performants avec le PMC. Cependant, l'entraînement du réseau PNN s'avère plus simple puisqu'il ne nécessite la détermination que d'un seul paramètre du modèle. Ils s'avèrent être deux outils intéressants et performants pour développer des systèmes de classification des matériaux géotechniques.

Enfin, les RNA sont « conditionnés » par les données utilisées lors de la calibration du modèle. C'est ainsi que les RNA développés dans le cadre de cette étude peuvent être étalonnés à nouveau avec une base de données élargie lorsque l'on disposera de nouvelles données et afin de s'adapter aux conditions locales. Toutefois, cette tâche n'a pas pu être abordée par manque de temps.

Il faut également noter que les outils et la méthodologie développés pourront être adaptés et élargis aux données issues du pénétromètre Panda3 et/ou de la géo-endoscopie qui fournissent plusieurs paramètres sur les sols traversés permettant ainsi d'améliorer notablement l'efficacité des RNA et ouvrant des perspectives très intéressantes pour cette technique en géotechnique.

On a présenté un deuxième cas d'application pour la classification de cinq mélanges de matériaux granulaires composés de pellets et de concassé de bentonite. Malgré le faible nombre de données pour certains mélange, une bonne performance du réseau RNA de type PMC a été obtenue.

En résumé, quand les essais de pénétration sont réalisés sur des matériaux homogènes en conditions "contrôlées", l'utilisation des techniques d'intelligence artificielle s'avère très performante. Cependant, la réalité des sols in situ est plus complexe et l'application de cette approche à la reconnaissance géotechnique des sols in situ à partir de la résistance de pointe seule est plus compliquée à mettre en œuvre en conservant le même taux de réussite. L'incorporation d'autres mesures ou données présente des atouts intéressants pour le développement et l'application fiable de ces approches. Un outil de reconnaissance tel que le Panda3 ou la possibilité d'incorporer des mesures issus de l'analyse d'image par la géondoscopie sont des pistes d'étude très intéressantes et prometteuses à développer.

Quatrième partie

Modélisation de la variabilité spatiale
des sols de surface en vue d'obtenir un
modèle du terrain

15. État de l'art et cadre théorique

15.1 Introduction

Maintenant que des outils de traitement du signal pénétrométrique en vue de reconnaître les unités de sol homogènes et de classer les sols traversés, ont été développés dans les deux parties précédentes, on s'intéresse dans cette partie au développement d'une méthodologie intégrant ces informations à l'échelle d'un site en vue de définir un modèle de distribution de la résistance de pointe q_d à l'échelle d'un terrain. Pour ce faire le modèle doit intégrer la variabilité spatiale du site tant verticale qu'horizontale. Avant de présenter la démarche mise en œuvre, nous revenons d'abord l'état de l'art concernant l'évaluation et la prise en compte de la variabilité spatiale.

15.2 Modélisation probabiliste de la variabilité spatiale

15.2.1 Introduction

Depuis le travail pionnier de (Lumb (1966)) la modélisation probabiliste de la variabilité spatiale repose principalement sur 3 techniques mathématiques : la théorie des champs aléatoires, la théorie des variables régionalisées (géostatistique) et les approches fractales (Jaksa (2007)). L'utilisation du formalisme des champs aléatoires et des variables régionalisées a été plus répandu et accepté dans le domaine géotechnique tandis que les exemples d'application des approches fractales pour la modélisation des données géotechniques sont encore limités. Un des rares exemples est donné par (Fenton (1999b)) qui applique une modélisation par ondelettes aux mesures de résistance de pointe q_c issues des essais CPT.

15.2.2 Théorie des champs aléatoires

Généralités

La théorie des champs aléatoires est une extension n-dimensionnelle de l'analyse des séries temporelles (Vanmarcke (1983)). Le champ aléatoire est ainsi décrit comme une famille de variables aléatoires définies dans le temps ou dans l'espace. Lorsqu'on analyse la variabilité spatiale des paramètres géotechniques, le domaine temporel est remplacé par le domaine spatial.

Soit $V(X)$ une variable géotechnique quelconque définie dans l'espace X du domaine R^p ($p = 1, 2$ ou 3). Elle est considérée comme aléatoire dans le domaine spatial d'étude, et

l'ensemble des variables constitue le champ aléatoire ou stochastique. Ce champ peut être décrit par :

- i. une mesure de tendance centrale, l'espérance μ ;
- ii. une mesure de dispersion sous la forme de variance σ^2 , ou sa racine carrée l'écart-type σ ou son coefficient de variation CV ;
- iii. sa fonction de répartition ou sa densité de probabilité.

Ces grandeurs sont définies ci-dessous :

$$\text{Espérance} \quad \mu_V(X) = E[V(X)] \quad (15.1)$$

$$\text{Variance} \quad \sigma_V^2(X) = Var[V(X)] = E[(V(X) - \mu_V)^2] \quad (15.2)$$

$$\text{Coefficient de variation} \quad CV(X) = \frac{\sigma_V(X)}{E[V(X)]} \quad (15.3)$$

$$\text{Fonction de répartition} \quad F_V(v, X) = P[V(X) \leq v] \quad (15.4)$$

La dépendance spatiale de la variable d'étude peut être caractérisée par la covariance (ou autocovariance), la densité spectrale ou la fonction de variance. Dans ce manuscrit, on décrira également la fonction d'autocorrélation employée dans la démarche proposée.

La fonction d'autocorrélation est fonction du moment statistique d'ordre 2 dont la version centrée est donnée par l'intermédiaire de la fonction d'autocovariance :

$$\begin{aligned} C_V(X_i, X_j) &= cov[V(X_i), V(X_j)] = E[(V(X_i) - \mu_V(X_i))[(V(X_j) - \mu_V(X_j))] = \\ &= E[(V(X_i)V(X_j)] - \mu_V(X_i)\mu_V(X_j) \end{aligned} \quad (15.5)$$

L'autocovariance représente le degré de dépendance linéaire entre les valeurs de la propriété étudiée en deux points différents X_i et X_j . Lorsque la covariance est normalisée par l'écart type de la variable, on obtient un coefficient adimensionnel d'autocorrélation :

$$\rho_V(X_i, X_j) = \frac{C_V(X_i, X_j)}{\sigma_V(X_i)\sigma_V(X_j)} \quad (15.6)$$

par définition $\rho \in [-1, 1]$ et $\rho(0) = 1$. On appelle ainsi fonction d'autocorrélation, la covariance normée. Un résultat important est que la valeur moyenne d'une propriété $V(X)$ dans un domaine A (segment, surface, volume) est donnée par :

$$V_A = \frac{1}{A} \int_A V(X) dX \quad (15.7)$$

La variance de cette moyenne (éq.15.7) est :

$$Var[V_A] = \frac{\sigma_V^2(X)}{A^2} \int_A \int_A \rho_V(X_i, X_j) dX_i dX_j \quad (15.8)$$

L'équation (éq.15.8) représente alors ce qui a été défini comme l'effet de réduction de variance (Auvinet (2002), Favre (2004)) : la variance de la valeur moyenne d'une propriété aléatoire dans un domaine A diminue avec la taille du domaine considéré. La variance ponctuelle σ_V^2 d'une propriété peut alors être reliée à la variance de la moyenne spatiale $Var[V_A]$ par

l'intermédiaire du facteur de réduction de variance (ou fonction de variance) Γ_A^2 comme l'a indiqué (Vanmarcke (1983)) :

$$\text{Var}[V_A] = \Gamma_A^2 \sigma_V^2 \quad (15.9)$$

Plusieurs expressions analytiques (exponentielles et gaussiennes) de la fonction de variance ont été introduites également par (Vanmarcke (1983)).

Processus stationnaire et transformation des données

La stationnarité au sens strict postule que la loi de probabilité du processus est invariante par translation. En pratique, cette hypothèse s'avère trop exigeante et on se limite à l'hypothèse de *stationnarité d'ordre 2*. Un champ est dit ainsi statistiquement homogène quand les deux premiers moments (espérance mathématique et variance), existent et peuvent être considérés constants dans le domaine d'étude, et la fonction d'autocovariance ne dépend que de l'écart qui sépare les couples de données considérés et pas de leur implantation particulière. Cette hypothèse est aussi connue dans la littérature sous le terme de *stationnarité au sens large* (Auvinet (2002), Christian and Baecher (2003)). On a ainsi :

$$E[V(X)] = \mu_V \quad (15.10)$$

$$C_V[V(X_i), V(X_j)] = C_V(X_i - X_j) = C_V(\tau) \quad (15.11)$$

à l'origine, cette fonction est égale à la variance de $V(X)$ et on appelle corrélogramme sa représentation graphique.

L'homogénéité statistique s'avère donc un préalable important vis-à-vis des analyses à mener. Lorsque la moyenne ou la variance ne sont pas constantes, on dit qu'il existe une tendance ou dérive. Ceci est courant dans le domaine de la géotechnique, où par exemple on peut observer souvent des augmentations systématiques des propriétés mécaniques le long d'un sondage par l'effet de la consolidation. Dans ce cas, des techniques de transformation des données peuvent être utilisées pour transformer les observations en données stationnaires, telles que la soustraction d'une tendance et des techniques de transformation de la variable (transformation logarithmique, transformation de Box-Cox) pour des distributions qui présentent une forte asymétrie. Le lecteur est invité à consulter (Jaksa (1995)) pour une description détaillée de ces techniques.

La technique de transformation la plus utilisée en géotechnique est la décomposition par dichotomie (Magnan (1982), Uzielli et al. (2006)). La variable d'étude est ainsi séparée en deux composantes additives : une fonction déterministe représentant la tendance, et une composante aléatoire de moyenne nulle constituée par l'ensemble des résidus. On signale que le modèle mathématique à retenir pour prendre en compte cette dérive relève d'un choix purement méthodologique (Christian and Baecher (2003), Uzielli et al. (2006)). Pour la résistance de pointe pénétrométrique, des modèles linéaires ((ex. Campanella et al. (1987), Popescu et al. (1998), Fenton (1999b), Phoon et al. (2003), Uzielli (2004), Uzielli et al. (2005), Nie et al. (2015)) et quadratiques ((Jaksa (1995), Jaksa (2007))) ont été retenus dans la littérature. L'ajustement de la tendance est fait par régression sur les données expérimentales (Uzielli et al. (2006)).

Enfin, l'hypothèse de stationnarité peut être vérifiée à l'aide de plusieurs tests proposés dans la littérature. On distingue deux approches principales : les tests paramétriques (une

distribution a priori des données est supposée) et les tests non paramétriques (aucune hypothèse sur la distribution sous-jacente des données). Le test non paramétrique du τ de Kendall (Jaksa (1995)), a été souvent utilisé en géotechnique pour vérifier l'indépendance statistique des données et donc la stationnarité relativement au modèle de tendance choisi par rapport aux données expérimentales (eg. Jaksa (1995), Salgado et al. (2015)). Une procédure plus rigoureuse pour des données en présence de corrélation spatiale est la version modifiée du test de Bartlett (Phoon et al. (2003)) décrit dans la partie III de ce manuscrit.

Champs aléatoires gaussiens

Afin d'appliquer le formalisme des champs aléatoires, il est nécessaire de retenir une fonction de densité de probabilité. Une hypothèse couramment employée est celle du champ gaussien dont la fonction de densité de probabilité est donnée par l'équation suivante :

$$f_V(v) = \frac{1}{\sigma_v \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{v-\mu_v}{\sigma_v} \right)^2} \quad (15.12)$$

L'intérêt d'un champ aléatoire gaussien est qu'il admet une caractérisation simple à partir de son espérance et de sa covariance. De plus, par définition, un champ aléatoire est gaussien si toute combinaison linéaire finie de ses coordonnées suit une loi normale sur le domaine d'étude.

Cependant, la loi de Gauss n'est pas toujours la mieux adaptée pour modéliser des propriétés du sol, dont la fourchette des valeurs est strictement positive. La distribution la plus simple pour modéliser des propriétés non négatives est la distribution log-normale (Fenton and Griffiths (2008), Phoon and Ching (2014)). Lorsqu'une variable aléatoire v suit une loi log-normale avec espérance λ et écart type ξ , son logarithme suit une loi normale de paramètres λ et ξ . La loi log-normale admet pour densité de probabilité :

$$f(v) = \frac{1}{v \xi \sqrt{2\pi}} e^{-\frac{1}{2} \left(\frac{\ln(v)-\lambda}{\xi} \right)^2} \quad v > 0 \quad (15.13)$$

où

$$\lambda = \ln \left(\frac{\mu}{\sqrt{1 + CV^2}} \right) \quad \xi = \sqrt{\ln(1 + CV^2)} \quad (15.14)$$

avec μ et CV l'espérance et le coefficient de variation de v .

Toutefois, il est possible de démontrer qu'un champ aléatoire quelconque peut être transformé en un champ normal au moyen d'une anamorphose gaussienne. De même, l'inversion de cette anamorphose permet de replacer le jeu de données dans le domaine d'origine.

Afin de vérifier l'adéquation d'une loi de distribution théorique sur un jeu de données quelconque avec un niveau de signification ou confiance α , plusieurs tests de conformité existent. Parmi eux, certains sont spécifiques à loi normale (test de normalité), d'autres sont applicables aux autres modèles théoriques ou encore adaptés à une certaine taille d'échantillon. Dans cet esprit, (Magnan (1982)) reporte plusieurs tests de conformité statistique. Le test χ^2 pour des cas où la taille de l'échantillon est importante ($n > 30$). L'hypothèse de normalité des données peut être testée à l'aide des tests de *Shapiro-Wilk* ($n < 30$) et le test non paramétrique de *Kolmogorov-Smirnov* (K-S), très souvent utilisé dans la littérature. Bien entendu, pour déterminer si un échantillon suit une loi log-normale, on peut aussi appliquer les tests de normalité à la transformée logarithmique de l'échantillon.

Types de champs aléatoires

(Auvinet et al. (2000)) proposent considérer plusieurs types de champs aléatoires suite à l'information disponible pour estimer les paramètres qui caractérisent le champ :

- i. champs estimatifs : le nombre de mesures s'avère insuffisant pour mener des analyses statistiques représentatives. On peut alors avoir recours à une estimation largement subjective des propriétés des sols grâce à l'expertise de l'ingénieur ;
- ii. champs descriptifs : même si notre connaissance de la réalité est toujours limitée on dispose d'un nombre suffisant de mesures fiables permettant de mener un analyse structurale afin d'obtenir les paramètres statistiques du champ ;
- iii. champs mixtes : il s'agit du cas intermédiaire entre les deux précédents. Les opinions expertes peuvent être introduites comme une information a priori grâce à la statistique bayésienne qui permet d'obtenir les probabilités a posteriori compte tenu des résultats expérimentaux disponibles.

15.2.3 Théorie des variables régionalisées

Généralités

Un processus naturel qui se déploie dans l'espace à 1, 2 ou 3 dimensions et y présente une organisation (ou structure) est qualifié de *phénomène régionalisé*. La "géostatistique" appelée ainsi par son concepteur Georges Matheron, se définit comme l'étude des phénomènes régionalisés. Une description détaillée du développement théorique de la géostatistique et de ses différents outils est disponible dans (Matheron (1962), Matheron (1970), Journel and Huijbregts (1978), Isaaks and Srivastava (1989)) et sur son application au domaine de la géotechnique (Bacconnet (1991)). Ici on se limitera à en donner les grandes lignes (dans le cadre de la géostatistique linéaire), à présenter les notions de base et les outils utilisés dans ce travail. Afin d'être cohérent avec la notation employée auparavant, la variable régionalisée sera notée $V(X)$.

La description mathématique censée représenter et mesurer le phénomène régionalisé est une fonction numérique appelée *variable régionalisée*. Une variable régionalisée présente les caractéristiques suivantes :

- i. aspect erratique : au niveau local des fluctuations erratiques très irrégulières ne peuvent pas être représentées par une fonction mathématique déterministe ;
- ii. structurée : globalement ces fluctuations présentent une organisation spatiale.

La variable régionalisée est étudiée à l'intérieur d'un domaine borné \mathcal{D} appelé champ d'extension de la variable. Les modèles géostatistiques considèrent la valeur $v(X)$ de la variable régionalisée en un point X du champ \mathcal{D} comme une réalisation d'une variable aléatoire $V(X)$. Lorsque X parcourt \mathcal{D} , on obtient une famille de variables aléatoires $V(X)$ qui constitue une fonction aléatoire ou processus stochastique. Pourtant, l'ensemble des valeurs $v(X)$ est une réalisation particulière de la fonction aléatoire.

De même, la variable régionalisée peut être définie en un point, mais aussi sur une surface ou volume, appelé support. Cette notion est importante puisque la distribution des données et concrètement leur variance est fonction du support sur lequel est définie la variable régionalisée (effet de support) en considérant la moyenne de la variable sur le support donné.

Hypothèse intrinsèque

Une fonction aléatoire $V(X)$ est dite intrinsèque quand les accroissements $V(X + \tau) - V(X)$ sont stationnaires d'ordre deux, c'est-à-dire :

$$E[V(X + \tau) - V(X)] = \mu(\tau) \quad (15.15)$$

$$\text{cov}[V(X_i + \tau_i) - V(X_i), V(X_j + \tau_j) - V(X_j)] = C_V(X_i - X_j; \tau_i, \tau_j) \quad (15.16)$$

avec $m(\tau)$ la dérive de la fonction aléatoire intrinsèque indépendant de X . Suite à la propriété (éq.15.16) qui fait intervenir que deux points X_i et X_j et deux vecteurs d'accroissement τ_i et τ_j , on peut définir le variogramme γ qui ne dépend que de l'écartement entre les deux points τ :

$$\gamma(\tau) = \frac{1}{2} \text{Var}[V(X + \tau) - V(X)] \quad (15.17)$$

Dans ce travail, on ne s'intéresse qu'aux fonctions aléatoires intrinsèques d'ordre 1 plus simples à manipuler ($m(\tau) = 0$). On signale quand même que les fonctions aléatoires intrinsèques d'ordre k (FAI- k) permettent de prendre en compte une dérive de degré supérieur à 1. L'absence de dérive permet de formuler la variance des accroissements comme l'espérance de leur carré :

$$\gamma(\tau) = \frac{1}{2} E\{[V(X + \tau) - V(X)]^2\} \quad (15.18)$$

Pour certains cas, l'utilisation du variogramme au lieu de l'autocovariance peut être avantageuse. En effet, le variogramme est un estimateur non biaisé puisque ne dépendant pas de l'espérance ni de la variance. Le variogramme est plus général puisqu'il peut être défini en dehors de l'hypothèse de stationnarité d'ordre deux. On remarque que la stationnarité du second ordre implique l'hypothèse intrinsèque mais la réciproque n'est pas vraie. Lorsque le variogramme est borné, la relation entre l'autocovariance et le variogramme est donnée par l'équation suivante :

$$\gamma(\tau) = C_V(0) - C_V(\tau) \quad (15.19)$$

On verra par la suite comment sa connaissance permettra de résoudre les problèmes d'estimation.

L'estimation locale : le krigeage

L'estimation (locale ou globale) au sens classique de prédiction, constitue un des intérêts principaux de la géostatistique. L'estimation globale consiste à caractériser le domaine \mathcal{D} entier par une valeur unique. Néanmoins, l'estimation locale est basée généralement sur un nombre limité des données et ne s'intéresse qu'à un secteur de la zone d'étude. Cette technique d'estimation appelée *krigeage* porte le nom de l'ingénieur Daniel Krige, qui lors des années 50 avait développé une série de méthodes empiriques d'estimation minière.

Le krigeage permet donc d'estimer la valeur de la variable régionalisée en un point ou sur un volume (appelé bloc) à l'aide d'une combinaison linéaire pondérée des données mesurées. Il permet aussi de mesurer la précision de la valeur estimée par l'intermédiaire d'une variance de l'erreur d'estimation (aussi appelée variance d'estimation ou variance de krigeage). Sa mise en œuvre nécessite la définition d'un modèle théorique de covariance ou variogramme pour prendre en compte la structure spatiale de la variable régionalisée ainsi que le choix du

voisinage, domaine de l'espace autour du point à estimer qui contient les points échantillonnés utilisés dans l'estimation.

La résolution d'un problème d'estimation par krigeage peut être synthétisé en 4 étapes :

i. Contrainte de linéarité

L'erreur de krigeage doit être une combinaison linéaire pondérée de la fonction aléatoire à étudier. D'après cette contrainte, l'estimateur s'écrit comme une combinaison linéaire des valeurs mesurées V_i sur les points de l'échantillon :

$$V^*(X) = \sum_{i=1}^n \lambda_i V_i \quad (15.20)$$

où les poids λ_i sont les inconnues des équations du krigeage, fonction de la structure spatiale de la variable régionalisée et de la configuration d'échantillonnage des données.

ii. Contrainte d'autorisation

Il doit être possible de calculer l'espérance et la variance de l'erreur du krigeage. Cette contrainte est toujours vérifiée dans le cadre de l'hypothèse de stationnarité d'ordre 2. Par contre, sous l'hypothèse intrinsèque, il faut imposer que la somme des poids de toute combinaison linéaire soit nulle.

iii. Contrainte de non-biais

Cette condition assure que le krigeage est un estimateur sans biais, ce qui équivaut à imposer que l'espérance de l'erreur d'estimation est nulle :

$$E[V^*(X) - V(X)] = 0 \quad (15.21)$$

iv. Contrainte d'optimalité

Cette contrainte correspond à la minimisation de l'erreur quadratique moyenne, le krigeage fournit alors une estimation à variance minimale. L'estimateur de krigeage est un estimateur linéaire efficace.

Il existe différentes variantes des systèmes de krigeage suite aux hypothèses retenues sur la fonction aléatoire représentant le phénomène régionalisé (stationnarité d'ordre deux, avec moyenne connue ou non, ou hypothèse intrinsèque). Dans le cadre de la géostatistique linéaire les deux principaux types de krigeage sont le *krigeage simple (KS)* à moyenne connue, le *krigeage ordinaire (KO)* à moyenne inconnue. Selon les textes consultés le krigeage universel (la variable régionalisée comporte une dérive déterministe) peut faire partie de la géostatistique linéaire ([Journal and Huijbregts \(1978\)](#)) ou non linéaire ([Emery \(2001\)](#)). Il peut être démontré que la variance d'estimation du krigeage simple est inférieure à celle du krigeage ordinaire. Toutefois, le krigeage simple retient comme hypothèse que la fonction aléatoire est stationnaire du second ordre. En revanche, les résultats issus du krigeage ordinaire sous l'hypothèse intrinsèque sont équivalents à ceux du krigeage simple dans le cas stationnaire d'ordre 2.

Krigeage à moyenne inconnue : krigeage ordinaire (KO)

La démonstration théorique des équations du krigeage ordinaire est donnée en annexe ([Annexe B](#)). Le système de $N + 1$ équations du krigeage ordinaire en notation matricielle

peut être écrit comme suit :

$$\begin{pmatrix} \gamma(X_1 - X_1) & \cdots & \gamma(X_1 - X_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(X_n - X_1) & \cdots & \gamma(X_n - X_n) & 1 \\ 1 & \cdots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ v \end{pmatrix} = \begin{pmatrix} \gamma(X_1 - X) \\ \vdots \\ \gamma(X_n - X) \\ 1 \end{pmatrix} \quad (15.22)$$

La matrice carrée, appelée matrice de krigeage, ne dépend que de la configuration des sondages et du variogramme. Le second membre de l'équation matricielle est fonction de la position relative du point à estimer par rapport aux sondages. D'où l'expression de la variance du krigeage ordinaire, qui fait intervenir le paramètre de Lagrange (Matheron (1970)), qu'on peut écrire en fonction du variogramme :

$$\sigma_{KO}^2(X) = \sum_{i=1}^n \lambda_i \gamma(X_i - X) + v \quad (15.23)$$

ou de la covariance :

$$\sigma_{KO}^2(X) = Var[V(X)] - \sum_{i=1}^n \lambda_i C_V(X_i - X) - v \quad (15.24)$$

Remarques sur le krigeage et les algorithmes d'interpolation

Tous les algorithmes d'interpolation fournissent l'estimation d'une variable comme une somme pondérée des échantillons voisins. Généralement, le poids attribué aux échantillons est contrôlé par une fonction mathématique et il décroît lorsque l'écart entre les points augmente. Le krigeage présente plusieurs avantages sur les techniques d'interpolation déterministes : méthodes de l'inverse des distances, splines, interpolations par les plus proches voisins, régression polynomiale, etc. (Li and Heap (2008)). D'une part, les techniques de krigeage fournissent une estimation précise sans biais de la grandeur recherchée, en tenant compte du nombre et de la configuration des points de mesures mais aussi de la structure spatiale du phénomène grâce au modèle variographique. En outre, il permet d'apprécier quantitativement la précision de l'estimation, à l'aide d'une variance d'estimation. Cependant le krigeage reste toujours un algorithme d'interpolation et vis-à-vis de l'estimation il fournira des résultats très similaires aux algorithmes d'interpolation déterministes dans certains cas (Isaaks and Srivastava (1989)) :

- lorsque le nombre de points de mesure est important et s'ils sont structurés dans une maille régulière, il est possible d'obtenir de bonnes estimations quelque soit l'algorithme d'estimation ou la technique d'interpolation employée ;
- lorsque les données forment des petits groupes isolés, on obtiendra des estimations peu fiables indépendamment de l'algorithme utilisé ;
- la plupart des algorithmes d'interpolation produisent un effet de "lissage", les valeurs élevées sont sous-estimées et les valeurs les plus faibles sont sur-estimées ;
- la carte des valeurs estimées par krigeage est plus lisse que celle des valeurs ponctuelles réelles. Cet effet de lissage est une caractéristique commune aux algorithmes d'interpolation car ils ne peuvent pas inventer les détails qui n'ont pas été observés

dans les données. Lorsque le but est de reproduire la variabilité spatiale du phénomène à étudier, il est nécessaire d'avoir recours aux techniques de simulations qu'on présente par la suite.

15.2.4 Méthodes de simulation

Les simulations sont des réalisations numériques, indépendantes les unes des autres, de variable aléatoire dont le but est de restituer les traits structuraux de la réalité. On distingue principalement les simulations conditionnelles ou non. Les simulations *non conditionnelles* sont des réalisations numériques respectant les deux premiers moments du processus aléatoire. Les simulations *conditionnelles* sont celles des réalisations qui respectent les données expérimentales.

Plusieurs techniques existent dans la littérature permettant d'obtenir des réalisations plausibles du champ ou fonction aléatoire stationnaire. Ces méthodes se distinguent par la loi spatiale de la fonction aléatoire, leur plus ou moins grande rigueur, leur rapidité de mise en œuvre et leurs limitations particulières. Les principales d'entre elles d'après (Fenton and Griffiths (2008)) sont décrites succinctement par la suite, en détaillant la technique de diagonalisation classique qui sera employée dans ce travail :

- 1) Diagonalisation ou décomposition de la matrice de covariance : soit \underline{C} la matrice de covariance $N \times N$ laquelle est définie positive par définition. Il existe donc une matrice triangulaire inférieure \underline{L} calculée par la technique de décomposition de Cholesky telle que $\underline{L} \underline{L}^T = \underline{C}$. Soit G un vecteur de N variables aléatoires normales réduites indépendantes $G \sim \mathcal{N}(0, 1)$, on définit le vecteur w :

$$\omega = L \cdot G \tag{15.25}$$

alors la variable aléatoire $W(X)$ correspondant à ω est une simulation non conditionnelle au sens qu'elle reproduit le modèle de covariance théorique à simuler.

Cette technique est la plus efficace lorsqu'on doit simuler de petits champs puisqu'il suffit de calculer la matrice triangulaire une fois et puis de réaliser des tirages aléatoires à partir d'une loi gaussienne standard pour obtenir plusieurs réalisations du champ à simuler. Un autre grand avantage est que c'est une méthode exacte qui peut être appliquée à n'importe quel modèle théorique de covariance admissible.

La principale contrainte est qu'elle est applicable tant que le nombre de points de la grille à simuler et la puissance informatique de l'ordinateur permettent l'inversion de la matrice de covariance par l'algorithme de Cholesky. Afin de résoudre le problème de la complexité du calcul informatique de l'algorithme, des approches basées sur l'utilisation d'une fenêtre glissante ont été proposées à l'époque (Alabert (1987)).

Enfin, on signale que cette méthode permet de réaliser presque aussi aisément des simulations conditionnelles que non-conditionnelles. (Davis (1987)) décrit la procédure des simulations conditionnelles par le biais de cette méthode dans une grille quelconque. D'après (Chiles (1995)), il s'agit d'un cas particulier de la méthode de simulation séquentielle gaussienne qu'on présentera après.

- 2) Méthode de moyennes mobiles : proposée par (Gersch and Yonemoto (1977)), elle consiste à appliquer un lissage par fenêtre glissante un nombre de fois ξ (qui peut être déterminé arbitrairement) aux nœuds d'une grille générés à partir d'une loi normale standard $\mathcal{N}(0, 1)$.

- 3) Méthodes basés sur la transformée de Fourier discrète (TFD) : pour le cas à une dimension, il est possible de recourir au formalisme des transformées de Fourier discrètes pour simuler un processus stationnaire $V(X)$ qui suit une loi normale. D'après (Chiles (1995)) la méthode spectrale discrète peut sans difficulté être étendue à 2D et 3D, et l'auteur cite (Pardo-Iguzquiza and Chica-Olmo (1993)) pour une description en détail des algorithmes nécessaires.
- 4) Méthodes basés sur la transformée rapide de Fourier (FFT) : cette approche proposée par (Cooley and Tukey (1965)), présente l'avantage d'utiliser directement la fonction de densité spectrale et s'adapte facilement au cas anisotrope, aux distributions non gaussiennes et à plusieurs dimensions. Par ailleurs, l'utilisation de la transformée rapide de Fourier permet d'accélérer les calculs par rapport au cas discret.
- 5) Méthode de bandes tournantes (TBM) : cette approche proposée par Matheron (Matheron (1970)) dérive d'une méthode spectrale qui permet de construire des simulations dans un espace à plusieurs dimensions 2D ou 3D à partir de simulations sur des droites 1D. Bien qu'il s'agit d'une des techniques les plus anciennes pour la simulation des champs, quand le nombre de droites ou lignes devient assez grand c'est une méthode qui s'avère plus efficace que la plupart des autres méthodes existantes. En effet, dans ces conditions, elle permet de reproduire avec grande précision la structure spatiale du champ.

(Emery and Lantuéjoul (2006)) rassemblent un certain nombre de recommandations concernant le nombre de lignes nécessaires pour obtenir un bon résultat. Pour simuler un champ en 2D :

(Mantoglou and Wilson (1982)) suggèrent de 4 à 16 lignes. (Chilès (1977)) recommande l'utilisation des 180 lignes et (Gneiting (1999)) trouve que 64 lignes permettent une excellente reproduction du modèle visé de covariance.

Par rapport aux simulations numériques en 3D :

(Lantuéjoul (2013)) a montré que des centaines de lignes sont nécessaires pour une simulation correcte des processus gaussiens. Ceci a été confirmé par les expériences menées par plusieurs auteurs (Tompson et al. (1989)), Freulon and Fouquet (1991)). Sur la base des considérations antérieures (Emery and Lantuéjoul (2006)) proposent un nombre de lignes égal à 1000.

Cependant, le fait de considérer un grand nombre de lignes ne représente pas une grande contrainte vis-à-vis de cette méthode, puisque la simulation sur une dimension est très rapide. La contrainte principale est qu'elle exige de connaître un modèle de covariance équivalent dans une dimension.

- 6) Simulation par subdivision en moyenne locale (LAS) : elle consiste à diviser le domaine d'étude en 2^i $i = 1, \dots, L$ sous domaines récursivement. À chaque étape et sur chaque sous domaine le champ est simulé en considérant la moyenne et l'autocorrélation des domaines supérieurs. On pourra se référer à (Fenton and Vanmarcke (1990)) pour plus de détails sur l'algorithme.

On signale que les méthodes de simulation citées auparavant, ne permettent pas d'obtenir des simulations conditionnelles d'une manière directe à l'exception de la méthode de diagonalisation de la matrice de covariance. Une autre alternative pour construire des simulations conditionnelles directement est la méthode de simulation séquentielle gaussienne qui permet de simuler un champ aléatoire d'une manière séquentielle grâce au krigeage de la

moyenne et l'écart-type des points successifs (Gómez-Hernández and Cassiraga (1994)). Elle peut être appliquée tant qu'il est possible de résoudre numériquement le système de krigeage, sinon il faut passer par des procédures avec une fenêtre glissante. Sinon, on doit construire une simulation non conditionnelle puis la conditionner aux données à l'aide de l'algorithme classique qui consiste à ajouter au krigeage la simulation de l'erreur de krigeage. On obtient alors une simulation conditionnelle du champ V_{cs} comme suit :

$$V_{cs}(X) = V_{kd}(X) + [V_{us}(X) - V_{kus}(X)] \quad (15.26)$$

Le principe est le suivant :

- i. tirage d'une simulation non conditionnelle en tout point du domaine d'étude V_{us} ;
- ii. krigeage aux points de simulation en utilisant les valeurs observées V_{kd} ;
- iii. krigeage aux points de simulation en utilisant les valeurs de la simulation non conditionnelle aux points observés V_{kus} .

Pour introduire l'incertitude sur les propriétés du sol, il s'avère nécessaire de passer de la représentation du champ continu à un nombre limité de variables aléatoires soit aux nœuds du maillage déterministe, soit aux milieux des mailles du maillage du modèle de calcul. Il s'agit donc de méthodes de discrétisation des champs aléatoires. Une revue exhaustive de ces méthodes est fournie dans (Sudret and Der Kiureghian (2000)), qui distingue trois groupes :

- 1) Méthodes de représentation aux points : telle que la méthode du point milieu (*Mid-point MP method*), qui est une des méthodes les plus simples. Elle a été introduite par (Der Kiureghian and Ke (1987)), et consiste à discrétiser le champ en assimilant la valeur sur un élément à celle de son centre de gravité ;
- 2) Méthodes de discrétisation par lissage : par exemple la méthode par subdivision en moyenne locale (LAS) citée auparavant.
- 3) Méthodes d'expansion en séries : plusieurs approches sont incluses dans cette classe. Parmi elles la représentation de Karhunen-Loève (Ghanem and Spanos (1991)), basée sur la décomposition spectrale de la fonction d'autocovariance à partir de la résolution de l'équation de intégrale de Fredholm. Très succinctement, il s'agit de trouver les valeurs et fonctions propres de cette intégrale. La solution analytique de cette intégrale n'a été démontrée que pour un champ homogène avec un modèle de covariances exponentiel ou triangulaire dans un domaine rectangulaire (Spanos and Ghanem (1989)). Sinon il faut avoir recours à des méthodes numériques telles que les méthodes de projection (polynômes de Galerkin) ou les méthodes de Nyström. Pour plus de détails, on renvoie à (Betz et al. (2014)) qui ont étudié la performance de ces algorithmes. D'autres approches sont celles de l'expansion en séries orthogonales (OSE) (Zhang and Ellingwood (1994)) et la discrétisation par estimation linéaire optimale (*Expansion Optimale Linear Estimation Method* EOLE) (Li and Der Kiureghian (1993)). Il s'agit d'une extension de la théorie de l'estimation linéaire optimale¹ au moyen d'une décomposition spectrale de la matrice de covariances.

1. Cette approche est basée sur la méthode de représentation linéaire optimale (*Optimal Linear Estimation Method* (OLE)) présentée aussi dans (Li and Der Kiureghian (1993)). Il faut noter que la méthode OLE est équivalente à la technique de krigeage puisqu'elle repose sur les mêmes contraintes.

15.3 Corrélation spatiale des propriétés du sol

15.3.1 Introduction

Les cadres théoriques exposés auparavant s'appuient sur l'identification des paramètres d'un champs aléatoire censé représenter le plus fidèlement possible les variations observées dans la nature. Parmi ces paramètres, la longueur de corrélation a été définie de façon très différente et avec des méthodes d'évaluation tout aussi différentes. Les praticiens ont eux mêmes essayé de donner des ordres de grandeur de ce paramètre pour les grandeurs géotechniques usuelles au gré des études de cas rapportées dans la littérature. De sorte qu'il est nécessaire à ce stade de ce mémoire, de s'arrêter sur les formes attachées à cette notion de longueur de corrélation.

Dans le cadre de la géostatistique développée par Matheron, c'est la portée, le paramètre qui définit la limite du voisinage de corrélation. Lorsque Vanmarcke développe la théorie des champs, il définit une nouvelle grandeur afin de caractériser cette distance caractéristique, l'échelle de fluctuation, qu'on notera par θ dans ce texte.

15.3.2 Notion d'échelle de fluctuation ou longueur de corrélation

D'une manière générale, l'échelle de fluctuation θ peut être définie comme la distance au-delà de laquelle les paramètres mesurés ne présentent pas de corrélation. Sous l'hypothèse d'une variance finie σ^2 , la variance sur un volume est notée par σ_v^2 . Ainsi, $v\sigma_v^2$ tend vers $\theta\sigma^2$ lorsque v tend vers l'infini. L'échelle de fluctuation est ainsi définie comme :

$$\theta = \lim_{v \rightarrow \infty} v \frac{\sigma_v^2}{\sigma^2} \quad (15.27)$$

Les valeurs faibles de θ impliquent des fluctuations rapides autour de la moyenne, tandis que les grandes valeurs témoignent d'une plus grande continuité autour de la valeur moyenne. Une valeur de θ égale à l'infini est équivalent à supposer une corrélation parfaite entre les propriétés des sols à étudier, en d'autres termes une valeur parfaitement constante en tout point de l'espace. Par contre, quand θ est nulle, on tombe sur une indépendance spatiale, en d'autres termes, une valeur parfaitement aléatoire en tout point, sans lien avec ses voisins.

Théoriquement, cette échelle correspond à l'intégrale de la fonction d'autocorrélation (Vanmarcke (1977a)) ou à deux fois la région sous la partie positive de la fonction d'autocorrélation qui est une fonction paire, ce qui revient au même :

$$\theta = \int_{-\infty}^{+\infty} \rho(\tau) d\tau = 2 \int_0^{+\infty} \rho(\tau) d\tau \quad (15.28)$$

Souvent, on parle de distance, longueur de corrélation ou échelle de fluctuation sans faire de distinction (Fenton and Griffiths (2008)). Cependant, il convient de mentionner qu'il existe d'autres définitions disponibles dans la littérature. (Diaz Padilla Guerrero (1974)) ont défini la longueur de corrélation comme la distance nécessaire pour que la fonction d'autocorrélation soit égale à e^{-1} (0.368). Elle se rapporte à une technique de calcul. Cette définition est aussi préconisée par d'autres auteurs (ex. DeGroot and Baecher (1993)).

La notion d'échelle de fluctuation sera donc ici retenue pour caractériser la structure spatiale des paramètres à étudier. Plusieurs approches ont été proposées dans la littérature

TABLE 15.1 – Fonctions d'autocorrélation théoriques (source : Vanmarcke (1977a), Vanmarcke (1983))

Modèle	Fonction d'autocorrélation	Échelle de fluctuation θ
Exponentiel (SNX)	$\rho(\tau) = e^{- \tau /L_1}$	$\theta = 2L_1$
Gaussien (SQX)	$\rho(\tau) = e^{-(\tau/L_2)^2}$	$\theta = \sqrt{\pi}L_2$
Cosinus (CSX)	$\rho(\tau) = \cos\left(\frac{\tau}{L_3}\right) e^{- \tau /L_3}$	$\theta = L_3$
Markov d'ordre 2 (SMK)	$\rho(\tau) = \left(1 + \frac{ \tau }{L_4}\right) e^{- \tau /L_4}$	$\theta = 2L_1$

pour déterminer expérimentalement la valeur de l'échelle de fluctuation. On présente ici celle basée sur la fonction d'autocorrélation théorique qui est la plus employée. Les autres méthodes sont données en annexe (Annexe B).

Vanmarcke (Vanmarcke (1978), Vanmarcke (1983)) propose que l'échelle de fluctuation soit déduite de la fonction d'autocorrélation grâce aux relations fournies dans le tableau (Tab. 15.1). Par conséquent, θ peut être estimée en ajustant un modèle théorique à l'autocorrélogramme empirique par le biais de diverses techniques (ex. régression par moindres carrés, méthode de maximum de vraisemblance). L'expression générale pour estimer la fonction de corrélation empirique est $\hat{\rho}(\tau)$:

$$\hat{\rho}(\tau_j) = \frac{1}{\hat{\sigma}^2(n-j)} \sum_{i=1}^{n-j+1} (V_i - \hat{\mu})(V_{i+j} - \hat{\mu}) \quad (15.29)$$

où $\hat{\mu}$ et $\hat{\sigma}$ les moyennes et écart-types empiriques, la distance entre données espacées régulièrement $\tau_j = j\Delta\tau$ et $j = 1, 2, \dots, n$ étant n le nombre d'observations.

Le modèle exponentiel est le plus utilisé en géotechnique (ex. Vanmarcke (1977a), Auvinet and Gonzalez (2000), Fenton and Griffiths (2002), Jaksa et al. (2005), Lloret-Cabot et al. (2014), Juárez-Camarena et al. (2016)). La base de données des lois d'autocorrélation théoriques utilisées en littérature géotechnique élaborée par (Huber (2013)) montre aussi que ce choix est le plus répandu. (Rackwitz (2000)) propose une revue exhaustive des différents modèles théoriques d'autocorrélation.

Enfin, l'ajustement d'une loi d'autocorrélation théorique est la procédure la plus souvent utilisée, les longueurs de corrélation des propriétés du sol apparaissant comme des paramètres des lois ajustées (Auvinet (2002), Uzielli et al. (2006)).

15.3.3 Notion de portée

Sous l'hypothèse stationnaire d'ordre 2, le variogramme apparaît comme l'image miroir de la covariance (éq.15.1). Dans ce cas, la portée a est la valeur de la distance au-delà de laquelle les accroissements ($V(X+\tau) - V(X)$) sont sans corrélation (ou négligeables) (Matheron (1970)). La portée est aussi la distance à partir de laquelle le demi-variogramme atteint sa valeur limite, qui est la variance a priori de la variable aléatoire. Bien entendu, la définition géostatistique de la portée est liée à l'existence du palier, puisque si le variogramme n'est pas borné, il ne peut pas avoir une portée finie. Toutefois, cette situation ne serait pas compatible avec l'hypothèse stationnaire et la fonction de covariance ne pourrait pas exister.

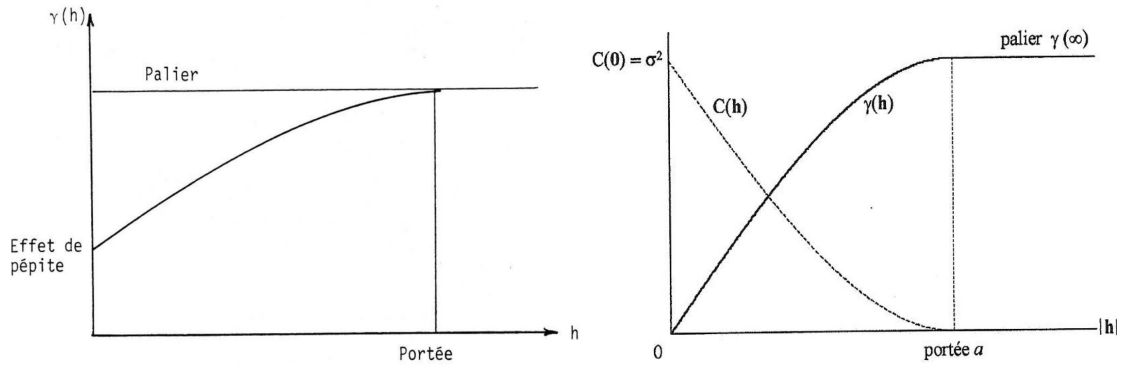


FIGURE 15.1 – À gauche : exemple de variogramme à palier (source : [Bacconnet \(1991\)](#)). À droite : relation entre variogramme et covariogramme (source : [Emery \(2001\)](#))

Dans la figure ([Fig.15.1](#)) on définit également l'effet pépité. Il s'agit d'une discontinuité à l'origine du variogramme liée à l'échelle d'observation. Il peut intégrer plusieurs facteurs comme les erreurs de mesure, l'absence naturelle de structuration spatiale, ou encore la variabilité à échelle "microscopique" qui n'est pas détectable car plus faible que la distance entre les échantillons.

La portée est une mesure de l'échelle à laquelle le phénomène est visible. La portée ainsi définie est une portée pratique qui fixe le domaine d'autocorrélation ([Bacconnet \(1991\)](#)). On peut retrouver la grandeur équivalente à l'échelle de fluctuation θ , appelée portée intégrale dans le cadre de la géostatistique. Une comparaison entre ces deux grandeurs pour trois modèles de variogramme stationnaires souvent utilisés est présentée dans le tableau ([Tab.15.2](#)), où le palier est noté par C . La valeur de la portée pratique est donc une limite purement conventionnelle et souvent supérieure à la portée intégrale. Par ailleurs, les points qui sont éloignés de plus de cette distance ont une faible influence dans l'élaboration des estimateurs de krigeage.

En ce qui concerne l'estimation du variogramme empirique, elle est souvent réalisée au moyen de la méthode classique des moments proposé par Matheron, qui pour une distance τ a l'expression suivante :

$$\hat{\gamma}(\tau) = \frac{1}{2N(\tau)} \sum_{i=1}^{N(\tau)} [V(X + \tau) - V(X)]^2 \quad (15.30)$$

avec $N(\tau)$ le nombre de couples de points distants de τ qu'il est possible d'extraire de l'échantillon des données.

15.3.4 Anisotropies

La continuité spatiale d'une grandeur physique à étudier n'est pas nécessairement la même dans toutes les directions. Au sens général, un système est isotrope si aucune de ses propriétés ne présente une dépendance directionnelle, à l'opposé d'un système anisotrope.

Il est bien connu que la variabilité spatiale des caractéristiques mécaniques ou physiques du sol est structurée à différentes échelles et qu'il existe la plupart du temps une anisotropie importante entre les directions horizontales et verticales. Ces structures de variabilité sont

TABLE 15.2 – Principaux modèles de variogramme (source : [Bacconnet \(1991\)](#))

Modèle	Variogramme à palier	Portée pratique	θ
Sphérique	$\gamma(\tau) = C \left(\frac{3\tau}{2a} - \frac{1}{2} \left(\frac{\tau}{a} \right)^3 \right)$ pour $\tau < a$ $\gamma(\tau) = C$ pour $\tau \geq a$	a	$\frac{3}{4}a$
Exponentiel	$\gamma(\tau) = C \left(1 - e^{-\frac{\tau}{a}} \right)$	$3a$	$2a$
Gaussien	$\gamma(\tau) = C \left(1 - e^{-\left(\frac{\tau}{a} \right)^2} \right)$	$a\sqrt{3}$	$a\sqrt{\pi}$

TABLE 15.3 – Rapports d’anisotropie courants suite au processus géologique d’après ([Journal and Huijbregts \(1978\)](#), [Koltermann and Gorelick \(1996\)](#), [Deutsch and Journal \(1998\)](#), [Chiles and Delfiner \(1999\)](#), [Deutsch \(2002\)](#), [Bakhtiari \(2011\)](#), [Huber \(2013\)](#), [Wackernagel \(2013\)](#))

Processus géologique	Rapport d’anisotropie θ_H/θ_V		
	10 : 1	100 : 1	1000 : 1
Barre de méandre	█		
Vallée fluviale		█	
Processus d’érosion éolienne		█	
Estuaire		█	
Dépôt sédimentaire sous marin		█	
Dépôt deltaïque		█	
Carbonates de plateforme continentale			█

principalement liées aux modes de dépôts (eg. [Bacconnet \(1991\)](#), [Jaksa and Fenton \(2002\)](#)). Bien qu’une structure de corrélation isotrope soit souvent retenue dans la littérature afin de simplifier les analyses (ex. [Salgado and Kim \(2013\)](#), [Gong et al. \(2016\)](#)), on étudie l’anisotropie de manière séparée. On peut distinguer ainsi une distance d’auto-corrélation horizontale θ_H et une distance d’auto-corrélation verticale θ_V qui présente des valeurs souvent plus faibles par rapport à la direction horizontale. Un proportion de 1 à 10 pour ces longueurs de corrélation est courante d’après ([Christian and Baecher \(2003\)](#)). Cet ordre de grandeur est cohérent avec les valeurs reportées par d’autres auteurs : ex. 9 ([Vanmarcke \(1977b\)](#)), 10 ([Soulie et al. \(1990\)](#)) ou 12 ([Popescu \(1995\)](#)). L’ordre de grandeur de ce rapport d’anisotropie a été mis en relation avec divers processus géologiques par plusieurs auteurs ([Tab.15.3](#)).

Vis-à-vis de la caractérisation de la loi spatiale, on distingue deux types d’anisotropie : anisotropie géométrique et zonale. Par définition, les variogrammes des différentes directions auront le même palier mais pas la même portée lorsqu’on parle d’anisotropie géométrique. Une expression isotrope du variogramme peut être donnée par transformation linéaire des coordonnées spatiales. L’anisotropie zonale se produit quand il existe une direction de l’espace qui exhibe un niveau de variabilité différent par rapport aux autres directions.

([Zhu and Zhang \(2013\)](#)) présentent la formulation théorique des structures de corrélation sous la forme de fonctions d’autocorrélation pour cinq cas d’anisotropie à l’échelle macroscopique en 2D : isotropie et anisotropie transversale ([Fig.15.2](#)), anisotropie en rotation, anisotropie générale et anisotropie générale en rotation. Par ailleurs, les processus géologiques à l’origine de ces formes d’anisotropie sont mis en correspondance ([Tab.15.4](#)).

TABLE 15.4 – Correspondance entre les quatre formes d’anisotropie et les formations naturelles et anthropiques (source : [Zhu and Zhang \(2013\)](#)).

Formes	Cas
Anisotropie transversale	Mode de dépôt horizontal, processus fluviaux, compactage
Anisotropie en rotation	Dépôts affectés par glissement
Anisotropie générale	Dépôts affectés par glissement, colluvions, piles de mines
Anisotropie générale en rotation	Dépôts affectés par glissement

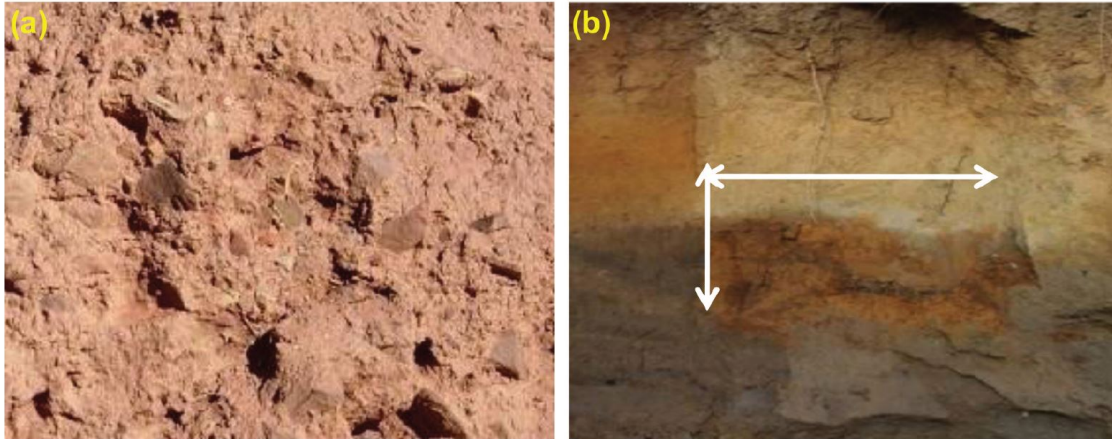


FIGURE 15.2 – (a) Formation isotrope et (b) anisotrope transverse (source : [Zhu and Zhang \(2013\)](#)).

Compte tenu de ces constats, dans la pratique courante on assume que l’échelle de fluctuation est constante mais anisotrope et plusieurs auteurs suggèrent donc d’utiliser un modèle de corrélation orthotrope ou séparable (ex. [Vanmarcke \(1977a\)](#), [Auvinet and Gonzalez \(2000\)](#)). D’ailleurs, la fonction de décroissance exponentielle dont l’expression théorique est donnée ci-dessous, s’avère le choix le plus commun :

$$\rho(x, y, z) = \exp \left[-2 \left(\frac{\tau_V}{\theta_V} + \frac{\tau_H}{\theta_H} \right) \right] \quad (15.31)$$

avec θ_V l’échelle de fluctuation selon la direction verticale, θ_H l’échelle de fluctuation selon le plan horizontal, τ_V la distance relative selon la direction verticale et τ_H la distance relative dans le plan horizontal calculé comme suit :

$$\tau_H = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2} \quad (15.32)$$

15.3.5 Longueurs de corrélation de la résistance de pointe rapportées dans la littérature

À défaut de mesures, on dispose à ce jour de nombreuses études sur la variabilité spatiale des paramètres géotechniques (surtout selon la direction verticale) sur différents types de sol. Grâce aux résultats reportés dans la littérature, il est parfois possible de fournir une

estimation de l'ordre de grandeur des distances de corrélation, à condition que les conditions géotechniques du site étudié soient vraiment similaires à celles du site de référence (Auvinet et al. (2000)). Le tableau (Tab.15.6) rassemble des ordres de grandeur de référence pour les longueurs de corrélation verticale et horizontale telles qu'elles apparaissent dans l'Eurocode 7 (Favre (2004)).

Des valeurs d'échelles de fluctuation de la résistance à la pénétration, selon la direction verticale et horizontale sont rapportées dans la littérature (Tab.15.5). Les échelles de variabilité sont métriques en direction verticale et décamétriques en direction horizontale. Toutefois, certaines études fournissent des valeurs supérieures à la centaine de mètres (Fenton (1999a), Liu and Chen (2010)). Il s'agit d'analyses de campagnes géotechniques réalisées sur de grandes surfaces, où l'échelle d'enquête est beaucoup plus grande à cause du grand écartement entre essais. En effet, la notion de longueur de corrélation est paradigmatique puisqu'elle est fonction du pas de mesure (DeGroot and Baecher (1993)) et des dimensions du domaine d'étude. (Vanmarcke (1983)) indique que pour que l'échelle de fluctuation estimée soit vraiment représentative, le pas d'échantillonnage doit être au moins inférieur à la moitié de sa valeur. (Magnan (1982)) statue qu'on ne doit pas accorder une signification à la longueur de corrélation que si elle reste inférieure au quart de la plus grande distance étudiée et au moins égale au plus petit pas d'observation des données. On discutera quand même, dans la section suivante, d'autres facteurs ayant une influence dans le calcul de la longueur de corrélation.

On peut remarquer que la plupart des grandeurs proposées dans la littérature ont été estimées à l'aide de l'essai de pénétration statique (CPT). Pour l'essai Panda, on ne dispose que des informations pour un ouvrage en terre, une digue argileuse (Deplagne and Bacconnet (1993)). La seule étude réalisée en terrains naturels est celle de (Athapaththu et al. (2007)). Les auteurs ont conduit une campagne d'essais Panda sur une maille régulière espacée 0.25, 1, 2.5 et 5 m pour caractériser la variabilité spatiale des sols dans une pente naturelle au Japon. Après une régularisation des profils pénétrométriques pour un pas égal à 5 cm, l'analyse structurale selon la direction horizontale a été réalisée à six profondeurs différentes, entre 0.1 et 2.5 mètres. Les échelles de fluctuation déduites à partir des modèles de variogramme sphériques varient entre 8 et 23 m.

15.3.6 Stratégie d'échantillonnage et représentativité de la longueur de corrélation

Même si une campagne de reconnaissance géotechnique est planifiée afin de répondre à un ou plusieurs objectifs, la distribution dans le plan des essais réalisés peut suivre trois schémas théoriques d'échantillonnage (Fig.15.3) d'après (Bacconnet (1991)) :

- i. échantillonnage systématique à maille régulière : très peu utilisé dans la pratique courante. Toutefois, il s'agit de la seule configuration permettant une inférence correcte de la loi spatiale ;
- ii. échantillonnage à maille aléatoire structurée : quand le sondage est supposé implanté au hasard dans une maille d'un réseau régulier ;
- iii. échantillonnage aléatoire pur : il pourrait correspondre à la plupart des reconnaissances géotechniques pour lesquelles l'implantation des sondages est en fait guidée par la géométrie et la nature du problème, même si toute implantation non systématique peut être interprétée comme le résultat d'un processus aléatoire.

TABLE 15.5 – Échelles de fluctuation de la résistance à la pénétration reportées dans la littérature (source : Huber (2013) et élaboration propre)

Paramètre	Nature	θ_V (m)	θ_H (m)	Référence
qc (CPT)	Argile	2	-	Alonso and Krizek (1975)
qc (CPT)	Sable	2.2	-	Alonso and Krizek (1975)
qc (CPT)	Argile	1.3	-	Vanmarcke (1977a)
N (SPT)	Sable	2.4	-	Vanmarcke (1977a)
qc (CPT)	Argile Offshore	-	60	Tang (1979)
qc (CPT)	Sable Offshore	5	-	Baecher et al. (1980)
qc (CPT)	Sable (dépôt alluvial)	0.1 - 2.6	-	Ximenez and Romana (1983)
qc (CPT)	Sable et graves	1.2	-	Soulie (1984)
qc (CPT)	Sable	3.2	-	Kulatilake and Ghosh (1988)
qc (CPT)	Sable offshore	-	28 - 76	Keaveny et al. (1989)
qc (CPT)	Sable	0.3	-	Wickremesinghe (1989)
qc (CPT)	Sable	2	20	Soulie et al. (1990)
qc (CPT)	Sable argileux	0.38	15	Bacconnet et al. (1987)
qc (CPT)	Argile plastique	-	7.5	Bacconnet et al. (1987)
qc (CPT)	Sable dense	0.8	35	Bacconnet et al. (1987)
qc (CPT)	Argile molle	2	40	Honjo and Kuroda (1991)
qd (Pda)	Digue argileux	0.1-0.5	10	Deplagne and Bacconnet (1993)
qc (CPT)	Argiles	0.1 - 2.8	15-17	Magnan (1993)
qc (CPT)	Sable fin	-	42-65	Castro and Zalzman (1994)
qc (CPT)	Argile sensible	4.0	-	Chiasson et al. (1995)
qc (CPT)	Argile sur-consolidé	0.15	2	Jaksa (1995)
qc (CPT)	Sable Offshore	1.9	24.2	Popescu (1995)
qc (CPT)	Argile limoneuse	2.0	10-24	Lacasse and de Lamballerie (1995)
N (SPT)	Alluvial	-	34	DeGroot (1996)
qc (CPT)	Sable avec graves	5	95	Kreuter (1996)
qc (CPT)	Argile sableuse	2.5 - 4.5	20 - 30	Kreuter (1996)
qc (CPT)	Argile	-	19.2	Lacasse and Nadim (1997)
qc (CPT)	Sable dense	-	75	Lacasse and Nadim (1997)
qc (CPT)	Argile	0.5-5	20-200	Fenton (1999a)
qc (CPT)	Argile	1.2	-	Tantalla et al. (2001)
qc (CPT)	Argile	0.29-0.40	-	Cafaro and Cherubini (2002)
qc (CPT)	Argile	0.6 - 6.4	2 - 25	Akkaya and Vanmarcke (2003)
qc (CPT)	Sable	0.2 - 10	2.5 - 30	Akkaya and Vanmarcke (2003)
qc (CPT)	Argile limoneuse	5.2	32	Assimaki et al. (2003)
Q_{t1} (CPT)	Argile, sable	0.13-1.11	-	Uzielli et al. (2005)
qd (Pda)	Arènes granitiques	-	8.25-22.5	Athapaththu et al. (2007)
qc (CPT)	Argile limoneuse	0.8 - 6.1	-	Haldar and Sivakumar Babu (2009)
qc (CPT)	Sable (offshore)	0.62	-	Van den Eijnden (2010)
qc (CPT)	Sable limoneux	0.1-2	>100	Liu and Chen (2010)
qc (CPT)	Sable	0.4 - 2.4	-	Bakhtiari (2011)
Q_{t1} (CPT)	Sable	0.1-0.45	1.2-2	Firouzianbandpey et al. (2014)
qc (CPT)	Sable (offshore)	0.4	1.8 -15.8	Lloret-Cabot et al. (2014)

TABLE 15.6 – Valeurs typiques de distances d’auto-corrélation d’après (source : Favre (2004))

Sol	L_H	L_V
Forte auto-corrélation	15 m	2 m
Auto-corrélation courante	10 m	1 m
Faible auto-corrélation	5 m	0.5 m

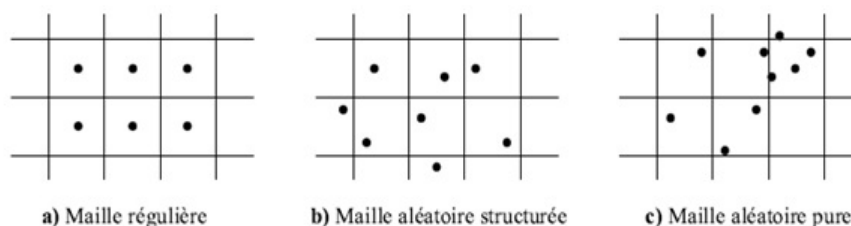


FIGURE 15.3 – Les trois configurations d’échantillonnage (source : Bacconnet (1991))

Le calcul expérimental du variogramme ou de l’autocorrélation doit également respecter certaines règles d’échantillonnage. Il est souhaitable d’avoir un pas de mesure régulier, sinon il faut grouper les données par intervalles de classes de distance comme le suggère (Christian and Baecher (2003)). De même, plusieurs règles empiriques basées sur le nombre d’observations n ont été fournies dans la littérature :

- (Box et al. (2015)), dans le cadre de l’analyse des séries temporelles, suggèrent un nombre minimale d’effectif égal à 50 ;
- (Magnan (1982)) conseille un nombre minimal de couples de valeurs analysées égal à 25 ;
- (Journel and Huijbregts (1978)) préconisent un nombre de couples supérieur à 30 pour estimer les points du variogramme expérimental en pratique minière ;
- (Bacconnet (1991)) statue que dès que le nombre de couples est inférieur à 10, il est déconseillé de considérer ce point ;
- pour une caractérisation de la variabilité spatiale en trois dimensions (Lumb (1975)) suggère un nombre de mesures au moins de l’ordre de 10^4 , quantité prohibitive même pour un projet de recherche spécifique. En conséquence, d’après cet auteur, le mieux que l’on puisse réaliser dans la pratique est d’étudier la variabilité dans une direction de l’espace, verticale ou horizontale, grâce à un nombre de mesures de l’ordre de 20 à 100 ;
- (Baroth et al. (2011)) recommandent des distances d’échantillonnage maximales pour une estimation correcte de la distance de corrélation pour un sol :
 - de l’ordre de $\Delta V = \theta_V/2$ dans la direction verticale ;
 - entre $\theta_H/8 < \Delta H < \theta_H/4$ dans la direction horizontale.
- suite aux simulations numériques réalisées par (Nie et al. (2015)), les auteurs suggèrent que pour une estimation précise, il est nécessaire d’avoir un pas d’échantillonnage au moins 10 fois plus faible et une taille du domaine d’étude au moins 100 fois plus grande que l’échelle de fluctuation.

Cependant, il est rare que, dans les études géotechniques, les essais soient réalisés à des

profondeurs régulièrement espacées et selon des dispositions régulières en plan. La réalité sur la plupart des sites, est que l'on dispose, à des profondeurs quelconques, d'ensembles de valeurs plus ou moins nombreuses de caractéristiques physiques et mécaniques des sols, provenant rarement des mêmes sondages et de profondeurs identiques. Il faut alors faire preuve d'une certaine tolérance si l'on veut disposer de données (Magnan (1993)).

Enfin, il convient de remarquer que les longueurs de corrélation ne sont pas des caractéristiques intrinsèques du sol. En effet, on a vu qu'il existe une grande panoplie de facteurs qui peuvent jouer un rôle important vis-à-vis de leur calcul. En ce qui concerne la fonction d'autocorrélation la soustraction de la dérive réduit fortement la distance d'autocorrélation et la ramène à des valeurs courantes dans les sols, cependant la distance de corrélation ainsi estimée est fonction de la procédure numérique retenue. On cite comme exemple (Cafaro and Cherubini (2002)) qui étudie les erreurs lors de l'estimation de l'échelle verticale pour des formations argileuses en fonction du modèle et de la technique de dérive choisie, mais aussi en fonction de l'intervalle des mesures. La fourchette des erreurs relatives varie entre 1 – 38%. Sinon, les modèles de transformation semblent avoir une faible influence vis-à-vis de l'estimation des longueurs de corrélation. On cite comme exemple (Van den Eijnden (2010)), qui estime l'angle de dilatance pour un remblai constitué de sable à partir d'un ensemble de 140 essais CPT grâce à un modèle de transformation basé sur plus d'une cinquantaine d'essais triaxiaux. L'échelle de fluctuation calculée pour la résistance au cône q_c (0.620 m) est presque identique à celle estimée pour l'angle de dilatance (0.627 m).

15.4 Homogénéité d'un site

Intuitivement le degré d'hétérogénéité du terrain apparaît rapidement à l'ingénieur grâce à l'examen des informations fournies par la campagne de reconnaissance. Cette incertitude naturelle peut avoir divers degrés et donc entraîner différents moyens d'investigation et techniques d'analyse (Favre (2000)) :

- i. terrains très erratiques à fortes hétérogénéités avec des fractures. Pour ces cas, Favre signale que la modélisation est essentiellement déterministe et dichotomique. La reconnaissance géophysique couplée avec des sondages sur site s'avère vitale afin de détecter les accidents et la théorie des ensembles flous peut s'avérer un outil intéressant d'aide à la décision ;
- ii. terrains désordonnés présentant une stratification relativement régulière avec, pour chaque couche, des propriétés régulières aussi. Cette situation représente un cadre d'application idéal pour l'application des méthodes statistiques et probabilistes ;
- iii. enfin, des terrains anthropiques (des barrages en terre, remblais, etc.). Dans ce cas le contrôle de fabrication et les analyses statistiques sont couramment employés.

(Paikowsky et al. (2004)) suggèrent que la variabilité à l'échelle d'un site peut être déterminée par l'expertise de l'ingénieur ou se baser sur la valeur du coefficient de variation CV des paramètres de sol mesurés au sein des différentes unités géotechniques. La formation à étudier peut être alors classée comme suit :

- homogène ($CV < 25\%$) ;
- niveau d'hétérogénéité moyen ($25 \leq CV < 40\%$) ;
- très hétérogène ($CV \geq 40\%$).

(Harr (1987)) proposait ainsi une classification similaire à la précédente en fonction de la

valeur du CV . Par l'intermédiaire des tests de comparaison de variances et du test d'égalité de moyennes, la statistique descriptive peut aussi conférer une certaine objectivité à l'étude de l'homogénéité d'un site (Cassan (2000)). Il donne des exemples d'application dans le cas de projets ferroviaires, sur la LGV Méditerranée ou le tramway de Strasbourg. Cependant la notion de site n'est pas forcément attachée à la dimension totale du terrain à ausculter. Un site peut être juste une petite zone de la surface totale où la structure sera construite, c'est l'analyste qui fixe l'échelle.

Une méthode plus récente est celle proposée par (Salgado et al. (2015)). Elle consiste à calculer l'indice de variabilité du terrain à étudier à partir de l'analyse des profils pénétrométriques issus des essais CPT. Cette approche est articulée en quatre étapes :

- i) détermination de la stratigraphie suite à l'analyse de chaque sondage grâce aux abaques de classification de Tumay (Tumay (1985)) ou Robertson (Robertson (1990));
- ii) calcul de l'indice de variabilité verticale (VVI);
- iii) calcul de l'indice de variabilité horizontale (HVI);
- iv) calcul de l'indice de variabilité du site, *site variability rating* (SVR).

Ces indices sont normalisés dans un intervalle $[0, 1]$ et trois niveaux d'hétérogénéité sont considérés en fonction de sa valeur :

- i) site d'hétérogénéité faible (0 – 33%);
- ii) site d'hétérogénéité moyen (33 – 66%);
- iii) site très hétérogène (66 – 100%).

On remarque qu'une profondeur d'étude doit être choisie et tous les sondages doivent être menés au moins jusqu'à cette profondeur. Le calcul de l'indice de variabilité verticale nécessite la connaissance de la nature du sol qui est déduite au moyen des abaques de classification spécifiques pour le CPT utilisé dans la méthode. Toutefois, on décrit la procédure pour le calcul de l'indice de variabilité horizontale. Il s'agit d'une grandeur plus générale pouvant donc être calculée pour les profils de résistance dynamique q_d mesurés à l'aide de l'essai Panda.

Pour chaque couple de sondages, deux paramètres sont estimés :

- 1) la différence en valeur absolue ($|\Delta q_{c,avg}|$) entre les modèles de tendance selon la profondeur (polynômes de premier ou deuxième degré) calés sur les profils pénétrométriques q_c . Cet écart est calculé comme la moyenne arithmétique des différences estimées à chaque mètre de profondeur, et ensuite normalisé par une constante ($|\Delta q_{c,avg}|_{max}$) en fonction de la profondeur du sondage. La valeur de la constante de normalisation est déduite à l'aide d'une fonction empirique (Fig.15.4). Elle a été calculée au moyen de la différence entre deux profils idéalisés, un constitué par une argile molle et l'autre par un sable de densité relative très dense. On obtient ainsi un coefficient normalisé ρ_0 qui mesure l'écart entre les tendances :

$$\rho_0 (|\Delta q_{c,avg}|) = \frac{|\Delta q_{c,avg}|}{|\Delta q_{c,avg}|_{max}} \quad (15.33)$$

- 2) le coefficient de corrélation (ρ) pour q_c entre les deux sondages. Afin qu'il soit normalisé sur un intervalle $[0, 1]$ la transformation suivante est appliquée :

$$f_1 (\rho) = \frac{\rho + 1}{2} \quad (15.34)$$

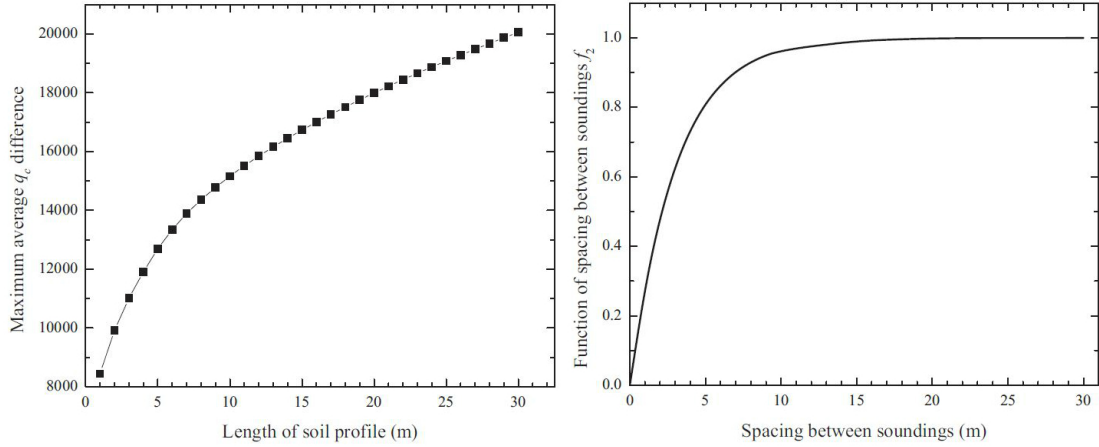


FIGURE 15.4 – À gauche, différence maximale de la moyenne de q_c en fonction de la longueur du profil analysé. À droite, facteur de correction en fonction de la distance relative entre sondages (source : [Salgado et al. \(2015\)](#))

Ensuite, on calcule une somme pondérée de ces deux paramètres. Le coefficient de corrélation (ρ) est affecté par un coefficient égal à 20% et la différence entre dérivés $|\Delta q_{c,avg}|$ reçoit un poids plus important, 80%. Afin de tenir compte de la distance entre les sondages, cette somme est multipliée par une fonction exponentielle $f_2(s)$ (Fig.15.4) définie ci-après :

$$f_2(s) = 1 - e^{-\frac{s}{3}} \quad (15.35)$$

où s est l'écart entre les deux sondages analysés. On voit que cette équation pénalisera les indices calculés pour des sondages écartés de moins de 9 m environ². Autrement dit, on pourrait interpréter que cette méthode suppose une sorte de distance de corrélation θ_H de l'ordre de 6 m (2a). L'expression complète de la fonction de variabilité horizontale est la suivante :

$$f(|\Delta q_{c,avg}|, \rho, s) = [0.8(1 - \rho_0(|\Delta q_{c,avg}|)) + 0.2 \cdot f_1(\rho)] \cdot f_2(s) \quad (15.36)$$

Enfin l'indice de variabilité horizontale HVI est calculé à partir de la valeur moyenne de la fonction de variabilité horizontale $f(|\Delta q_{c,avg}|, \rho, s)$ calculée pour tous les couples possibles de sondages. L'indice HVI varie entre 0 et 1 et son expression mathématique est :

$$HVI = 1 - \frac{\sum_{i=1}^n f(|\Delta q_{c,avg}|_{ij}, \rho_{i-1,i}, s_{i-1,i})}{n} \quad (15.37)$$

On ajoute que l'auteur recommande un minimum de 5 sondages sur une surface de $1000m^2$ afin de caractériser d'une manière raisonnable la variabilité que présente le site d'étude.

2. Le modèle exponentiel atteint le 95% du palier à la valeur de $3a$, équivalente à la portée pratique.

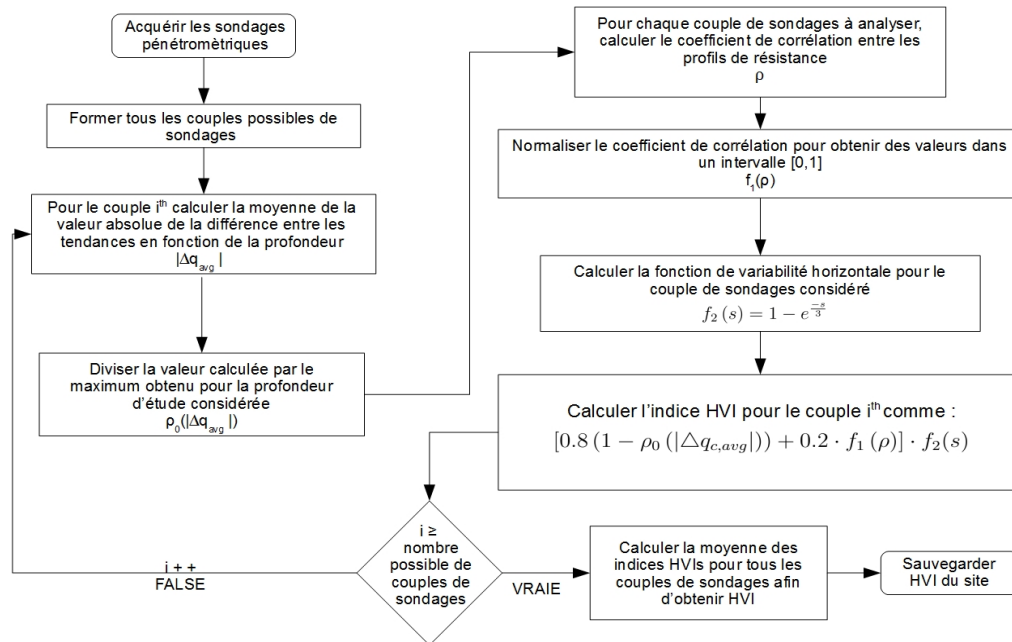


FIGURE 15.5 – Organigramme de calcul de l'indice de variabilité horizontale du site HVI (source : Salgado et al. (2015))

15.5 Description probabiliste de l'incertitude : la statistique bayésienne

15.5.1 Introduction

L'approche bayésienne permet d'introduire sous forme de contrainte le jugement subjectif d'un expert ou de l'information a priori afin d'obtenir une meilleure estimation des paramètres inconnus. Un autre atout intéressant de la procédure bayésienne est la possibilité de combiner des informations fournies par plusieurs sources (mesures in situ, résultats des essais au laboratoire, expertise de l'ingénieur, données issues de la littérature...). Grâce à ces qualités, les méthodes bayésiennes constituent un sujet de recherche d'intérêt et d'actualité. En effet, la mise en relation entre la méthode observationnelle et la mise à jour bayésienne (*bayesian updating*) est une des 10 questions essentielles qui demeurent toujours ouvertes d'après (Christian and Baecher (2011)). Les auteurs signalent que la statistique bayésienne répond à la même logique décrite dans un des derniers travaux de Terzaghi (Terzaghi (1961) cité par (Christian and Baecher (2011))). Une brève description de l'analyse bayésienne en mettant l'accent sur son application au domaine géotechnique est réalisée par la suite. On renvoie le lecteur aux livres de (Sivia and Skilling (2006), et Gelman et al. (2014)), ouvrages de référence sur la théorie bayésienne et son application pour l'analyse de données. De plus, le travail de (Straub and Papaioannou (2014)) aborde dans une perspective très pédagogique l'implémentation pratique des procédures bayésiennes à plusieurs problèmes géotechniques. Un document plus récent (Zhang et al. (2017)) présente une discussion assez complète sur

la théorie et les méthodes bayésiennes, une synthèse de plusieurs travaux en géotechnique et une série de suggestions sur l'application des méthodes bayésiennes à l'Eurocode 7.

15.5.2 Généralités et techniques numériques

Le noyau du raisonnement bayésien est le théorème de Bayes qui permet de mettre à jour la probabilité d'un événement ou d'une hypothèse, compte tenu des observations réalisées :

$$P(\text{hypothèses}|\text{données}) \propto P(\text{hypothèses}) P(\text{données}|\text{hypothèses}) \quad (15.38)$$

Les incertitudes dans la conception géotechnique peuvent être modélisées par le biais des variables aléatoires continues représentées par des fonctions de densité de probabilité. On notera par Θ le paramètre ou vecteur des paramètres qu'on cherche à estimer à partir d'un échantillon de données y . La transformation de l'opinion a priori des données (la fonction de vraisemblance) en une opinion a posteriori peut se faire à l'aide du théorème de Bayes comme suit :

$$f''_{\Theta}(\Theta|y) = aL(y|\Theta) f'_{\Theta}(\Theta) \quad (15.39)$$

avec :

$f'_{\Theta}(\Theta)$: la probabilité *a priori* qui résume l'information disponible sur le paramètre d'intérêt avant le recueil des données ;

$L(y|\Theta)$: vraisemblance, c'est l'information issue de l'expérience. Elle est égale à la probabilité d'observer les données pour des paramètres du modèle retenu ;

$f''_{\Theta}(\Theta|y)$: la probabilité *a posteriori*, c'est à dire la loi conditionnelle suite à l'intégration de ces deux informations ;

a : constante de normalisation qui garantit que l'intégrale de la fonction de densité a posteriori soit égale à 1.

L'expression mathématique de la constante a , est donnée ci-dessous :

$$a = \frac{1}{\int_{\Theta} L(y|\Theta) f'_{\Theta}(\Theta) d\Theta} \quad (15.40)$$

La statistique Bayésienne obéit au principe de vraisemblance. Le choix de la loi a priori constitue une étape fondamentale dans l'analyse bayésienne. On peut distinguer entre deux types de lois a priori : loi informative ou non informative (même si une loi n'est jamais totalement non informative) aussi connue comme vague, neutre ou peu informative. L'absence d'information est prise en compte au moyen d'une loi peu informative en minimisant donc son rôle sur l'inférence. Un choix courant est celle de la loi uniforme, donnant la même vraisemblance à chaque valeur du paramètre à estimer. Sinon, la loi de Jeffreys basée sur la maximisation d'un critère de l'information apportée par les données ([Jeffreys \(1961\)](#)) est une alternative valable pour le cas unidimensionnel.

La distribution a posteriori du paramètre à estimer dépend alors de la relation entre les observations et de la loi a priori retenue :

- la loi a posteriori sera proche de la vraisemblance si la loi a priori est peu informative ou vague ou dès que le nombre de données est important ;
- la loi a priori aura plus d'influence si elle est informative ou si le nombre de données est faible.

Une autre difficulté de l'approche bayésienne est le calcul de la loi a posteriori définie dans l'équation (eq.15.39) qui est facilitée lorsque la loi a priori et la loi a posteriori ont la même forme. Dans ce cas, on parle de loi a priori conjuguée, et l'estimation de la loi a posteriori peut être abordée grâce aux expressions analytiques. Une liste complète des lois a priori conjuguées se trouve dans (Gelman et al. (2014)). En littérature géotechnique française, (Favre (2004)) récapitule la famille des lois conjuguées naturelles. Lorsque le nombre de variables considérées dans le problème est faible on peut appliquer les méthodes de *quadrature* de Gauss-Legendre souvent utilisées en analyse numérique et la méthode des éléments finis (ex. Christian and Baecher (1999)). Les approximations asymptotiques de Laplace sont un autre choix possible lorsque la dimension du problème est encore réduite. À titre d'exemple, dans la démarche bayésienne pour estimer l'angle de frottement à partir des essais CPT proposée par (Wang et al. (2010)), la loi a posteriori est calculée à l'aide de l'approximation asymptotique de Laplace. Cependant, l'estimation bayésienne demande de calculer des intégrales complexes pour obtenir les distributions a posteriori, et le calcul de la constante de normalisation pose des problèmes. D'ailleurs, lorsque le nombre de paramètres à estimer est élevé et que la loi a posteriori à estimer présente un caractère multimodal, il faut avoir recours aux méthodes de simulation de Monte Carlo.

En fait, la forte progression de l'analyse bayésienne est essentiellement due à l'essor des techniques de Monte Carlo. Ces techniques peuvent être divisées en deux groupes : les méthodes de Monte-Carlo par chaînes de Markov (MCMC pour *Monte Carlo Markov Chain* en anglais) et les méthodes de Monte Carlo séquentielles (SMC pour *les méthodes de Monte Carlo séquentielles*). Les méthodes MCMC sont des techniques d'échantillonnage de variables aléatoires en construisant une chaîne de Markov. Selon la théorie des chaînes de Markov, la mesure empirique associée à la chaîne converge vers la densité cible. Les méthodes SMC sont une alternative aux techniques MCMC. Elles sont basées sur la simulation d'un grand nombre d'échantillons selon une série de distributions de probabilités et procèdent par échantillonnage/ré-échantillonnage.

L'idée originelle des méthodes MCMC est publiée dans les années 50 par Metropolis (Metropolis et al. (1953)). Ensuite elle est étendue par Hastings, (Hastings (1970)) donnant naissance à l'algorithme de Metropolis-Hastings (MH). Un cas particulier de l'algorithme (MH) est l'échantillonneur de Gibbs (Geman and Geman (1984)). Une contrainte de l'algorithme (MH) est qu'il ne permet pas d'approcher des lois a posteriori multimodales. Cette contrainte est résolue par la version modifiée de l'algorithme MH proposée par (Beck and Au (2002)), mais son application n'est envisageable que pour des problèmes avec un nombre de paramètres relativement faible. Pour une discussion détaillée sur la performance de ces algorithmes d'échantillonnage on pourra consulter (Au and Beck (2003)), qui expliquent leurs caractéristiques, avantages et limitations. Une version plus récente dite TMCMC (*Transitional Markov Chain Monte Carlo Method*) (Ching and Chen (2007)) présentant plusieurs avantages par rapport à la version initiale est rapportée par (Ching and Wang (2016)). Il montre l'application de l'algorithme TMCMC pour la caractérisation de l'incertitude statistique de la variabilité spatiale des paramètres géotechniques estimés à partir des données issues de l'essai CPT. D'autres approches numériques avancées sont utilisées dans la littérature géotechnique comme l'approche (SI) (*System Identification*) spécialement développée pour la mise à jour des paramètres d'un modèle mécanique suite aux données observées, ou la technique BUS (*Bayesian updating with structural reliability methods*) fondée sur des principes de fiabilité structurale ainsi que d'autres méthodes (Straub and Papaioannou (2014))

et Zhang et al. (2017)).

15.5.3 Bilan sur l'application de la statistique bayésienne en géotechnique

(Ching et al. (2015b)) fournissent une comparaison entre les méthodes classiques issues de la statistique fréquentielle et les techniques bayésiennes vis-à-vis de l'estimation des paramètres du champ aléatoire pour la caractérisation du profil géotechnique. La conclusion principale de cette étude est que le raisonnement bayésien à l'aide des méthodes de Monte Carlo par chaînes de Markov s'avère la méthode la plus performante, surtout dès que le nombre de données disponible est réduit.

Parmi les différentes applications de la théorie bayésienne on distingue principalement (Ching and Chen (2007)) :

- i. *Bayesian Model Averaging* (BMA) : c'est une solution au problème d'inférence en présence de plusieurs modèles en compétition. On dispose au départ de N modèles possibles (emboîtés ou non) $\mathcal{M} = \{\mathcal{M}_1, \dots, \mathcal{M}_N\}$. Le but est alors de sélectionner les modèles les plus pertinents et ensuite pondérer les modèles retenus ;
- ii. *Bayesian Model Class Selection* (BMC) : l'approche bayésienne peut être aussi appliquée pour choisir le meilleur modèle qui est recherché en maximisant la probabilité a posteriori $P(\mathcal{M}_i|\text{données})$ du modèle sachant les données. Il est utilisé dans les méthodes stratigraphiques citées dans la partie précédente (Cao (2012), Wang et al. (2013)), afin de choisir le "meilleur modèle" en fonction du nombre d'unités géotechniques. Dans (Cao and Wang (2014)), une procédure BMC est aussi retenue afin de choisir le meilleur modèle théorique de fonction d'autocorrélation pour représenter la variabilité spatiale des paramètres géotechniques grâce au formalisme des champs aléatoires.
- iii. *Bayesian Model Updating* (BMU) : il s'agit de faire évoluer l'estimation de l'incertitude dès qu'une nouvelle information est disponible. Un aperçu de la mise à jour bayésienne dans le contexte de la géotechnique est montré dans la figure (Fig.15.6). Des procédures bayésiennes ont été souvent employées en géotechnique pour étudier l'incertitude du modèle suite à la performance observée ainsi que les nouvelles données recueillies. (Zhang et al. (2009)) ont proposé un procédure BMU pour étudier l'incertitude des 4 méthodes à l'équilibre limite pour évaluer la stabilité des talus, en prenant en compte l'incertitude des paramètres d'entrée et de sortie des modèles. Une démarche similaire est suivie par (Zhang et al. (2012)) pour l'étude de l'incertitude du modèle vis-à-vis de l'estimation de la capacité portante d'une fondation superficielle et la détermination du coefficient de poussée des terres au repos. Le lien entre la mise à jour bayésienne et la méthode observationnelle³ postulé par Terzaghi est mise en oeuvre dans le travail de (Kelly and Huang (2015)) pour un problème de consolidation unidimensionnelle durant la construction d'un remblai sur une formation des sols mous. L'étude montre que la confiance sur le tassement calculé augmente rapidement à partir des 3 ou 4 mesures.

3. Schématiquement, la méthode observationnelle consiste à ausculter un ouvrage géotechnique pendant sa construction pour renforcer le dimensionnement si les mesures en montrent la nécessité. L'E7 la conseille (NF P 94-251-1 (2005)) lorsque il s'avère compliqué de prévoir le comportement géotechnique d'un ouvrage.

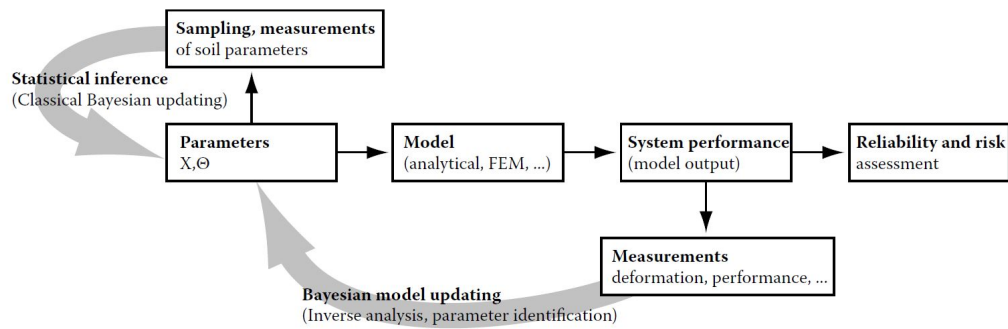


FIGURE 15.6 – Aperçu de la mise à jour bayésienne en géotechnique (source : (Straub and Papaioannou (2014)))

Les approches BMU sont aussi souvent proposées vis-à-vis d'une caractérisation géotechnique probabiliste du site. (Wang et al. (2016b)) proposent une démarche bayésienne permettant d'évaluer la propagation des incertitudes lors des différentes phases du projet. D'ailleurs l'approche envisage la fusion des données mesurées sur le terrain avec les mesures issues des essais de laboratoire. Les auteurs suggèrent de réaliser l'inférence statistique des paramètres géotechniques au moyen des techniques MCMC car ils s'avèrent plus robustes dans le cas où le nombre de données est faible. Dans la même optique, (Wang et al. (2016a)) ont développé une application sur Excel basée sur des simulations MCMC (algorithme MH) (*Bayesian Equivalent Sample Toolkit* (BEST)) permettant l'inférence statistique des propriétés des sols et roches à partir d'un nombre d'observations limité. (Papaioannou and Straub (2016)) appliquent des analyses bayésiennes pour coupler les informations mesurées in situ avec les données de laboratoire. Le calcul de la capacité portante pour une fondation superficielle dans un sol limoneux montre que la modélisation de l'angle de frottement comme une variable aléatoire peut entraîner une sous-estimation de l'indice de fiabilité par rapport à la prise en compte de la variabilité spatiale grâce à une modélisation via les champs aléatoires.

De plus, on a déjà évoqué la construction des lois a priori comme un des points délicats de la démarche bayésienne. Le problème de comment traduire les informations disponibles (incluant l'expertise de l'ingénieur) en loi a priori et traduire la qualité de cette information est abordée par (Cao et al. (2016)) vis-à-vis de l'estimation des paramètres géotechniques. Une approche inspirée d'un questionnaire de loterie a été proposée par l'auteur afin d'apporter une aide au géotechnicien pour la construction des lois a priori. (Zhang et al. (2017)) proposent un bilan plus détaillé sur d'autres travaux d'applications de la statistique bayésienne en géotechnique.

Finalement les simulations MCMC peuvent être utilisées pour simuler l'incertitude lithologique. (Li et al. (2016c)) appliquent la modélisation des différentes formations rencontrées dans les forages (sables, limons et argiles) comme variables aléatoires discrètes. Une démarche similaire est utilisée par (Li et al. (2016a)) afin d'étudier l'influence de différentes configurations lithologiques et du nombre de forages pour un problème de stabilité de talus.

15.6 Bilan d'ouvrages sur la modélisation spatiale des paramètres géotechniques

Les outils issus de la géostatistique ont été souvent employés pour l'établissement de modèles géologiques et géotechniques. On peut distinguer principalement deux échelles (Dubost (2009)) :

- L'échelle régionale ou urbaine

L'objectif est généralement de disposer d'une connaissance générale sur un territoire pour en gérer l'aménagement. Dans le cadre du projet RIVIERA en France, (Dominique et al. (2010)) ont construit un modèle sur le centre de Bordeaux par krigeage en utilisant principalement les données litho-stratigraphiques des sondages géotechniques. Dans le cadre du même projet, (Marache et al. (2009)) utilisent des méthodes de simulations non-gaussiennes issues de la géostatistique non linéaire (simulation séquentielle d'indicatrices) pour modéliser la distribution spatiale des classes de module pressiométrique (0, 4, 12 et 36 MPa) pour la ville de Pessac. Sinon, (Folle et al. (2008)) appliquent le krigeage universel par bloc pour cartographier en 3D la résistance du sol dans un milieu urbain au moyen des mesures issues de l'essai SPT.

D'autres travaux ont utilisé des outils géostatistiques afin de fournir des cartes d'estimation des paramètres de résistances du sol pour de grandes surfaces où le nombre de données disponibles est importante. (Samui and Sitharam (2007)) utilisent le krigeage ordinaire et le krigeage disjonctif pour modéliser la variabilité spatiale des données N_{SPT} en 3 dimensions. Pour la base de données utilisée, laquelle comporte plus de 700 sondages sur une surface de 220 km^2 , le krigeage disjonctif a fourni les meilleures estimations (variance d'estimation plus faible et valeurs estimées plus proches des valeurs réelles). Le krigeage ordinaire a été aussi employé par (Mendes and Lorandi (2008)) utilisant un total de 204 sondages SPT afin de fournir des cartes d'estimation en 2D pour une surface de terrain de 60 km^2 .

Une contribution importante pour la caractérisation géotechnique des dépôts argileux d'origine lacustre de la Vallée de Mexico au moyen des outils géostatistiques a été réalisée grâce à plusieurs travaux menés par l'Université Nationale de Mexico (UNAM) (ex. Auvinet et al. (1995), Medina (2001), Auvinet (2002), Jiménez (2007), Auvinet-Guichard (2011), Juárez-Camarena et al. (2014)). Dans cet esprit, un système d'information géographique (SIG) a été développé (Auvinet et al. (1995)) ainsi que des logiciels qui présentent l'ensemble des techniques géostatistiques conçus pour l'application aux données géotechniques : SAAG (Système d'aide à l'analyse géostatistique) (Auvinet (2002)) et MESCAL (Juárez-Camarena et al. (2016)).

- L'échelle locale ou de projet

La superficie étudiée est plus restreinte et le modèle est généralement destiné à une application spécifique. Dans, (Soulie et al. (1990)) l'application du krigeage ordinaire en 2D pour estimer la résistance au cisaillement non drainé a permis de mettre en évidence une zone de faiblesse dans le dépôt d'argile marine étudié. (Chiasson et al. (1995)) ont utilisé aussi des techniques géostatistiques pour la caractérisation de la variabilité spatiale d'une formation argileuse légèrement surconsolidée sur la base des mesures recueillies au piezocône et au scissomètre, avec une covariance expérimentale très similaire. Dans le cadre de la conception d'une fondation superficielle (Lacasse

et al. (1996)) fournissent une cartographie 3D de la résistance de pointe au cône q_c comme une superposition des cartes en 2D pour chaque mètre de profondeur à l'aide du krigeage. Le krigeage peut être utilisé aussi pour mesurer la qualité de l'échantillonnage (Bacconnet (1991)), mettant en évidence que le fait de négliger la variabilité spatiale conduit à surestimer fortement les variances d'échantillonnages. (Młynarek et al. (2005)) présentent une étude comparative entre les estimations en 2D à différentes profondeurs fournies par le krigeage ordinaire et une technique d'interpolation linéaire juste à partir de 10 sondages CPTU. Ils n'ont pas constaté de différences significatives entre les deux techniques. Une application plus récente est donnée par (Firouzianbandpey et al. (2015)), qui utilisent le krigeage ordinaire pour estimer la résistance de pointe normalisée mesurée à l'aide de l'essai CPT. Compte tenu du faible nombre d'essais, deux valeurs de la portée intégrale ont été prises en compte à 5 et 10 m pour un modèle de covariance exponentiel.

Le krigeage ordinaire en 2D a été aussi appliqué aux mesures de résistance dynamique q_d mesurées à l'aide du Panda en (Athapaththu et al. (2007)). Des modèles de variogramme sphériques ont été utilisés, cependant une anisotropie zonale en fonction de la profondeur a été prise en compte.

Le krigeage peut être abordé dans un contexte bayésien. Des exemples dans la littérature sont issues des travaux de Omre (Omre (1987), Omre and Halvorsen (1989)). On peut regarder le krigeage bayésien, comme une variante du krigeage capable de traiter l'incertitude qui entoure la fonction spatiale (covariance ou variogramme) (Handcock and Stein (1993)). Pour des renseignements sur l'implémentation pratique du krigeage bayésien, on se reportera à l'article de (Pilz and Spöck (2008)). Vis-à-vis de l'application de ces techniques dans le domaine géotechnique, (Uzielli et al. (2010)), appliquent le krigeage bayésien (considérant la tendance comme incertaine) aux mesures de vitesse des ondes de cisaillement issues de 15 sondages SCPT. (Wang et al. (2016c)) quant à lui aboutit à une carte d'isovaleurs du module de Young E dans le plan horizontal grâce au couplage de mesures directes fournies par les essais à l'oedomètre, incorporées sous la forme de loi a priori et la transformation mécanique des mesures de résistance recueillies par le CPT.

Cependant, on remarque que l'utilisation de la géostatistique pour modéliser la variabilité des sols dans les calculs géotechniques est moins répandue que celle de la théorie des champs. Le travail pionnier de (Vanmarcke (1977b)) a consisté à mettre en œuvre une modélisation par la théorie des champs de la cohésion non drainée vis-à-vis d'un problème de stabilité de talus. Suite à Vanmarcke, plusieurs auteurs ont ainsi appliqué la théorie de champs pour l'analyse de plusieurs problèmes géotechniques comme par exemple :

- stabilité des ouvrages en terre : en 2D (Griffiths and Fenton (2000)), (Hicks and Samy (2002)), (Salgado and Kim (2013)) et en 3D (Auvinet and Gonzalez (2000)), (Hicks and Spencer (2010)).
- fondations superficielles : (Griffiths and Fenton (2001)), (Fenton and Griffiths (2002)), (Fenton and Griffiths (2003)) et (Popescu et al. (2005a)).
- liquéfaction sismique : en 3D (Popescu et al. (2005b)).

En outre, la simulation par champs aléatoires a été aussi proposée (Jaksa et al. (2003)) dans le cadre de la caractérisation du site afin d'optimiser la campagne de reconnaissance géotechnique et maximiser le ratio bénéfice/coût. Cette démarche a été mise en œuvre par (Jaksa et al. (2005)) en simulant des modèles de terrain en 3D pour étudier les effets de

la variabilité spatiale des paramètres géotechniques et des divers plans d'implantation de sondages dans la conception d'une fondation profonde.

On remarque aussi que les approches utilisées dans les travaux cités auparavant sont basées sur des simulations non conditionnelles. Cependant, ces méthodes ne s'avèrent pas optimales dès qu'on dispose de mesures sur le terrain. En effet, plusieurs réalisations non conditionnelles peuvent ne pas honorer les données mesurées, entraînant un plus large éventail de réponses vis-à-vis du calcul géotechnique à mener. Les simulations conditionnelles peuvent alors combler cet écart. Malgré leur grande popularité dans le domaine minier (ex. [Chiles and Delfiner \(1999\)](#)), leur utilisation en géotechnique s'avère encore limitée.

Un cas d'application notable est celui proposé dans ([Gaouar \(1997\)](#)) pour une approche fiabiliste afin d'évaluer la stabilité des barrages en terre en 2D au moyen d'une simulation par la méthode de bandes tournantes conditionnées par krigeage simple. Le modèle est alimenté à partir des données pénétrométriques recueillies par l'essai Panda qui ont montré une forte corrélation avec la résistance à la compression simple. D'autres études plus récentes sont celles issues du travail de thèse de ([Van den Eijnden \(2010\)](#)). Les champs sont générés au moyen de la méthode de simulation par subdivision en moyenne locale et conditionnés à l'aide du krigeage ordinaire en 2D, utilisant un modèle de corrélation théorique exponentiel. Les simulations conditionnelles sont utilisées pour modéliser la variabilité naturelle de l'angle de dilatance d'un île artificielle d'une plateforme offshore grâce aux sondages CPT. Cette méthode a été aussi employée dans les travaux ([Lloret Cabot et al. \(2012\)](#), [Lloret-Cabot et al. \(2013\)](#) et [Lloret-Cabot et al. \(2014\)](#)) afin d'étudier la variabilité spatiale d'un île artificielle constituée par un sable dans une plateforme offshore. L'extension de l'algorithme au domaine 3D a été introduit par ([Li et al. \(2016b\)](#)) et appliqué à deux problèmes théoriques de stabilité de talus. La conclusion principale est que l'utilisation des simulations conditionnelles permettrait un dimensionnement plus rentable par rapport au cas non conditionnel.

15.7 Conclusions

Il devient évident d'après la revue bibliographique réalisée que la théorie des champs aléatoires et la géostatistique fournissent un cadre théorique valable pour quantifier et estimer la variabilité spatiale des propriétés géotechniques. En résumé ces deux approches théoriques sont importantes puisqu'elles permettent :

- générer des modèles numériques du terrain grâce à l'inférence statistique des données mesurées ou à partir des valeurs rapportées dans la littérature, et optimiser les stratégies d'échantillonnage ;
- prendre en compte la structure spatiale des propriétés du sol dans les modèles de calcul géotechniques et les calculs fiabilistes vis-à-vis des états limites.

La géostatistique permet de traiter des problèmes non stationnaires, et les techniques de krigeage permettent d'estimer les propriétés du sol aux points non échantillonnés. Le krigeage est un interpolateur exact puisque tout krigeage d'une valeur mesurée redonne cette valeur avec une variance nulle. De plus, il permet de mesurer la précision de la valeur estimée par l'intermédiaire d'une variance d'estimation ainsi que le conditionnement de la matrice de covariance du champ étudié. Toutefois, l'emploi de la géostatistique en géotechnique reste peu répandu (ex. [Jaksa \(1995\)](#), [Uzielli et al. \(2006\)](#)).

Par ailleurs, il a été démontré dans la littérature que le fait de négliger la variabilité

spatiale entraîne la plupart du temps une surestimation de la variance des propriétés du sol, ce qui déboucherait sur une estimation prudente de la probabilité de ruine dans le cadre d'une analyse de fiabilité. Cependant, dans la pratique courante de la géotechnique, la caractérisation de cette variabilité s'avère compliquée à mettre en œuvre faute de données en quantité suffisante. C'est pourquoi, dans ce travail, on a choisi de s'appuyer sur un essai tel que le Panda afin de pouvoir réaliser une analyse structurale des données de résistance du sol.

Enfin, on remarque que la variabilité spatiale des caractéristiques mécaniques ou physiques du sol est structurée à différentes échelles avec une anisotropie entre la direction horizontale et verticale due essentiellement au mode de dépôt. Les sondages au pénétromètre fournissent un nombre de mesures important qui rend possible l'analyse de la variabilité le long du sondage. Néanmoins, l'étude de la variabilité selon le plan horizontal s'avère toujours un sujet problématique compte tenu de la pratique courante. La valeur de la longueur de corrélation horizontale doit alors être supposée avec un ordre de grandeur plutôt décimétrique.

16. Démarche proposée

16.1 Introduction

Contexte

L'objectif de ce chapitre concerne ce que nous appellerons la spatialisation, c'est-à-dire, proposer une répartition dans l'espace 3D (x, y, z) de la résistance de pointe dynamique mesurée au pénétromètre dynamique Panda. On cherche donc à proposer une stratégie de modélisation capable de prendre en compte les variations des propriétés que l'on observe dans l'espace. On a vu que ces variations sont attribuées principalement à la variabilité naturelle intangible des matériaux, même si l'on sait qu'il existe aussi des causes épistémiques qui peuvent expliquer une partie de cette variabilité.

Plusieurs méthodes comme la géostatistique permettent de répondre à cet objectif et caractériser la structuration spatiale des propriétés du sol à différentes échelles lorsque la taille de l'échantillon de données le permet. Les campagnes comportant un grand nombre d'essais répartis de manière uniforme dans l'espace constituent le cadre idéal de ces techniques. Cependant, cette situation est très rare en géotechnique. Dans ce travail, on s'est placé dans un contexte plus réaliste vis-à-vis de la pratique courante, où le nombre d'essais est limité. On cherche à proposer une démarche constituée par un ensemble de procédures appliquées à l'essai de pénétration Panda, permettant d'aboutir à un modèle probabiliste qui fournit une évolution de l'incertitude dans l'espace. Le cadre d'application est la caractérisation des sols in situ à faible profondeur. Une profondeur d'étude maximale égale à 5 mètres est retenue. Cette démarche est appliquée au site expérimental de Castello d'Empúries en Espagne. La description de ce site est donnée en annexe ([Annexe C.1](#)).

Hypothèses de base

On retiendra comme hypothèse de base que la variable d'étude, q_d , peut être représentée par un champ aléatoire homogène en se limitant à l'information statistique du second ordre selon Vanmarcke ([Vanmarcke \(1978\)](#), [Vanmarcke \(1983\)](#)). Les hypothèses sur les deux premiers moments statistiques sont :

$$\text{Espérance du champ constante dans le domaine d'étude : } \mu_V(X) = cte \quad (16.1)$$

$$\text{Il existe aussi une variance finie considérée constante : } \sigma_V^2(X) = cte \quad (16.2)$$

On suppose que la distribution des mesures q_d suit un loi log-normale comme en témoignent les travaux de recherche antérieurs menés pour l'essai Panda ([Deplagne and Bacconnet \(1993\)](#), [Gaouar \(1997\)](#), [Lepetit \(1998\)](#), [Chaigneau \(2001\)](#), [Villavicencio \(2009\)](#)). Ceci

est équivalent à postuler que le logarithme népérien de la résistance de pointe $\ln(q_d)$ est normalement distribué.

La fonction de décroissance exponentielle sera retenue pour modéliser la fonction d'autocorrélation. On choisira un modèle orthotrope pour tenir compte de l'anisotropie entre les directions horizontale et verticale du sous-sol.

Résumé des étapes de la démarche

La démarche envisagée comporte plusieurs étapes pour lesquelles se pose un certain nombre de questions auxquelles on tentera de répondre. Ces différentes phases sont présentées ci-dessous et décrites par la suite :

- 1) Découpage vertical des sondages : on commence par un découpage vertical afin de définir des unités de terrain homogènes. Pour répondre à ce besoin, on a déjà étudié et proposé trois méthodes statistiques dans en Partie II permettant d'aboutir à un découpage par segmentation le long de chaque sondage Panda. On s'intéresse maintenant à l'exploitation des résultats fournis par ces analyses à l'échelle du site pour proposer une partition de l'espace en « couches » ;
- 2) Évaluation du degré d'homogénéité du terrain : une fois que le terrain d'étude a été divisé en plusieurs unités, le but est d'évaluer le degré d'homogénéité selon la direction horizontale en s'appuyant sur les données disponibles. On propose d'étudier ici l'indice de variabilité horizontale (HVI) proposé par Salgado ([Salgado et al. \(2015\)](#)) qui peut nous permettre de statuer sur l'hypothèse de ressemblance ou homogénéité entre sondages en fonction du modèle de découpage que l'on a défini. Cet indicateur dépend principalement de la différence en valeur absolue entre les modèles de tendance déduits pour les profils pénétrométriques. Premièrement, le but est d'étudier si on peut étendre cette méthode aux sondages Panda. Ensuite, si on peut orienter le choix des paramètres du champ par rapport aux résultats fournis par cet indice ;
- 3) Définition des paramètres du champ aléatoire : le passage de l'ensemble des mesures aux paramètres du champ aléatoire est abordé grâce à l'analyse structurale. Lors de cette phase, on vise à définir un modèle de dérive fonctionnelle. Dans ce travail on a privilégié l'utilisation d'une dérive ou tendance linéaire. Ensuite, on doit obtenir un modèle statistique des résidus et étudier leur organisation spatiale. L'objectif final est de définir séparément les échelles de fluctuation selon la direction verticale et horizontale pour proposer un modèle de fonction d'autocorrélation du champ ;
- 4) Simulation probabiliste conditionnelle de q_d : la dernière étape consiste à générer plusieurs simulations du modèle probabiliste grâce aux résultats de l'analyse structurale. Dans ce travail, on propose de conditionner les simulations par les sondages Panda disponibles sur le site grâce au krigeage. Le post-conditionnement du champ permet ainsi d'évaluer comment se propage l'incertitude du modèle par rapport à la configuration de la campagne d'essais pénétrométriques effectuée.

16.1.1 Découpage vertical

D'après plusieurs auteurs l'hypothèse de stationnarité au sens large des propriétés du sol est raisonnable lorsque les données sont analysées au sein d'un même dépôt géologique uniforme (ex. [Phoon and Kulhawy \(1999\)](#), [Christian and Baecher \(2003\)](#)). La détermination

de la structure géologique et géotechnique du site s'avère donc un travail habituel du géotechnicien préalable et nécessaire pour ensuite modéliser la variabilité par une loi statistique.

Les techniques statistiques de découpage vertical étudiées précédemment dans ce travail permettent une identification correcte de cette structure géologique mais aussi de zones statistiquement homogènes au sein d'une même formation géologique. Concrètement ces techniques permettent de proposer un découpage du profil de résistance sur la base des deux premiers moments statiques nécessaires pour définir une loi gaussienne. On propose donc d'utiliser ces méthodes pour proposer une division du terrain en zones homogènes selon un critère statistique et ensuite modéliser cette variabilité spatiale sur chacune de ces zones, comme illustré sur la figure (Fig.16.1).

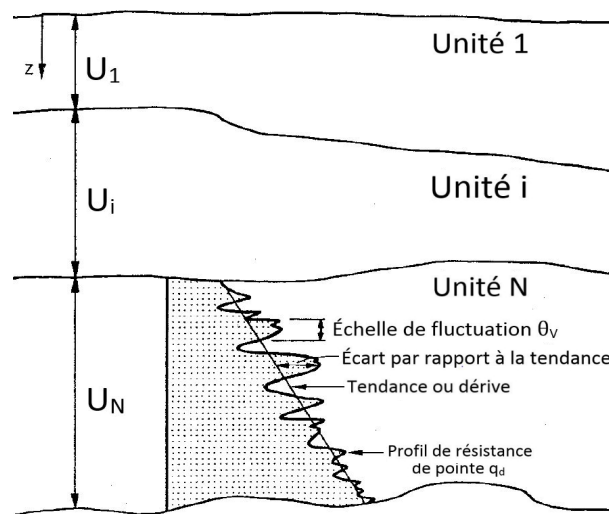


FIGURE 16.1 – Modélisation de la variabilité spatiale d'un profil de sol hétérogène (source : modifié d'après Phoon and Kulhawy (1999))

16.1.2 Critère d'homogénéité horizontale

Dans cette étape on évalue le niveau ou degré d'hétérogénéité selon le plan horizontal à l'échelle du site. Pour ce faire, le critère d'homogénéité retenu est l'indice de variabilité horizontale (HVI) issu de la méthode de Salgado (Salgado et al. (2015)). Cette méthode est en effet intéressante puisqu'elle est applicable même si on dispose d'un nombre restreint d'essais. Elle permet de refléter d'une manière simple l'idée visuelle qu'on peut se faire sur l'homogénéité du site grâce à l'analyse des profils pénétrométriques. Son application aux sondages Panda sera donc étudiée.

Le calcul de cet indice est basé sur la fonction de variabilité horizontale pour chaque couple de sondages (éq.15.33) qu'on peut synthétiser de la manière suivante :

$$F_{HVI} = [A \cdot \Delta tendace + B \cdot correlation] \cdot exp(distance) \quad (16.3)$$

avec A le coefficient pondérateur de l'écart des tendances et B le coefficient pondérateur du coefficient de corrélation entre sondages. Cette somme est pondérée par une fonction exponentielle qui dépend de l'écart entre les points du couple de sondages analysés. Les

valeurs de ces coefficients proposés dans la méthode par (Salgado et al. (2015)) sont ($A = 0.8$ et $B = 0.2$).

La valeur de cet indice peut pourtant guider le choix des paramètres du champ et valider la structure du modèle à couches retenu, puisque :

- il dépend fondamentalement des écarts entre les valeurs moyennes de résistances en fonction de la profondeur. Par conséquent, dans le cas où l'indice calculé présente une faible valeur, on pourrait valider le choix d'un modèle de dérive globale pour représenter la moyenne du champ. Dans le cas contraire, il faudrait s'interroger sur la nature de cette hétérogénéité :
 - o Dans le cas d'un terrain constitué de dépôts très hétérogènes, l'hypothèse d'une moyenne constante selon le plan horizontal n'est plus pertinente. Dans ce cas, on propose soit de définir une tendance selon les directions (x, y) , soit diviser le terrain d'étude en zones où on pourrait supposer homogènes les tendances des sondages réalisés ;
- le coefficient de corrélation entre sondages est aussi une autre mesure pertinente sur la structuration horizontale du terrain. Pour des terrains sédimentaires censés présenter une structuration horizontale avec des couches de faible pendage, ce paramètre devrait présenter une valeur positive relativement élevée. A contrario, quand on est en présence de terrains structurés en couches qui présentent un fort pendage, on devrait avoir des valeurs du coefficient de corrélation beaucoup plus faibles voire négatives à cause du « décalage » entre sondages. En théorie, l'indice HVI devrait dans ce cas là, également présenter des valeurs plus élevées signalant cette hétérogénéité au niveau horizontal.

16.1.3 Analyse structurale des données pénétrométriques

Pour un modèle de corrélation donné, un champ gaussien stationnaire d'ordre 2 $V(X)$ est caractérisé par son espérance μ_V , son écart-type σ_V et la valeur de son échelle de fluctuation θ . Sur la base du profil de mesures fourni par le pénétromètre Panda, on considère qu'il est possible d'estimer empiriquement ces deux premiers moments statistiques. De même, on suppose qu'il est possible de fournir une estimation de l'échelle de fluctuation verticale à partir de l'analyse de chacun des sondages. D'ailleurs, l'hypothèse d'un modèle de corrélation orthotrope permet d'analyser séparément la structure spatiale au sens horizontal et vertical.

Compte tenu de ces constats, un traitement statistique des sondages Panda a été proposé. Il comporte un certain nombre d'étapes résumées dans la figure (Fig.16.2) et décrites ci-dessous :

- (1) Lissage à l'aide d'une moyenne glissante (procédure qu'on avait appelé "régularisation") appliqué à chaque sondage¹. Le pas de régularisation proposé dans cette étude pour le traitement des sondages in situ est égal à 10 mm ;
- (2) Transformation logarithmique de la résistance de pointe dynamique q_d afin de travailler avec une variable aléatoire normale ;

1. En pratique, le calcul de la fonction d'autocorrélation suppose l'existence d'un pas régulier de mesure ce qui n'est pas le cas de l'essai Panda. La technique de régularisation permet d'homogénéiser le support de la mesure et de limiter les biais que cela pourrait engendrer.

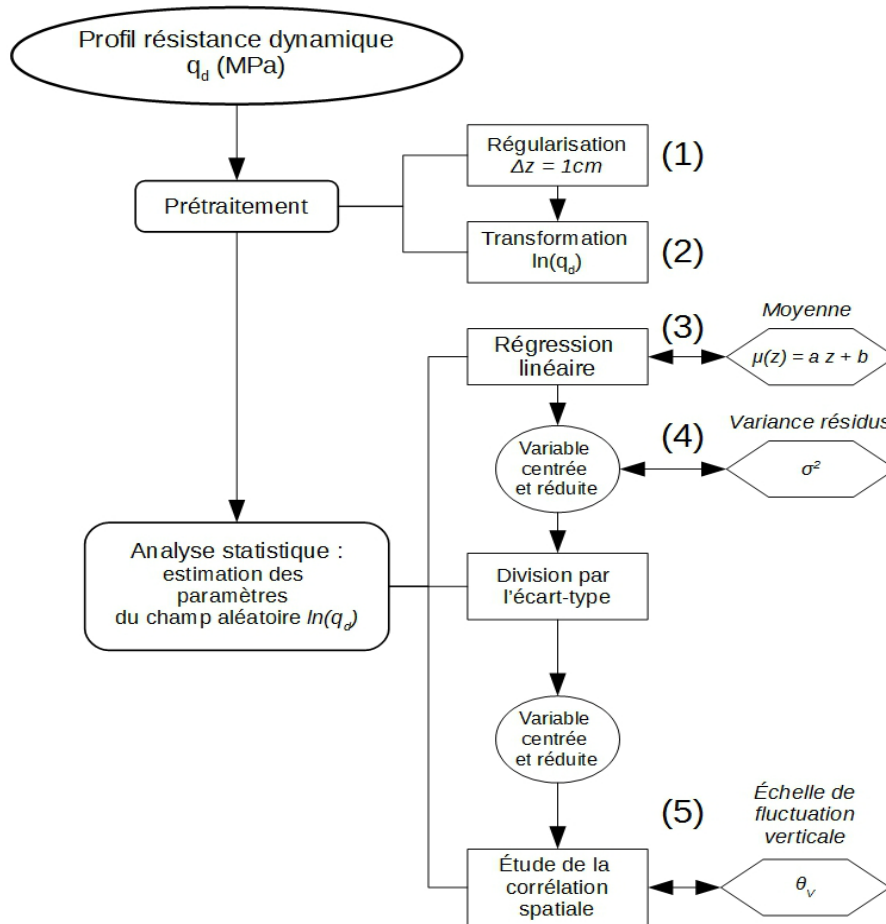


FIGURE 16.2 – Schéma du traitement statistique des sondages Panda

- (3*) Calcul de la tendance ou dérive linéaire de chaque unité selon la profondeur calculée par la technique de moindres carrés ordinaires ;
 - (4*) Calcul de la variance des résidus une fois la dérive enlevée ;
 - (5*) Estimation de la fonction d'autocorrélation empirique pour l'échantillon des résidus normalisés par leur écart-type et ajustement d'une loi d'autocorrélation théorique permettant de calculer la valeur de l'échelle de fluctuation verticale θ_V ;
- * Il faut noter que les étapes n°3, 4 et 5 sont réalisées indépendamment sur chacune des unités du modèle retenu.

A partir des résultats issus de l'analyse statistique de chacun des sondages Panda, on calcule les paramètres du champ, pour chaque unité ou couche du modèle comme suit :

- la moyenne du champ est représentée par un modèle linéaire. Il est calculé à l'aide d'une régression globale par moindres carrés ordinaires pour l'ensemble des sondages disponibles ;
- l'écart-type est calculé comme l'écart-type empirique de l'échantillon des résidus ;
- la valeur retenue pour l'échelle de fluctuation verticale est la moyenne des valeurs de θ_V déduites pour chaque sondage ;
- on estime l'échelle de fluctuation horizontale à partir de l'autocorrélogramme empirique de l'ensemble de sondages. Ceci s'avère le point le plus délicat puisque en général le nombre de sondages pénétrométriques dans une campagne de reconnaissance cou-

rante ne permet pas de définir les lois de variabilité horizontale des massifs de sol d'une manière précise.

16.1.4 Simulation du champ aléatoire conditionné par les sondages Panda

Les champs sont simulés grâce à la factorisation de Cholesky de la matrice de covariance. Il est possible de conditionner le champ aléatoire par les sondages Panda effectués sur site grâce au post-conditionnement par krigeage et obtenir ainsi des simulations conditionnelles (cf. Fig. 16.3). Les étapes principales de l'algorithme de post-conditionnement proposé sont données dans la figure (Fig. 16.4). On signale que le choix de la technique, krigeage simple (KS) ou ordinaire (KO), n'est pas trivial. La variance du KS est plus faible et pourtant l'incertitude du modèle sera plus petite. Par contre, l'hypothèse de moyenne nulle constante dans tout l'espace peut ne pas être valable lorsque l'on enlève un modèle de tendance globale aux profils de mesures. De même, l'espérance du champ peut changer en fonction du pas vertical utilisé pour le conditionnement. L'hypothèse de moyenne connue semble être très forte et le choix du KO s'avère donc plus pertinent.

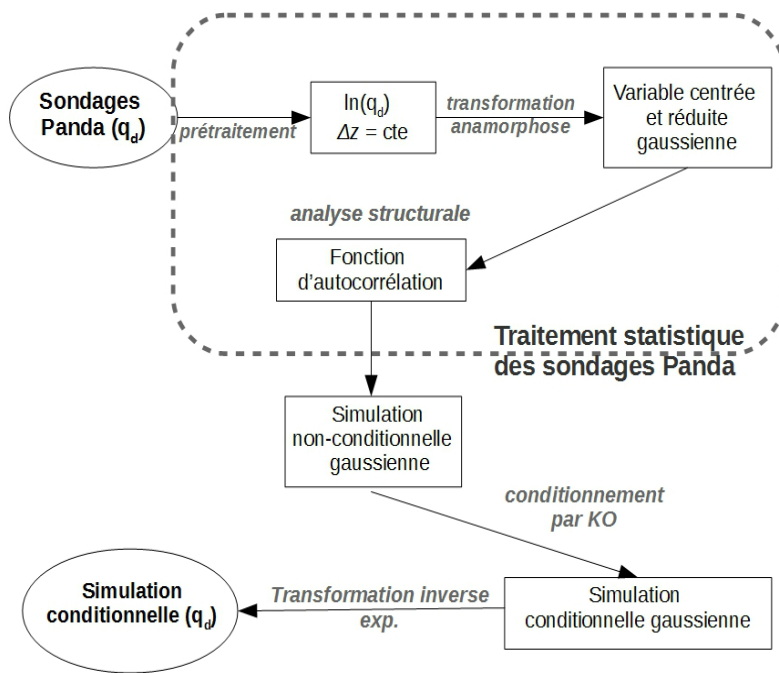


FIGURE 16.3 – Organigramme des différentes étapes de calcul des simulations conditionnelles

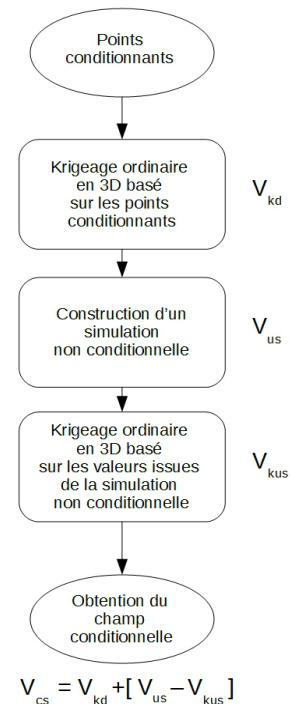


FIGURE 16.4 – Schéma de la procédure de post-conditionnement par krigeage

16.2 Cas d'étude : site expérimental de Castello d'Empúries

16.2.1 Résultats des essais de pénétration Panda

Les résultats bruts pour les huit sondages Panda jusqu'à 5 mètres sont fournis sous la forme de pénétrogrammes dans la figure (Fig.16.5 (a)). On peut observer des valeurs relativement fortes de q_d , entre 10 et 20 MPa pour le premier mètre en surface. Ensuite la résistance décroît et au dessous de 2m de profondeur la valeur de q_d devient constante, avec des résistances mécaniques comprises entre 1 et 2 MPa. Dans la même figure (Fig.16.5 (b)) on a représenté la distribution des mesures de q_d , qui présente d'une manière très claire une allure bimodale. La distribution des enfoncements enregistrés lors du battage pendant la réalisation des essais est également incluse dans la figure (cf. Fig.16.5 (c)). On signale que l'enfoncement moyen, dont la valeur exacte estimée est de 9.8 mm, est presque le même que le pas de régularisation proposé (10 mm).

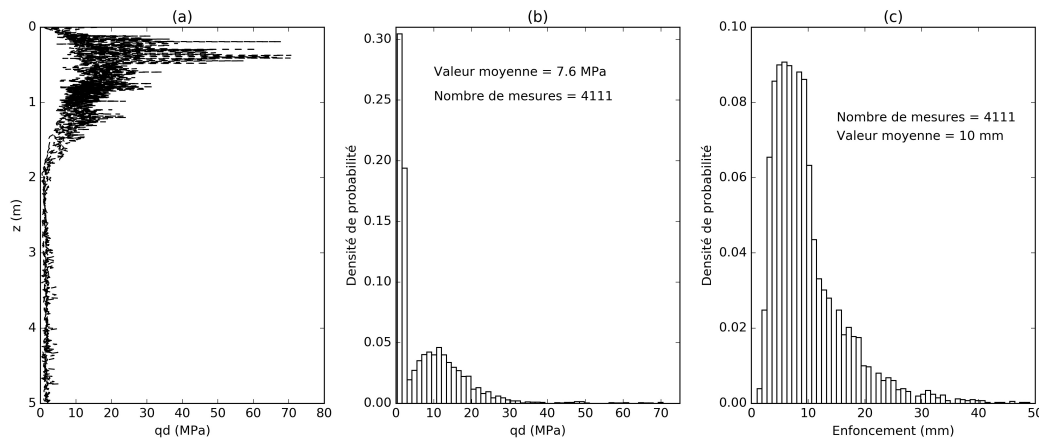


FIGURE 16.5 – (a) Pénétrogrammes bruts jusqu'à 5 mètres de profondeur. (b) Histogramme des mesures de résistance. (c) Histogramme des enfoncements enregistrés.

L'ensemble des pénétrogrammes régularisés est présenté sur la figure (Fig.16.6 (a)) et les caractéristiques statistiques des sondages régularisés et bruts sont regroupés dans un tableau donné en annexe (Annexe C.2). On n'observe que sur les pénétrogrammes régularisés, il n'y a plus de mesures de résistance supérieures à 50 MPa qui pouvaient être associées à des valeurs aberrantes ou à la présence de cailloutis plutôt qu'à une résistance de pointe représentative de la formation auscultée. L'évolution des deux premiers moments statistiques des pénétrogrammes en fonction de la profondeur (Fig. 16.6 (b) (c)) montre que le coefficient de variation de la résistance de pointe présente des variations comprises entre 15 et 50 % avec des fluctuations erratiques le long de la profondeur, surtout à partir de 3 mètres. Ces fluctuations sont dues certainement à des incertitudes de mesure qu'on a constaté durant la réalisation de certains sondages. Sinon, l'inspection du profil de résistance moyenne permet de mettre en évidence d'une manière plus claire, une formation plus compacte en surface sur les 80 premiers centimètres avec une résistance moyenne de 17 MPa. Suite à l'examen visuel et statistique, on peut remarquer une structuration horizontale évidente typique de

formations deltaïques.

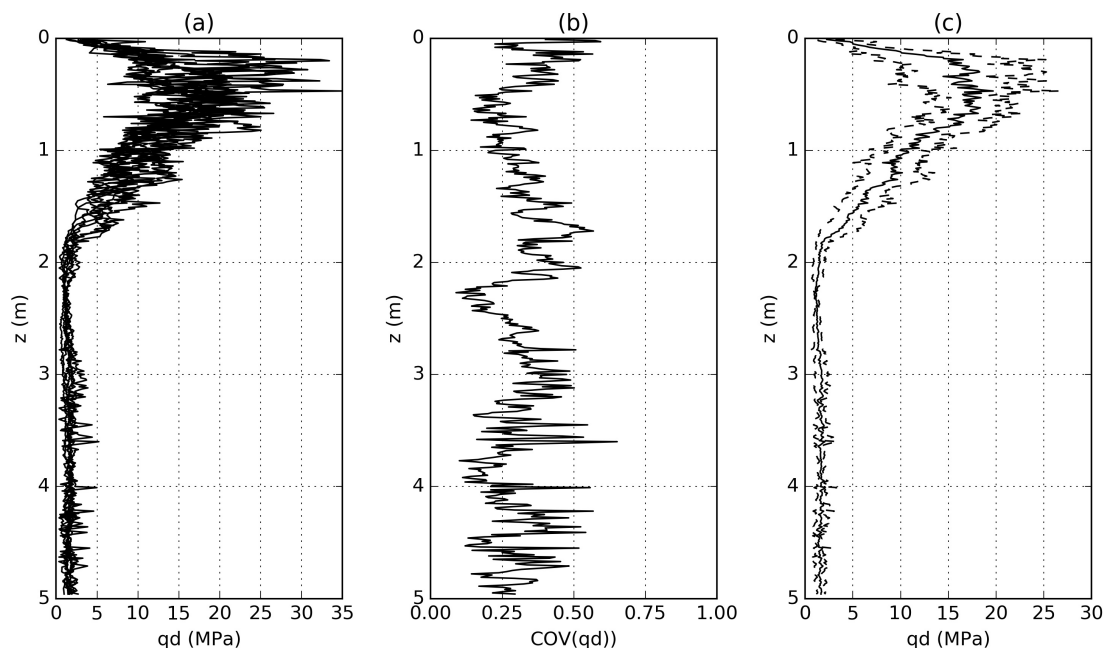


FIGURE 16.6 – Caractéristiques statistiques des huit sondages Panda : (a) Pénétrogrammes régularisés (b) coefficient de variation (c) moyenne et écart-type.

16.2.2 Application des techniques stratigraphiques pour la définition des unités statistiquement homogènes

Définition des couches

La stratigraphie de ce site a déjà été décrite dans le deuxième chapitre de ce manuscrit. Les techniques de découpage statistique ont été appliquées à l'ensemble des sondages régularisés en vue de définir les différentes unités de terrain. Des fenêtres de l'ordre d'un mètre ont été testées (1, 1.5 et 2 m). Les tests *Tratio*, *Intraclasse* ont été appliqués au $\ln(q_d)$ avec les valeurs seuil retenues auparavant et le test de *Bartlett* a été utilisé sur l'analyse de q_d .

Les profondeurs des limites de couches détectées avec les trois méthodes statistiques pour chacun des sondages sont présentées dans la figure (Fig.16.7). D'après cette figure, on observe que indépendamment de la taille de fenêtre choisie, les trois approches repèrent une limite de couche entre 1.5 et 2 mètres, qui correspond à la transition entre la formation sableuse plus compacte en surface et la formation de matériaux fins. De plus, lorsque on retient la taille de fenêtre la plus grande (2 m), seule cette interface est détectée, à l'exception des sondages S3 et S4 pour lesquels la méthode *Tratio* identifie également une limite à environ 3 mètres de profondeur. Avec une fenêtre égale à 1m, on repère une interface plus proche de la surface, autour d'un mètre, qui correspond avec la formation plus compacte signalée auparavant. Cette limite a été aussi identifiée quand la longueur de fenêtre est égale à 1.5 m sauf pour le pénétrogramme S4. Enfin, autour de 3m, une interface est aussi identifiée pour certains sondages.

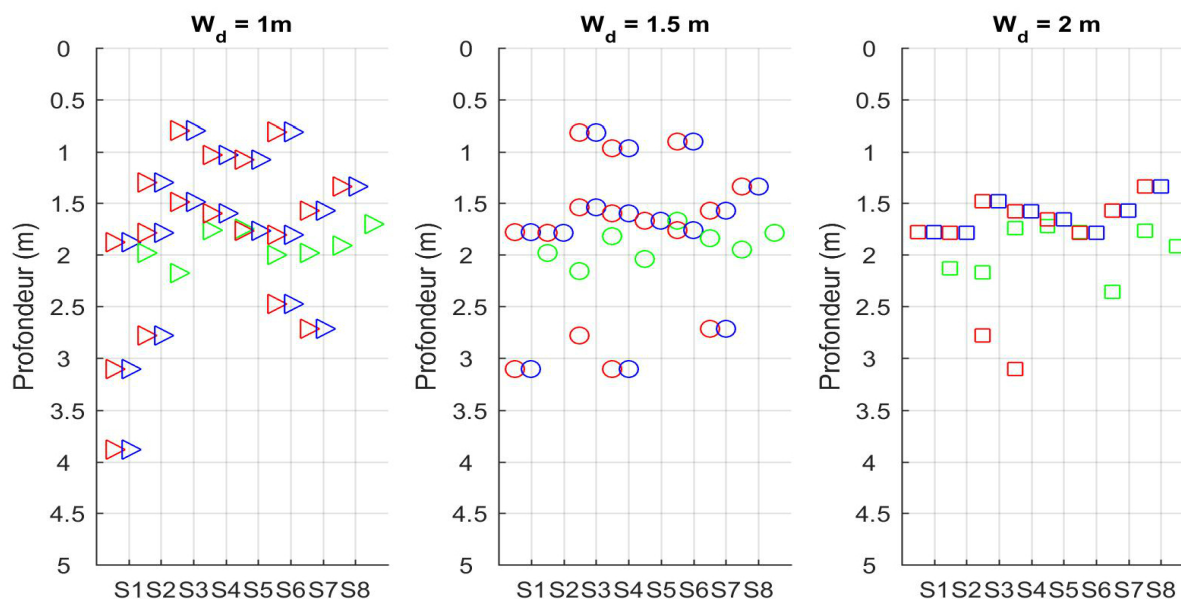


FIGURE 16.7 – Limites de couches détectées avec les méthodes statistiques. Code couleur : **Tratio**, **Intra-classe** et **Bartlett**. Avec W_d la longueur de la fenêtre testée.

Modèles de découpage choisis

Au regard des analyses réalisées, on a décidé d'étudier et de comparer 4 « modèles de découpage » possibles. On précise que la profondeur des interfaces a été calculée comme la valeur moyenne des interfaces estimées pour chacun des sondages. Les valeurs retenues pour le calcul sont celles issues de la méthode Tratio avec une longueur de fenêtre égale à 1 m, ce qui permet de distinguer jusqu'à 4 unités. Dans la figure (Fig.16.8) on fournit un schéma synthétique des quatre modèles retenus et une synthèse des statistiques de q_d pour les différents sondages est donnée dans les tableaux fournis en annexe (Annexe C.2). Ces quatre modèles sont décrits ci-dessous :

- i. **M1** : aucun découpage est considéré ;
- ii. **M2** : deux unités sont considérées. La limite entre ces deux couches a été estimée à 1.63 m. Ce modèle est cohérent avec la géologie du site, constituée par une couche de sol sableux en surface et une couche de matériaux fins jusqu'à 6 m. De même, on souligne que cette interface est la seule identifiée sur tous les sondages existants. Ce qui n'est pas le cas pour les autres modèles ;
- iii. **M3** : il comporte 3 unités. Une couche est considérée en surface à 0.9 m de profondeur détectée sur les sondages S3, S4, S5 et S6 ;
- iv. **M4** : quatre unités sont considérées dans ce modèle, en prenant en compte une unité supplémentaire à partir de 2.9 m de profondeur.

Bilan

Pour un terrain deltaïque qui présente une structuration horizontale, les méthodes étudiées ont montré une répétabilité satisfaisante quant à la profondeur des interfaces identifiées pour la campagne de sondages réalisée sur site. Ils sont donc peu sensibles à la variabilité naturelle du terrain pour une fenêtre de l'ordre du mètre et robustes.

Prof. (m)	M1	M2	M3	M4	
0	Unité 1	Unité 1	Unité 1	Unité 1	
0.2					Unité 1
0.4					
0.6					
0.8				0.90 ± 0.07 (m)	0.90 ± 0.07 (m)
1					
1.2				Unité 2	Unité 2
1.4					
1.6			1.63 ± 0.15 (m)	1.63 ± 0.15 (m)	1.63 ± 0.15 (m)
1.8					
2					
2.2				Unité 3	
2.4					
2.6					
2.8				2.93 ± 0.20 (m)	
3					
3.2		Unité 2	Unité 3		
3.4					
3.6					
3.8					
4				Unité 4	
4.2					
4.4					
4.6					
4.8					
5					

FIGURE 16.8 – Coupe schématique des 4 modèles de découpage étudiés avec la valeur moyenne et l'écart-type des limites entre couches.

Les méthodes Tratio et Intra-classe fournissent les mêmes résultats alors que le test de Bartlett ne distingue que les deux formations lithologiques du terrain à une profondeur légèrement supérieure par rapport aux deux autres méthodes.

Cette étape permet donc de proposer un découpage vertical du terrain sur la base d'unités homogènes. Un total de quatre modèles de découpage en fonction du nombre d'unités de terrain ont été retenus. Par la suite, on étudiera l'influence des géométries proposées vis-à-vis de la stratégie de modélisation adoptée.

16.2.3 Homogénéité horizontale du site

Critère d'homogénéité : indice de variabilité horizontale HVI

L'indice de variabilité horizontale (HVI) permet d'évaluer l'homogénéité horizontale à l'échelle du site. Afin d'étudier la pertinence de ce critère et son applicabilité au pénétromètre Panda, la stratégie suivante a été développée :

- i. étude de sensibilité de l'indice de variabilité horizontale (HVI) en fonction des valeurs des coefficients A et B pour chacun des modèles considérés (M1, M2, M3 et M4) ;
- ii. comparaison de l'indice de variabilité horizontale (HVI) obtenu pour les sondages Panda avec celui déterminé à partir des sondages CPT sur le même site ;
- iii. étude de l'homogénéité par unité : le coefficient de normalisation des tendances étant fonction de la profondeur du sondage, sa valeur augmente avec la profondeur d'étude choisie. C'est pourquoi, on propose d'appliquer cette méthode sur les unités de terrain prises isolément ;
- iv. finalement, comparaison des indices HVI calculés avec ceux issus de l'analyse de la

transformée logarithmique des pénétrogrammes.

On précise que pour l'application de cette méthode, on retiendra un modèle linéaire de dérive sur chaque unité pour chacun des sondages.

Influence des coefficients pondérateurs

Les résultats issus de l'étude de sensibilité pour HVI en fonction des modèles considérés (M1, M2, M3 et M4) sont regroupés dans le tableau (Tab.16.1). On peut noter que le site est homogène selon la direction horizontale puisque l'indice HVI est toujours plus petit que 10% donc largement inférieur à la valeur seuil de 33% fixée par la méthode. Les principales constatations et conclusions tirées de l'analyse sont :

- la variation des coefficients A et B n'entraîne pas de différences significative sur HVI. Ceci confirme la similarité entre sondages selon le plan horizontal ;
- la division en une ou plusieurs couches entraîne une réduction de l'indice HVI relativement faible. L'homogénéité horizontale du terrain est vérifiée pour la profondeur d'étude choisie indépendamment de la structuration en couches choisie. D'ailleurs, on signale que le coefficient de corrélation entre sondages est indépendant de la stratigraphie associé aux sondages. Par conséquent, l'influence du modèle de stratigraphie choisi décroît avec l'importance qu'on accorde à la corrélation entre sondages dans le calcul de cet indice ;
- quand on ne tient compte que des écarts entre sondages ($A = 1$) les modèles de terrain M3 et M4 sont ceux qui fournissent la valeur de HVI la plus faible et le terrain est plus homogène selon la direction horizontale. on obtient la même conclusion quand on retient les valeurs des coefficients A et B proposés dans la formule originale. D'ailleurs, il n'existe pas de différences significatives entre l'indice HVI calculé pour le modèle M3 ou M4 (0.1%). Sur la base de cette analyse, on pourrait proposer de retenir le modèle de découpage M3 qui donne la plus faible valeur de HVI.

TABLE 16.1 – Indice de variabilité horizontale HVI pour les différents modèles testés

Coeff. A	1	0.8	0.6	0.5	0.4	0.2	0
Coeff. B	0	0.2	0.4	0.5	0.6	0.8	1
M1	6.8%	6.9%	6.9%	6.9%	6.9%	7.0%	7.0%
M2	7.3%	7.2%	7.2%	7.1%	7.1%	7.0%	7.0%
M3	5.3%	5.7%	6.0%	6.2%	6.3%	6.7%	7.0%
M4	5.2%	5.6%	5.9%	6.1%	6.3%	6.6%	7.0%

Comparaison avec les sondages CPT

Afin de comparer les résultats obtenus à partir de l'analyse des sondages Panda, l'indice HVI a aussi été calculé sur la base des deux sondages CPTU réalisés sur ce site (cf. Annexe A.2). Pour ce faire, on a procédé de manière similaire au cas précédent. On a tout d'abord considéré l'intégralité du sondage, sans réaliser aucun découpage. On a ensuite suivi le procédé suggéré par Salgado (cf. Robertson (1990)). Les limites des couches identifiées ainsi que leur nature sont présentées dans la figure (Fig.16.9). L'indice de variabilité horizontale est calculé pour les deux cas suivants :

- aucune découpage est pris en compte, $HVI = 9.8\%$;
- prise en compte de la stratigraphie du site selon l'abaque normalisé de Robertson, $HVI = 5.7\%$;

Bien que le calcul de cet indice n'est basé que sur une seule paire de sondages CPTU, le résultat obtenu est très similaire à celui calculé à partir des 8 sondages Panda et confirme que le site est homogène selon la direction horizontale d'après cette méthode.

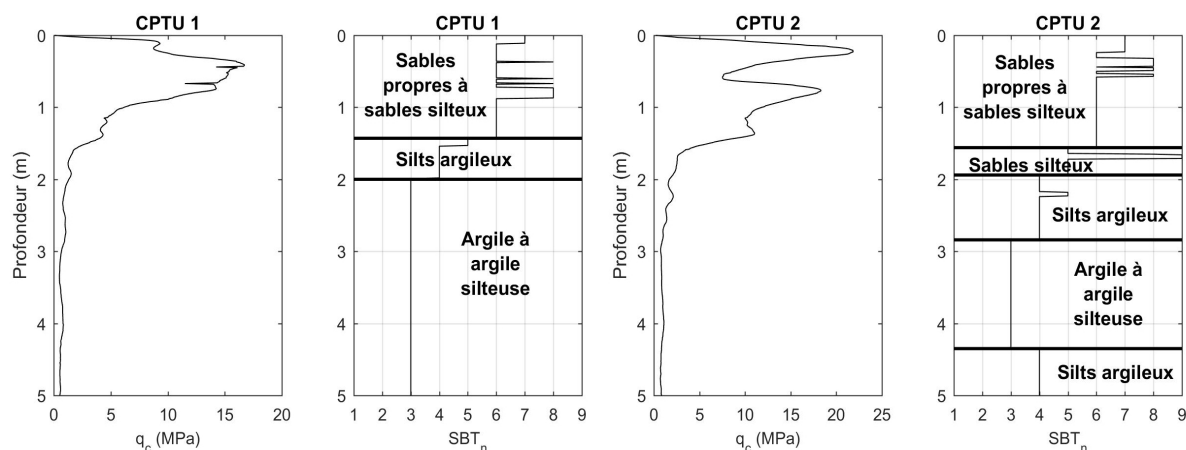


FIGURE 16.9 – Profils de résistance de pointe statique q_c et stratigraphies interprétées selon l'abaque de classification de Robertson (Robertson (1990))

Variabilité horizontale de la couche d'argile

Le calcul du HVI est relancé à nouveau mais cette fois sur la seule couche de matériaux fins après 1.5 à 2 m de profondeur. La méthode n'est donc appliquée qu'aux pénétrogrammes entre 2 et 5 mètres (cf. Fig. 16.10), c'est-à-dire, pour une "profondeur" d'étude égale à 3m. La valeur du coefficient de normalisation de la différence entre tendances pour une profondeur d'étude de 3 m est égal à 11 MPa alors que pour une profondeur de 5 m sa valeur est égale 13 MPa.

Les résultats des indices HVI calculés sont regroupés dans le tableau (Tab. 16.2), en fonction des valeurs des coefficients pondérateurs. On peut noter que :

- les valeurs par défaut de la méthode ($A = 0.8$ et $B = 0.2$) donnent une valeur de HVI égale à 12%, similaire bien que supérieure à celle calculée précédemment. Néanmoins avec ces valeurs, le site serait toujours classé comme « homogène » ;
- a contrario, si on ne tient compte que de l'écart entre tendances, on constate une valeur de HVI plus faible par rapport au calcul issu de la totalité du sondage. Ceci est cohérent avec ce qu'on observe, les mesures de q_d au sein de cette unité sont plus centrées autour d'une valeur moyenne ;
- une augmentation systématique de la valeur de HVI est observée lorsqu'on augmente la valeur du coefficient B. La corrélation entre sondages pour cette épaisseur de terrain est plus faible que celle calculée lorsque on prend en compte le sondage complet. Les erreurs de mesures qu'on voit apparaître peuvent avoir une influence.

Les résultats fournis par cette méthode sont similaires, même si la profondeur d'étude change ou si l'on étudie de manière isolée une unité lithologique.

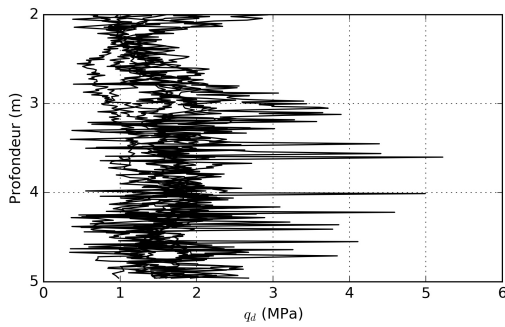


FIGURE 16.10 – Pénétrogrammes regularisés au sein de la couche argileuse

HVI	Coeff. A	Coeff. B
3.6%	1	0
12.1%	0.8	0.2
20.6%	0.6	0.4
24.8%	0.5	0.5
29.1%	0.4	0.6
37.5%	0.2	0.8
46.0%	0	1

TABLE 16.2 – Indices HVI calculés pour les pénétrogrammes au sein de la couche argileuse

Influence de la transformation logarithmique

La méthode de Salgado est maintenant appliquée aux profils du $\ln(q_d)$. Pour cela, une transformation logarithmique a été appliquée aux constantes de normalisation déduites à l'aide de la fonction empirique (cf. Fig.15.4).

Comme précédemment les indices de variabilité horizontale (HVI) pour le site sont calculés en fonction des différents modèles de terrain. Ces nouveaux indices sont alors comparés à ceux calculés pour les profils de q_d sur les graphiques (Fig.16.11). On peut noter que les résultats fournis par cette méthode sont très similaires indépendamment de la variable analysée (q_d ou $\ln(q_d)$), mais la valeur de HVI est légèrement supérieure quand on découpe les sondages en plusieurs unités (M2, M3, M4) et qu'on analyse le logarithme de la résistance. On peut noter que lorsque l'on ne tient compte que de la corrélation entre sondages (coefficient A est nulle), la valeur de HVI est identique. L'écart entre la valeur de HVI pour q_d et $\ln(q_d)$ augmente avec l'importance qu'on accorde au coefficient A (pondérateur de l'écart entre tendances).

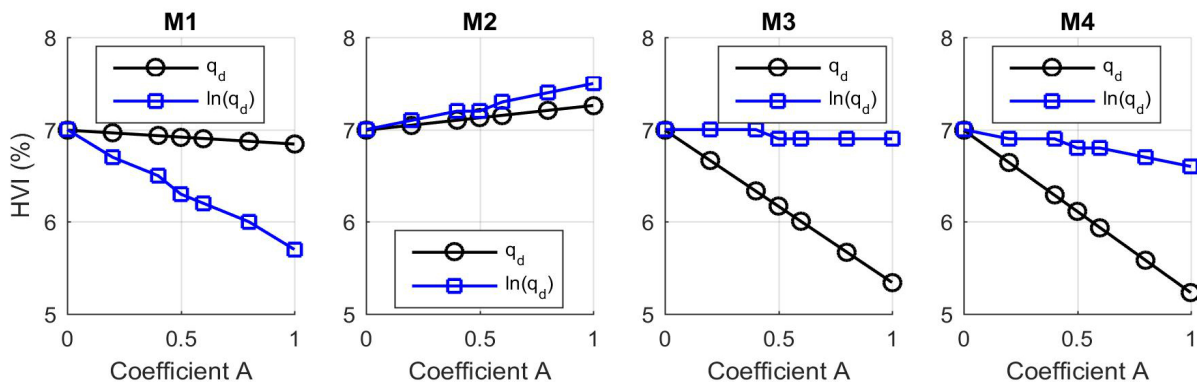


FIGURE 16.11 – Comparaison entre les indices de variabilité issus de l'analyse de q_d avec ceux calculés pour $\ln(q_d)$ en fonction du modèle de découpage

De la même manière, on compare la valeur des indices HVI issus de l'analyse de la couche d'argile pour $\ln(q_d)$ et q_d . Les nouvelles valeurs des indices de variabilité sont regroupés dans le tableau (Tab.16.3). Un graphique comparatif est donné en (Fig.16.12) à partir duquel

on peut tirer des conclusions très similaires que celles obtenues auparavant. Cependant, compte tenu que pour cette formation on a constaté des valeurs de résistances faibles, même inférieures à 1 MPa, la transformation logarithmique peut entraîner des valeurs négatives. L'écart absolu entre modèles de tendances peut donc augmenter. Pour ce cas, on l'observe puisque l'écart entre l'indice HVI issu de l'analyse q_d et l'indice HVI issu de l'analyse $\ln(q_d)$ est légèrement plus important (de l'ordre de 5% quand $A = 0.8$) par rapport à l'analyse antérieure.

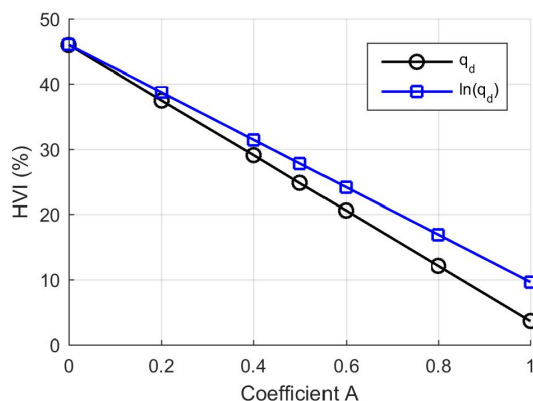


FIGURE 16.12 – Comparaison entre les indices HVI

HVI	Coeff. A	Coeff. B
9.6%	1	0
16.9%	0.8	0.2
24.2%	0.6	0.4
27.8%	0.5	0.5
31.5%	0.4	0.6
38.7%	0.2	0.8
46.0%	0	1

TABLE 16.3 – Indices HVI calculés pour $\ln(q_d)$ au sein de la couche argileuse

Bilan

La méthodologie d'analyse des sondages CPT proposée par Salgado a été retenue afin de fournir un critère pour caractériser la variabilité horizontale du site. Les résultats ont permis de mettre en évidence que :

- d'après cette approche le site est homogène selon la direction horizontale, indépendamment des sondages analysés (CPT ou Panda). Ceci est cohérent avec la géologie d'un site deltaïque.
- la similarité entre les résultats issus de l'analyse de sondages Panda et ceux obtenus pour la couple de sondages CPT semblent indiquer qu'on peut étendre l'application de cette technique aux données recueillies par le Panda ;
- l'hétérogénéité horizontale est plus faible (valeur de HVI plus petite) quand on prend en compte un modèle de découpage M3 ou M4 ;
- on obtient des résultats très similaires lorsqu'on analyse les profils logarithmiques de résistance. Cependant l'indice de variabilité du site peut varier légèrement (cette variation augmente avec la profondeur des sondages) quand on analyse des formations avec des résistances mécaniques très faibles. Vis-à-vis de l'application de cette méthode pour des essais menés à faible profondeur (jusqu'à 5m), le résultat final ne devrait pas être altéré.

Compte tenu des résultats fournis par cette méthode, qui indiquent une homogénéité selon la direction horizontale, on retiendra le choix d'une dérive unique ou globale pour $\ln(q_d)$ pour chacune des unités des différents modèles.

16.2.4 Analyse structurale

Dans ce paragraphe on présente l'analyse structurale des données de résistance à travers le paramètre $\ln(q_d)$. Pour réaliser cette étude on a choisi de retenir le modèle M2 constitué par deux unités qui présentent un lien avec la géologie du site. Par conséquent, on traite séparément la couche des matériaux sablo-limoneux et argileux. Les différentes étapes de cette analyse sont :

- i. définition du modèle de dérive ou tendance et caractéristiques statistiques des résidus ;
- ii. étude de la corrélation spatiale des résidus selon la direction verticale ;
- iii. étude sur la corrélation spatiale des résidus selon la direction horizontale.

Choix et estimation de la dérive

Dans notre démarche, on a préconisé le choix pragmatique d'un modèle de dérive ou tendance linéaire en fonction de la profondeur $z(m)$. On présente les résultats issus de la caractérisation statistique pour les deux formations considérées. Une figure avec les profils du $\ln(q_d)$ et les modèles de tendances linéaires ajustés pour chaque essai, au sein de chacune des unités lithologiques est donnée en annexe (Annexe C.4.1). Les équations des modèles de dérives linéaires en fonction de la profondeur sont regroupées dans le tableau (Tab.16.4). On a également inclus les statistiques des résidus définies pour ces dérives ou tendances, notamment la variance σ_r^2 et l'écart-type σ_r .

TABLE 16.4 – Modèles de dérives et caractéristiques statistiques des résidus de la formation sableuse

Essais	Unité 1			Unité 2		
	Dérive	σ_r^2	σ_r	Dérive	σ_r^2	σ_r
S1	$\ln(q_d) = 2.808 - 0.317z$	0.165	0.406	$\ln(q_d) = -0.051 + 0.066z$	0.242	0.492
S2	$\ln(q_d) = 2.351 - 0.037z$	0.179	0.424	$\ln(q_d) = 1.281 - 0.270z$	0.144	0.379
S3	$\ln(q_d) = 3.136 - 1.076z$	0.213	0.462	$\ln(q_d) = 0.220 + 0.054z$	0.078	0.279
S4	$\ln(q_d) = 2.972 - 0.638z$	0.329	0.573	$\ln(q_d) = 0.036 + 0.142z$	0.087	0.294
S5	$\ln(q_d) = 2.882 - 0.805z$	0.162	0.403	$\ln(q_d) = 0.395 - 0.017z$	0.112	0.335
S6	$\ln(q_d) = 2.687 - 0.512z$	0.247	0.497	$\ln(q_d) = 0.370 + 0.015z$	0.124	0.352
S6	$\ln(q_d) = 2.500 - 0.326z$	0.158	0.397	$\ln(q_d) = 0.677 - 0.034z$	0.081	0.285
S8	$\ln(q_d) = 2.578 - 0.468z$	0.275	0.525	$\ln(q_d) = 0.612 - 0.004z$	0.086	0.293

Précédemment, on a déterminé que le site est horizontalement homogène d'après le critère issu de la méthode de Salgado. Par conséquent, on cherche à calculer une dérive globale pour le site et sur chacune des unités retenues. Les résultats issus d'une régression globale pour les deux formations sont donnés dans la figure (Fig.16.13). Une analyse similaire est donnée en annexe (Annexe C.4.2) pour les trois autres modèles de découpage (M1, M3 et M4). On peut voir dans le graphique placé à gauche l'ensemble des sondages en échelle logarithmique et en rouge les deux modèles de dérive linéaire calculés. Les équations de ces deux tendances sont données en (éq.16.4) et (éq.16.5) pour la première et deuxième unité respectivement.

$$\text{Unité 1} \quad \ln(q_d) = 2.739 - 0.513z \quad (16.4)$$

$$\text{Unité 2} \quad \ln(q_d) = 0.434 - 0.005z \quad (16.5)$$

Le coefficient directeur du modèle linéaire pour la première unité présente une valeur négative, indiquant une décroissance des résistances mécaniques avec la profondeur. Cependant, la pente calculée pour la deuxième unité a une valeur très faible, presque nulle, donc l'espérance du champ est presque constante pour cette formation. Dans la figure (Fig.16.13), on fournit aussi les histogrammes des résidus pour les deux formations lithologiques une fois que la tendance globale a été enlevée. De même, on peut remarquer un calage satisfaisant avec la loi normale malgré une légère asymétrie positive pour la première formation. Ceci aurait pu être contourné en négligeant les 20 cm premiers de terre végétale ou en calculant un modèle de dérive quadratique pour mieux s'ajuster à cette augmentation de la résistance dans les premiers centimètres en surface. Toutefois, quand on applique aux résidus le test de *Kolmogorov-Smirnov* afin de tester l'hypothèse de normalité des résidus, les *P-valeurs*² calculées sont presque égales à 0, ce qui conduirait à rejeter cette hypothèse, malgré le bon calage visuel entre les données expérimentales et la loi théorique.

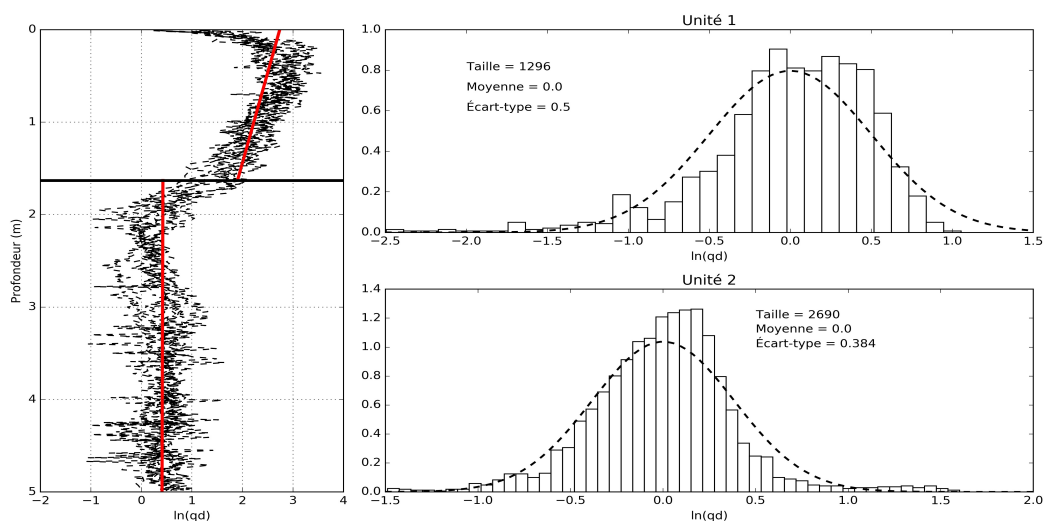


FIGURE 16.13 – À gauche : tendances linéaires issues d’une régression globale des profils de $\ln(q_d)$. À droite : histogrammes des résidus pour les deux unités et ajustement d’une loi normale

Autocorrélation verticale des résidus

Afin de caractériser la variabilité spatiale selon la direction verticale, on analyse séparément chacune des unités du modèle de découpage afin de calculer la fonction d’autocorrélation empirique. Un modèle théorique est ensuite ajusté pour déterminer l’échelle de fluctuation verticale pour chacun des sondages disponibles.

Les résultats des analyses pour les 8 sondages sont présentés dans deux figures : (Fig.16.17) pour les mesures au sein de la formation sableuse et (Fig.16.18) pour la deuxième couche argileuse. Ces figures sont structurées en 3 colonnes. Les résidus du $\ln(q_d)$ normalisés par l’écart-type présentés dans la première colonne sont définis par les dérivées linéaires calculées pour chacun des sondages et dont les équations ont été données dans le tableau (Tab.16.4).

2. La *valeur-P* ou *P-valeur* peut être interprétée comme le plus petit seuil de signification α pour lequel l’hypothèse nulle est acceptée. Couramment, on suppose une valeur de α égale à 5% qui est équivalent à supposer un niveau de confiance égal à 95%

Dans la deuxième colonne, on montre les autocorrélogrammes expérimentaux avec les 4 modèles d'autocorrélation théoriques ajustés et les limites de Bartlett, représentées par les lignes en pointillés noirs. On a suivi la même notation que celle donnée dans le tableau (Tab.15.1) pour les lois d'autocorrélation théoriques.

Pour la première formation sableuse plus compacte en surface, les autocorrélogrammes empiriques présentent un comportement à l'origine semblables, à l'exception du sondage S2 pour lequel la corrélation décroît plus lentement. Les valeurs des échelles de fluctuation verticales θ_V estimées à partir des fonctions d'autocorrélation théoriques sont reportées dans le tableau (Tab.16.5). La fourchette des valeurs varie entre 8 et 23 cm à l'exception du sondage S2. Pour cet essai, la valeur de θ_V est plus forte, indépendamment du modèle théorique choisi. On remarque des valeurs pour θ_V similaires indépendamment de la loi de corrélation. Cependant, les distances de corrélation les plus élevées sont issues du modèle exponentiel, de l'ordre de 0.2 m, et les valeurs plus faibles sont calculées pour le modèle cosinus avec une valeur moyenne égale à 0.14 m.

Pour la formation argileuse, on voit apparaître d'une manière très claire les fluctuations à faible échelle (pouvant être liées à des erreurs de mesure) qu'on avait évoquées précédemment pour les sondages S4, S5 et S8. L'inspection visuelle des autocorrélogrammes empiriques pour ces essais semble indiquer que ces singularités dans le signal pénétrométrique entraînent des structures parasites comme des effets de trou. Malgré ces valeurs extrêmes, les histogrammes des résidus semblent s'ajuster assez bien à la loi normale. Les effets de trou³ ont été repérés pour les sondages S2, S5, S6 et S7. Cependant on observe des comportement plus ou moins similaires à l'origine de la fonction d'autocorrélation. Les échelles de fluctuation verticales θ_V estimés pour cette formation sont récapitulés dans le tableau (Tab.16.5).

TABLE 16.5 – Échelles de fluctuation verticales estimées à l'aide des différents modèles d'autocorrélation théoriques sur chaque unité

Essais	Unité 1				Unité 2			
	θ_{SNX} (m)	θ_{SQX} (m)	θ_{CSX} (m)	θ_{SMK} (m)	θ_{SNX} (m)	θ_{SQX} (m)	θ_{CSX} (m)	θ_{SMK} (m)
S1	0.209	0.190	0.159	0.193	0.182	0.151	0.149	0.165
S2	0.404	0.343	0.281	0.357	0.284	0.259	0.211	0.264
S3	0.135	0.115	0.094	0.120	0.142	0.101	0.120	0.125
S4	0.165	0.140	0.114	0.146	0.126	0.109	0.107	0.115
S5	0.238	0.196	0.162	0.207	0.164	0.151	0.135	0.152
S6	0.234	0.207	0.168	0.213	0.300	0.271	0.218	0.277
S7	0.136	0.117	0.107	0.124	0.369	0.336	0.269	0.342
S8	0.097	0.087	0.071	0.088	0.264	0.216	0.225	0.241
Moyenne (m)	0.202	0.174	0.144	0.181	0.229	0.199	0.179	0.210
Écart-type (m)	0.096	0.081	0.066	0.084	0.09	0.08	0.06	0.08
CV	0.47	0.46	0.45	0.47	0.38	0.42	0.33	0.39

Pour les deux formations étudiées, on présente l'ensemble des autocorrélogrammes empiriques calculés ainsi que l'autocorrélogramme moyen dans la figure (Tab.16.15). On montre aussi l'échelle de fluctuation verticale déduite à partir du modèle exponentiel ajusté à l'autocorrélogramme moyen. On peut vérifier que les distances de corrélation ainsi calculées, de

3. L'effet de trou se manifeste lorsque la fonction d'autocorrélation (ou variogramme) n'est pas monotone mais présente une ou plusieurs oscillations.

respectivement 0.18 et 0.2 m pour la première et deuxième formation, sont très similaires aux valeurs moyennes déduites de la procédure antérieure.

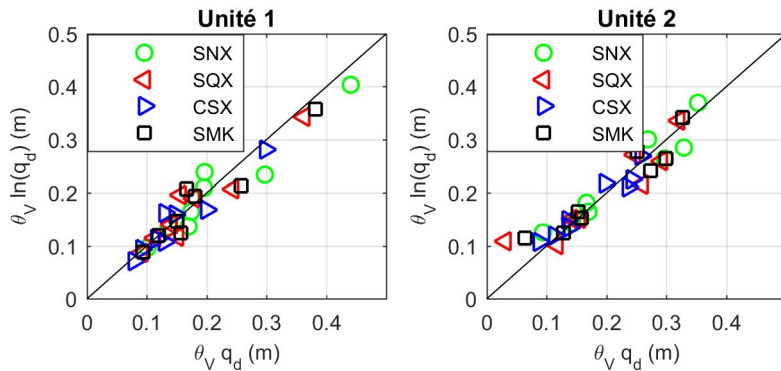


FIGURE 16.14 – Comparaison entre les échelles de fluctuation verticales estimées pour la résistance de pointe q_d en abscisse et pour la transformée logarithmique $\ln(q_d)$.

On s'intéresse maintenant à l'influence de la transformation logarithmique vis-à-vis des distances de corrélations estimées à l'aide des différents modèles théoriques. On compare alors les longueurs de corrélation estimées pour $\ln(q_d)$ avec celles calculées pour q_d , après la soustraction d'une dérive linéaire ajustée pour chaque essai. Les résultats de cette analyse comparative sont donnés dans la figure (Fig.16.14) pour chacune des unités considérées. Il devient clair que la transformation des données n'a pas d'influence significative sur la valeur des échelles de fluctuation verticales calculées. Ceci met en évidence la robustesse du modèle retenu.

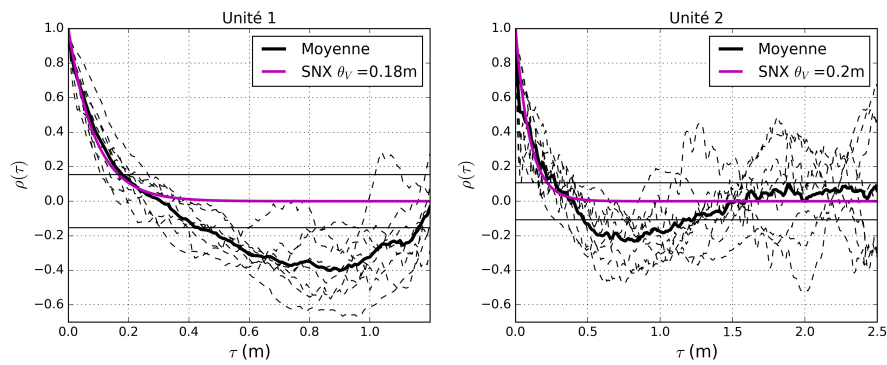


FIGURE 16.15 – Autocorrélogrammes moyens et ajustement du modèle exponentiel

Autocorrélation horizontale des résidus

La variabilité dans le plan horizontal est étudiée à travers la fonction d'autocorrélation pour les deux formations lithologiques retenues. Le fait de compter sur 8 sondages au pénétromètre représente un nombre relativement important par rapport à la dimension du terrain d'étude, et notamment par rapport à la pratique géotechnique courante. Il s'agit toutefois, d'un échantillon de données de taille restreinte vis-à-vis du calcul de la loi spatiale.

Les autocorrélogrammes empiriques calculés ainsi que les échelles de fluctuation estimées avec un modèle de corrélation exponentiel sont données dans la figure (Fig.16.16). Les deux

distances d'autocorrélation estimées sont très similaires, avec un ordre de grandeur décimétrique, de respectivement 13 et 11 m pour la formation sableuse et argileuse.

Cependant, il faut rester prudents sur la signification des distances de corrélation horizontale estimées. Pour la campagne menée, l'écart le plus faible entre sondages est de environ 8 mètres et on ne dispose que de deux couples de sondages (S3 et S7, S2 et S8). Ceci empêche l'aboutissement d'une description pertinente du comportement à l'origine de la fonction d'autocorrélation spatiale, pourtant très important pour calculer correctement le modèle d'autocorrélation.

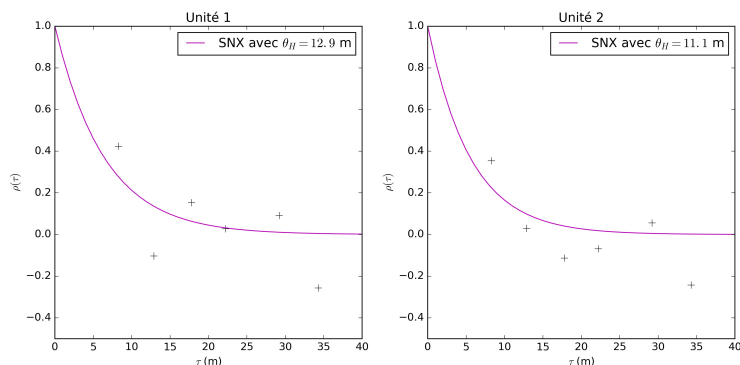


FIGURE 16.16 – Estimation de l'échelle de fluctuation horizontale du logarithme de la résistance de pointe pour les deux unités de terrain considérés

Bilan

L'analyse structurale des profils de $\ln(q_d)$ a été présentée pour le modèle de découpage M2, qui présente un lien direct avec la géologie du site. L'objectif était d'évaluer la pertinence du choix d'un modèle global et unique, représentatif de chacune des unités de terrains considérées. Il est constitué par un modèle de dérive linéaire avec la profondeur et de l'ensemble des résidus dont la corrélation spatiale peut être caractérisée à l'aide d'un modèle exponentiel.

Les analyses présentées ici ont permis de vérifier la validité du modèle. Les résidus définis par une tendance linéaire semblent pouvoir être modélisés correctement par une loi normale, bien que l'hypothèse de normalité ait été rejetée par le test de *Kolmogorov-Smirnov*. La construction des fonctions d'autocorrélation verticales des résidus a mis en évidence une périodicité spatiale de l'ordre de 20 cm. L'analyse visuelle des autocorrélogrammes empiriques semble indiquer que les résidus sont stationnaires a priori, malgré les effets de trou, plus importants pour la couche sableuse. L'hypothèse de préconiser une fonction de corrélation exponentielle ne semble pas avoir une influence significative vis-à-vis des distances d'autocorrélation estimées. D'ailleurs, l'analyse variographique des données q_d a fourni des valeurs des échelles de fluctuation verticales presque identiques à celles calculées pour $\ln(q_d)$,

Dans la direction horizontale, des échelles de fluctuation de l'ordre décimétrique ont été constatées, avec plus d'incertitude compte tenu du faible nombre de sondages et de leur écartement.

Cependant, on remarque à nouveau que les distances d'autocorrélation calculées ici sont fonction du modèle de dérive choisi.

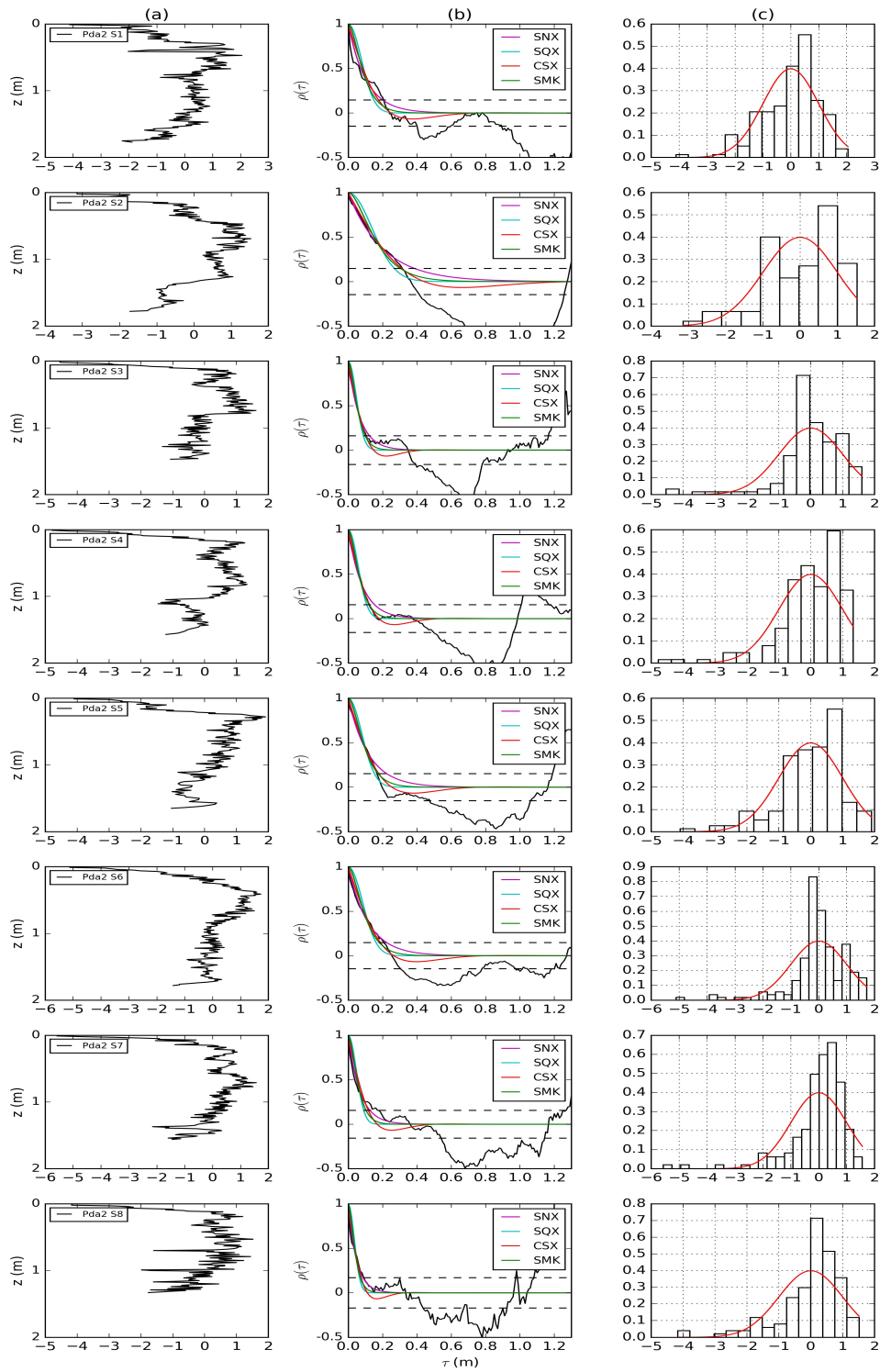


FIGURE 16.17 – Analyse structurale de la formation sableuse : (a) résidus normalisés après le passage au logarithme et soustraction de la dérive linéaire avec la profondeur (b) autocorrélogrammes empiriques et ajustement des modèles théoriques (c) distribution de résidus et ajustement de la loi normale standard

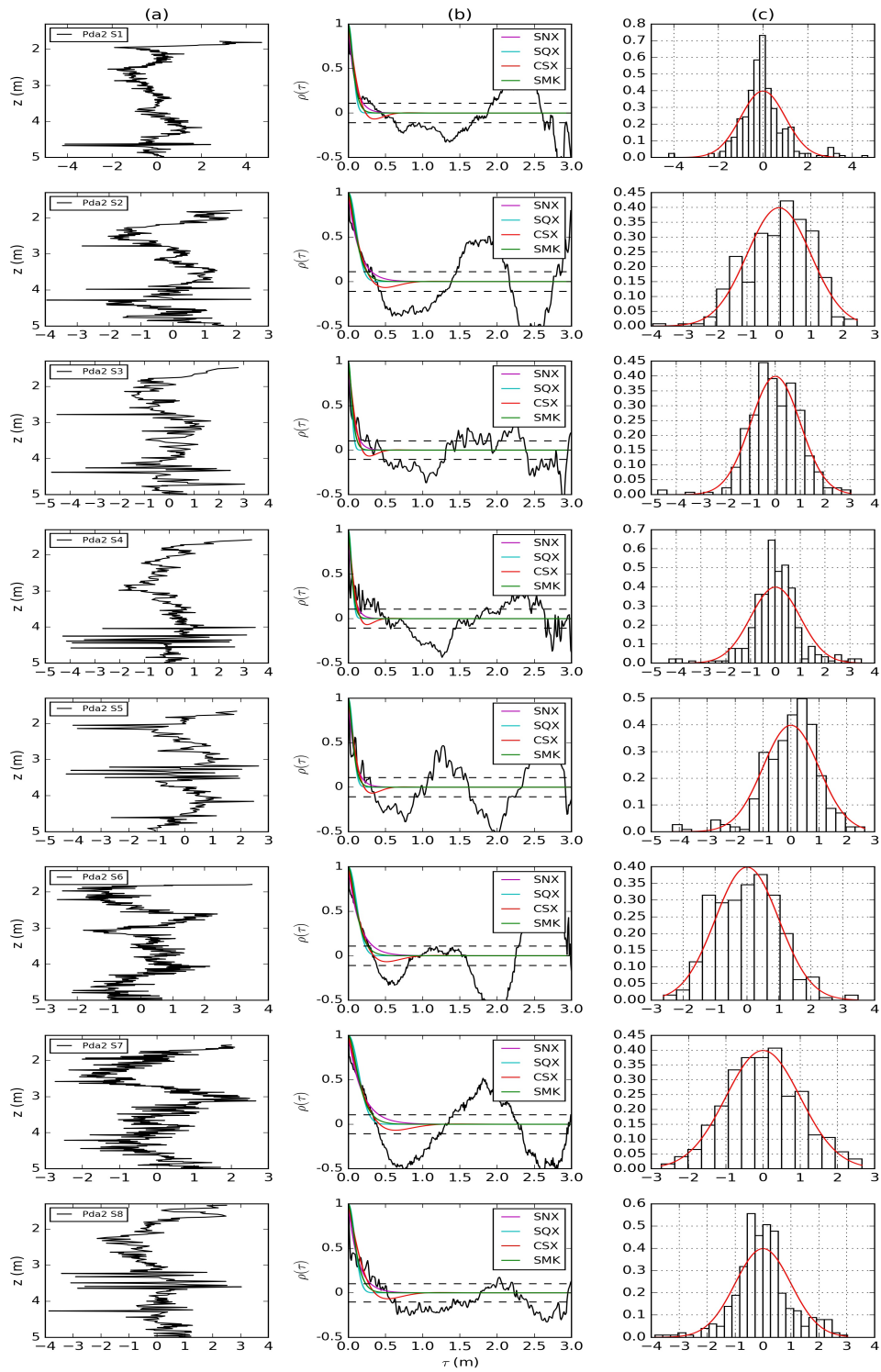


FIGURE 16.18 – Analyse structurale de la formation argileuse : (a) résidus normalisés après le passage au logarithme et soustraction de la dérive linéaire avec la profondeur (b) autocorrélogrammes empiriques et ajustement des modèles théoriques (c) distribution des résidus et ajustement de la loi normale standard

16.2.5 Simulation de q_d conditionnée par les sondages Panda

Mise en œuvre des simulations

Cette partie présente les résultats de la mise en œuvre des simulations conditionnelles afin de fournir un ensemble de réalisations dans l'espace 3D de la résistance de pointe dynamique q_d mesurée à l'aide de l'essai Panda.

Parmi l'ensemble de huit sondages réalisés sur terrain, six ont été retenus pour le calcul des paramètres du champ et le conditionnement par krigeage (Fig.16.19). Les deux essais restants, les sondages S3 et S8, seront utilisés pour valider l'approche proposée. De même, l'influence sur le choix des différents modèles de découpage (M1, M2, M3 et M4) sera évaluée grâce à ces sondages.

Pour la simulation des champs, on a retenu les caractéristiques suivantes :

- échelle de fluctuation horizontale θ_H : compte tenu du faible nombre de sondages disponibles sur le site pour fournir une estimation de l'échelle de fluctuation horizontale, on retient l'hypothèse d'une valeur globale de θ_H sur l'ensemble du site. Pour l'exemple développé, on fixe la valeur de la longueur de corrélation à 13 m dans le plan horizontal isotrope sur la base des résultats précédents.⁴
- nombre total de simulations : 500
- maillage : $\Delta_x = \Delta_y = 1m$ et $\Delta_z = 0.05m$

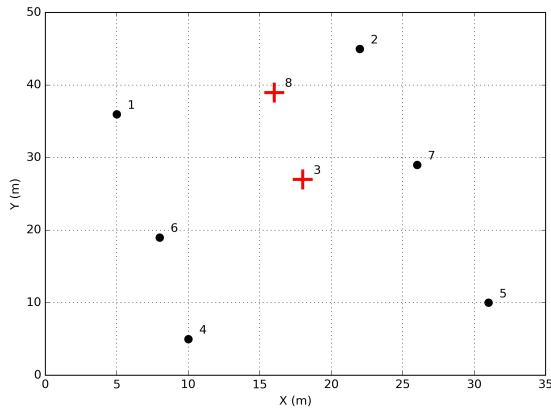


FIGURE 16.19 – Plan de la position relative des sondages. Cercle noir : sondages utilisés pour la simulation du champ. Croix rouge : sondages non pris en compte dans la simulation

Modèle	Unité	$\mu_{\ln(q_d)}$	σ_r^2	$\theta_V(\text{m})$
M1	U1	2.371-0.540z	0.445	0.700
	U2	2.700-0.426z	0.236	0.180
M2	U1	0.439-0.016z	0.160	0.220
	U2	2.169+0.919z	0.208	0.106
	U3	3.576-1.195z	0.083	0.061
M3	U3	0.439-0.016z	0.161	0.216
	U1	2.169+0.919z	0.208	0.106
	U2	3.576-1.195z	0.083	0.061
	U3	1.637-0.563z	0.209	0.158
M4	U4	0.501-0.024z	0.100	0.141

TABLE 16.6 – Paramètres du champ aléatoire selon le modèle de découpage choisi

Les paramètres du champ $\ln(q_d)$ sur chacune des unités des quatre modèles de découpage sont regroupés dans le tableau (Tab.16.6), avec :

- $\mu_{\ln(q_d)}$: moyenne du champ
- σ_{res}^2 : variance des résidus
- θ_V : échelle de fluctuation verticale

4. En toute rigueur chacune des unités peut présenter sa propre structure spatiale indépendante l'une de l'autre. D'ailleurs, la prise en compte d'une longueur de corrélation horizontale différente sur chacune des unités ne poserait pas de problème dans la procédure développée.

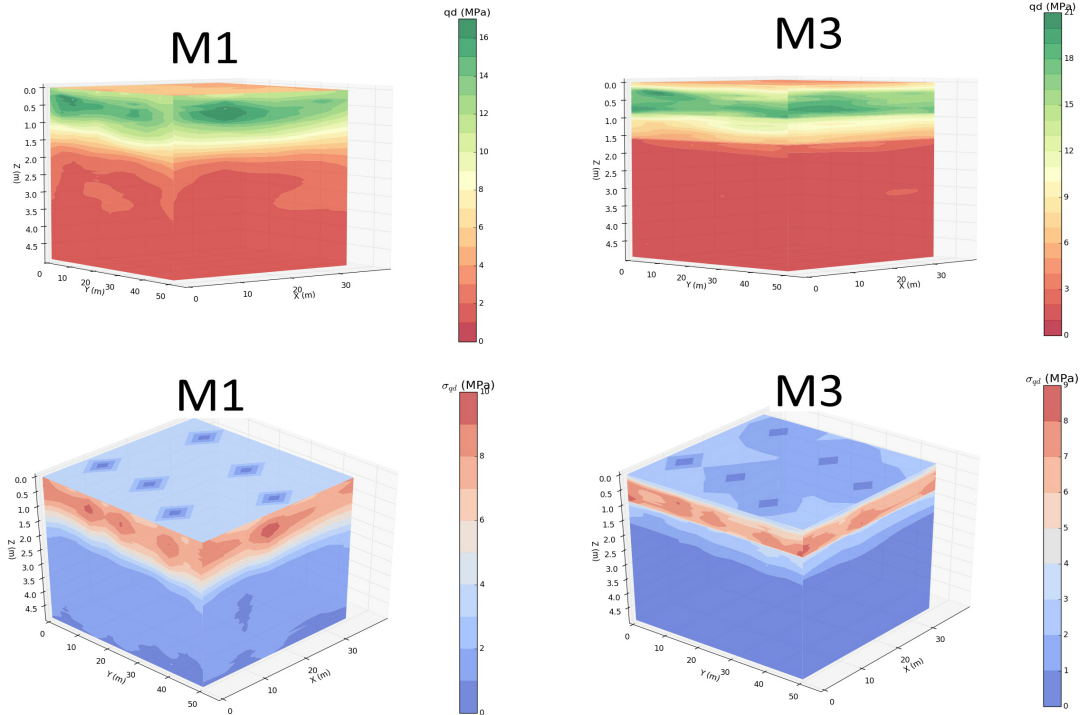


FIGURE 16.20 – En haut moyenne et en bas écart type du modèle 3D suite à 500 tirages

on signale que les deux sondages, S3 et S8 n'ont pas été pris en compte pour le calcul de ces paramètres.

À titre d'illustration, on présente dans la figure (Fig.16.20) plusieurs vues des modèles 3D obtenus quand aucune unité est considéré (modèle M1) et pour un modèle à 3 unités (M3). En haut, avec un code couleur variant du rouge (la plus faible résistance) au vert (la plus forte résistance) on montre l'estimation moyenne sur 500 tirages réalisés. De manière général, on observe des résultats similaires, un premier horizon jusqu'à 1 mètre de profondeur avec des résistances au delà de 10 MPa, ensuite une diminution graduelle de la valeur de résistance. Au dessous de 1.5 m de profondeur prédomine la couleur rouge indiquant des faibles valeurs de résistance. Cependant, pour le modèle M3 on remarque des valeurs de résistances plus élevées par rapport au modèle M1, et on peut distinguer visuellement les interfaces entre les trois « couches ». Dans la partie basse, on a représenté l'incertitude du modèle, sous la forme d'écart-type de krigeage du q_d calculé après les 500 simulations réalisées. Sur le plan horizontal, on peut distinguer clairement les 6 sondages utilisés pour le conditionnement par krigeage, où l'écart type du modèle doit être nulle en l'absence d'effet de pépité.

Ensuite, on peut visualiser dans la figures (Fig.16.21) une comparaison qualitative entre deux sondages virtuels simulés à la position des sondages de validation S3 et S8, et les sondages réels. La bonne correspondance tient autant du modèle que des sondages conditionnants. Le nombre de 6 sondages semble suffisant, dans ce cas, pour pouvoir proposer des versions plausibles d'un sondage non réalisé. On peut expliquer cette bonne correspondance entre les simulations et la réalité par :

- l'homogénéité du site qui permet d'obtenir des estimations représentatives des deux premiers moments du champ grâce aux données expérimentales, abondantes d'ailleurs

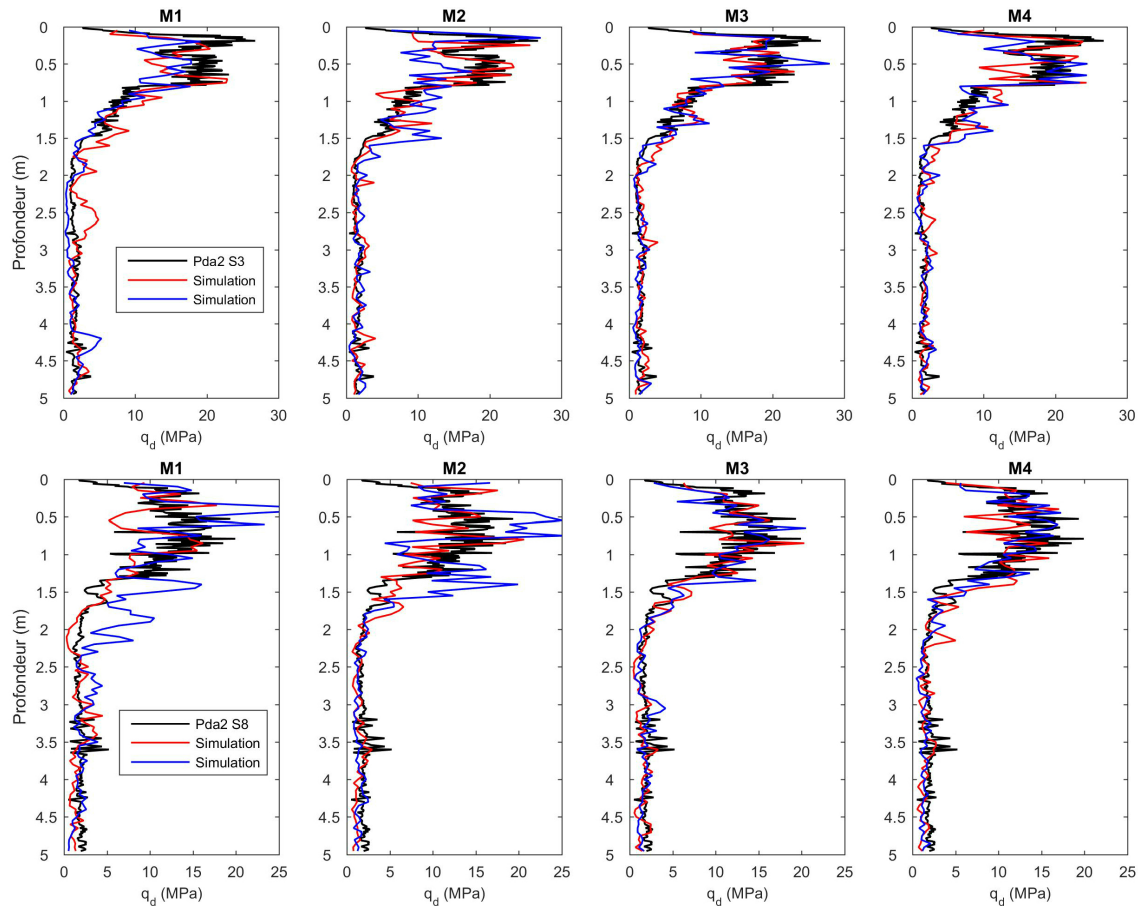


FIGURE 16.21 – Comparaison entre les pénétrogrammes mesurés et ceux simulés pour différents modèles de découpage

pour chacun des essais Panda réalisés ;

- le fait de prendre en compte plusieurs unités qui permet un meilleur calage sur les données expérimentales de la loi théorique et de s'affranchir au mieux de l'hypothèse de stationnarité au sens large. C'est pourquoi, on observe une meilleure similarité des sondages simulés avec les sondages réels pour les modèles M2, M3 ou M4 par rapport au modèle M1 ;
- la bonne correspondance avec les sondages conditionnants. Les simulations issues du modèle M1 représentent pour cela un exemple clair.

Influence du modèle de découpage

Afin d'évaluer l'influence des différents modèles géométriques d'une manière quantitative, on propose de calculer la racine carrée de l'erreur quadratique moyenne (RMSE) entre les pénétrogrammes S3 et S8 et les pénétrogrammes simulés suite à chaque tirage :

$$RMSE(X) = \sqrt{\frac{1}{N} \sum_{i=1}^N [q_{d_i}(X) - \hat{q}_{d_i}(X)]^2} \quad (16.6)$$

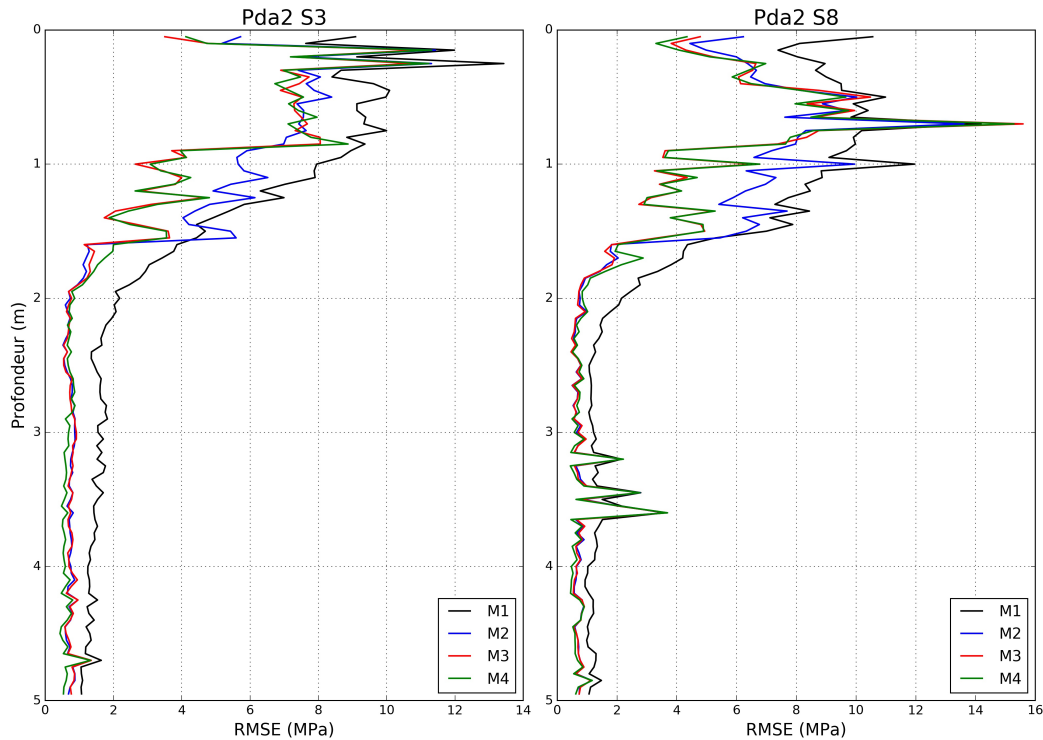


FIGURE 16.22 – RMSE calculée selon la profondeur pour les pénétrogrammes S3 et S8 après 500 simulations

avec q_{d_i} la valeur de résistance de pointe dynamique mesurée au point X , \hat{q}_{d_i} la valeur simulée en X et N le nombre de simulations réalisées.

Les erreurs calculées tous les 5 cm (pas d'échantillonnage vertical des simulations) selon la profondeur sont données dans la figure (Fig.16.22). Le fait de diviser le modèle en plusieurs couches entraîne clairement une réduction de l'erreur. Pour une profondeur inférieure à 1.5 m, le fait de considérer plus de deux couches n'entraîne pas de diminution de l'erreur apparente. On remarque juste les fortes valeurs RMSE entre 3 et 4 m indépendamment du modèle considéré, qui correspondent aux erreurs de mesures qui ne sont pas reproductibles par les simulations. En revanche, on observe une diminution de l'erreur entre 1 et 1.5 m de profondeur lorsque 3 unités sont retenues.

Le calcul de l'erreur relative obtenue pour chaque modèle par rapport au modèle M1, où une seule unité est prise en compte, peut permettre de mieux apprécier l'avantage de découper le terrain en plusieurs unités :

$$\varepsilon(\%) = \left(1 - \frac{\varepsilon_i}{\varepsilon_1}\right) \cdot 100\% \quad (16.7)$$

avec ε_i l'erreur RMSE moyenne du modèle M_i .

Les résultats sont donnés sous la forme de tableau et graphique dans (Fig et Tab.16.7). Des conclusions similaires sont déduites pour les deux sondages. Le fait de considérer deux couches (modèle M2), modélisant séparément la formation sableuse et argileuse entraîne presque 30% de réduction de l'erreur. De même, en ajoutant une troisième unité depuis la surface jusqu'à 1 mètre de profondeur, le taux de réduction de l'erreur approche les 40%.

Modèle	Sondages	
	Pda 2 S3	Pda 2 S8
M2	29.3%	26.4%
M3	38.2%	36.4%
M4	38.5%	36.9%

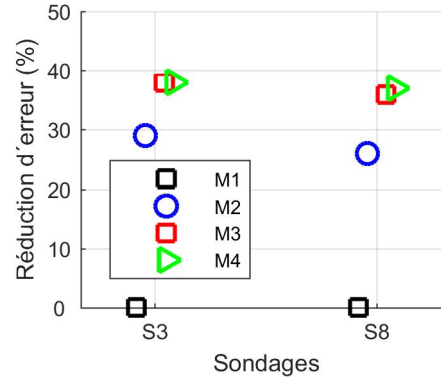


TABLE 16.7 – Réduction de l'erreur (%) par rapport à ne considérer qu'une seule couche homogène

Enfin, le modèle M4 ne semble pas présenter d'avantages par rapport au modèle M3. On peut alors conclure que le modèle M3 s'avère un choix optimal compte tenu des hypothèses réalisées pour aboutir à une meilleure caractérisation du site.

16.2.6 Post-conditionnement du champ par krigeage. Réduction de l'incertitude et application au plan d'échantillonnage

Cadre théorique

Le krigeage est un estimateur linéaire et non biaisé qui assure la variance d'estimation minimale. Le fait de conditionner un champ aléatoire par le biais de cette technique permet d'aboutir à un champ dont la variance est réduite par rapport au cas non conditionnel. L'incertitude est ainsi réduite. Cette réduction peut être calculée théoriquement à partir du rapport entre la variance du krigeage et la variance du champ.

La variance du krigeage, qu'on notera par σ_{KO}^2 vis-à-vis du krigeage ordinaire, n'est pas constante. La variance de krigeage est la variance de l'erreur d'échantillonnage dans l'espace ou domaine d'étude \mathcal{D} . Par définition (contrainte de non-biais), la variance ponctuelle de krigeage sera nulle aux positions où les sondages conditionnants ont été implantés et elle augmente au fur et à mesure qu'on s'en éloigne. La valeur maximale de la variance du krigeage ordinaire peut être déduite grâce à l'équation (éq.15.24). Elle est égale à la valeur de la covariance à l'origine (la variance théorique du champ) plus la valeur du multiplicateur de Lagrange comme suit :

$$\sigma_{KO_{max}}^2 = C(0) + v_{KO} = \sigma^2 + v_{KO} \quad (16.8)$$

où la valeur du multiplicateur de Lagrange v_{KO} est aussi fonction de l'échantillonnage spatial. Mieux les observations sont placées géométriquement, plus nombreuses sont elles, et plus faible est la valeur de v_{KO} . Pour quantifier l'incertitude totale u , du champ conditionné sur le domaine d'étude complet \mathcal{D} , on peut approcher le rapport de la variance du krigeage σ_{KO}^2 et la variance du champ σ^2 par le biais de l'intégrale suivante :

$$u = \frac{\sigma_{KO}^2}{\sigma^2} \cong \frac{1}{\sigma^2 \mathcal{D}} \int_{\mathcal{D}} \sigma_{KO}^2(X) dX \quad (16.9)$$

La réduction de l'incertitude du champ peut être exprimée en pourcentage grâce à l'indice I défini ci-dessous :

$$I = (1 - u) \cdot 100\% \quad (16.10)$$

Sous l'hypothèse d'un champ gaussien stationnaire au sens large et isotrope selon le plan horizontal, cette réduction de l'incertitude (fuseau des simulations) ne dépend que des trois facteurs :

- i. le nombre de points (sondages) conditionnants utilisés dans le système de krigeage ;
- ii. l'échelle de fluctuation horizontale θ_H et le modèle théorique de variogramme ou covariogramme choisi ;
- iii. le plan d'implantation des sondages.

Cas d'étude pratique sur le site de Castello

A partir du calcul de l'indice I (éq.16.10) on tente de répondre aux questions suivantes :

- (A) quel est l'apport du conditionnement avec les 8 sondages disponibles sur site par rapport aux 6 utilisés précédemment ?
- (B) quelle est la configuration des essais permettant de minimiser l'erreur d'échantillonnage ?
- (C) quelle est la réduction de l'incertitude du modèle qu'on peut atteindre en fonction à l'implantation d'un nouveau sondage ?

Pour cette étude théorique, on a aussi retenu l'hypothèse d'un modèle de corrélation exponentiel.

(A) Réduction de l'incertitude en fonction du nombre de sondages et θ_H

L'évolution de la réduction de l'incertitude du modèle en fonction de la valeur de la distance de corrélation horizontale a été calculée pour le conditionnement avec les 6 essais utilisés dans le paragraphe précédent (§16.2.5). Le même calcul a été réalisé lorsque les sondages n°3 et 8 sont aussi pris en compte et le conditionnement est réalisé avec les 8 sondages. Les résultats obtenus sont présentés dans la figure (Fig.16.23). Pour une valeur de θ_H de l'ordre de 10 m, une réduction de l'incertitude de respectivement 12.5% et 15.8% serait atteinte pour 6 et 8 sondages. Quand θ_H est égale à 30 m, l'incertitude est réduite de presque 50% pour les deux cas. On peut aussi vérifier que lorsque on suppose une faible valeur pour la distance de corrélation horizontale ($\theta_H \leq 5m$), la réduction de l'erreur est inférieure à 5% et le fait de conditionner devient donc moins intéressant. Dans la figure à droite (Fig.16.24), on a représenté la différence de la réduction d'incertitude calculée avec 8 sondages par rapport à 6 ($\Delta I\%$). L'apport de ces deux sondages supplémentaires est relativement faible. Sa valeur maximale est atteinte pour une valeur de θ_H comprise entre 10 et 20 m.

Le tableau qui regroupe les résultats numériques issus de ces analyses est donné en annexe (Annexe C.5). Enfin, il est difficile d'isoler le poids relatif de θ_H vis-à-vis des analyses présentées ici. Les calculs réalisés sont influencés non seulement par la taille du domaine d'étude, mais aussi par l'espacement entre sondages.

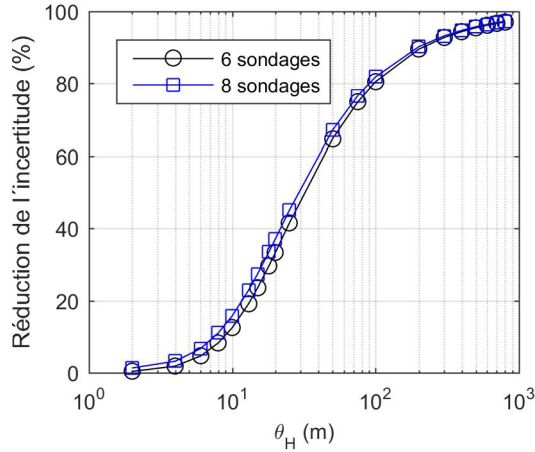


FIGURE 16.23 – Réduction de l'incertitude en fonction de θ_H

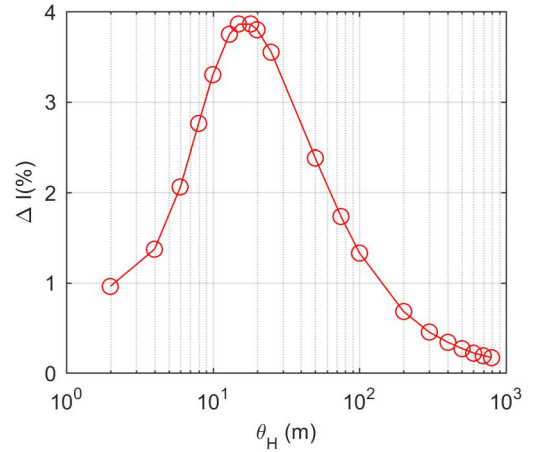


FIGURE 16.24 – Apport du conditionnement par les 8 sondages

(B) Plan d'échantillonnage optimale

L'influence du plan d'implantation des sondages est étudié vis-à-vis de la réduction de l'incertitude atteinte grâce au krigeage. Le but final sera de chercher la meilleure configuration possible pour les 6 sondages considérés suivant le schéma d'implantation conçu initialement : deux profils de trois sondages orientés selon l'axe Y. Sous l'hypothèse d'isotropie dans le plan horizontal, on peut étudier l'écartement optimal entre sondages selon l'axe X et Y séparément.

Espacement de deux sondages selon l'axe X

L'objectif est de calculer quelle est l'espacement Δx optimal entre deux sondages centrés le long d'un profil dont l'extension ou longueur L est égale à 40 m. La réduction de l'incertitude I en fonction de cet écartement sur l'ensemble du profil de 40 m a été calculée pour plusieurs valeurs θ_H . Les résultats sont présentés dans la figure (Fig.16.25). Elle est scindée en deux graphiques.

Celui de gauche représente les courbes de gain en réduction d'incertitude en fonction de l'espacement des sondages. Il apparaît clairement que l'incertitude du modèle décroît avec l'importance de la longueur de corrélation.

Dans le deuxième graphique, on a tenté de rendre le problème adimensionnel. Pour cela, on a normalisé l'écart Δx et l'extension du domaine L par θ_H . En rouge on a représenté l'union des maxima locaux des différentes courbes, qui représentent la plus grande réduction d'incertitude qu'il est possible d'atteindre pour une valeur θ_H quelconque. On peut noter que l'écart optimal entre sondages ne dépend pas de la distance de corrélation du champ. En effet, $\Delta x/\theta_H$ et L/θ_H présentent une relation linéaire dont la constante de proportionnalité est 2. Par conséquent, le conditionnement par krigeage sera optimal lorsque les deux sondages sont écartés d'une distance égale à la moitié de l'extension du domaine L , dans notre cas 20 m :

$$\Delta x_{opt} = L/2 \quad (16.11)$$

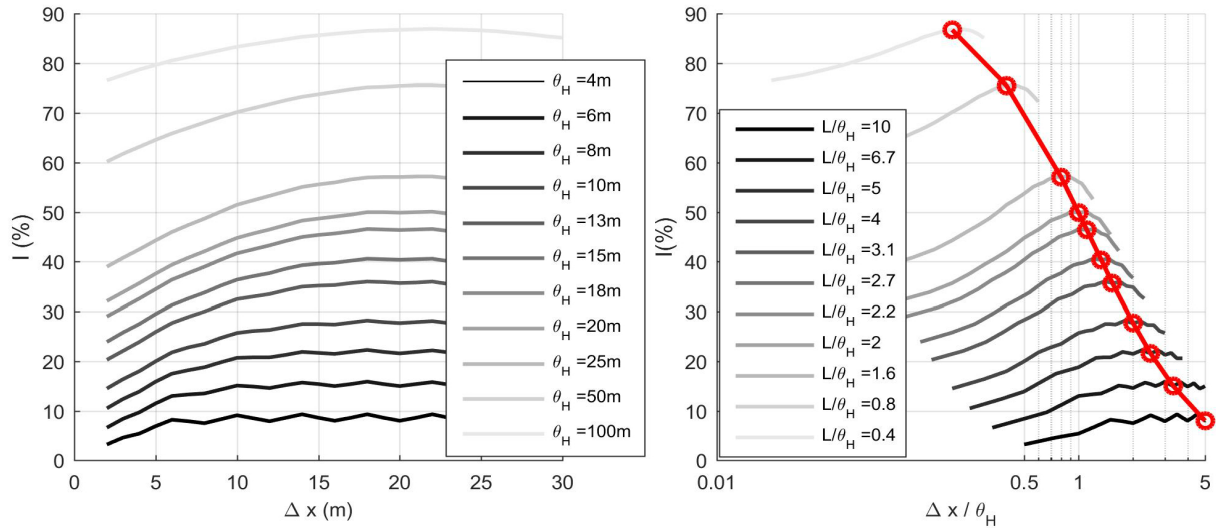


FIGURE 16.25 – Influence de la distance de corrélation θ_H et de l'écart Δx sur l'implantation optimale de deux sondages le long d'un profil d'extension L

Espacement de trois sondages selon l'axe Y

Dans ce cas, on cherche à implanter trois sondages repartis selon l'axe Y. Un sondage est positionné au centre du profil d'étude. On écartera les autres deux sondages une distance Δy par rapport à celui du centre. On a répété des analyses analogues à celle présentée auparavant et les résultats sont donnés dans la figure (Fig.16.26).

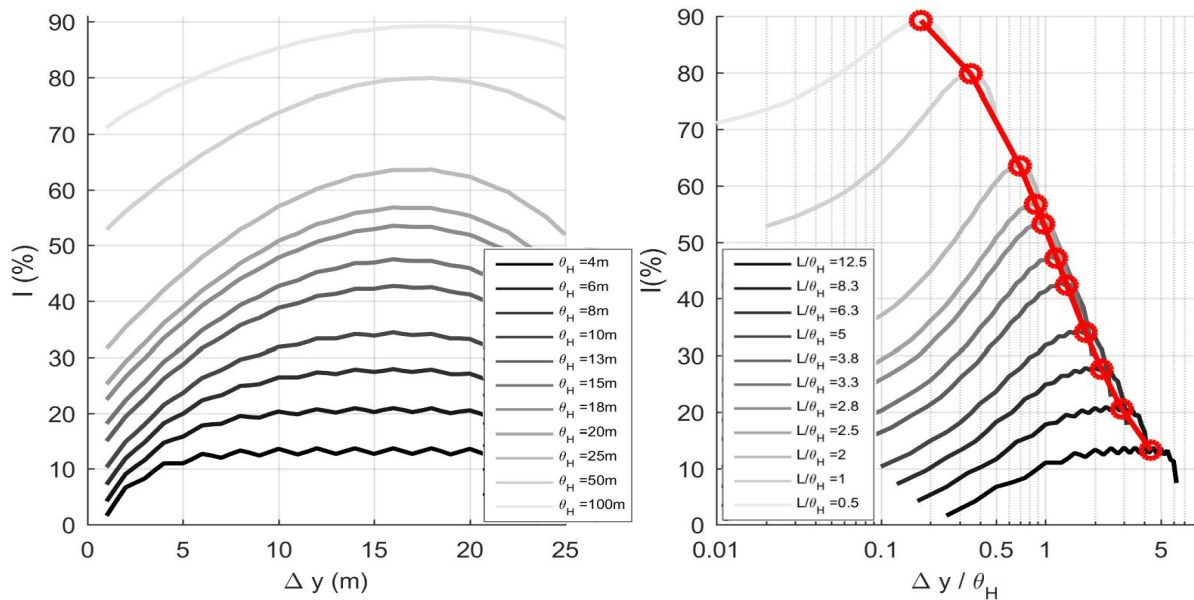


FIGURE 16.26 – Influence de la distance de corrélation θ_H et de l'écart Δy sur l'implantation optimale de trois sondages le long d'un profil d'extension L

On constate à nouveau que la valeur de l'échelle de fluctuation ne règle que la valeur du pourcentage total de l'incertitude du champ. D'ailleurs, quand Δy et L sont normalisés par la distance de corrélation, ils présentent une relation linéaire. Comme dans le cas précédent, la

séparation optimale peut être déduite en fonction de la longueur du profil L par l'expression suivante :

$$\Delta y_{opt} = 0.35 \cdot L \quad (16.12)$$

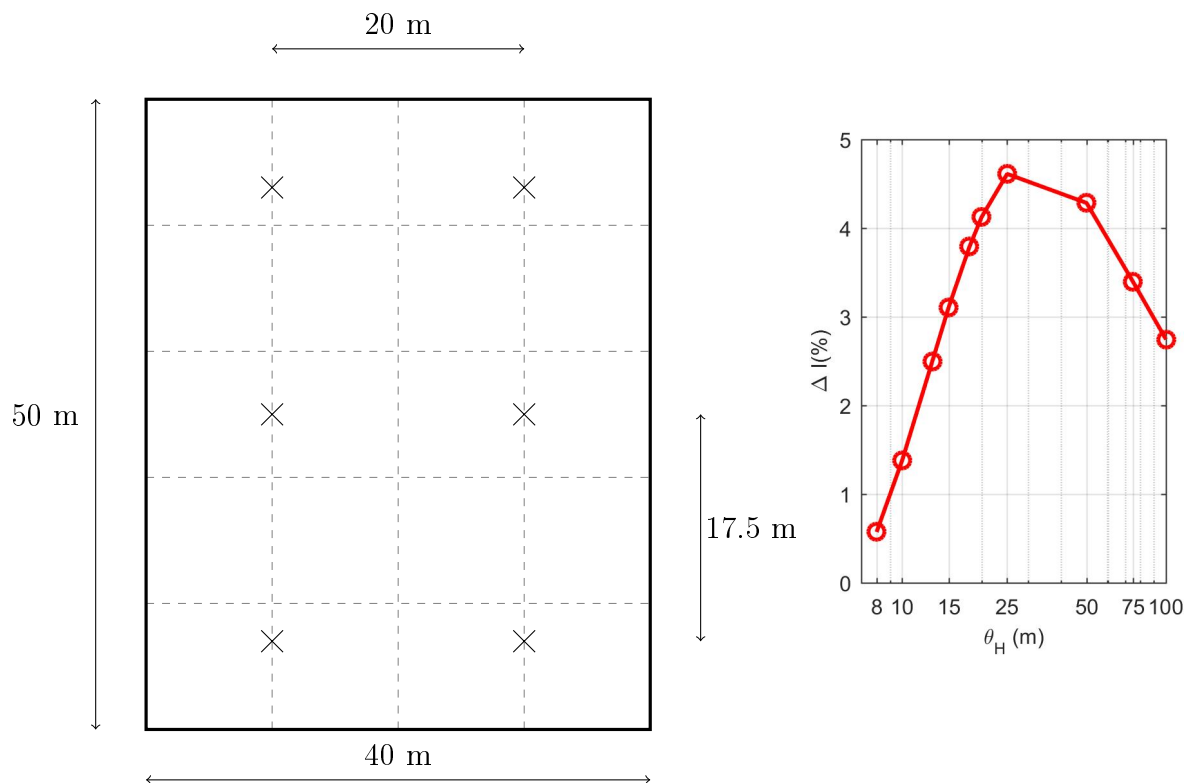


FIGURE 16.27 – À gauche, plan d'implantation optimal des 6 sondages. À droite, gain du nouveau schéma par rapport à l'implantation originale en fonction de la distance de corrélation

Plan d'implantation optimal

Sur la base des analyses antérieures, la configuration optimale pour implanter les 6 sondages est donnée dans le schéma à gauche de la figure (Fig.16.27). On calcule la différence entre la réduction d'incertitude atteinte avec cette nouvelle configuration de sondages par rapport au plan d'implantation original. L'apport du nouveau schéma d'implantation en fonction de l'échelle de fluctuation est représenté dans le graphique à droite de la figure (Fig.16.27). Cet apport est maximal quand la distance de corrélation est égale à 25 m. C'est apport est d'environ 4.5% en terme de réduction d'incertitude, une valeur supérieure à celle obtenue en prenant en compte les 8 sondages effectués sur le terrain.

Apport d'un nouveau sondage

Quand on recherche la position optimale d'un nouveau sondage, la méthode dite du « point fictif » (Delhomme (1976) cité par Bacconnet (1991)) permet d'associer fictivement un sondage à ceux existants de manière à calculer la variance d'estimation correspondante. Ce même principe peut être utilisé pour cartographier une fonction de gain en termes de réduction supplémentaire de l'incertitude du modèle en fonction de la position du nouvel essai.

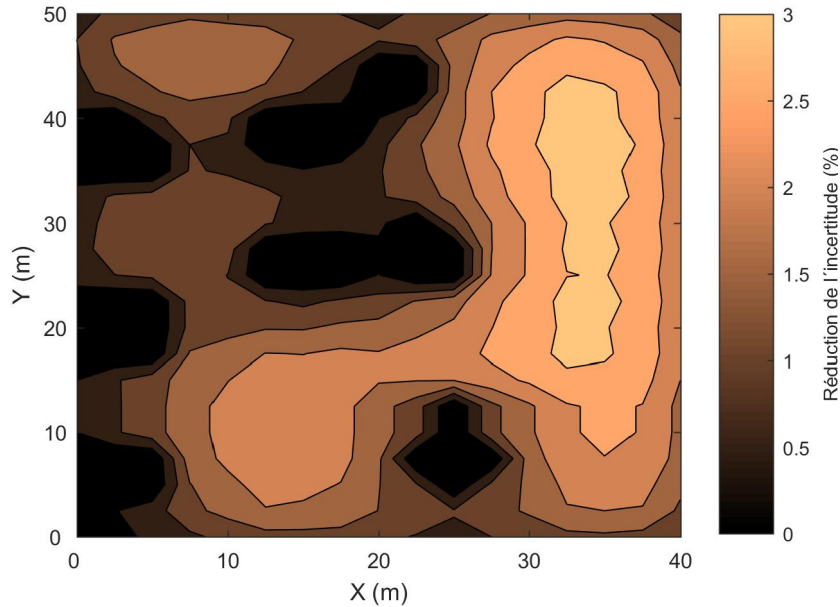


FIGURE 16.28 – Apport à la réduction de l’incertitude en fonction de la position d’implantation d’un nouveau sondage pour $\theta_H = 13m$.

Cette carte est donnée dans la figure (Fig.16.28), pour une valeur de θ_H de 13 m. Elle est basée sur l’implantation des 8 sondages effectués lors de la campagne géotechnique dans le site d’étude. Il est clair que la zone convenant pour l’implantation d’un nouvel essai serait celle placée dans l’axe de coordonnée $X = 35$ m, qu’on aurait pu choisir par intuition. Cependant, l’intérêt de cette carte est la possibilité d’évaluer le gain de réduction de l’incertitude associée, qui dans ce cas est égal à 3%.

Bilan

Dans cette section une étude théorique sur le conditionnement par krigeage ordinaire du champ aléatoire avec un modèle de corrélation exponentiel a été présenté. Le fait de conditionner le champ grâce aux sondages disponibles permet de réduire la variance du champ et donc l’incertitude du modèle. L’influence de la distance de corrélation horizontale semble être beaucoup plus importante vis-à-vis de la réduction de l’erreur.

16.2.7 Conclusion

La démarche proposée pour bâtir un modèle probabiliste du terrain à faible profondeur (5m) en 3D a été appliquée sur un site expérimental deltaïque. Des différentes étapes abordées, on peut déduire les conclusions suivantes :

- les techniques statistiques pour définir un découpage vertical grâce à l’analyse des sondages Panda ont montré une bonne répétabilité en termes d’unités et d’interfaces identifiées. Leur intérêt pour l’application à l’échelle du site géotechnique est donc confirmé. Leur automatisation est aussi faisable. Même si les stratigraphies proposées sont cohérentes avec la lithologie du terrain, les données géologiques sont complémentaires aux résultats fournis par ces procédures ;

- une campagne d'essais à l'aide d'un pénétromètre de type Panda présente une structure très particulière. On dispose d'une grande quantité de données le long de la profondeur répartis souvent en quelques points du site d'étude. Malgré le faible nombre de sondages, la taille de l'échantillon de mesures est très importante. Ceci permet une estimation aisée des premiers moments statistiques de la loi de probabilité et une caractérisation en détail des structures verticales de la formation d'étude.
- la méthode proposée par Salgado pour évaluer l'homogénéité horizontale d'un site vis à vis de la similarité entre sondages est intéressante, puisque elle permet de valider d'une manière simple, la représentativité qu'on peut accorder aux statistiques issues de chaque sondage pour la mise en œuvre d'un modèle global de la couche (ou de l'unité étudiée), et ce quelle que soit la définition des couches que l'on retient. De même, le principal intérêt de cette approche est qu'elle peut être appliquée même quand on dispose d'un faible nombre d'essais. Les analyses effectuées montrent qu'elle est applicable à une campagne d'essais Panda et pour analyser les profils de $\ln(q_d)$;
- une étude structurale détaillée des données de résistance au sein des deux principales formations lithologiques du site expérimental étudié a été présentée. Le choix pragmatique d'une tendance linéaire selon la profondeur a été retenu. On a remarqué une bonne répétabilité quant aux distances de corrélation verticales des résidus, indépendamment du modèle de corrélation théorique utilisé. De même, le fait d'analyser les données de résistance en échelle logarithmique ne change pas les valeurs des longueurs de corrélation calculées et l'hypothèse d'un modèle exponentiel semble être cohérente ;
- la variabilité spatiale de la résistance de pointe du sol a été modélisée au moyen d'un champ aléatoire log normal au sein de chacune des unités du modèle de terrain retenu. Les champs ont été conditionnés par les sondages Panda effectués sur site grâce au krigeage ordinaire. La comparaison qualitative et quantitative entre les sondages simulés et effectués a permis de valider la pertinence de l'approche proposée et de statuer sur le modèle de découpage optimal compte tenu des hypothèses faites ;
- le krigeage permet de mesurer, par sa variance, la qualité de l'échantillonnage. Le post-conditionnement par krigeage du champ permet donc de réduire l'incertitude par rapport à un champ non conditionnel. Le rapport de la variance du krigeage et de la variance du champ permet de mesurer l'apport du conditionnement. Dans le cadre d'une campagne de reconnaissance courante, formé par quelques essais, c'est la distance de corrélation horizontale qui gouverne principalement la réduction d'incertitude que l'on peut atteindre. Par ailleurs, l'optimisation de l'implantation des essais peut toujours être abordée dans un cadre théorique grâce aux outils de la géostatistique.

17. Apports supplémentaires

17.1 Introduction

La qualité de la modélisation de la variabilité spatiale horizontale repose largement sur la taille de l'échantillon disponible. Dans cette section, on s'intéresse à l'apport de données supplémentaires qui pourront permettre d'améliorer éventuellement la stratégie de modélisation proposée. Pour cela, on s'appuie sur un second site d'étude constitué d'une parcelle agricole située dans la plaine de la Limagne. La description du site et des campagnes de reconnaissance menées sont données en annexes ([Annexe D.1](#)) et ([Annexe D.1.3](#)). En résumé, on dispose d'une première campagne de reconnaissance constituée de 8 essais Panda comme sur le site de Castello. Ont été ajoutés trois profils de tomographie électrique et une soixantaine d'essais Panda répartis suivant une croix de 30x30 sondages. Répartis régulièrement avec un pas de 0.5 m, cette seconde campagne permet une analyse variographique dans le plan de façon plus rigoureuse. Ainsi on aborde les questions suivantes :

- quel est l'apport d'une croix de sondages par rapport à une campagne de sondages pénétrométriques de taille plus modeste ?
- l'analyse variographique des panneaux électriques réalisés peut-elle apporter des informations pertinentes sur la structure de variabilité horizontale ?
- dans le cas où on dispose d'une information a priori, peut-on améliorer nos estimations sur les distances de corrélation grâce aux procédures bayésiennes ?

17.2 Apport d'une croix de sondages de type géostatistique

17.2.1 Analyse structurale de la première campagne de 8 essais

Résultats des essais de pénétration

Les résultats issus de la première campagne d'essais sont synthétisés dans la figure ([Fig.17.1](#)) rassemblant les pénétrogrammes bruts (a), l'histogramme des mesures de résistances (b) et celui des enfoncements mesurés (c). L'enfoncement moyen calculé est égal à 4mm, valeur inférieure au pas de régularisation proposé dans ce travail (10mm).

Suite à l'inspection visuelle des pénétrogrammes, on observe une première formation avec des résistances mécaniques inférieures à 5 MPa, à l'exception des deux pics locaux autour de 30 et 60 cm de profondeur. À partir de deux mètres, on constate une augmentation systématique de la résistance et une plus grande dispersion des mesures. Deux unités lithologiques

sont présentes jusqu'à 4 m, une première couche de limons argileux et à partir d'environ 2m une couche d'argile compacte. On signale la bonne correspondance de cette interface lithologique avec les limites identifiées à l'aide des méthodes statistiques par fenêtre glissante. Une synthèse de traitement de deux campagnes de sondages et des résultats calculés est également fournie en annexe ([Annexe D.1.4](#)).

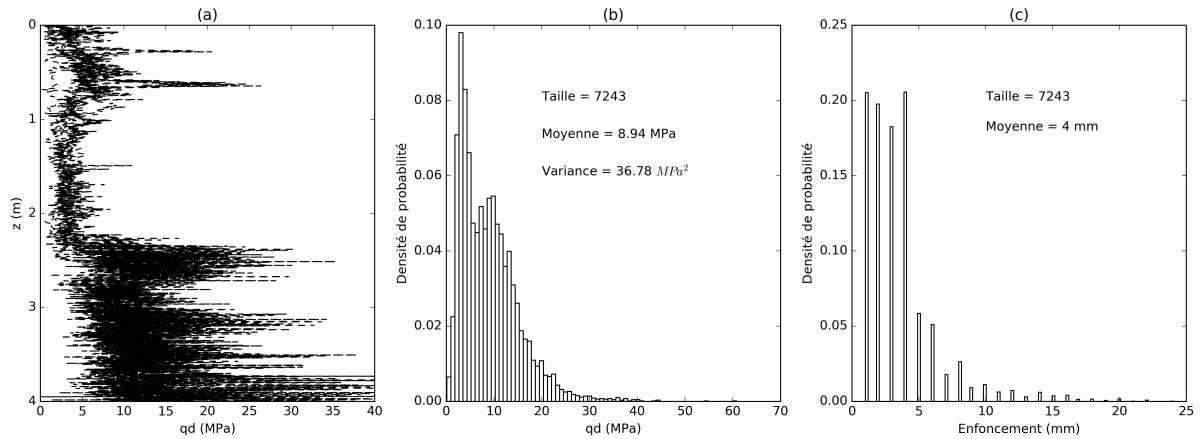


FIGURE 17.1 – Résultats des mesures de la première campagne

Les pénétrogrammes obtenus suite à une régularisation avec un pas de 10mm sont donnés dans la figure ([Fig.17.2](#)), ainsi que les histogrammes des résistances au sein de chacune des formations lithologiques.

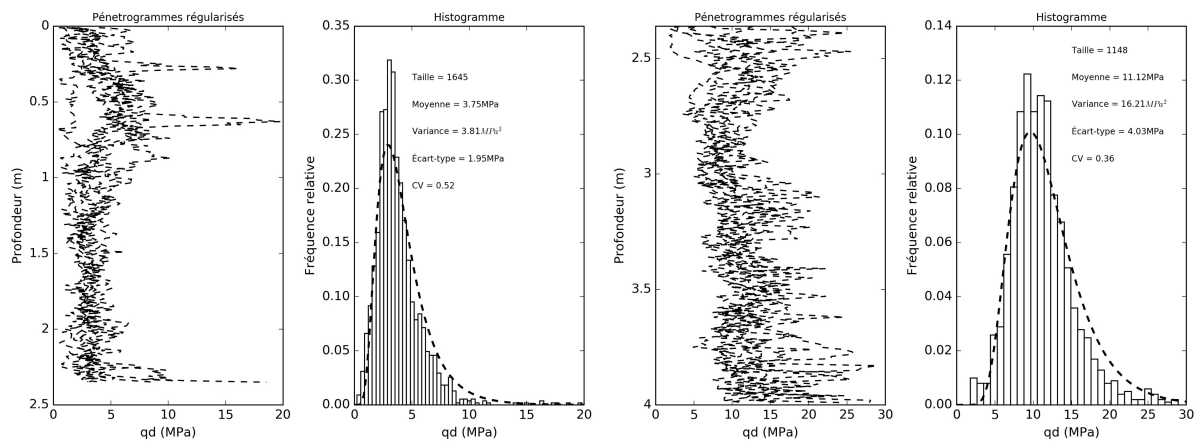


FIGURE 17.2 – Pénétrogrammes régularisés et histogrammes des résistances de pointe sur la couche de limons et argiles

17.2.2 Analyse structurale de la croix de sondages

Variabilité spatiale verticale

On présente ici les résultats de l'analyse de la variabilité spatiale verticale le long de la profondeur pour les 60 essais réalisés dans la croix de sondages. Ces résultats seront comparés

à ceux issus de l'analyse des sondages effectués dans la première campagne. L'étude porte sur chacune des deux formations lithologiques du terrain.

L'analyse comparative de la variabilité spatiale des résistances de pointe le long de la profondeur présentée ici ne concerne que la couche de limons. La profondeur atteinte pour les sondages de la deuxième campagne ($z < 3.4m$) ne permet pas d'ausculter de manière complète la seconde unité. Deux procédures ont été suivies pour cette analyse comparative :

- A) premièrement, on analyse les variogrammes verticaux de q_d . Le modèle de variogramme exponentiel, celui préconisé dans notre travail, sera utilisé pour calculer les distances de corrélations verticales ;
- B) deuxièmement, on étudie à nouveau la structuration verticale de q_d , mais en enlevant une tendance linéaire selon la profondeur. Cette dérive est calculée pour chacun des sondages analysés. Les distances de corrélation verticales sont recalculées à l'aide d'une fonction d'autocorrélation exponentielle ajustée aux résidus. On remarque que cette procédure est analogue à celle retenue dans la démarche de modélisation proposée.

A) Variogrammes verticaux

Les variogrammes verticaux moyens sur la couche de limons sont tracés dans la figure (Fig.17.3) pour les deux campagnes d'essais. Pour la deuxième campagne d'essais, les essais dans l'axe X ont été analysés indépendamment de ceux situés dans l'axe Y. Les trois variogrammes présentent un caractère stationnaire sans effet de pépité. Les deux variogrammes obtenus pour les essais de la deuxième campagne présentent un palier avec une valeur très similaire, alors que les portées présentent un rapport de valeur proche à 2 (0.42 m dans l'axe X et 0.24 m dans l'axe Y). Le variogramme de la première campagne présente une portée similaire à celle de l'axe X de la deuxième campagne. Par contre, le palier atteint une valeur trois fois plus forte. Une explication possible de cette observation peut être lié à l'influence des conditions hydriques différentes entre les deux campagnes altérant les valeurs fortes des premiers 80 cm, on constate par ailleurs une différence significative entre les moyennes et la variance des deux échantillons.

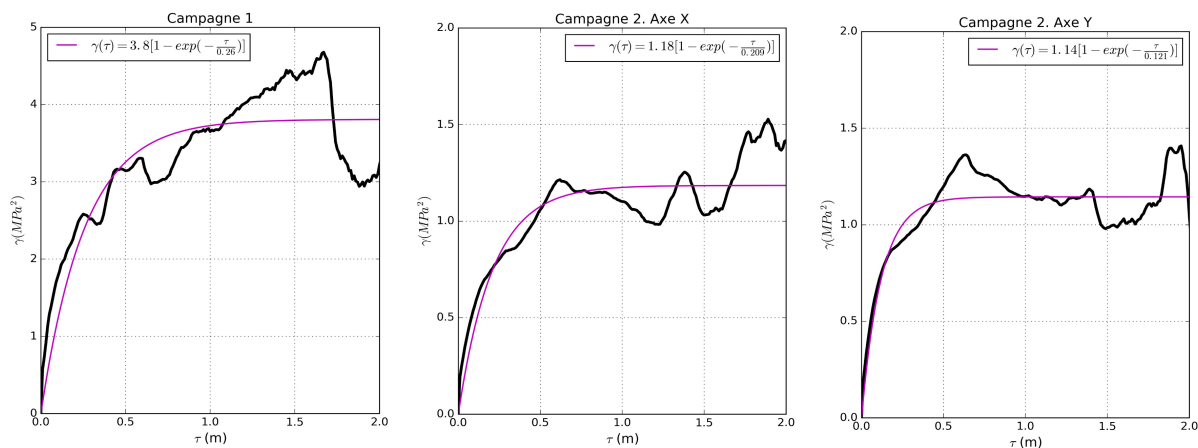


FIGURE 17.3 – Variogrammes verticaux de q_d sur la couche de limons

B) Fonction d'autocorrélation des résidus pour une dérive linéaire

Le fuseau des fonctions d'autocorrélation empiriques et l'autocorrélogramme moyen ont été calculés pour les deux campagnes de sondages, sur les résidus après avoir enlevé une dérive linéaire aux profils de résistances. L'ensemble des résultats concernant la couche de

limons est donné dans la figure (Fig.17.4). On remarque une différence notable en termes de résultats par rapport aux analyses précédentes, dans lesquels la dérive n'avait pas été prise en compte (cf.Fig.17.3). Les résidus semblent être stationnaires. Les distances de corrélation verticales estimés à l'aide du modèle exponentiel pour les trois cas étudiés ont des valeurs plus faibles. D'ailleurs, on constate une répétabilité des valeurs calculées de l'échelle de fluctuation verticale (0.20 m), indépendamment de l'échantillon analysé.

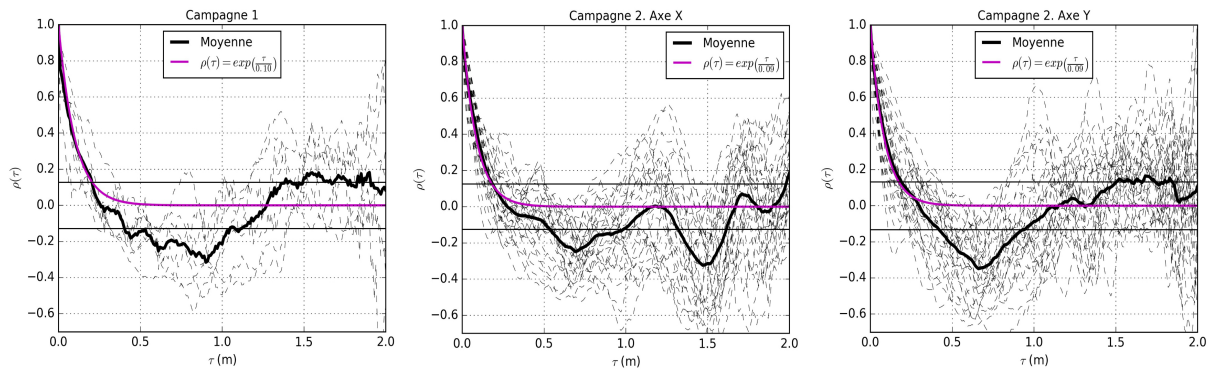


FIGURE 17.4 – Fonctions d'autocorrélation verticales de la couche de limons

Étude de la variabilité spatiale horizontale

Le faible écartement (0.5 m) entre les sondages effectués lors de la deuxième campagne permet de mener une étude variographique dans le plan horizontal. Les résultats suivant l'axe Y sont donnés dans la figure (Fig.17.5).

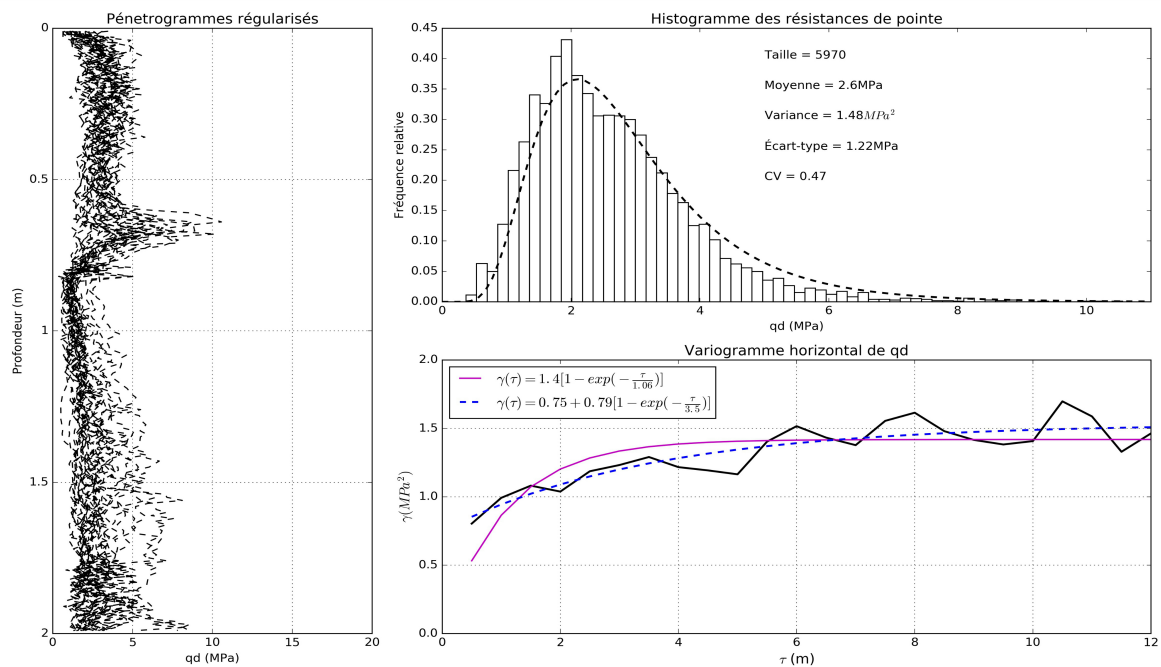


FIGURE 17.5 – Analyse structurale de q_d pour les sondages placés dans l'axe Y

Les statistiques et l'histogramme avec l'ajustement d'une loi lognormale de l'échantillon des mesures de q_d sont aussi fournis sur ce graphique. Le variogramme de la résistance de pointe montre un caractère stationnaire mais aussi un effet de pépite. Deux modèles exponentiels ont été ajustés avec et sans effet de pépite.

Par ailleurs, on présente sous le même format les résultats concernant l'axe X (Fig.17.6). Pour la couche de limon, notée par U1 on peut remarquer une valeur moyenne très similaire à celle déduite des sondages dans l'axe Y, alors que le coefficient de variation est légèrement plus faible. La stationnarité dans le plan horizontal semble donc être pertinente. L'ensemble des points issus du variogramme expérimental montre un effet de pépite notable, de la même manière que l'analyse des mesures selon l'axe Y. Le variogramme exponentiel s'ajuste bien aux données quand l'effet de pépite est pris en compte, avec une valeur de θ_H égale à 7m. L'analyse des mesures au sein de la couche d'argile (U2) est possible suivant l'axe X uniquement. Elle met en évidence la plus forte dispersion de l'échantillon des données bien que le coefficient de variation présente une valeur similaire à celui de la couche de limons. Un modèle gigogne ou additif permettrait de reproduire beaucoup mieux la montée (partie initiale croissante du variogramme). La somme de deux modèles exponentiels avec un effet de pépite de 7.8 MPa^2 a été proposée.

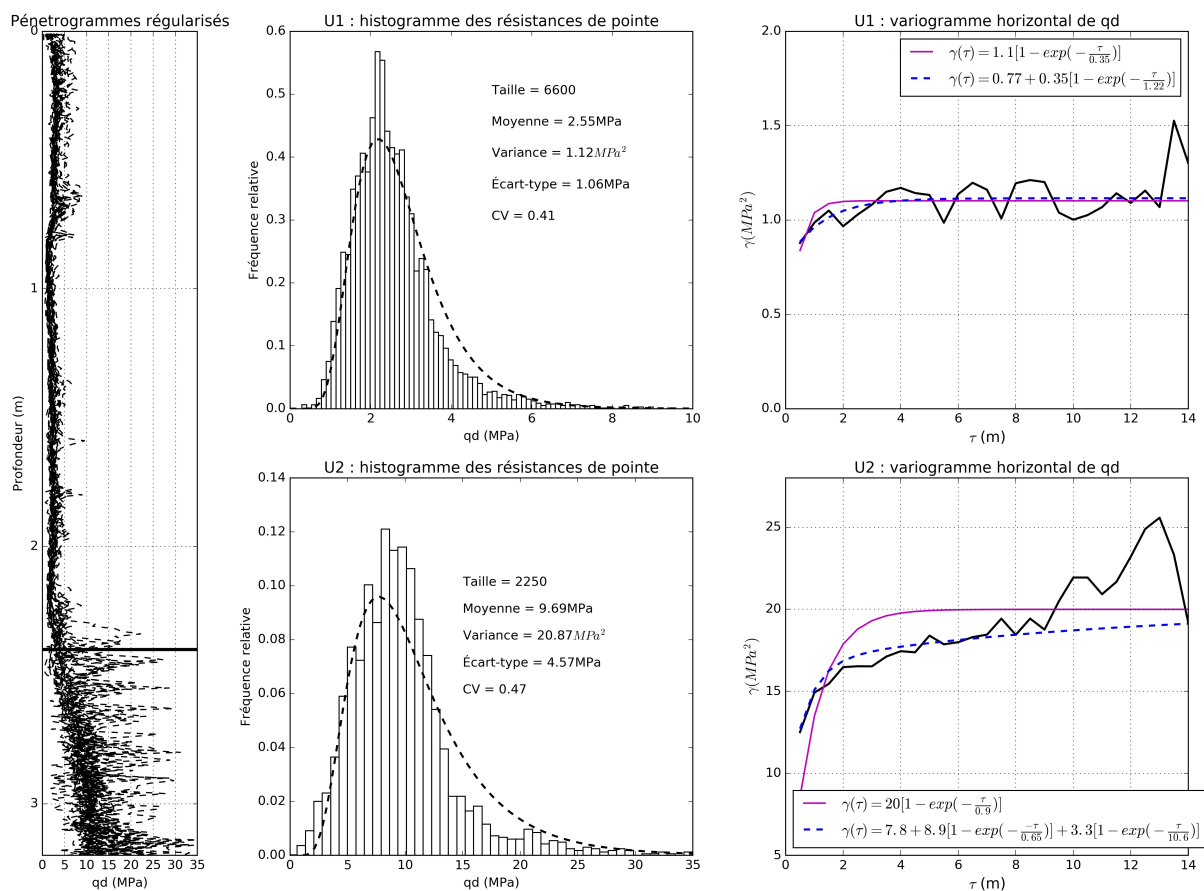


FIGURE 17.6 – Analyse structurale de q_d pour les sondages placés dans l'axe X

Enfin, on a tenté d'estimer les distances de corrélation horizontales à partir des sondages effectués dans la première campagne, malgré l'incertitude qui pèse sur cette analyse compte

tenu du faible nombre de sondages disponibles. Les coefficients de corrélation empiriques ainsi que le modèle exponentiel ajusté sont tracés dans la figure (Fig.17.7). Les échelles de fluctuation horizontales estimées présentent des valeurs très similaires pour les deux unités analysées, 4.5 et 5.8 m, proches du plus petit espacement entre sondages, 5 m.

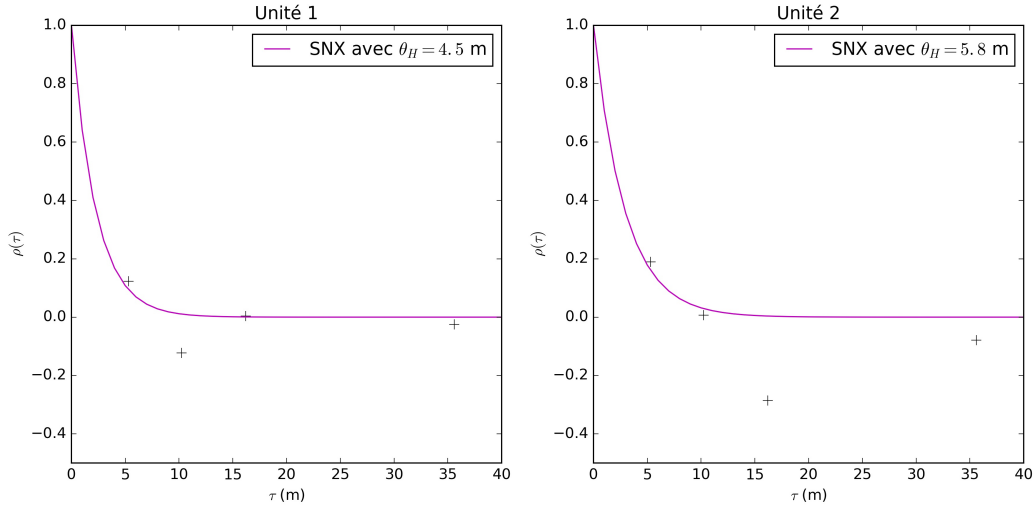


FIGURE 17.7 – Estimation de l'échelle de fluctuation horizontale de q_d pour les deux unités de terrain considérées

17.2.3 Critère d'homogénéité horizontale

Première campagne d'essais

Premièrement, l'indice de variabilité horizontale (HVI) a été calculé pour l'ensemble des sondages de la première campagne d'essais afin d'évaluer l'homogénéité du site selon le plan horizontal. La profondeur d'étude est égale à 4m. On présente les résultats pour deux cas d'étude :

- quand aucune couche ou unité est prise en compte et on cale la dérive linéaire sur la totalité du sondage (modèle M1) ;
- quand on calcule la dérive pour chacune des unités lithologiques repérées (modèle M2).

Les valeurs de l'indice HVI sont récapitulées dans le tableau (Tab.17.1).

TABLE 17.1 – Indices de variabilité horizontale calculés pour la première campagne d'essais. Profondeur d'étude égale à 4 m.

Coeff. A	1	0.8	0.6	0.5	0.4	0.2	0
Coeff. B	0	0.2	0.4	0.5	0.6	0.8	1
M1	13.8%	14.5%	15.2%	15.5%	15.8%	16.5%	17.2%
M2	13.2%	14.0%	14.8%	15.2%	15.6%	16.4%	17.2%

On observe qu'elles sont toujours inférieures au seuil de 33% fixé pour les terrains homogènes. L'ordre de grandeur des valeurs de l'indice HVI est quand même le double par

rapport aux indices de variabilité calculés sur le terrain de Castello. D'ailleurs, on voit que quand on prend en compte deux unités (modèle M2), les résultats calculés ne présentent pas de variations significatives. par ailleurs, lorsqu'on réduit la valeur du coefficient A, qui pondère la différence entre la valeur moyenne des tendances, la variabilité horizontale augmente. Autrement dit, la corrélation entre sondages apparaît plus pénalisante dans ce cas.

Deuxième campagne d'essais

Analyse avec la méthode de Salgado

On observe ici des résultats assez différents par rapport à ceux issus de l'analyse de la première campagne. Les résultats pour les 30 essais dans l'axe X sont donnés dans le tableau (Tab.17.2). Ils indiquent un niveau d'hétérogénéité moyen du terrain ($HVI > 33\%$). Selon l'axe Y ces valeurs sont encore plus élevées (Tab.17.3). Cette augmentation systématique des indices HVI ne semble pas être cohérente avec l'allure des nouveaux pénétrogrammes mesurés, très similaire à celle des sondages initiaux, alors que le terrain était homogène d'après l'analyse de la première campagne.

TABLE 17.2 – Indices de variabilité horizontale calculés pour les essais dans l'axe X. Profondeur d'étude égale à 3.4 m.

Coeff. A	1	0.8	0.6	0.5	0.4	0.2	0
Coeff. B	0	0.2	0.4	0.5	0.6	0.8	1
M1	33.7%	35.34%	36.9%	37.7%	38.5%	40.2%	41.8%
M2	33.1%	34.8%	36.6%	37.4%	38.3%	40.0%	41.8%

TABLE 17.3 – Indices de variabilité horizontale calculés pour les essais dans l'axe Y. Profondeur d'étude égale à 2 m.

Coeff. A	1	0.8	0.6	0.5	0.4	0.2	0
Coeff. B	0	0.2	0.4	0.5	0.6	0.8	1
M1	34.63%	38.63 %	42.63%	44.63%	46.63%	50.64%	54.64%

Modification proposée

Pour tenir compte de l'écartement entre sondages, noté s , le calcul de l'indice de variabilité horizontale incorpore une fonction exponentielle (éq.15.35) et la notion de distance de corrélation horizontale, fixée de manière forfaitaire à 6m.

Cependant, l'analyse de la variabilité spatiale horizontale de la croix de sondages montre que l'ordre de grandeur des distances de corrélation est beaucoup plus faible. C'est pourquoi, on propose de modifier la valeur du paramètre a .

Pour les sondages dans l'axe X, suite aux résultats de l'ajustement des modèles exponentiels aux variogrammes empiriques, on propose une valeur de a calculée comme la somme pondérée par l'épaisseur des formations analysées, c'est-à-dire :

$$a = \frac{0.35 \cdot 2.4 + 0.9 \cdot 1}{3.4} = 0.5 \quad (17.1)$$

Pour les sondages dans l'axe Y, on retiendra une valeur de a égale à 1.06.

Une analyse comparative des indices HVI obtenus selon la campagne d'essais et la formulation retenue est donnée dans la figure (Fig.17.8). Les nouvelles valeurs de HVI sont beaucoup plus proches de celles résultant de l'analyse de la première campagne d'essais. Cependant, l'analyse de la corrélation entre sondages de la deuxième campagne (valeur de HVI pour $A=0$) met en évidence une différence d'hétérogénéité entre les directions X et Y.

Par conséquent, il faut être prudent sur les résultats de cette méthode lorsque les sondages sont espacés d'une faible distance. La méthode fournit un critère qui permet de classer les niveaux d'hétérogénéité entre eux. Il semble pertinent d'intégrer dans la méthode la valeur de la distance de corrélation dès que l'on dispose de cette information.

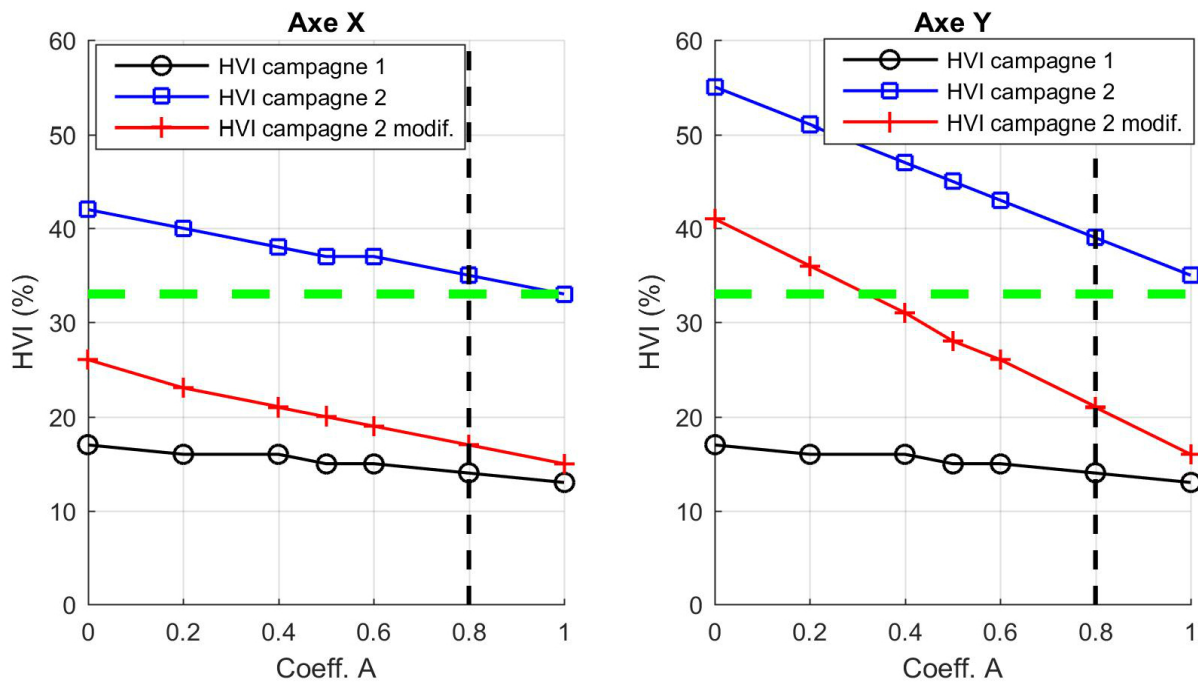


FIGURE 17.8 – Comparaison entre les indices HVI calculés en fonction de la campagne analysée et la modification proposée.

17.2.4 Bilan

L'analyse structurale des deux campagnes de sondages a permis de mettre en évidence que :

- pour la caractérisation des formations relativement homogènes à l'échelle d'un site géotechnique, un analyse structurale basée sur quelques sondages Panda permet d'obtenir une caractérisation suffisamment précise de la variabilité naturelle selon la profondeur. Dans ce sens, le fait de réaliser une planche avec un grand nombre d'essais ne représente pas d'avantage significatif a priori ;
- par contre, une croix de sondages de type géostatistique permet une définition plus aisée des structures horizontales qu'on ne peut pas atteindre grâce à l'analyse d'une campagne d'essais conventionnelle ;
- la prise en compte d'une dérive présente un influence notable vis-à-vis de la caractérisation des structures de variabilité naturelles. Les distances de corrélation ainsi

estimées ne sont plus un paramètre intrinsèque du sol mais dépendent aussi du modèle déterministe de tendance retenu. Par contre, on a constaté une plus grande répétabilité pour les distances de corrélation verticales quand elles sont calculées sur les résidus après avoir enlevé une tendance linéaire ;

- les modèles stationnaires de type exponentiel sont aptes à représenter les structures verticales rencontrées. Des effet pépites notables, de l'ordre du 50% de la variance empirique des données, ont été rencontrés lors de la construction des variogrammes horizontaux. Il s'agit bien d'un ordre de grandeur courant pour des structures horizontales en géotechnique dont l'origine provient des microstructures et des erreurs de mesure ;
- on a trouvé des résultats très différents lorsque on a appliqué la méthode de Salgado aux sondages issus de la première et de la deuxième campagne d'essais, alors qu'a priori les pénétrogrammes mesurés avaient une allure similaire. C'est pourquoi, une mise à jour de la formulation a été proposée dans la méthode de Salgado en intégrant la distance d'autocorrélation horizontale connue. Les résultats ainsi calculés semblent être plus répétables pour le cas étudié. Cependant, quand on dispose d'une planche d'essais à pas constant, la grande quantité de mesures disponibles permet une analyse structurale détaillée et cette méthode perd de son intérêt.

17.3 Variabilité horizontale des valeurs de la résistivité électrique

Le nombre et l'espacement des sondages d'une campagne géotechnique courante ne permet pas une définition précise des lois de variabilité horizontale des massifs de sol. Les essais géophysiques, et notamment la tomographie électrique, permettent un échantillonnage plus continu dans le sens horizontal avec une bonne finesse de résolution. Ces essais présentent l'avantage de fournir un nombre de mesures avec un pas d'échantillonnage plus faible. Il est fonction de l'écartement entre électrodes mais décroît avec la profondeur. Cela représente un cadre d'application idéal pour réaliser un étude sur la variabilité spatiale du sous-sol dans le plan horizontal, même si on est conscient que les propriétés diélectriques analysées ici n'ont pas de correspondance directe avec un paramètre mécanique telle que la résistance de pointe.

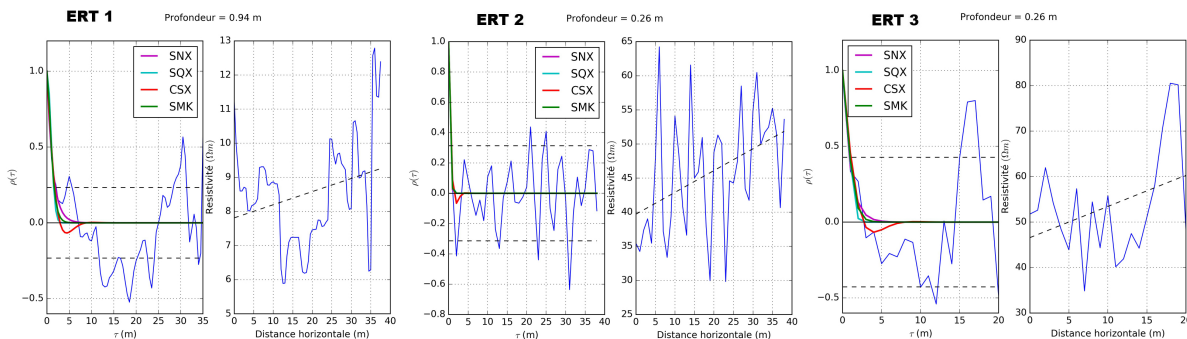


FIGURE 17.9 – Profils de variation de la résistivité électrique et ajustement de modèles théoriques d'autocorrélation.

Les caractéristiques des trois dispositifs électriques mis en œuvre ainsi que les pseudo-sections 2D calculées sont données en annexes ([Annexe D.1.5](#)). La variabilité spatiale des mesures de résistivité a été étudiée à l'aide de la fonction d'autocorrélation. Une tendance linéaire a été enlevée aux profils longitudinaux des mesures électriques. On signale que la principale contrainte de la mise en œuvre de la méthode électrique est que la résolution diminue avec la profondeur lorsque la mesure est réalisée en surface. Dans le même temps, le support de la mesure n'est pas homogène et la distance de corrélation horizontale du profil électrique est calculée à partir des valeurs issues de l'inversion à chaque profondeur de façon séparée. Dans la figure ([Fig.17.9](#)), on présente quelques exemples de l'analyse et des calculs effectués.

Les résultats de l'analyse de la corrélation spatiale des 3 panneaux électriques ainsi que le nombre de mesures (*Nb.*), la moyenne de la résistivité (μ_ρ) et le coefficient de variation (CV) ont été récapitulés dans les tableaux ([Tab.17.4](#)), ([Tab.17.5](#)) et ([Tab.17.6](#)).

TABLE 17.4 – Résultats de l'analyse de variabilité spatiale pour le panneau ERT1

Profondeur (m)	Nb.	$\mu_\rho(\Omega m)$	CV	$\theta_{SNX}(m)$	$\theta_{SQX}(m)$	$\theta_{CSX}(m)$	$\theta_{SMK}(m)$
0.17	78	58.48	0.26	0.79	0.71	0.69	0.75
0.54	78	34.94	0.26	2.22	1.98	1.63	2.05
0.94	76	8.53	0.18	2.61	2.23	1.87	2.33
1.38	72	8.35	0.16	2.70	2.26	1.90	2.39
1.86	68	12.91	0.18	6.50	5.11	5.78	6.24
2.39	64	15.07	0.17	4.34	3.17	2.76	3.48
2.98	58	15.45	0.09	1.41	1.08	1.18	1.28
Moyenne (m)				2.94	2.36	2.26	2.65
Écart-type (m)				1.93	1.46	1.68	1.81
CV				0.66	0.62	0.74	0.68

TABLE 17.5 – Résultats de l'analyse de variabilité spatiale pour le panneau ERT2

Profondeur (m)	N°	$\mu_\rho(\Omega m)$	CV	$\theta_{SNX}(m)$	$\theta_{SQX}(m)$	$\theta_{CSX}(m)$	$\theta_{SMK}(m)$
0.26	39	45.79	0.19	0.53	1.00	0.80	0.76
0.80	39	14.90	0.37	1.25	1.46	1.04	1.40
1.40	37	9.63	0.22	1.61	1.88	1.41	1.82
2.06	35	12.60	0.34	3.94	3.23	3.01	3.55
2.79	31	18.06	0.38	3.16	3.11	2.45	3.12
3.59	27	16.46	0.26	2.09	2.36	1.75	2.29
Moyenne (m)				2.1	2.17	1.74	2.16
Écart-type (m)				1.26	0.9	0.85	1.05
CV				0.60	0.41	0.49	0.49

On constate que les échelles de fluctuation horizontales estimées sont similaires indépendamment du panneau analysé, avec des valeurs relativement faibles, qui varient entre 0.5

TABLE 17.6 – Résultats de l’analyse de variabilité spatiale pour le panneau ERT3

Profondeur (m)	N ^o	$\mu_\rho(\Omega m)$	CV	$\theta_{\text{SNX}}(\text{m})$	$\theta_{\text{SQX}}(\text{m})$	$\theta_{\text{CSX}}(\text{m})$	$\theta_{\text{SMK}}(\text{m})$
0.26	21	53.44	0.22	1.96	1.83	1.69	2.00
0.80	21	8.90	0.22	0.07	0.03	0.03	0.13
1.40	19	8.97	0.13	1.65	1.85	1.57	1.87
2.06	17	18.05	0.35	3.50	3.36	2.63	3.41
Moyenne (m)				1.79	1.77	1.56	1.85
Écart-type (m)				1.41	1.36	0.93	1.34
CV				0.78	0.77	0.59	0.72

et 6 m. Même si la loi théorique choisie ne semble pas avoir de forte influence vis-à-vis des résultats calculés, on voit que le modèle exponentiel est celui qui fournit en général les valeurs les plus élevées. On constate néanmoins que les fonctions d’autocorrélation empiriques des données du troisième panneau sont plus difficiles à modéliser. Ceci est dû au fait que c’est le panneau qui a été réalisé avec le moins d’électrodes et donc un nombre de mesures disponibles plus faible.

Les analyses variographiques des données de résistivité ne sont pas courantes. Cependant, on peut citer (Salloum (2015)), qui a conduit des analyses similaires pour les profils de tomographie électrique réalisés dans le plaine alluvial de Nahr-Beyrouth au Liban. Le quadripole de type Wenner-Schlumberger et un espacement entre électrodes de 1.5 m ont été utilisés pour réaliser la campagne de mesures. La longueur du profil électrique analysé était de 94.5 m. Des échelles de fluctuation horizontales entre 3.8 et 10.4 m selon la profondeur ont été rapportées, valeurs supérieures à celles calculées ici.

Bilan

La variabilité spatiale horizontale du terrain a été étudiée au moyen de l’analyse des mesures de résistivité fournies par la tomographie électrique. L’outil pour caractériser cette variabilité a été la fonction d’autocorrélation, en prenant en compte une dérive ou tendance linéaire. La valeur des distances de corrélation horizontales calculée est faible, autour de 2 m. Les résultats tels qu’ils ont été calculés montrent un ordre de grandeur métrique. Les distances de corrélation restent proches de celles issues des analyses de q_d pour la couche de limons en surface.

17.4 Application de l’inférence bayésienne

17.4.1 Cadre théorique et inférence

On présente ici l’ajustement d’un modèle de corrélation exponentiel au moyen de la statistique bayésienne. Le seul paramètre inconnu à estimer est l’échelle de fluctuation horizontale θ_H . Le problème posé est alors une régression non linéaire décrite par l’équation suivante :

$$\hat{\rho}(\tau) = \rho(\tau|\theta_H) + \varepsilon \quad (17.2)$$

avec $\hat{\rho}$ la fonction d'autocorrélation empirique, $\rho(\tau|\theta_H)$ le modèle d'autocorrélation (exponentiel dans notre cas), et ε le terme d'erreur que l'on suppose normalement distribué autour de 0 ($\varepsilon \sim \mathcal{N}(0, \sigma)$).

La fonction de densité de probabilité de la vraisemblance suit une loi normale si on suppose l'indépendance entre les N observations disponibles (Gelman et al. (2014)). La fonction de vraisemblance peut être alors écrite comme suit :

$$L(\hat{\rho}_1, \dots, \hat{\rho}_N | \sigma_\varepsilon^2, \theta_H) = \prod_{i=1}^N \frac{1}{\sqrt{2\pi}\sigma_\varepsilon} \exp \left[-\frac{1}{2} \left(\frac{\rho(\tau_i) - \hat{\rho}(\tau_i|\theta_H)}{\sigma_\varepsilon} \right)^2 \right] \quad (17.3)$$

sur la base du principe bayésien, l'information a priori est donnée par la loi jointe des paramètres du modèle, soit θ_H) et l'écart-type de l'erreur du modèle $f'(\sigma_\varepsilon, \theta_H)$. Dans le cadre bayésien, les paramètres du modèle que l'on cherche à estimer sont traités comme des variables aléatoires. Le but est d'obtenir la loi a posteriori pour chacun d'entre eux grâce au théorème de Bayes. On actualise ainsi la loi a priori vers la loi a posteriori $f''(\sigma_\varepsilon, \theta_H)$, dont l'équation mathématique à résoudre a été donnée auparavant (éq.15.39).

La revue bibliographique a montré que les techniques numériques les plus robustes sont celles basées sur les méthodes de simulations par chaînes de Markov. Par conséquent, l'inférence est faite à l'aide d'un échantillonneur MCMC en utilisant le module PyMC (Patil et al. (2010)) implémenté dans le langage de programmation Python et l'algorithme NUTS (*no-u-turn sample*) (Hoffman and Gelman (2014)). Un des grands avantages de cet algorithme est l'absence de période de chauffe (*burn-in period*), à la différence de l'algorithme de Metropolis-Hasting (MH)¹.

Par rapport à l'information a priori à utiliser, on étudiera deux cas :

- loi a priori vague ou peu informative
- loi a priori informative

L'écart-type de l'erreur du modèle σ_ε est modélisé par une loi demi-normale, d'écart-type égal à 1. De manière générale, quand une variable x suit un loi demi-normale, sa fonction de densité de probabilité a l'expression suivante :

$$f_x(x; \sigma) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp \left(-\frac{x^2}{2\sigma^2} \right) \quad x \geq 0 \quad (17.4)$$

17.4.2 Loi a priori vague

Quand on ne dispose pas avec certitude de l'information au préalable sur le phénomène à modéliser la loi uniforme peut être retenue. Dans notre problème, et à titre d'étude théorique, on peut utiliser les données publiées dans la littérature comme information a priori. Suite aux valeurs de référence pour l'échelle de fluctuation horizontale données en (Favre (2004)), 15 m est la valeur retenue comme limite supérieure pour θ_H . Comme borne inférieure, on choisit la plus petite valeur reportée dans la littérature θ_H (Tab.15.5) est égal 1 m. La loi

1. Pour la mise en œuvre des algorithmes de type MH le choix des conditions initiales peut affecter la convergence de l'algorithme. Il est alors nécessaire de ne pas tenir compte des premières itérations, mais il est difficile de déterminer la durée de cette période de chauffe, surtout en grandes dimensions.

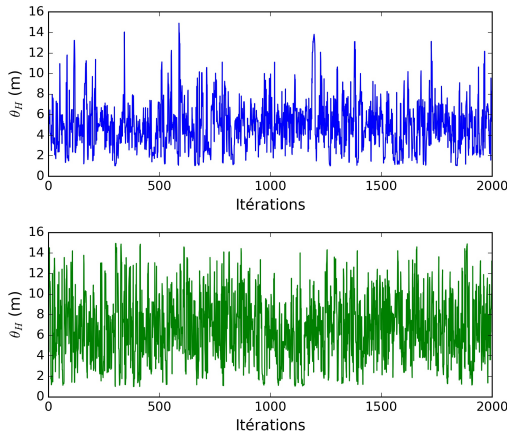


FIGURE 17.10 – Chaînes de Markov simulées

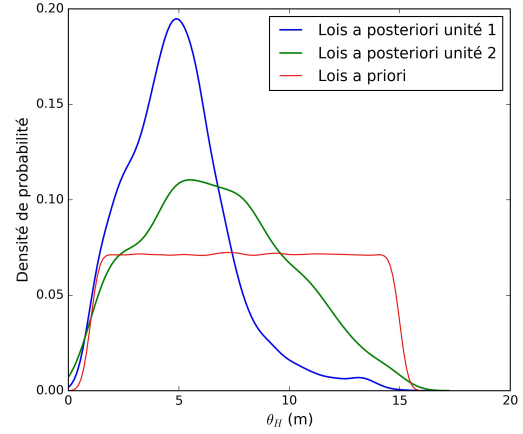


FIGURE 17.11 – Loi a posteriori marginale de θ_H

TABLE 17.7 – Résultats de l'inférence bayésienne pour θ_H avec une loi a priori uniforme peu informative

	Lois a priori		Lois a posteriori			
	a	b	μ_{θ_H} (m)	σ_{θ_H} (m)	CV_{θ_H}	Ic 95% (m)
Unité 1	1	15	4.97	2.31	0.47	1.01 - 9.2
Unité 2	1	15	6.72	3.26	0.49	1.00 - 12.36

marginale a priori de l'échelle de fluctuation est alors :

$$f(\theta_H) = \begin{cases} \frac{1}{b-a} & \text{pour } a \leq \theta_H \leq b, \\ 0 & \text{sinon} \end{cases} \quad (17.5)$$

avec $b = 15m$ et $a = 1m$.

Un total de 2000 itérations ont été réalisées. Dans le graphique de gauche de la figure (Fig.17.10) on a représenté les chaînes de Markov simulées de θ_H , où l'on peut vérifier qu'elles ne semblent pas autocorrélées. Les lois a posteriori marginales de θ_H pour les deux unités étudiées sont aussi données dans le graphique de droite dans la figure (Fig.17.11). On fournit les estimations ponctuelles de θ_H en forme de moyenne, écart-type et coefficient de variation ainsi que l'intervalle de confiance ou crédibilité à 95%. Ces statistiques sont regroupées dans le tableau (Tab.17.7). La valeur moyenne calculée de θ_H est égale à 5 m pour la première unité. Une valeur supérieure de presque 7 m a été estimée pour la deuxième unité, dont l'écart type et l'intervalle de confiance sont aussi plus larges, montrant une plus grande incertitude.

Finalement, la moyenne et l'écart-type estimés pour θ_H en fonction de l'écart-type de l'erreur du modèle σ_ε sont fournis dans la figure (Fig.17.12). D'après ce graphique, on observe que cette valeur n'a pas une influence significative sur les statistiques de la loi a posteriori. Par contre, les différences tout en restant faibles sont plus marquées quand σ_ε atteint une valeur très faible (0.1).

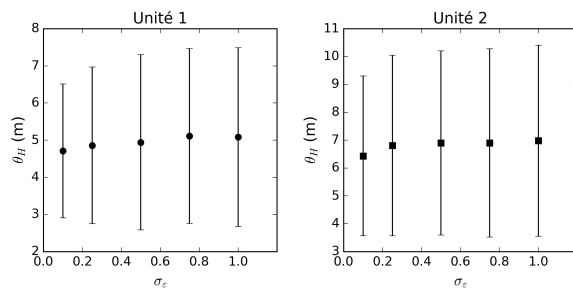


FIGURE 17.12 – Influence de l'écart-type de l'erreur sur l'estimation de la valeur moyenne de θ_H pour une loi a priori informative

17.4.3 Loi a priori informative

Les panneaux électriques réalisés dans ce terrain permettent de mettre en évidence les variations latérales et en profondeur des formations limoneuses. Pour cela, on propose d'utiliser les distances de corrélations des mesures de résistivité calculées à partir des analyses des pseudo-sections obtenues comme information a priori. Une loi lognormale a ainsi été ajustée aux valeurs de θ_H (Fig.17.13), avec une moyenne $\mu_{SH} = 2.55$ m et un coefficient de variation $COV_{\theta_H} = 68\%$. Suite aux 2000 itérations, les lois a posteriori marginales des distances de corrélation horizontales estimées sont fournies (Fig.17.14). On constate que leurs allures sont très similaires à celle de la loi a priori utilisée.

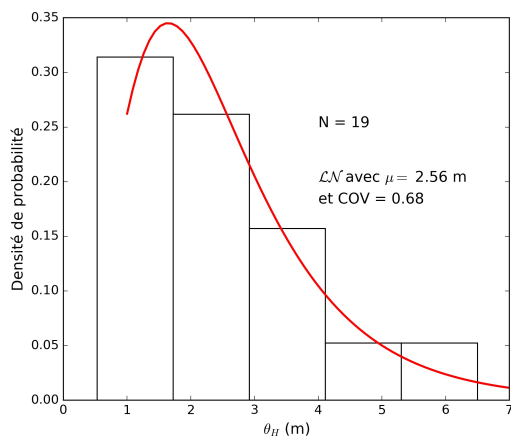


FIGURE 17.13 – Ajustement de la loi lognormale aux valeurs de θ_H estimées à partir de la tomographie électrique

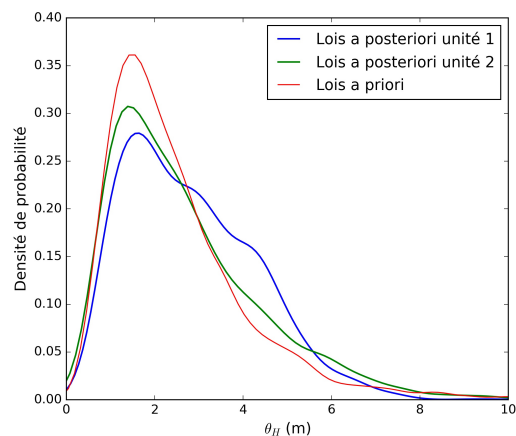


FIGURE 17.14 – Lois a posteriori marginales de θ_H

Les statistiques des lois a posteriori marginales de θ_H sont regroupées dans le tableau (Tab.17.8). Les estimations ponctuelles a posteriori calculées sont très similaires pour les deux unités de terrain étudiées. Par ailleurs, les estimations de la moyenne de θ_H sont très proches mais légèrement supérieures à celles de la loi a priori utilisée. Enfin, comme pour le cas précédent, on a constaté que l'influence du terme d'erreur est négligeable.

TABLE 17.8 – Résultats de l'inférence bayésienne pour θ_H avec une loi informative

	Lois a priori θ_H		Lois a posteriori θ_H			
	μ_{θ_H}	CV_{θ_H}	μ_{θ_H} (m)	σ_{θ_H} (m)	CV_{θ_H}	Ic 95% (m)
Unité 1	2.55	0.68	2.82	1.47	0.52	0.5 - 5.47
Unité 2	2.55	0.68	2.7	1.73	0.64	0.46 - 6.11

17.4.4 Bilan

On fournit une synthèse des valeurs de θ_H (cf. Tab.17.9) estimées selon les différentes techniques.

TABLE 17.9 – Synthèse des valeurs de θ_H estimées

Approche	Unité 1		Unité 2		
	θ_H (m)	Ic 95%	θ_H (m)	Ic 95%	
MCO	<i>Rec. 1</i>	4.5	-	5.8	-
	<i>Rec. 2</i>	$\theta_{H_x} = 0.7$ et $\theta_{H_y} = 2.12$	-	$\theta_{H_x} = 1.8$	-
Bayes	<i>Rec.1 (Vague)</i>	4.97	1.0 - 9.2	6.72	1.0 - 12.36
	<i>Rec.1 (Informative)</i>	2.82	0.5 - 5.47	2.7	0.46 - 6.11

On note *MCO* la procédure des moindres carrés ordinaire, *Rec.1* les mesures de q_d issues de la première campagne de reconnaissance, *Rec.2* les mesures de q_d issues de la deuxième campagne de reconnaissance (croix de 60 sondages). De ce tableau comparatif, on peut tirer les conclusions suivantes :

- avec la loi a priori vague utilisée, les estimations de θ_H sont proches mais plus grandes que les valeurs calculées grâce à une régression par les moindres carrés classique, particulièrement pour la première unité ;
- la loi a posteriori pour θ_H présente un coefficient de variation plus faible lorsque la loi a priori informative est préconisée. En effet l'intervalle de confiance est réduit à presque la moitié pour l'analyse de la deuxième unité par rapport au cas de la loi peu informative. Comme le nombre d'observations est relativement faible, la loi a priori joue un rôle important dans l'inférence ;
- les distances de corrélation horizontales estimées à partir des mesures de résistivité avaient des valeurs très faibles. Par conséquent, si on retient comme valeurs de référence les échelles de fluctuation estimées à partir des données de la deuxième campagne d'essais grâce au modèle exponentiel, le scénario où on utilise la loi informative fournit les valeurs les plus proches.

17.5 Conclusion sur les apports supplémentaires

On a présenté plusieurs tentatives pour tenter de définir plus finement la variabilité spatiale du sous-sol et voir leur influence sur la modélisation réalisée.

Premièrement, l'étude structurale d'une croix de sondages utilisée classiquement lors des études de type géostatistique a été présentée et comparée aux analyses d'une campagne d'essais classique. Il a été mis en évidence que le faible pas d'échantillonnage et la grande quantité de mesures fournies par le pénétromètre dynamique de type Panda permet d'aboutir à une caractérisation aisée de la variabilité verticale et ce, à partir de quelques essais. Cette affirmation prend tout son sens pour les cas ici étudiés, des terrains peu désordonnés présentant une stratification relativement régulière, avec pour chaque couche, des propriétés relativement régulières aussi. Donc pour la caractérisation de la variabilité verticale, il n'est pas nécessaire de multiplier le nombre de sondages pour obtenir une bonne caractérisation de celle-ci. Par contre, l'étude des données issues de la croix sondages réalisés durant la deuxième campagne d'essais Panda, a permis via le calcul des variogrammes empiriques, une définition détaillée des structures horizontales des deux formations étudiées fournissant une amélioration notable de la précision de la modélisation. Des effets de pépite importants ont été systématiquement observés pour les variogrammes horizontaux malgré le faible écartement entre sondages (0.5 m), ce qui suggère l'existence de micro-structures inférieures à cet écartement. Toutefois, il est important de garder à l'esprit que les contraintes budgétaires des projets de construction ne permettent pas la réalisation d'une planche d'essais de ce type dans le cadre des campagnes de reconnaissance géotechniques courantes. Le problème de la caractérisation de la variabilité spatiale dans le plan horizontal à partir d'un nombre restreint de sondages reste donc une question persistante.

Pour tenter d'y répondre, une seconde tentative a été menée à partir de mesures géophysiques. L'analyse de la corrélation spatiale des mesures géophysiques semble intéressante, puisque ces essais permettent de surmonter les contraintes du nombre de données et d'obtenir un profil de mesures avec un faible pas d'échantillonnage. Dans cette logique, on s'est intéressé à l'analyse des essais de tomographie électrique. Les profils analysés sont exploitables pour le calcul d'une loi de variation spatiale malgré les inconvénients inhérents aux méthodes géophysiques (diminution de la résolution des mesures avec la profondeur et inexistence d'un modèle unique après l'inversion). Néanmoins, le calcul d'un variogramme ou d'une fonction d'autocorrélation moyenne n'est pas faisable puisqu'on ne dispose pas d'un pas d'échantillonnage constant. Par ailleurs, la pertinence d'utiliser les distances de corrélation déduites grâce à l'analyse des propriétés diélectriques pour définir les lois spatiales d'un paramètre mécanique tel que la résistance de pointe reste à démontrer.

Finalement, nous avons évalué l'intérêt de l'inférence bayésienne à l'aide des algorithmes d'échantillonnage par chaînes de Markov, qui s'avère un outil d'estimation très puissant. Cette technique offre la possibilité d'incorporer de l'information a priori dans l'analyse. C'est pourquoi, on a proposé une démarche bayésienne vis-à-vis du calcul de la fonction d'autocorrélation. Deux cas théoriques ont été présentés, pour une loi a priori peu informative, on obtient des résultats similaires à ceux issus d'une régression non linéaire par moindres carrés ordinaire. Cependant, pour un jeu de données de faible taille, la loi informative a beaucoup plus de poids dans l'estimation et apporte une amélioration indéniable. On pourrait étendre la procédure présentée pour modéliser également des modèles de variogrammes et la géologie pourrait être le fil conducteur pour fixer l'étendue des distances de corrélation.

18. Conclusion sur la modélisation de la variabilité spatiale

Sur la base d'une campagne relativement modeste d'essais au pénétromètre Panda, nous avons défini une démarche permettant de modéliser la variabilité spatiale de la résistance de pointe q_d en vue d'obtenir un modèle de terrain. Nous avons montré qu'avec cette méthode, il est possible d'obtenir des modèles plausibles réalistes de q_d du terrain en 3D. Pour cela, l'apport des techniques de découpage du signal pénétrométrique en unités homogènes a été mis en évidence. La méthodologie de caractérisation de l'homogénéité horizontale du site proposée par Salgado, a été adaptée aux signaux pénétrométriques Panda et incorporée à la méthode. Cette caractérisation s'avère intéressante pour valider quantitativement l'hypothèse d'homogénéité selon le plan horizontal mais peut être très pénalisante pour des sondages écartés d'une faible distance. L'étape suivante de la méthodologie repose sur des simulations conditionnelles. Le conditionnement du champ aléatoire par les sondages Panda a fait l'objet d'une étude théorique sur les apports du krigeage à la réduction de l'incertitude du modèle mais également pour optimiser le plan d'échantillonnage de la reconnaissance. Cette étude a montré tout l'intérêt de cette technique et a mis en évidence le rôle prépondérant de la distance de corrélation horizontale.

Pour tenter de mieux définir la variabilité spatiale sur le plan horizontal, plusieurs approches ont été testées. Si l'apport d'une croix de sondage à pas régulier est intéressant, cette solution n'est pas envisageable pour une reconnaissance géotechnique courante. C'est pourquoi, l'analyse des variations spatiales sur la base d'essais géophysiques a été étudiée. L'analyse de mesures de résistivité a montré de faibles valeurs d'autocorrélation horizontale et ce, quelle que soit la fonction théorique utilisée. Ces faibles valeurs montrent une certaine cohérence avec celles déduites grâce à l'analyse des campagnes pénétrométriques. L'analyse des lois de variations spatiales des essais géophysiques s'avère être une voie d'étude intéressante, puisque l'échantillonnage spatial de ces essais permet d'étudier les variations des paramètres physiques du sol dans l'espace. Toutefois, il est encore difficile d'en déduire des conclusions définitives concernant leur application puisque la pertinence de l'approximation des longueurs de corrélation de paramètres mécaniques par celles déduites de l'analyse des propriétés géophysiques reste à démontrer. Finalement, les techniques d'inférence bayésiennes permettant d'incorporer de l'information a priori pour la modélisation de la loi de variation spatiale ont été proposées. L'utilisation de ces techniques s'avère prometteuse et d'un grand intérêt notamment pour pouvoir intégrer et combiner des informations fournies par d'autres essais.

Conclusion générale et perspectives

Dans un contexte d'optimisation de la conception et de la justification des ouvrages où les approches probabilistes se généralisent, pouvoir définir un modèle de terrain fournissant une estimation probabiliste des caractéristiques physico-mécaniques des propriétés du sol en tout point de celui-ci, la plus réaliste et la plus fiable possible est une des problématiques actuelles de la géotechnique.

Dans ce cadre, ce travail s'est focalisé sur la reconnaissance des sols à faible profondeur et sur l'exploitation des données de résistance de pointe dynamique recueillies à l'aide de l'essai de pénétration dynamique à énergie variable, Panda®. L'objectif principal était d'étudier et de proposer un ensemble d'approches au sein d'une méthode globale permettant d'exploiter les mesures issues d'une campagne de sondages Panda afin de bâtir un modèle géotechnique du terrain. La démarche envisagée s'est basée sur les étapes suivantes :

- 1) identification automatique des unités homogènes de sol au sein des sondages ;
- 2) pour chaque formation homogène détectée à l'étape précédente, identification de la nature des matériaux traversés à partir de leur classification ;
- 3) établissement d'un modèle de terrain probabiliste en 3D fournissant la répartition de la résistance de pointe en intégrant l'incertitude spatiale ;
- 4) développement des approches permettant le passage du modèle de terrain au modèle géotechnique.

Ce travail s'est focalisé sur les étapes 1 à 3 de la méthodologie et l'étape 4 n'a pas été abordée.

Après un rappel sur le contexte actuel de la reconnaissance des sols et les différents outils d'investigation, nous nous sommes focalisé sur une description détaillée de l'essai de pénétration dynamique Panda. Parmi ses différents avantages, la forte résolution des mesures dans la direction verticale et sa rapidité de mise en œuvre et d'exécution, permettant de multiplier les essais, font que cet essai présente un intérêt indéniable pour évaluer la variabilité des sols et mettre en œuvre des méthodes statistiques de traitement de données en vue d'élaborer un modèle de terrain. La suite du travail a consisté à élaborer des outils et procédures de traitement des données issues de l'essai Panda afin d'accomplir les différentes étapes de la méthodologie générale d'élaboration d'un modèle de terrain présentée auparavant. Les principales conclusions obtenues pour chacune des étapes et les pistes de développement et d'amélioration envisagées pour la suite de ces travaux sont présentées ci-dessous.

1) Identification automatique des formations homogènes du sol

Après avoir démontré les limites en terme de subjectivité et de variabilité des analyses expertes des pénétrogrammes Panda, il est apparu nécessaire et intéressant de développer un outil d'analyse automatique des sondages. A partir de la définition de « couche » retenue dans ce travail (unité de sol géologiquement et mécaniquement homogène), nous avons développé différentes procédures basées sur la notion de test d'une hypothèse paramétrique à l'aide d'une fenêtre glissante. Cela afin de constituer une aide à l'ingénieur et d'objectiver l'analyse des pénétrogrammes. Ces procédures ont été testées et validées sur des sondages réalisés en conditions maîtrisées en laboratoire et sur des sondages réalisés in situ.

Étude en laboratoire en conditions maîtrisées

L'étude des limites et de la robustesse des procédures mises en place sur des matériaux granulaires étudiés en laboratoire a permis de :

- valider les techniques proposées dans le cadre de la détection des couches de faible épaisseur (jusqu'à 25 cm de d'épaisseur) ;
- démontrer l'influence du pas de régularisation permettant d'obtenir un échantillonnage régulier, de la taille de la fenêtre W_d d'analyse et de la forme de la variable étudié (q_d ou son logarithme népérien $\ln(q_d)$).

Le tableau (Tab.18.1) synthétise les conclusions principales déduites pour chacune des techniques. En conditions de laboratoire, il n'a pas pu être étudié l'efficacité des méthodes pour des matériaux fins ou l'influence de la consolidation.

TABLE 18.1 – Récapitulatif des conclusions sur les différentes méthodes

Méthodes	Pas de régularisation	Taille de la fenêtre	Variable d'étude
<i>Tratio</i>	enfoucement moyen	0.4 m	$\ln(q_d)$
<i>Intraclasses</i>	pas d'influence	0.4 m	$\ln(q_d)$
<i>Barlett</i>	enfoucement moyen	0.4 m	q_d

Étude sur des essais in situ

L'efficacité des méthodes statistiques a été évaluée pour des essais réels effectués sur des sites naturels. Ces études ont montré que :

- les méthodes développées permettent d'identifier correctement les interfaces entre les matériaux fins et sableux ;
- malgré la cohérence entre les faciès géologiques et les couches mécaniques identifiées à l'aide de ces approches statistiques, la correspondance n'est pas biunivoque et par conséquent ces deux informations sont complémentaires ;
- pour l'analyse de sondages in situ, une taille de fenêtre de l'ordre de 1 m semble bien adaptée afin de ne pas être trop influencés par les fluctuations à faible échelle causées par la variabilité naturelle du sol ;
- la méthode basée sur le test du *Tratio* (avec un seuil à 10) semblait la plus adaptée à notre problématique ;
- qu'il existait une bonne correspondance avec les essais CPT pour l'identification des couches homogènes.

Bilan sur les techniques étudiées

Finalement, les méthodes étudiées s'avèrent être des procédures objectives permettant de rationaliser la procédure de reconnaissance des sols et d'identifier les différents horizons homogènes mécaniquement. La taille de la fenêtre n'est pas un paramètre « intrinsèque », elle doit être adaptée à la profondeur du sondage et au cas d'étude. Les résultats issus du test de *Barlett* permettent d'identifier d'une manière évidente des couches de sol avec des caractéristiques mécaniques très distinctes, mais ces horizons sont en général facilement détectables de manière visuelle. Les profils statistiques *Tratio* et *Intraclasses* montrent des allures presque identiques. Ils sont plus sensibles aux variations de la résistance moyenne et peuvent fournir ainsi un découpage plus précis. Ces deux techniques seront à privilégier car contrairement au test de *Barlett*, elles peuvent être automatisées. Enfin, il est important de rappeler que la coupe du terrain déduite à partir de l'étude géologique reste complémentaire aux stratigraphies mécaniques obtenus à l'aide de ces techniques.

2) Classification de la nature des matériaux traversés

L'objectif de cette partie était d'étudier la possibilité d'obtenir une classification de la nature du matériau sondé à partir de l'analyse du signal Panda. Étant conscient que la mesure de q_d seule est insuffisante pour situer qualitativement la nature du sol, au lieu de tenter d'établir une corrélation entre les mesures de résistance de pointe avec le type de sol traversé, les potentialités des techniques d'intelligence artificielle (approches basées sur des algorithmes d'apprentissage automatique ou techniques de *machine learning* (ML)) ont été explorées. Suite à un état de l'art sur ces techniques, ce sont les algorithmes de classification automatique basés sur les réseaux de neurones artificiels (RNA) qui ont été privilégiés dans ce travail. La démarche d'étude a comporté la création d'une base de données de sondages pénétrométriques, la définition des entrées et sorties du modèle, l'apprentissage du système et la validation du modèle. Cette méthode a été étudiée pour deux problématiques distinctes :

- a) la classification de sols naturels à partir d'une banque de données composée d'essais maîtrisés en laboratoire et d'essais in situ ;
- b) la classification d'un matériau granulaire argileux (bentonite) formé par un mélange de billes sèches et de matériau concassé.

a) Classification de sols naturels

Quatre catégories de sols (sols fins, sables avec fines, sables fins et graves) ont été définies et deux architectures de RNA ont été testées, le perceptron multicouche (PMC) et le réseau de neurones probabiliste (PNN). Suite aux résultats obtenus pour le premier échantillon de données, il se dégage que :

- les deux architectures testées fournissent des résultats très prometteurs avec une excellente performance, mais le PMC permet d'obtenir une classification légèrement plus précise ;
- les classes intermédiaires (sables avec fines et sables propres) sont les types de sols les plus difficiles à classer correctement. Mais cela peut être dû au fait que ce sont aussi les classes de sol les moins représentées dans la base d'apprentissage.

La robustesse des deux architectures proposées a été étudiée sur des essais effectués sur de nouveaux matériaux et sur des sondages en conditions réelles ne faisant pas partie de la base de données originale. De manière générale, il se dégage que :

- la performance des RNA est beaucoup plus faible que celle obtenue lors des résultats initiaux ;
- les résultats fournis par le réseau PMC s'avèrent être légèrement plus précis que ceux issus du réseau PNN ;
- les matériaux fins et grossiers sont relativement bien identifiés mais les matériaux intermédiaires sont là encore, difficiles à identifier.

b) Classification d'un matériau granulaire argileux (bentonite)

Dans le cadre d'une application plus simple où il s'agissait de reconnaître 5 mélanges de matériau comportant deux phases granulaires différentes (billes sèches et matériau concassé), un RNA de type PMC a été étudié. Le nombre de paramètres d'entrée (7) pour définir le réseau a été réduit par rapport au cas d'étude précédent et les résultats ont mis en évidence l'excellente performance du réseau et l'intérêt d'une telle approche pour aborder des problèmes de classification moins complexes.

3) Modélisation probabiliste de la variabilité spatiale en vue d'obtenir un modèle de terrain

Nous avons proposé à la fin de ce travail une méthodologie pour obtenir une modélisation probabiliste dans l'espace 3D des valeurs de q_d mesurées à l'aide de l'essai Panda. A partir de l'hypothèse que la résistance de pointe dynamique peut être représentée par un champ aléatoire homogène au sein de chacune des unités homogènes du terrain, la démarche proposée comporte les étapes suivantes :

- a) découpage vertical des sondages afin de définir les unités de terrain homogènes ;
- b) évaluation de l'homogénéité du site sur le plan horizontal à partir de l'indice de variabilité proposée pour l'essai au pénétromètre statique CPT ;
- c) analyse structurale des sondages afin d'obtenir les paramètres du champ aléatoire ;
- d) simulations stochastiques du champ en 3D, conditionnées par les sondages Panda effectués.

La méthode a été appliquée sur un site réel en Espagne. Des études complémentaires en vue de raffiner cette démarche ont été réalisées pour un deuxième site expérimental. Les conclusions principales de cette méthodologie sont présentées par la suite :

- les méthodes de découpages des sondages en unités homogènes développés en partie II sont bien adaptées à l'étude d'un site complet ;
- l'indice de variabilité horizontale (HVI) défini par ([Salgado et al. \(2015\)](#)) peut s'adapter aux sondages pénétrométriques Panda et permet de mesurer l'homogénéité du site en s'appuyant sur la similarité entre les couples de sondages d'une campagne d'essais au pénétromètre ;
- de l'analyse structurale des sondages, il est possible de modéliser la résistance de pointe q_d par un champ aléatoire lognormal au sein de chacune des unités du modèle retenu. Même s'il est difficile de se prononcer sur le type de lois probabiliste à laquelle obéissent les caractéristiques géotechniques, ce choix a été motivé par des travaux de recherche antérieurs menés pour l'essai Panda ([Deplagne and Bacconnet \(1993\)](#), [Gaouar \(1997\)](#), [Lepetit \(1998\)](#), [Chaigneau \(2001\)](#), [Villavicencio \(2009\)](#)). La moyenne du champ est représentée par un modèle de tendance linéaire selon la profondeur au sein de chacune des unités du terrain et la variabilité spatiale est alors étudiée sur l'échantillon des résidus de moyenne nulle. On a trouvé une bonne répétabilité des structures spatiales selon la direction verticale, de l'ordre de 20 cm, étudiées à l'aide de la fonction d'autocorrélation renforçant ainsi l'hypothèse qu'un modèle de corrélation exponentiel semble être pertinent ;
- pour effectuer la simulation du champ aléatoire, l'hypothèse simplificatrice d'une profondeur constante des limites entre les couches de terrain détectées a été retenue. On a proposé de conditionner le modèle du champ aléatoire par les sondages Panda effectués sur site au moyen du krigeage ordinaire.

Nous avons montré que, pour le cas d'application étudié ici :

- le nombre de 6 essais semble suffisant, pour pouvoir proposer des versions plausibles d'un sondage "virtuel" en n'importe quel point du terrain ;
- quand on divise le terrain en unités homogènes grâce aux procédures statistiques, on obtient un meilleur calage sur les données expérimentales de la loi théorique en s'affranchissant au mieux de l'hypothèse de stationnarité au sens large ;
- il est possible de définir un critère pour évaluer de manière quantitative quel est le

nombre d'unités optimale à prendre en compte dans le modèle de sol, sur la base des hypothèses retenues dans la modélisation. Il est fondé sur l'écart entre les pénétrogrammes virtuels et ceux mesurés ;

- la réduction de l'incertitude du modèle grâce au conditionnement avec les sondages disponibles est plus sensible à la distance de corrélation horizontale qu'au nombre de sondages ou qu'à leur plan d'implantation pour le cas d'une campagne de reconnaissance dont le nombre de sondages est faible.

De manière générale, en géotechnique le nombre de sondages et leur écartement ne permet pas d'inférer correctement les lois de variabilité horizontale des massifs de sols. C'est pourquoi, on s'est intéressé à l'apport des données supplémentaires qui pourraient permettre d'améliorer la connaissance des structures horizontales afin de raffiner le modèle de terrain proposé :

- l'apport d'une croix de sondages utilisée classiquement en géostatistique a montré que cela pouvait permettre de mieux estimer les distances d'autocorrélation. Toutefois, la portée d'un tel outil reste limitée car les contraintes budgétaires des projets de construction ne permettent pas la réalisation d'une planche d'essais de ce type dans le cadre des campagnes de reconnaissance géotechniques courantes ;
- l'apport de résultats d'essais géophysiques paraît prometteur et peut permettre de mieux cerner les distances d'autocorrélation horizontales. Néanmoins la diminution de la résolution des mesures avec la profondeur, et l'inexistence d'un modèle unique après l'inversion restent des limites de ces méthodes ;
- finalement, dans le cas où l'on dispose d'informations a priori, les méthodes d'inférence bayésiennes où l'on peut supposer une loi informative a priori sont intéressantes et permettent de raffiner la simulation.

Malgré les pistes ici étudiées et développées, la question sur l'obtention d'une estimation plus fiable de la distance de corrélation horizontale, dans le cadre d'une campagne de reconnaissance géotechnique courante, reste encore un problème ouvert.

Perspectives

Suite aux travaux réalisés plusieurs pistes de recherche et de développement sont envisageables.

- a) Les méthodes stratigraphiques étudiées dans cette partie portent sur l'analyse d'un seul paramètre, la résistance de pointe q_d mesurée à l'aide de l'essai Panda. L'essai Panda3®[®], permet de mesurer plusieurs paramètres du sol à la même profondeur et pour un même coût. Pour cet outil, l'application des analyses par clusters s'avère très intéressante pour déterminer une stratigraphie sur la base des paramètres fournis par l'essai.
- b) la contrainte principale des méthodes statistiques est que l'analyse est réalisée sur la base d'une taille de fenêtre fixe. C'est pourquoi, la possibilité de réaliser une analyse multi-échelle grâce aux méthodes par ondelettes devrait être regardée.
- c) L'approche par RNA mise en œuvre pour la classification des sols naturels à l'aide de l'essai de pénétration Panda a montré ses limites. Malgré les intérêts et potentialités de la démarche proposée, des informations complémentaires s'avèrent nécessaires pour pouvoir la rendre plus efficace. Dans cette optique, l'application des RNA à un essai

tel que le Panda3 s'avère très prometteuse. De même, le couplage avec d'autres techniques, tel que la géondoscopie permettrait de compléter avantageusement les données et d'améliorer significativement l'efficacité de la procédure.

- d) Le grand intérêt des simulations conditionnelles est la possibilité de générer des valeurs des paramètres du sol pour le calcul probabiliste des ouvrages. La poursuite de ces travaux devrait permettre d'étudier les possibilités d'incorporer la variabilité spatiale des caractéristiques mécaniques du sol déterminé à l'aide du Panda, en vue de proposer des méthodes de calcul fiables. Le domaine des terres compactées, tel que barrages, digues et remblais, où la réalisation d'un planche d'essais pénétrométrique de type Panda est envisageable, s'avère un autre cadre idéal d'application, tel que démontré précédemment (Gaouar (1997));
- e) L'estimation d'une valeur de distance de corrélation horizontale « représentative » des données géotechniques s'avère encore un problème à résoudre. C'est pourquoi, à ce jour, plusieurs études présentées dans la littérature tentent de définir le cas le plus défavorable de distance de corrélation (Li et al. (2017)) en fonction de la géométrie de l'ouvrage à calculer (largeur de la fondation, hauteur du mur de soutènement...). Adopter une approche similaire pour en déduire la valeur de cette distance, permettrait de mener des calculs du côté de la sécurité. L'étude des essais géophysiques et leur couplage avec les données de sondage pour mieux estimer cette distance, semble être une piste d'étude intéressante ;
- f) Les sites expérimentaux supports de ce travail étaient des terrains homogènes selon le plan horizontal et le traitement des sites présentant une forte hétérogénéité horizontale s'avère être une autre question à traiter. Dans le cas où les couches du sous-sol présentent un fort pendage, la mise en œuvre des algorithmes d'interpolation pour caractériser le toit des différents substrats s'avère nécessaire.
- g) Enfin, la grande panoplie des possibilités offertes par les approches basées sur la statistique bayésienne n'ont pas pu être exploitées dans ce travail. Le raisonnement bayésien représente un formalisme excellent en vue de proposer des approches pour transformer les données de résistance de pointe en paramètres géomécaniques du modèle de calcul. En effet, la statistique bayésienne permettrait d'incorporer des paramètres mesurés par le biais d'autres essais (in situ et/ou en laboratoire) ainsi que le raisonnement d'expert en vue de fournir une estimation probabiliste des valeurs de calcul géotechniques. Dans le cadre de la reconnaissance de sols de surface, les tassements différentiels constituent l'une des causes principales de défaillance des fondations superficielles. Dans ce cas, la statistique bayésienne peut être intéressante pour améliorer l'estimation du module de déformation à partir d'un essai à la rupture tel qu'un pénétromètre dynamique Panda, tout en incorporant les mesures de déformation recueillies par l'essai pressiométrique par exemple comme information a priori. Ce même exemple est tout à fait réalisable avec un appareil comme le Panda3.

Bibliographie

- Ahmadi, M. and Robertson, P. (2005). Thin-layer effects on the cpt qc measurement. *Canadian Geotechnical Journal*, 42(5) :1302–1317.
- Akkaya, A. D. and Vanmarcke, E. H. (2003). Estimation of spatial correlation of soil parameters based on data from the texas a&m university nges. *Probabilistic site characterization at the National Geotechnical Experimentation Sites (eds GA Fenton and EH Vanmarcke)*, pages 29–40.
- Alabert, F. (1987). The practice of fast conditional simulations through the lu decomposition of the covariance matrix. *Mathematical Geology*, 19(5) :369–386.
- Alonso, E. E. and Krizek, R. J. (1975). Consolidation of randomly heterogeneous clay strata. *Transportation Research Record*, (548).
- Anderson, J. A., Pellionisz, A., and Rosenfeld, E. (1993). *Neurocomputing*. MIT press.
- Arroyo, M., Butlanska, J., Gens, A., Calvetti, F., and Jamiolkowski, M. (2011). Cone penetration tests in a virtual calibration chamber. *Géotechnique*, 61(6) :525–531.
- Arroyo, M., Pineda, J., Sau, N., Devincenzi, M., and Pérez, N. (2015). Sample quality examination on silty soils.
- Assimaki, D., Pecker, A., Popescu, R., and Prevost, J. (2003). Effects of spatial variability of soil properties on surface ground motion. *Journal of earthquake engineering*, 7(spec01) :1–44.
- ASTM (2010). Standard practice for classification of soils for engineering purposes (unified soil classification system). In *Annual Book of ASTM Standards*. ASTM, International West Conshohocken, PA.
- Athapaththu, A., Tsuchida, T., and Kano, S. (2015). A new geotechnical method for natural slope exploration and analysis. *Natural Hazards*, 75(2) :1327–1348.
- Athapaththu, A., Tsuchida, T., Suga, K., Nakai, S., and Takeuchi, J. (2007). Evaluation of in-situ strength variability of masado slopes. *Doboku Gakkai Ronbunshuu C*, 63(3) :848–861.
- Au, S. and Beck, J. (2003). Important sampling in high dimensions. *Structural safety*, 25(2) :139–163.

- Auvinet, G. (2002). Incertidumbre en geotecnia. In *XVI conferencia Nabor Carrillo, SMMS, Querétaro, México*, pages 73–111.
- Auvinet, G. and Gonzalez, J. (2000). Three-dimensional reliability analysis of earth slopes. *Computers and Geotechnics*, 26(3) :247–261.
- Auvinet, G., Juárez, M., Méndez, E., and Ovando, E. (1995). Sistema de información geográfica para sondeos geotécnicos. In *Proceedings, Xth Pan-American Conference on Soil Mechanics and Foundation Engineering*, volume 1, pages 312–324.
- Auvinet, G., Mellah, R., Masrouri, F., and Rodriguez, J. (2000). La méthode des éléments finis stochastiques en géotechnique. *Revue française de géotechnique*, (93) :67–79.
- Auvinet-Guichard, G. (2011). Modelo geoestadístico y geotécnico de la generación de emanaciones de hidrocarburos en el fondo marino de la sonda de campeche, méxico geostatistical and geotechnical model of hydrocarbon leaking in the seabed of sonda de campeche, mexico. *Ingeniería Investigación y Tecnología*, 12(2) :223–234.
- Bacconnet, C. (1991). *Géostatistique et géotechnique- Application à la reconnaissance des sols*. PhD thesis.
- Bacconnet, C., Azzouz, R., and Faugeras, J. (1987). Analyse géostatistique d’une campagne de reconnaissance au pénétromètre statique. In *Proceedings of the 5th International Conference on Applications of Statistics and Probability in Soil and Structural Engineering, Vancouver*, pages 821–828.
- Baecher, G. B., Chan, M., Ingra, T., Lee, T., and Nucci, L. (1980). Geotechnical reliability of offshore gravity platforms.
- Baguelin, F. and Kovarik, J.-B. (2000). Une méthode de détermination des valeurs caractéristiques des paramètres géotechniques. *Revue française de géotechnique*, (93) :35–41.
- Bakhtiari, S. (2011). *Stochastic finite element slope stability analysis*. PhD thesis, University of Manchester.
- Banton, O., Cimon, M.-A., and Seguin, M.-K. (1997). Mapping field-scale physical properties of soil with electrical resistivity. *Soil Science Society of America Journal*, 61(4) :1010–1017.
- Baroth, J., Schoefs, F., and Breysse, D. (2011). *Fiabilité des ouvrages : Sûreté, variabilité, maintenance, sécurité*. Lavoisier.
- Bartlett, M. S. (1937). Properties of sufficiency and statistical tests. *Proceedings of the Royal Society of London. Series A, Mathematical and Physical Sciences*, pages 268–282.
- Baud, J. (2005). Analyse des résultats pressiométriques ménard dans un diagramme spectral [$\log(p \text{ lm})$, $\log(e \text{ m/p lm})$] et utilisation des regroupements statistiques dans la modélisation d’un site. *Gambin, Magnan, Mestat (eds) ISP5—PRESSIO*, 1 :167–174.
- Beck, J. L. and Au, S.-K. (2002). Bayesian updating of structural models and reliability using markov chain monte carlo simulation. *Journal of Engineering Mechanics*, 128(4) :380–391.

- Benz-Navarrete, M.-A. (2009). *Mesures dynamiques lors du battage du pénétromètre Panda 2*. PhD thesis, Université Blaise Pascal-Clermont-Ferrand II.
- Betz, W., Papaioannou, I., and Straub, D. (2014). Numerical methods for the discretization of random fields by means of the karhunen–loève expansion. *Computer Methods in Applied Mechanics and Engineering*, 271 :109–129.
- Bhattacharya, B. and Solomatine, D. P. (2006). Machine learning in soil classification. *Neural networks*, 19(2) :186–195.
- Bishop, C. M. (1995). *Neural networks for pattern recognition*. Oxford university press.
- Bond, A. and Harris, A. (2008). *Decoding eurocode 7*. CRC Press.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., and Ljung, G. M. (2015). *Time series analysis : forecasting and control*. John Wiley & Sons.
- Breul, P. (1999a). *Caractérisation endoscopique des milieux granulaires couplée à lessai de pénétration*. PhD thesis, Université Blaise Pascal. Clermont-Ferrand II.
- Breul, P. (1999b). Caractérisation endoscopique des milieux granulaires couplée à lessai de pénétration. *Revue Française de Géotechnique*.
- Cafaro, F. and Cherubini, C. (2002). Large sample spacing in evaluation of vertical strength variability of clayey soil. *Journal of geotechnical and geoenvironmental engineering*, 128(7) :558–568.
- Campanella, R., Wickremesinghe, D., and Robertson, P. (1987). Statistical treatment of cone penetrometer test data. In *Proceedings of the 5th International Conference on Application of Statistics and Probability, AA Balkema, Rotterdam*, pages 1011–1019.
- Cao, Z. (2012). Probabilistic approaches for geotechnical site characterization and slope stability analysis.
- Cao, Z. and Wang, Y. (2013). Bayesian approach for probabilistic site characterization using cone penetration tests. *Journal of Geotechnical and Geoenvironmental Engineering*, 139(2) :267–276.
- Cao, Z. and Wang, Y. (2014). Bayesian model comparison and selection of spatial correlation functions for soil parameters. *Structural Safety*, 49 :10–17.
- Cao, Z., Wang, Y., and Li, D. (2016). Quantification of prior knowledge in geotechnical site characterization. *Engineering Geology*, 203 :107–116.
- Cassan, M. (1978). *Les essais in situ en mécanique des sols : Applications et méthodes de calcul*. Eyrolles.
- Cassan, M. (2000). Utilisation de la statistique descriptive en géotechnique. *Revue française de géotechnique*, (93) :21–34.

- Castro, G. and Zalzman, S. (1994). Estimator for evaluating spatial correlation of soil properties using unequally spaced data. In *Proceedings of the 8th Conference on Computers Methods and Advances in Geomechanics*, pages 369–374.
- Chaigneau, L. (2001). Caractérisation des milieux granulaires de surface à l'aide d'un pénétromètre. *Doctor of Philosophy dissertation. University Blaise Pascal, Clermont Ferrand, France.*
- Cherubini, C., F, C., and G, D. V. (2002). Variability-based parameters for the soil zonation. *PARAM2002*, 2-3(11) :457–463.
- Chiasson, P., Laffleur, J., Soulié, M., and Law, K. T. (1995). Characterizing spatial variability of a clay by geostatistics. *Canadian Geotechnical Journal*, 32(1) :1–10.
- Chiles, J. and Delfiner, A. (1999). Geostatistics : Modelling spatial uncertainty : Wiley interscience. *New York.*
- Chilès, J.-P. (1977). *Géostatistique des phénomènes non stationnaires :(dans le plan)*. PhD thesis.
- Chiles, J.-P. (1995). Quelques méthodes de simulation de fonctions aléatoires intrinseques. *Cahiers de Géostatistique*, 5 :97–112.
- Ching, J. and Chen, Y.-C. (2007). Transitional markov chain monte carlo method for bayesian model updating, model class selection, and model averaging. *Journal of engineering mechanics*, 133(7) :816–832.
- Ching, J. and Wang, J.-S. (2016). Application of the transitional markov chain monte carlo algorithm to probabilistic site characterization. *Engineering Geology*, 203 :151–167.
- Ching, J., Wang, J.-S., Juang, C. H., and Ku, C.-S. (2015a). Cone penetration test (cpt)-based stratigraphic profiling using the wavelet transform modulus maxima method. *Canadian Geotechnical Journal*, 52(12) :1993–2007.
- Ching, J., Wu, S.-S., and Phoon, K.-K. (2015b). Statistical characterization of random field parameters using frequentist and bayesian approaches. *Canadian Geotechnical Journal*, 53(2) :285–298.
- Christian, J. and Baecher, B. (2003). Reliability and statistics in geotechnical engineering.
- Christian, J. T. and Baecher, G. B. (1999). Point-estimate method as numerical quadrature. *Journal of Geotechnical and Geoenvironmental Engineering*, 125(9) :779–786.
- Christian, J. T. and Baecher, G. B. (2011). Unresolved problems in geotechnical risk and reliability. In *Geo-Risk 2011 : Risk Assessment and Management*, pages 50–63.
- Cooley, J. W. and Tukey, J. W. (1965). An algorithm for the machine calculation of complex fourier series. *Mathematics of computation*, 19(90) :297–301.

- Correia, A. G., Cortez, P., Tinoco, J., and Marques, R. (2013). Artificial intelligence applications in transportation geotechnics. *Geotechnical and Geological Engineering*, 31(3) :861–879.
- Cosenza, P., Marmet, E., Rejiba, F., Cui, Y. J., Tabbagh, A., and Charlery, Y. (2006). Correlations between geotechnical and electrical data : A case study at garchy in france. *Journal of Applied Geophysics*, 60(3) :165–178.
- Davis, M. W. (1987). Production of conditional simulations via the lu triangular decomposition of the covariance matrix. *Mathematical Geology*, 19(2) :91–98.
- DeGroot, D. J. (1996). Analyzing spatial variability of in situ soil properties. In *Uncertainty in the geologic environment : From theory to practice*, pages 210–238. ASCE.
- DeGroot, D. J. and Baecher, G. B. (1993). Estimating autocovariance of in-situ soil properties. *Journal of Geotechnical Engineering*, 119(1) :147–166.
- Delhomme, J.-P. (1976). *Applications de la théorie des variables régionalisées dans les sciences de l'eau*. PhD thesis, Université Pierre et Marie Curie.
- Demuth, H. B., Beale, M. H., De Jess, O., and Hagan, M. T. (2014). *Neural network design*. Martin Hagan.
- Denis, A., Elachachi, S. M., and Niandou, H. (2011). Effects of longitudinal variability of soil on a continuous spread footing. *Engineering Geology*, 122(3) :179–190.
- Deplagne, F. (1991). Analyse structurale d'essais de pénétration.
- Deplagne, F. and Bacconnet, C. (1993). Analyse structurale d'une digue en argile. *Cahiers de Géostatistique, Fascicule 3*, pages 181–188.
- Der Kiureghian, A. and Ke, J.-B. (1987). The stochastic finite element method in structural reliability. In *Stochastic structural mechanics*, pages 84–109. Springer.
- Deutsch, C. (2002). Geostatistical reservoir modeling : Oxford university press, new york, 376 p. *Google Scholar*.
- Deutsch, C. V. and Journel, A. G. (1998). Gslib. *Geostatistical software library and user's guide*, 369.
- Díaz, J. and Ercilla, G. (1993). Holocene depositional history of the fluviá—muga prodelta, northwestern mediterranean sea. *Marine Geology*, 111(1-2) :83–92.
- Diaz Padilla Guerrero, J. (1974). *Settlement of structures on shallow foundations : a probabilistic analysis*. PhD thesis, Massachusetts Institute of Technology.
- Domingos, P. (2012). A few useful things to know about machine learning. *Communications of the ACM*, 55(10) :78–87.
- Dominique, S., Marache, A., and Blancard, P. R.-S. (2010). Apport de la géostatistique à la connaissance de la morphologie naturelle du site de bordeaux (france). *ArcheoSciences. Revue d'archéométrie*, (34) :25–37.

- Douglas, B. and Olsen, R. (1981). Soil classification using electric cone penetrometer, symposium on cone penetration testing and experience, geo-technical engineering division, asce, st. Louis, Oct.
- Dubost, J. (2009). *Variabilité et incertitudes en géotechnique : de leur estimation à leur prise en compte*. PhD thesis, Bordeaux 1.
- Emery, X. (2001). Géostatistique linéaire, école des mines de paris.
- Emery, X. and Lantuéjoul, C. (2006). Tbsim : A computer program for conditional simulation of three-dimensional gaussian random fields via the turning bands method. *Computers & Geosciences*, 32(10) :1615–1628.
- Escobar, E. J. (2015). *Mise au point et exploitation d'une nouvelle technique pour la reconnaissance des sols : le PANDA 3*. PhD thesis, Université Blaise Pascal-Clermont-Ferrand II.
- Facciorusso, J. and Uzielli, M. (2004). Stratigraphic profiling by cluster analysis and fuzzy soil classification from mechanical cone penetration tests. *Proceedings of ISC-2 on Geotechnical and Geophysical Site Characterization, Porto, Millpress, Rotterdam*, pages 905–912.
- Favre, J.-L. (2000). Les différents types d'erreurs et leur prise en compte dans les calculs géotechniques. *Revue française de géotechnique*, (93) :11–20.
- Favre, J.-L. (2004). *Sécurité des ouvrages, risques : modélisation de l'incertain, fiabilité, analyse des risques*. Ellipses.
- Fayyad, U., Piatetsky-Shapiro, G., and Smyth, P. (1996). The kdd process for extracting useful knowledge from volumes of data. *Communications of the ACM*, 39(11) :27–34.
- Fenton, G. A. (1999a). Estimation for stochastic soil models. *Journal of Geotechnical and Geoenvironmental Engineering*, 125(6) :470–485.
- Fenton, G. A. (1999b). Random field modeling of cpt data. *Journal of Geotechnical and Geoenvironmental Engineering*, 125(6) :486–498.
- Fenton, G. A. and Griffiths, D. (2002). Probabilistic foundation settlement on spatially random soil. *Journal of Geotechnical and Geoenvironmental Engineering*, 128(5) :381–390.
- Fenton, G. A. and Griffiths, D. (2003). Bearing-capacity prediction of spatially random c φ soils. *Canadian geotechnical journal*, 40(1) :54–65.
- Fenton, G. A. and Griffiths, D. V. (2008). *Risk assessment in geotechnical engineering*. Wiley.
- Fenton, G. A. and Vanmarcke, E. H. (1990). Simulation of random fields via local average subdivision. *Journal of Engineering Mechanics*, 116(8) :1733–1749.

- Firouziandbandpey, S., Griffiths, D., Ibsen, L. B., and Andersen, L. V. (2014). Spatial correlation length of normalized cone data in sand : case study in the north of denmark. *Canadian Geotechnical Journal*, 51(8) :844–857.
- Firouziandbandpey, S., Ibsen, L. B., Griffiths, D., Vahdatirad, M., Andersen, L. V., and Sørensen, J. D. (2015). Effect of spatial correlation length on the interpretation of normalized cpt data using a kriging approach. *Journal of Geotechnical and Geoenvironmental Engineering*, 141(12) :04015052.
- Folle, D., Costa, J. F. C. L., Renard, D., Koppe, J. C., and Zingano, A. C. (2008). 3-d soil-resistance maps in the presence of a strong vertical trend. *Engineering Geology*, 102(1) :46–53.
- Freulon, X. and Fouquet, C. (1991). Remarques sur la pratique des bandes tournantes à trois dimensions. *Cahiers de géostatistique, Fascicule 1*, pages 101–117.
- Gachon, L. (1963). *Contribution à l'étude du quaternaire récent de la grande Limagne marno-calcaire : morphogénèse et pédogénèse*. Institut national de la recherche agronomique.
- Galinié, H., Laurent, A., Rodier, X., Breyse, D., Niandou, L. H., and Breul, P. (2003). Utilisation du pénétromètre dynamique de type panda en milieu urbain pour l'évaluation et la caractérisation du dépôt archéologique. *Revue d'archéométrie*, 27(1) :15–26.
- Gaouar, E. M. (1997). *Approche fiabiliste de la stabilité des barrages en terre par simulation de champs aléatoires*. PhD thesis.
- Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., and Rubin, D. B. (2014). *Bayesian data analysis*, volume 2. CRC press Boca Raton, FL.
- Geman, S. and Geman, D. (1984). Stochastic relaxation, gibbs distributions, and the bayesian restoration of images. *IEEE Transactions on pattern analysis and machine intelligence*, (6) :721–741.
- Gersch, W. and Yonemoto, J. (1977). Parametric time series models for multivariate eeg analysis. *Computers and Biomedical Research*, 10(2) :113–125.
- Ghanem, R. G. and Spanos, P. D. (1991). Spectral stochastic finite-element formulation for reliability analysis. *Journal of Engineering Mechanics*, 117(10) :2351–2372.
- Gneiting, T. (1999). The correlation bias for two-dimensional simulations by turning bands. *Mathematical geology*, 31(2) :195–211.
- Goh, A. T. and Goh, S. (2007). Support vector machines : their use in geotechnical engineering as illustrated using seismic liquefaction data. *Computers and Geotechnics*, 34(5) :410–421.
- Gómez-Hernández, J. J. and Cassiraga, E. F. (1994). Theory and practice of sequential simulation. In *Geostatistical simulations*, pages 111–124. Springer.

- Gong, W., Tien, Y.-M., Juang, C. H., Martin II, J. R., and Luo, Z. (2016). Optimization of site investigation program for improved statistical characterization of geotechnical property based on random field theory. *Bulletin of Engineering Geology and the Environment*, pages 1–15.
- Gourvès, R. (1991). Le panda–pénétrromètre dynamique léger à énergie variable. *LERMES CUST, Université Blaise Pascal, Clermont-Ferrand*.
- Gourves, R. and Barjot, R. (1995). The panda ultralight dynamic penetrometer. *Proceeding of ECSMFE. Copenhagen, Denmark*, pages 83–88.
- Griffiths, D. and Fenton, G. A. (2000). Influence of soil strength spatial variability on the stability of an undrained clay slope by finite elements. In *Slope stability 2000*, pages 184–193.
- Griffiths, D. and Fenton, G. A. (2001). Bearing capacity of spatially random soil : the undrained clay prandtl problem revisited. *Geotechnique*, 51(4) :351–360.
- Guerrero, O. (2014). *Comparaison et couplage de méthodes géophysiques pour l’amélioration des reconnaissances des sols dans les projets géotechniques en milieu périurbain*. PhD thesis, Université de Bordeaux.
- Haddani, Y. (2004). *Caractérisation et classification des milieux granulaires par géoendoscopie*. PhD thesis, Clermont-Ferrand 2.
- Haldar, S. and Sivakumar Babu, G. (2009). Design of laterally loaded piles in clays based on cone penetration test data : a reliability-based approach. *Geotechnique*, 59(7) :593–607.
- Han, J., Pei, J., and Kamber, M. (2011). *Data mining : concepts and techniques*. Elsevier.
- Handcock, M. S. and Stein, M. L. (1993). A bayesian analysis of kriging. *Technometrics*, 35(4) :403–410.
- Harr, M. E. (1987). Reliability-based design in civil engineering.
- Hastings, W. K. (1970). Monte carlo sampling methods using markov chains and their applications. *Biometrika*, 57(1) :97–109.
- Hecht-Nielsen, R. (1989). Kolmogorov’s mapping neural network existence theorem. In *Proceedings of the International Joint Conference in Neural Networks*, volume 3, pages 11–14.
- Hegazy, Y. A. (1998). *Delineating geostratigraphy by cluster analysis of piezocone data*. PhD thesis, Georgia Institute of Technology.
- Hegazy, Y. A. and Mayne, P. W. (2002). Objective site characterization using clustering of piezocone data. *Journal of Geotechnical and Geoenvironmental Engineering*, 128(12) :986–996.
- Hicks, M. A. and Samy, K. (2002). Influence of heterogeneity on undrained clay slope stability. *Quarterly Journal of Engineering Geology and Hydrogeology*, 35(1) :41–49.

- Hicks, M. A. and Spencer, W. A. (2010). Influence of heterogeneity on the reliability and failure of a long 3d slope. *Computers and Geotechnics*, 37(7) :948–955.
- Hoffman, M. D. and Gelman, A. (2014). The no-u-turn sampler : adaptively setting path lengths in hamiltonian monte carlo. *Journal of Machine Learning Research*, 15(1) :1593–1623.
- Honjo, Y. and Kuroda, K. (1991). A new look at fluctuating geotechnical data for reliability design. *Soils and Foundations*, 31(1) :110–120.
- Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5) :359–366.
- Hsein Juang, C., Chen, C. J., and Tien, Y.-M. (1999). Appraising cone penetration test based liquefaction resistance evaluation methods : artificial neural network approach. *Canadian Geotechnical Journal*, 36(3) :443–454.
- Hsu, C.-W., Chang, C.-C., Lin, C.-J., et al. (2003). A practical guide to support vector classification.
- Huber, M. (2013). Soil variability and its consequences in geotechnical engineering.
- Imanzadeh, S. (2013). *Effets des incertitudes et de la variabilité spatiale des propriétés des sols et des structures sur le dimensionnement des semelles filantes et des conduites enterrées*. PhD thesis, Université Sciences et Technologies-Bordeaux I.
- Isaaks, E. H. and Srivastava, R. M. (1989). Applied geostatistics.
- ISO, E. (2005). 22476-3 : 2005. *Geotechnical Investigation and Testing*.
- Jacquard, C. and Boutet, C. (2016). Correlations a partir du penetrometre statique. Journées Nationales de Géotechnique et de Géologie de l'Ingénieur.
- Jaffard, S. (1991). Pointwise smoothness, two-microlocalization and wavelet coefficients. *Publicacions Matemàtiques*, 35(1) :155–168.
- Jain, A. K., Mao, J., and Mohiuddin, K. M. (1996). Artificial neural networks : A tutorial. *Computer*, 29(3) :31–44.
- Jaksa, M. (2007). Modeling the natural variability of over-consolidated clay in adelaide, south australia. In *Proceedings of the Second International Workshop on Characterisation and Engineering Properties of Natural Soils*. London : Taylor and Francis, pages 2721–2751.
- Jaksa, M. and Fenton, G. (2002). Assessment of fractal behavior of soils.
- Jaksa, M., Goldsworthy, J., Fenton, G., Kaggwa, W., Griffiths, D., Kuo, Y., and Poulos, H. (2005). Towards reliable and effective site investigations.
- Jaksa, M., Kaggwa, W., and Brooker, P. (1999). Experimental evaluation of the scale of fluctuation of a stiff clay. In *Proc. 8th Int. conf. on the application of statistics and probability, Sydney*, volume 1, pages 415–422.

- Jaksa, M., Kaggwa, W., Fenton, G., and Poulos, H. (2003). A framework for quantifying the reliability of geotechnical investigations. In *9th International Conference on the Application of Statistics and Probability in Civil Engineering*, pages 1285–1291.
- Jaksa, M. B. (1995). *The influence of spatial variability on the geotechnical design properties of a stiff, overconsolidated clay*. PhD thesis.
- Jardine, R. (1995). One perspective of the pre-failure deformation characteristics of some geomaterials. In *Pre-failure deformation of geomaterials. Proceedings of the international symposium, 12-14 September 1994, Sapporo, Japan. 2 Vols.*
- Jeffreys, H. (1961). *Theory of probability* (3rd ed.) oxford university press. *MR0187257*.
- Jiménez, O. (2007). Caracterización geoestadística del subsuelo de la zona poniente del valle de México.
- Journel, A. G. and Huijbregts, C. J. (1978). *Mining geostatistics*. Academic press.
- Juang, C., Jiang, T., and Christopher, R. (2001). Three-dimensional site characterisation : neural network approach. *Geotechnique*, 51(9) :799–809.
- Juang, C. H., Yuan, H., Lee, D.-H., and Lin, P.-S. (2003). Simplified cone penetration test-based method for evaluating liquefaction resistance of soils. *Journal of geotechnical and geoenvironmental engineering*, 129(1) :66–80.
- Juárez-Camarena, M., Auvinet-Guichard, G., and Méndez-Sánchez, E. (2016). Geotechnical zoning of Mexico valley subsoil. *Ingeniería, Investigación y Tecnología*, 17(3) :297–308.
- Juarez-Camarena, M., Martinez, C., Auvinet, G., and Tenorio, A. (2014). Herramienta computacional para el análisis geoestadístico en geotecnia. *Función*, 5(2) :3.
- Keaveny, J., Nadim, F., and Lacasse, S. (1989). Autocorrelation functions for offshore geotechnical data. In *Structural Safety and Reliability* :, pages 263–270. ASCE.
- Kelly, R. and Huang, J. (2015). Bayesian updating for one-dimensional consolidation measurements. *Canadian Geotechnical Journal*, 52(9) :1318–1330.
- Kiang, M. Y. (2003). A comparative assessment of classification methods. *Decision Support Systems*, 35(4) :441–454.
- Koltermann, C. E. and Gorelick, S. M. (1996). Heterogeneity in sedimentary deposits : A review of structure-imitating, process-imitating, and descriptive approaches. *Water Resources Research*, 32(9) :2617–2658.
- Kotsiantis, S. B., Zaharakis, I. D., and Pintelas, P. E. (2007). Machine learning : a review of classification and combining techniques. *Artificial Intelligence Review*, 26(3) :159–190.
- Kreuter, H. (1996). *Ingenieurgeologische Aspekte geostatistischer Methoden*. Institutes für Bodenmechanik und Felsmechanik der Universität Fridericiana in Karlsruhe.

- Kulatilake, P. and Ghosh, A. (1988). An investigation into accuracy of spatial variation estimation using static cone penetrometer data. In *Proceedings of the 1st International Conference on Penetration Testing, Orlando, Fla*, pages 815–821.
- Kurup, P. U. and Griffin, E. P. (2006). Prediction of soil composition from cpt data using general regression neural network. *Journal of Computing in Civil Engineering*, 20(4) :281–289.
- Lacasse, S. and de Lamballerie, J. (1995). Statistical treatment of cpt data. In *Proceedings of the International Symposium on Cone Penetration Testing*, pages 4–5.
- Lacasse, S., Hoeg, K., Liu, Z., and Nadim, F. (2013). An homage to wilson tang : Reliability and risk in geotechnical practice—how wilson led the way. *Geotechnical Safety and Risk IV, Hong Kong, 4-6 Dec*, pages 3–26.
- Lacasse, S. and Nadim, F. (1997). Uncertainties in characterising soil properties. *Publikasjon-Norges Geotekniske Institutt*, 201 :49–75.
- Lacasse, S., Nadim, F., Shackelford, C., Nelson, P., and Roth, M. (1996). Uncertainty in the geologic environment : From theory and practice.
- Lantuéjoul, C. (2013). *Geostatistical simulation : models and algorithms*. Springer Science & Business Media.
- Lebon, P., Bosgiraud, J., Foin, R., and Armand, G. (2013). The full scale seal experiment—a seal industrial prototype for cigéo–13106. WM Symposia, 1628 E. Southern Avenue, Suite 9-332, Tempe, AZ 85282 (United States).
- Lepetit, L. (1998). Campagne d’essai sur les digues du rhône étude de faisabilité.
- Lepetit, L. (2002). *Etude d’une méthode de diagnostic de digues avec prise en compte du risque de liquéfaction*. PhD thesis, Université Blaise Pascal-Clermont-Ferrand II.
- Leroueil, S. and Hight, D. (2003). Behaviour and properties of natural soils and soft rocks. *Characterisation and engineering properties of natural soils*, 1 :29–254.
- Li, C.-C. and Der Kiureghian, A. (1993). Optimal discretization of random fields. *Journal of engineering mechanics*, 119(6) :1136–1154.
- Li, D.-Q., Cao, Z.-J., Ching, J., Dasaka, S. M., Huang, J., Jaksa, M., Nishimura, S.-i., Stuedlein, A., and Vessia, G. (2017). Incorporating spatial variability into geotechnical reliability-based design. *TC205/TC304 Discussion Groups, ISSMGE*.
- Li, D.-Q., Qi, X.-H., Cao, Z.-J., Tang, X.-S., Phoon, K.-K., and Zhou, C.-B. (2016a). Evaluating slope stability uncertainty using coupled markov chain. *Computers and Geotechnics*, 73 :72–82.
- Li, J. and Heap, A. D. (2008). A review of spatial interpolation methods for environmental scientists.

- Li, Y., Hicks, M., and Vardon, P. (2016b). Uncertainty reduction and sampling efficiency in slope designs using 3d conditional random fields. *Computers and Geotechnics*, 79 :159–172.
- Li, Z., Wang, X., Wang, H., and Liang, R. Y. (2016c). Quantifying stratigraphic uncertainties by stochastic simulation techniques based on markov random field. *Engineering Geology*, 201 :106–122.
- Liao, T. and Mayne, P. W. (2007). Post processing of cone penetration to assess seismic ground hazards, with specific application to the new madrid seismic zone. Technical report, Mid-America Earthquake Center.
- Liu, C. N. and Chen, C.-H. (2010). Spatial correlation structures of cpt data in a liquefaction site. *Engineering Geology*, 111(1) :43–50.
- Lloret-Cabot, M., Fenton, G. A., and Hicks, M. A. (2014). On the estimation of scale of fluctuation in geostatistics. *Georisk : Assessment and Management of Risk for Engineered Systems and Geohazards*, 8(2) :129–140.
- Lloret-Cabot, M., Hicks, M., and Nuttall, J. (2013). Investigating the scales of fluctuation of an artificial sand island. In *Proceedings of the International Conference, Geotechnical Installations in Geotechnical Engineering*, pages 192–197.
- Lloret Cabot, M., Hicks, M. A., and Van den Eijnden, A. P. (2012). Investigation of the reduction in uncertainty due to soil variability when conditioning a random field using kriging. *Geotechnique Letters (Online)*, 2 (July-September), 2012.
- Loke, M. (2002). Res2dinv ver. 3.50. *Rapid 2-D resistivity and IP inversion using the least square method*.
- Lumb, P. (1966). The variability of natural soils. *Canadian Geotechnical Journal*, 3(2) :74–97.
- Lumb, P. (1975). Spatial variability of soil properties. In *Proceedings of the second international conference on application of statistics and probability in soil and structural engineering, Aachen, Germany*, pages 397–421.
- Lunne, T., Robertson, P., and Powell, J. (1997). Cone penetration testing. *Geotechnical Practice*.
- Magnan, J.-P. (1982). *Les méthodes statistiques et probabilistes en mécanique des sols*. Presses ENPC.
- Magnan, J.-P. (1993). *Corrélations entre les propriétés des sols*. Ed. Techniques Ingénieur.
- Magnan, J.-P. (1996). Les applications des statistiques et probabilités en géotechnique au lcpc. *Bulletin des Laboratoires des Ponts et Chaussées*, (202) :25–35.
- Magnan, J.-P. (1997). *Description, identification et classification des sols*. Ed. Techniques Ingénieur.

- Magnan, J.-P. (2000). Quelques spécificités du problème des incertitudes en géotechnique. *Revue française de géotechnique*, (93) :3–9.
- Magnan, J.-P. (2002). L'organisation du travail en géotechnique : développement, normalisation et artisanat. *Revue française de géotechnique*, (99) :73–80.
- Magnan, J.-P. (2010). L'étude géotechnique et son adaptation aux grands projets.
- Mallat, S. G. (1989). A theory for multiresolution signal decomposition : the wavelet representation. *IEEE transactions on pattern analysis and machine intelligence*, 11(7) :674–693.
- Mantoglou, A. and Wilson, J. L. (1982). The turning bands method for simulation of random fields using line generation by a spectral method. *Water Resources Research*, 18(5) :1379–1394.
- Marache, A., Breyse, D., Piette, C., and Thierry, P. (2009). Geotechnical modeling at the city scale using statistical and geostatistical tools : The pessac case (france). *Engineering Geology*, 107(3) :67–76.
- Matheron, G. (1962). *Traité de géostatistique appliquée. 1 (1962)*, volume 1. Editions Technip.
- Matheron, G. (1970). La théorie des variables généralisées et ses applications. *Les Cahiers du Centre de Morphologie Mathématiques de Fontainebleau*, 5.
- Mayne, P. W., Coop, M. R., Springman, S. M., Huang, A.-B., and Zornberg, J. G. (2009). Geomaterial behavior and testing. In *Proc., 17th Int. Conf. on Soil Mechanics and Geotechnical Engineering*, volume 4, pages 2777–2872. IOS.
- Medina, Z. (2001). *Interpretación Geoestadística de Campañas de Reconocimiento del Sub-suelo*. PhD thesis, Tesis de Maestría.
- Mendes, R. M. and Lorandi, R. (2008). Analysis of spatial variability of spt penetration resistance in collapsible soils considering water table depth. *Engineering Geology*, 101(3) :218–225.
- Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equation of state calculations by fast computing machines. *The journal of chemical physics*, 21(6) :1087–1092.
- Młynarek, Z., Wierzbicki, J., and Wąyęski, W. (2007). Efficiency of selected statistical criteria in determination of geotechnical parameters from cptu. *Studia Geotechnica et Mechanica*, 29(1-2) :137–149.
- Młynarek, Z., Wierzbicki, J., and Wołyński, W. (2005). Use of interpolation methods for geotechnical profiling. *Studia geotechnica et mechanica*, 27(3-4) :5–13.
- Monnet, J. and Broucke, M. (2012). The use of a cluster analysis in a menard pressure-meter survey. *Proceedings of the Institution of Civil Engineers-Geotechnical engineering*, 165(6) :367–377.

- Monnet, J., Chapeau, C., and Godard, G. (2002). Caractérisation des sols pour le tunnel de la rocade nord de grenoble. *Revue française de géotechnique*, (101) :57–70.
- Monnet, J., Ploto, P., and Kadiri, I. (2007). Méthodes de reconnaissance croisées pour l’analyse de stabilité des digues soumises a érosion interne. *18ème Congrès Français de Mécanique (Grenoble 2007)*.
- Mora, R. (2013). Uso de métodos estadísticos para la identificación de capas de suelos volcánicos con el ensayo del cono de penetración en los terrenos de la universidad de costa rica, montes de oca, san josé, costa rica. *Revista Geológica de América Central*, (49).
- Murthy, S. K. (1998). Automatic construction of decision trees from data : A multi-disciplinary survey. *Data mining and knowledge discovery*, 2(4) :345–389.
- Nasreddine, K. (2004). *Effet de la rotation des contraintes sur le comportement des sols argileux*. PhD thesis, École nationale des ponts et chaussées (France).
- NF EN ISO 22476-1 (2010). Reconnaissance et essais géotechniques - essais en place . partie 1- essai au pénétromètre statique pointe électrique.
- NF EN ISO 22476-2 (2005). Reconnaissance et essais géotechniques - essais en place. partie 2 : Essai de pénétration dynamique.
- NF P 11-300 (1992). Exécution des terrassements-classification des matériaux utilisables dans la construction des remblais et des couches de forme d’infrastructures routières.
- NF P 94-056 (1996). Sols : Reconnaissance et essais - analyse granulométrique - méthode par tamisage à sec après lavage.
- NF P 94-114 (1990). Géotechnique - sols : Reconnaissance et essais - essai de pénétration dynamique type a.
- NF P 94-115 (1990). Géotechnique - sols : Reconnaissance et essais - sondage au pénétromètre dynamique type b.
- NF P 94-116 (1991). Sols : Reconnaissance et essais - essai de pénétration au carottier (spt).
- NF P 94-119 (1995). Sols : Reconnaissance et essais - essai au piézocône.
- NF P 94-251-1 (2005). Calcul géotechnique - partie 1 : Règles générales.
- NF P 94-252 (2007). Calcul géotechnique - partie 2 : Reconnaissance des terrains et essais.
- NF P 94-261 (2013). Justification des ouvrages géotechniques - norme d’application nationale de l’eurocode 7 - fondations superficielles.
- NF P 94-262 (2012). Justification des ouvrages géotechniques - norme d’application nationale de l’eurocode 7 - fondations profondes.
- NF P 94-500 (2013). Missions d’ingénierie géotechnique : classification et spécifications.

- NF XP P 94-105 (2000). Sols : reconnaissance et essais - contrôle de la qualité du compactage - méthode au pénétromètre à énergie variable.
- Nie, X., Zhanga, J., Huang, H., Liu, Z., and Lacasse, S. (2015). Scale of fluctuation for geotechnical probabilistic analysis. *Proceedings of ISGSR2015, Rotterdam, The Netherlands, October*.
- Olsen, R. and Mitchell, J. (1995). Cpt stress normalization and prediction of soil classification. In *Proceedings of the International Symposium on Cone Penetration Testing, CPT*, volume 95, pages 257–262.
- Omre, H. (1987). Bayesian kriging—merging observations and qualified guesses in kriging. *Mathematical Geology*, 19(1) :25–39.
- Omre, H. and Halvorsen, K. B. (1989). The bayesian bridge between simple and universal kriging. *Mathematical Geology*, 21(7) :767–786.
- Orr, T. L. (2016). Defining and selecting characteristic values of geotechnical parameters for designs to eurocode 7. *Georisk : Assessment and Management of Risk for Engineered Systems and Geohazards*, 11(1) :103–115.
- Paikowsky, S., Birgisson, B., McVay, M., Nguyen, T., Kuo, C., Baecher, G., Ayyub, B., Steensen, K., O’Malley, K., Chernauskas, L., et al. (2004). Load and resistance factor design for deep foundations, nchrp report 507. *Transportation Research Board, Washington, DC*.
- Paliwal, M. and Kumar, U. A. (2009). Neural networks and statistical techniques : A review of applications. *Expert systems with applications*, 36(1) :2–17.
- Papaoiannou, I. and Straub, D. (2016). Learning soil parameters and updating geotechnical reliability estimates under spatial variability—theory and application to shallow foundations. *Georisk : Assessment and Management of Risk for Engineered Systems and Geohazards*, 11(1) :116–128.
- Pardo-Iguzquiza, E. and Chica-Olmo, M. (1993). The fourier integral method : an efficient spectral method for simulation of random fields. *Mathematical Geology*, 25(2) :177–217.
- Patil, A., Huard, D., and Fonnesbeck, C. J. (2010). Pymc : Bayesian stochastic modelling in python. *Journal of statistical software*, 35(4) :1.
- Pérez, N., Sau, N., Devicenzi, M., Arroyo, M., and Pineda, J. (2013). Pressiometric and non-pressiometric tools on a mediterranean deltaic deposit. In *Proceedings of the 18th International Conference on Soil Mechanics and Geotechnical Engineering, Paris*.
- Phoon, K.-K. and Ching, J. (2014). *Risk and reliability in geotechnical engineering*. CRC Press.
- Phoon, K.-K. and Kulhawy, F. H. (1999). Characterization of geotechnical variability. *Canadian Geotechnical Journal*, 36(4) :612–624.

- Phoon, K.-K., Quek, S.-T., and An, P. (2003). Identification of statistically homogeneous soil layers using modified bartlett statistics. *Journal of Geotechnical and Geoenvironmental Engineering*, 129(7) :649–659.
- Phoon, K.-K., Quek, S.-T., and An, P. (2004). Geostatistical analysis of cone penetration test (cpt) sounding using the modified bartlett test. *Canadian geotechnical journal*, 41(2) :356–365.
- Pilz, J. and Spöck, G. (2008). Why do we need and how should we implement bayesian kriging methods. *Stochastic Environmental Research and Risk Assessment*, 22(5) :621–632.
- Plumelle, C., Jun Cui, Y., Fabre, D., Fouché, O., Hirschauer, A., and Tabbagh, A. (2013). *Théorie et pratique de la géotechnique*. Le Moniteur.
- Popescu, R. (1995). *Stochastic variability of soil properties : Data analysis, digital simulation, effects on system behavior*. PhD thesis.
- Popescu, R., Deodatis, G., and Nobahar, A. (2005a). Effects of random heterogeneity of soil properties on bearing capacity. *Probabilistic Engineering Mechanics*, 20(4) :324–341.
- Popescu, R., Prevost, J., and Deodatis, G. (2005b). 3d effects in seismic liquefaction of stochastically variable soil deposits.
- Popescu, R., Prevost, J. H., and Deodatis, G. (1998). Spatial variability of soil properties : two case studies. In *Geotechnical Earthquake Engineering and Soil Dynamics III*, pages 568–579. ASCE.
- Rackwitz, R. (2000). Reviewing probabilistic soils modelling. *Computers and Geotechnics*, 26(3) :199–223.
- Reiffsteck, P., Benoit, J., Hamel, M., and Vaillant, J.-M. (2016). Proposition d’une méthode de classification basé sur les paramètres de forage. Journées Nationales de Géotechnique et de Géologie de l’Ingénieur.
- Reiffsteck, P., Lossy, D., and Benoît, J. (2012). *Forages, Sondages et Essais In Situ Géotechniques—les outils pour la reconnaissance des sols et des roches*. Presses des Ponts.
- Rétháti, L. (1988). *Probabilistic solutions in geotechnics*. Elsevier, Amsterdam.
- Reynolds, J. M. (2011). *An introduction to applied and environmental geophysics*. John Wiley & Sons.
- Robertson, P. (1990). Soil classification using the cone penetration test. *Canadian Geotechnical Journal*, 27(1) :151–158.
- Robertson, P. (2009). Interpretation of cone penetration tests—a unified approach. *Canadian Geotechnical Journal*, 46(11) :1337–1355.
- Robertson, P. and Cabal, K. (2015). Guide to cone penetration testing for geotechnical engineering. *Gregg Drilling and Testing Inc., USA,*, pages 6–15.

- Robertson, P. and Wride, C. (1998). Evaluating cyclic liquefaction potential using the cone penetration test. *Canadian Geotechnical Journal*, 35(3) :442–459.
- Robertson, P. K., Campanella, R., Gillespie, D., and Greig, J. (1986). Use of piezometer cone data. In *Use of in situ tests in geotechnical engineering*, pages 1263–1280. ASCE.
- Rudel, A. (1957). Découverte de galets à facettes en limagne de clermont. *Sci. nat. Auvergne*, 3 :47–50.
- Rumelhart, D. E., Hinton, G. E., Williams, R. J., et al. (1988). Learning representations by back-propagating errors. *Cognitive modeling*, 5(3) :1.
- Sabeti, M., Katebi, S., and Boostani, R. (2009). Entropy and complexity measures for eeg signal classification of schizophrenic and control participants. *Artificial Intelligence in Medicine*, 47(3) :263–274.
- Salgado, R. (2012). The mechanics of cone penetration : Contributions from experimental and theoretical studies. *Geotechnical and Geophysical Characterization*, 4 :131–135.
- Salgado, R. and Kim, D. (2013). Reliability analysis of load and resistance factor design of slopes. *Journal of Geotechnical and Geoenvironmental Engineering*, 140(1) :57–73.
- Salgado, R., Prezzi, M., and Ganju, E. (2015). Assessment of site variability from analysis of cone penetration test data.
- Salloum, N. (2015). *Evaluation de la variabilité spatiale des paramètres géotechnique du sol à partir de mesures géophysiques : application à la plaine alluviale de Nahr-Beyrouth (Liban)*. PhD thesis, Université Grenoble Alpes.
- Samouëlian, A., Cousin, I., Tabbagh, A., Bruand, A., and Richard, G. (2005). Electrical resistivity survey in soil science : a review. *Soil and Tillage research*, 83(2) :173–193.
- Samui, P. (2008). Support vector machine applied to settlement of shallow foundations on cohesionless soils. *Computers and Geotechnics*, 35(3) :419–427.
- Samui, P. (2012). Three-dimensional site characterization model of bangalore using support vector machine. *ISRN Soil Science*, 2012.
- Samui, P. and Sitharam, T. (2007). Spatial variability of spt data using ordinary and disjunctive kriging.
- Sanglerat, G. (1972). *The penetrometer and soil exploration*, volume 1. Elsevier.
- Santamarina, J. and Park, J. (2016). Geophysical properties of soil. *Journal and news of the Australian Geomechanics Society*, 51(4) :183.
- Schmertmann, J. H. (1978). Guidelines for cone penetration test.(performance and design). Technical report.
- Schmidhuber, J. (2015). Deep learning in neural networks : An overview. *Neural networks*, 61 :85–117.

- Schnaid, F. (2005). Geocharacterisation and properties of natural soils by in situ tests. In *Proceedings of the International Conference on Soil Mechanics and Geotechnical Engineering*, volume 16, page 3. AA BALKEMA PUBLISHERS.
- Schneider, H. R. and Schneider, M. A. (2012). Dealing with uncertainties in ec7 with emphasis on determination of characteristic soil properties. *Modern Geotechnical Design Codes of Practice (Arnold P, Fenton GA, Hicks MA and Schweckendiek T (eds))*. IOS Press, Rotterdam, the Netherlands, pages 87–101.
- Schneider, J. A., Randolph, M. F., Mayne, P. W., and Ramsey, N. R. (2008). Analysis of factors influencing soil classification using normalized piezocone tip resistance and pore pressure parameters. *Journal of geotechnical and geoenvironmental engineering*, 134(11) :1569–1586.
- Shahin, M., Maier, H., and Jaksa, M. (2005). Investigation into the robustness of artificial neural networks for a case study in civil engineering.
- Shahin, M. A., Jaksa, M. B., and Maier, H. R. (2001). Artificial neural network applications in geotechnical engineering. *Australian Geomechanics*, 36(1) :49–62.
- Shahin, M. A., Jaksa, M. B., and Maier, H. R. (2008). State of the art of artificial neural networks in geotechnical engineering. *Electronic Journal of Geotechnical Engineering*, 8 :1–26.
- Shahin, M. A., Jaksa, M. B., and Maier, H. R. (2009). Recent advances and future challenges for artificial neural systems in geotechnical engineering applications. *Advances in Artificial Neural Systems*, 2009 :5.
- Shahin, M. A., Maier, H. R., and Jaksa, M. B. (2003). Closure to “predicting settlement of shallow foundations using neural networks” by mohamed a. shahin, holger r. maier, and mark b. jaksa. *Journal of Geotechnical and Geoenvironmental Engineering*, 129(12) :1175–1177.
- Shahin, M. A., Maier, H. R., and Jaksa, M. B. (2004). Data division for developing neural networks applied to geotechnical engineering. *Journal of Computing in Civil Engineering*, 18(2) :105–114.
- Shahrour, I. and Gourvés, R. (2005). *Reconnaissance des terrains in situ*. hermes science.
- Sivia, D. and Skilling, J. (2006). *Data analysis : a Bayesian tutorial*. OUP Oxford.
- Smith, M. (1993). *Neural networks for statistical modeling*. Thomson Learning.
- Soulie, M. (1984). Geostatistical applications in geotechnics. *Geostatistics for natural resources characterization, Pt, 2* :703–730.
- Soulie, M., Montes, P., and Silvestri, V. (1990). Modelling spatial variability of soil parameters. *Canadian Geotechnical Journal*, 27(5) :617–630.
- Spanos, P. D. and Ghanem, R. (1989). Stochastic finite element expansion for random media. *Journal of engineering mechanics*, 115(5) :1035–1053.

- Specht, D. F. (1990). Probabilistic neural networks. *Neural networks*, 3(1) :109–118.
- Specht, D. F. (1991). A general regression neural network. *IEEE transactions on neural networks*, 2(6) :568–576.
- Straub, D. and Papaioannou, I. (2014). Bayesian analysis for learning and updating geotechnical parameters and models with measurements. *Risk and reliability in geotechnical engineering. Part II. Chapter 5*, 5 :221–264.
- Sudret, B. and Der Kiureghian, A. (2000). *Stochastic finite element methods and reliability : a state-of-the-art report*. Department of Civil and Environmental Engineering, University of California Berkeley.
- Sulewska, M. J. (2011). Applying artificial neural networks for analysis of geotechnical problems. *Computer Assisted Methods in Engineering and Science*, 18(4) :231–241.
- Tam, K. Y. and Kiang, M. Y. (1992). Managerial applications of neural networks : the case of bank failure predictions. *Management science*, 38(7) :926–947.
- Tang, W. H. (1979). Probabilistic evaluation of penetration resistances. *Journal of Geotechnical and Geoenvironmental Engineering*, 105(ASCE 14902 Proc Paper).
- Tantalla, J., Prevost, J.-H., and Deodatis, G. (2001). Spatial variability of soil properties in slope stability analysis : Fragility curve generation. *Structural Safety and Reliability : ICOSSAR'01*, page 2001.
- Terzaghi, K. (1961). Past and future of applied soil mechanics. *J. Boston Soc. Civ Engr*, 68 :110–139.
- Terzaghi, K., Peck, R. B., and Mesri, G. (1996). *Soil mechanics in engineering practice*. John Wiley & Sons.
- Tompson, A. F., Ababou, R., and Gelhar, L. W. (1989). Implementation of the three-dimensional turning bands random field generator. *Water Resources Research*, 25(10) :2227–2243.
- Torrence, C. and Compo, G. P. (1998). A practical guide to wavelet analysis. *Bulletin of the American Meteorological society*, 79(1) :61–78.
- Treadwell, D. D. (1976). *The influence of gravity, prestress, compressibility, and layering on soil resistance to static penetration*. University of California, Berkeley.
- Tumay, M. (1985). Field calibration of electric cone penetrometers in soft soil. executive summary. Technical report.
- Uzielli, M. (2004). *Variability of stress-normalized CPT parameters and application to seismic liquefaction initiation analysis*. PhD thesis, PhD dissertation, University of Florence.
- Uzielli, M. (2008). Statistical analysis of geotechnical data. *W : Huang A., Mayne PW Geotechnical and Geophysical Site Characterization*, pages 173–193.

- Uzielli, M., Facciorusso, J., and Mayne, P. (2010). Spatial characterization of vs at amherst nges site from scpt using bayesian kriging. In *Proceedings, 2nd Intl. Symposium on Cone Penetration Testing*, pages 177–184.
- Uzielli, M., Lacasse, S., Nadim, F., and Phoon, K. (2006). Soil variability analysis for geotechnical practice. *Characterization and engineering properties of natural soils*, 3 :1653–1752.
- Uzielli, M., Vannucchi, G., and Phoon, K. (2005). Random field characterisation of stress-normalised cone penetration testing parameters. *Geotechnique*, 55(1) :3–20.
- Van den Eijnden, A. (2010). Conditional simulation for characterising the spatial variability of sand state.
- Vanmarcke, E. (1983). Random fields : Analysis and synthesis. *MIT Press, Cambridge Massachussets*, page 185.
- Vanmarcke, E. H. (1977a). Probabilistic modeling of soil profiles. *Journal of the geotechnical engineering division*, 103(11) :1227–1246.
- Vanmarcke, E. H. (1977b). Reliability of earth slopes. *Journal of the Geotechnical Engineering Division*, 103(11) :1247–1265.
- Vanmarcke, E. H. (1978). Probabilistic characterization of soil profiles. In *Site Characterization & Exploration*, pages 199–219. ASCE.
- Villavicencio, G. (2009). Méthodologie pour evaluer la stabilité mécanique des barrages de résidus miniers. *Doctor of Philosophy Dissertation. University Blaise Pascal, Clermont Ferrand*.
- Vivitrat, V. (1978). *Cone penetration in clays*. PhD thesis.
- Wackernagel, H. (2013). *Multivariate geostatistics : an introduction with applications*. Springer Science & Business Media.
- Wang, Y., Akeju, O. V., and Cao, Z. (2016a). Bayesian equivalent sample toolkit (best) : an excel vba program for probabilistic characterisation of geotechnical properties from limited observation data. *Georisk : Assessment and Management of Risk for Engineered Systems and Geohazards*, 10(4) :251–268.
- Wang, Y., Au, S.-K., and Cao, Z. (2010). Bayesian approach for probabilistic characterization of sand friction angles. *Engineering Geology*, 114(3) :354–363.
- Wang, Y., Cao, Z., and Li, D. (2016b). Bayesian perspective on geotechnical variability and site characterization. *Engineering Geology*, 203 :117–125.
- Wang, Y., Cao, Z., and Li, D. (2016c). Bayesian perspective on geotechnical variability and site characterization. *Engineering Geology*, 203 :117–125.
- Wang, Y., Huang, K., and Cao, Z. (2013). Probabilistic identification of underground soil stratification using cone penetration tests. *Canadian Geotechnical Journal*, 50(7) :766–776.

- Waschkowski, E. (1983). Essais de penetration-le penetrometre dynamique. *BULL LIAISON LAB PONTS CHAUSS*, (125).
- Webster, R. (1973). Automatic soil-boundary location from transect data. *Mathematical geology*, 5(1) :27–37.
- Webster, R. and Beckett, P. (1968). Quality and usefulness of soil maps. *Nature*, 219 :680–682.
- Webster, R. and Wong, I. (1969). A numerical procedure for testing soil boundaries interpreted from air photographs. *Photogrammetria*, 24(2) :59–72.
- Wickremesinghe, D. and Campanella, R. G. (1993). Scale of fluctuation as a descriptor of soil variability. In *Proceedings of the Conference on Probabilistic Methods in Geotechnical Engineering. Edited by KS Li and S.-C. Lo. AA Balkema, Rotterdam*, pages 233–239.
- Wickremesinghe, D. S. (1989). *Statistical characterization of soil profiles using in situ tests*. PhD thesis, University of British Columbia.
- Ximenez, E. and Romana, R. (1983). The formulation of a soil model from penetrometer probabilistic information. application to differential settlement prediction. In *International conference on applications of statistics and probability in soil and structural engineering. 4*, pages 1601–1613.
- Zhang, G. P. (2000). Neural networks for classification : a survey. *IEEE Transactions on Systems, Man, and Cybernetics, Part C (Applications and Reviews)*, 30(4) :451–462.
- Zhang, J., Boothroyd, P., Calvello, M., Eddleston, M., Grimal, A. C., Iason, P., Luo, Z., Wang, Y., and Walter, H. (2017). Bayesian method : A natural tool for processing geotechnical information. *TC205/TC304 Discussion Groups, ISSMGE*.
- Zhang, J. and Ellingwood, B. (1994). Orthogonal series expansions of random fields in reliability analysis. *Journal of Engineering Mechanics*, 120(12) :2660–2677.
- Zhang, J., Tang, W. H., Zhang, L., and Huang, H. (2012). Characterising geotechnical model uncertainty by hybrid markov chain monte carlo simulation. *Computers and Geotechnics*, 43 :26–36.
- Zhang, J., Zhang, L., and Tang, W. H. (2009). Bayesian framework for characterizing geotechnical model uncertainty. *Journal of Geotechnical and Geoenvironmental Engineering*, 135(7) :932–940.
- Zhang, Z. and Tumay, M. (2003). Non-traditional approaches in soil classification derived from the cone penetration test. *Probabilistic site characterization at the national geotechnical experimentation sites*, pages 101–149.
- Zhang, Z. and Tumay, M. T. (1996). The reliability of soil classification derived from cone penetration test. ASCE.
- Zhou, S. (1997). *Caractérisation des sols de surface à l'aide du pénétromètre à énergie variable type «PANDA»*. PhD thesis.

Zhu, H. and Zhang, L. M. (2013). Characterizing geotechnical anisotropic spatial variations using random field theory. *Canadian Geotechnical Journal*, 50(7) :723–734.

Annexes

A. Annexes A

A.1 Essais de caractérisation au laboratoire

Dans cette annexe on présente les résultats des analyses granulométriques des matériaux granulaires utilisés pour la confection des éprouvettes au laboratoire. Les courbes granulométriques ont été obtenues par tamisage des grains en conformité à la norme (NF P 94-056 (1996)).

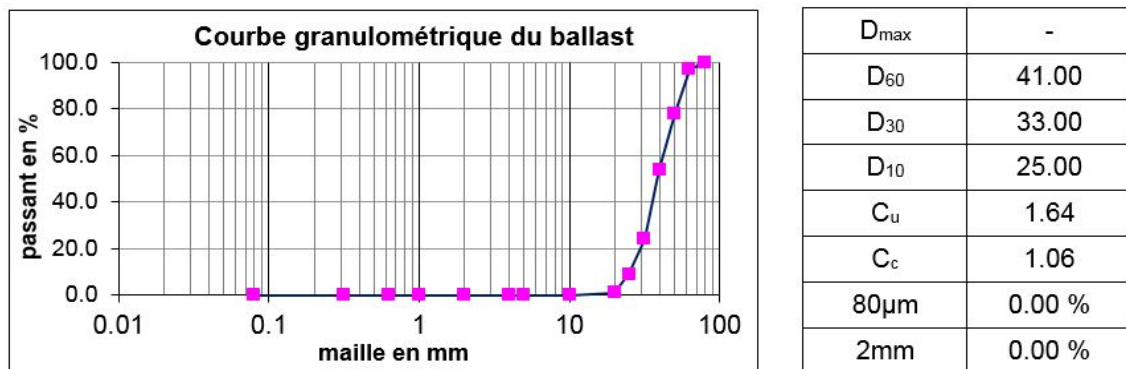


FIGURE A.1 – Analyse granulométrique par tamisage du ballast

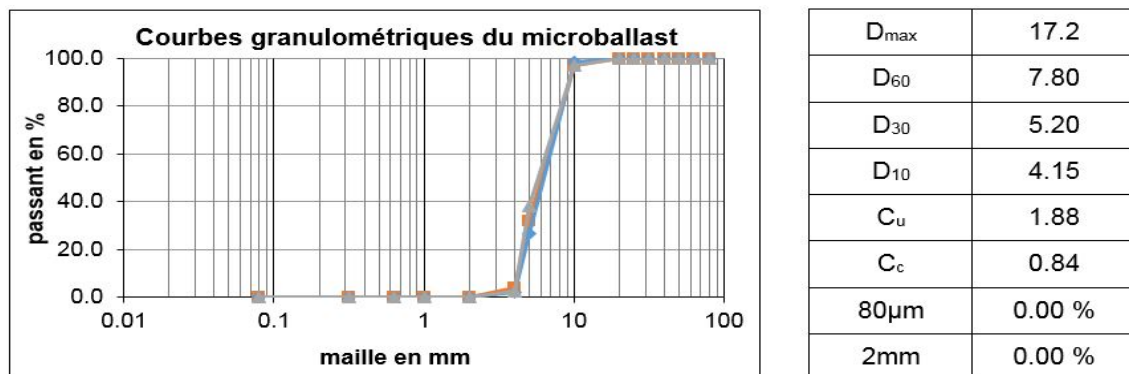
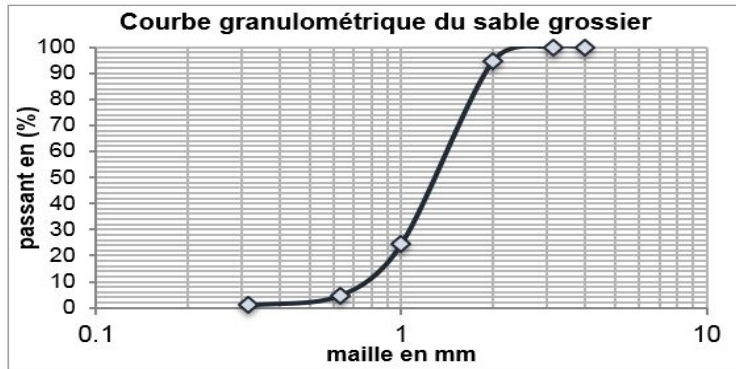
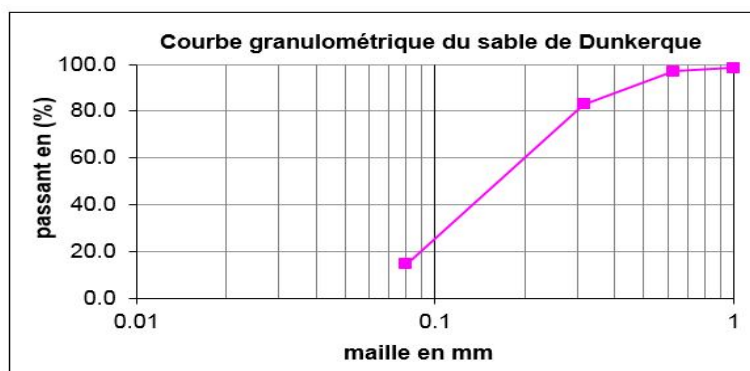


FIGURE A.2 – Analyse granulométrique par tamisage du microballast



D_{max}	17.2
D_{60}	1.6
D_{30}	1.20
D_{10}	0.75
C_u	2.1
C_c	1.2
80 μ m	0.00 %
2mm	94.5 %

FIGURE A.3 – Analyse granulométrique par tamisage du sable grossier



D_{max}	1
D_{60}	0.2
D_{30}	0.11
D_{10}	0.07
C_u	2.82
C_c	0.78
80 μ m	16.00 %
2mm	99.40 %

FIGURE A.4 – Analyse granulométrique par tamisage du sable de Dunkerque

A.2 Essais CPTU sur le site de Castelló d'Empúries

La figure (Fig. A.5) indique les résultats des deux essais au piezocône jusqu'à 5 m de profondeur réalisés par l'entreprise IGEOtest sur le site de Castello d'Empuries.

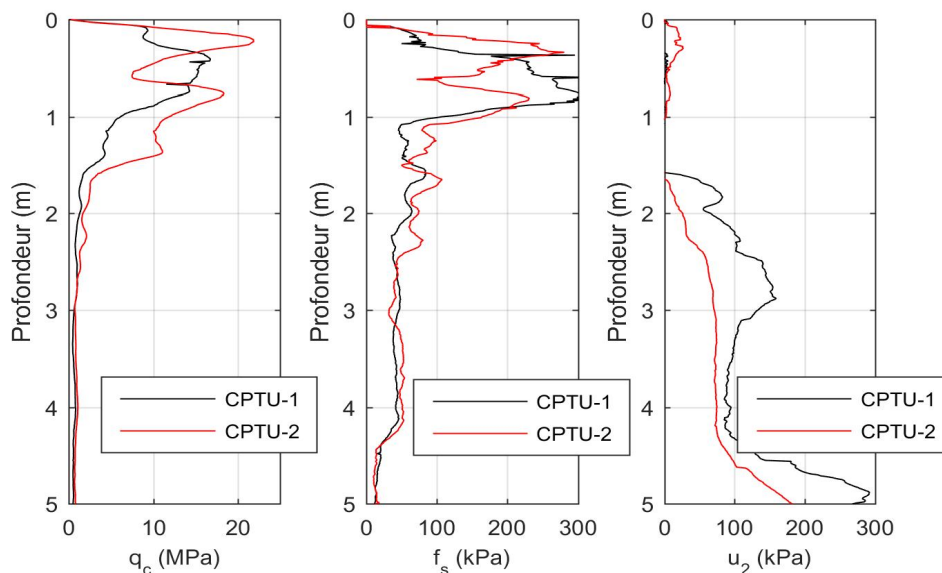


FIGURE A.5 – Résultats des essais CPTU sur le site de Castello

Ces essais ont été analysés selon l'abaque normalisé de Robertson (Fig.A.6). On observe que la plupart des mesures sont situées dans des sols surconsolidés avec des variations notables du type de sol (échantillons classés entre les zones 8 à 3 au sens de la classification de Robertson). Ceci est en accord avec la géologie du site, constituée par des dépôts deltaïques sur une plaine alluviale, avec des alternances entre dépôts sableux et argilo-limoneux.

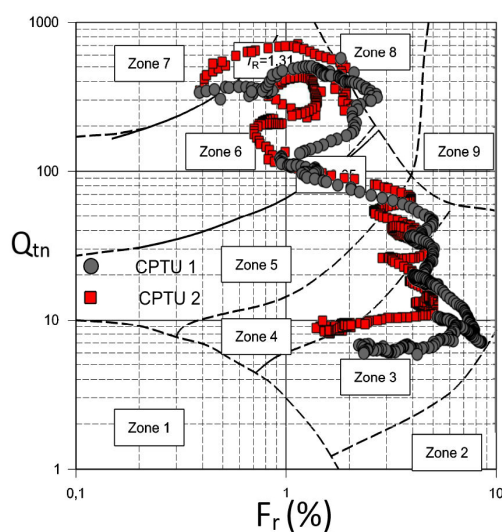


FIGURE A.6 – Classification des sols du site selon l'abaque de (Robertson (1990))

A.3 Transformée par ondelettes du profil de résistance dynamique q_d

Les ondelettes sont des outils de traitement du signal et de l'image, intégrant simultanément des paramètres de localisation et d'échelle. Ici on montre l'application de la transformée par ondelettes continue afin de déterminer la succession des couches à partir de l'analyse du profil de résistance dynamique q_d . Pour ce faire on a suivi une approche similaire à celle proposée par (Ching et al. (2015a)) avec les recommandations de (Torrence and Compo (1998)), en remplaçant le domaine temporel par le domaine spatial. Les principales étapes de la démarche suivie sont décrites ci-dessous :

- i. régularisation du pénétrogramme afin d'obtenir un pas d'échantillonnage constant ;
- ii. afin de faciliter le calcul de la transformée de Fourier rapide (FFT) on utilise la technique de complétion par des zéros (*zero-padding*). Elle consiste à augmenter le nombre de points d'analyse par des zéros pour affiner la précision spectrale. On obtient ainsi une séquence dyadique (nombre de points égal à une puissance de 2) ;
- iii. la transformée en ondelettes continue du profil de mesures de la résistance de pointe dynamique est décrite ci-dessous :

$$W_{q_d}(u, s) = \int_0^D q_d(z) \cdot \psi_{u,s}(z) dz \quad (\text{A.1})$$

où z est la profondeur, D la profondeur totale du sondage, $\psi_{u,s}$ la famille d'ondelettes positionnées en coordonnée u avec une échelle s . La transformée en ondelettes revient à réaliser le produit de convolution entre le profil de résistance dynamique q_d et des fonctions oscillantes (les ondelettes), tout en intégrant un aspect multiéchelle (ou multirésolution).

- iv. en ce que concerne le choix de l'ondelette mère, l'ondelette dérivée première de gaussienne ($m = 1$) est retenue, tel que suggéré par (Ching et al. (2015a)). Ce choix est basé sur le théorème proposé par (Jaffard (1991)), qui établit la relation entre une discontinuité (quantifiée par l'irrégularité de Lipschitz) avec $\log(s)$ et $\log[W(u, s)]$. Ainsi, le nombre de moments d'évanouissement m de l'ondelette mère doit être inférieur à la valeur de l'irrégularité de Lipschitz. Pour détecter les positions (profondeurs) où, pour une échelle d'observation donnée, se produit un changement ou une variation du profil de résistance dynamique, on ne s'intéresse qu'à la dérivée du premier ordre ;
- v. une fois que la transformée est calculée, il faut identifier les lignes de maxima du module de la transformée ($|W_{q_d}(u, s)|$) pour les différentes échelles s de l'ondelette. Finalement, ces lignes formées par les maxima locaux représentent des discontinuités sur plusieurs échelles, que l'on cherche à associer aux limites entre les couches.

Cette approche a été appliquée aux sondages réalisés en laboratoire présentés auparavant dans cette partie. Les résultats obtenus sur le pénétrogramme en pointe de 4 cm^2 réalisés dans le premier moule (HET1) sont illustrés par la figure (Fig.A.7), où est montré à droite le spectre des ondelettes. La ligne bleue en pointillés représente le cône d'influence, au delà duquel les résultats calculés ne sont plus fiables. Les lignes rouges représentent les lignes de maxima. On peut observer plusieurs lignes de maxima aux faibles échelles, équivalentes aux périodes inférieures à 25 cm. Or à partir de 50 cm on distingue très nettement 5 lignes de maxima dont trois signalent correctement les interfaces à identifier. Cependant, celle située

plus en surface ne correspond pas à une limite entre couche et la ligne de maxima située à environ 1 mètre de profondeur semble coïncider avec l'effet de fond de moule. Bien que des résultats similaires aient été trouvés pour les autres essais en laboratoire, on n'a pas pu formaliser le choix de ces lignes de maxima pour différencier les variations à faible échelle causées par la variabilité naturelle du matériau ou les erreurs de mesures, des changements dans la tendance du profil étudié qui seraient associées aux positions des interfaces entre couches. Pour palier à ce problème on pourrait étalonner ce choix à partir d'une base de données.

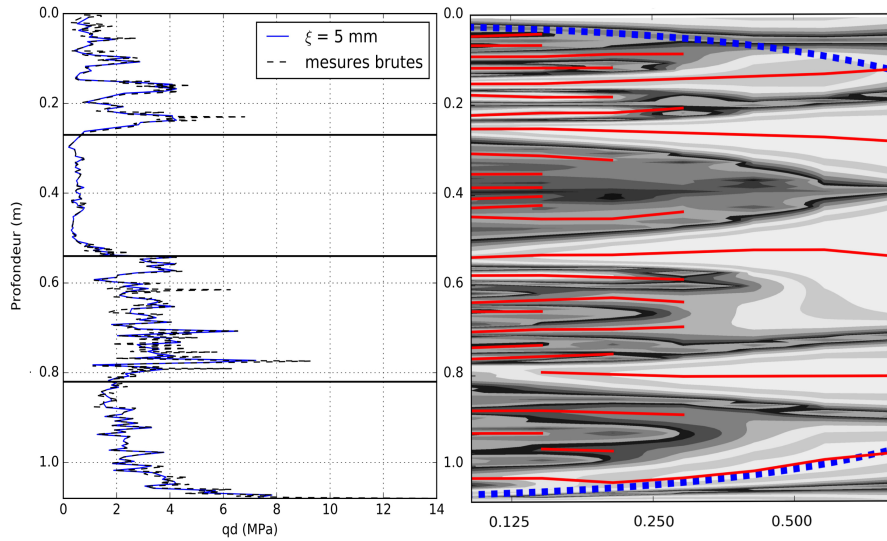


FIGURE A.7 – Résultats de l'analyse par la transformée par ondelettes du profil q_d issue du sondage effectué avec une pointe de 4 cm^2 dans le moule HET 1

Dans (Ching et al. (2015a)) une approche probabiliste a été proposée pour le choix des lignes de maxima correspondant aux limites entre couches. La probabilité qu'une ligne de maxima représente un changement dans le profil de mesures (l'indice de comportement I_c est le paramètre analysé) est donnée par l'équation suivante :

$$p_{jump} = \frac{1}{1 + \exp[-(4.398 \cdot L + 3.429 \cdot M + 5.025)]} \quad (\text{A.2})$$

avec :

$$L = \log(s_{max}) \quad (\text{A.3})$$

où s_{max} (m) est la valeur la plus grande de l'échelle utilisée. Le paramètre M est défini ci-dessous :

$$M = \frac{1}{n} \sum_{i=1}^n WI_c(u, s_i) \quad \text{avec} \quad s_n = s_{max} \quad (\text{A.4})$$

L'équation (éq.A.2) a été obtenue au moyen d'une régression logistique à partir d'une base de sondages constituée par un total de 50 essais CPT.

A.4 Courbes statistiques calculées pour les essais in situ analysés par les experts

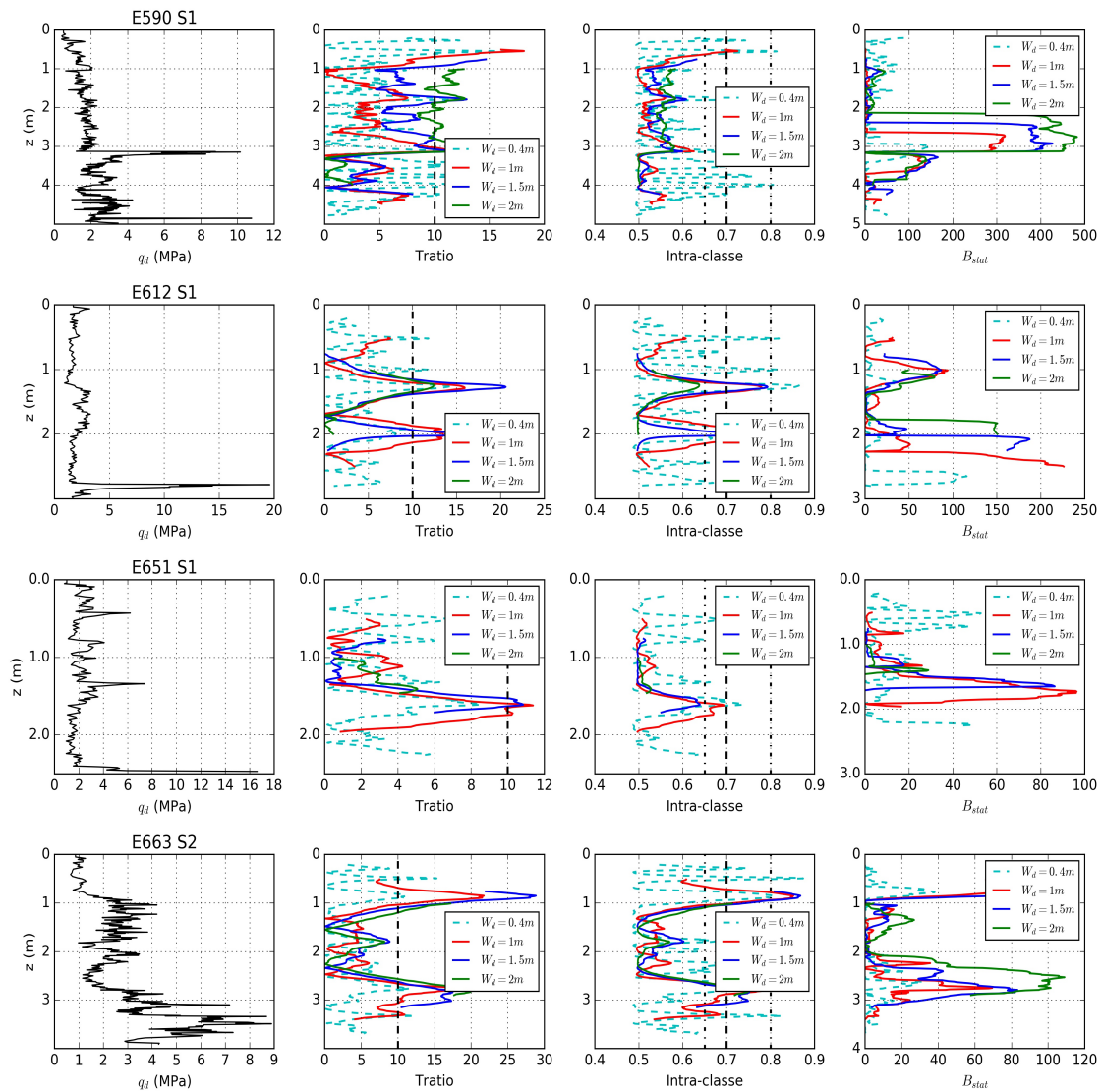


FIGURE A.8 – Courbes des tests statistiques selon la hauteur de la fenêtre utilisée pour les sondages in-situ analysés par les experts

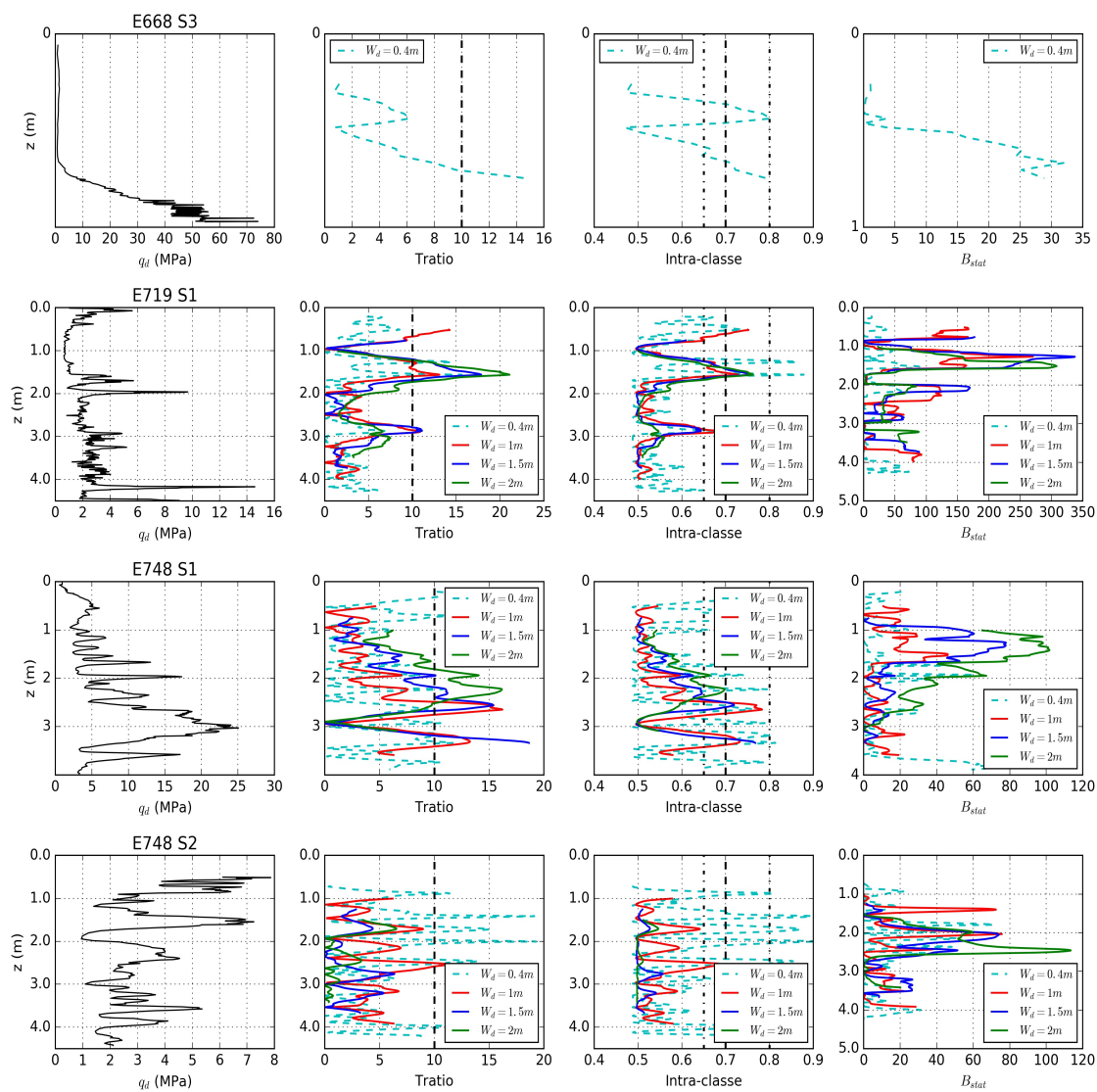


FIGURE A.9 – Courbes des tests statistiques selon la hauteur de la fenêtre utilisée pour les sondages in-situ analysés par les experts

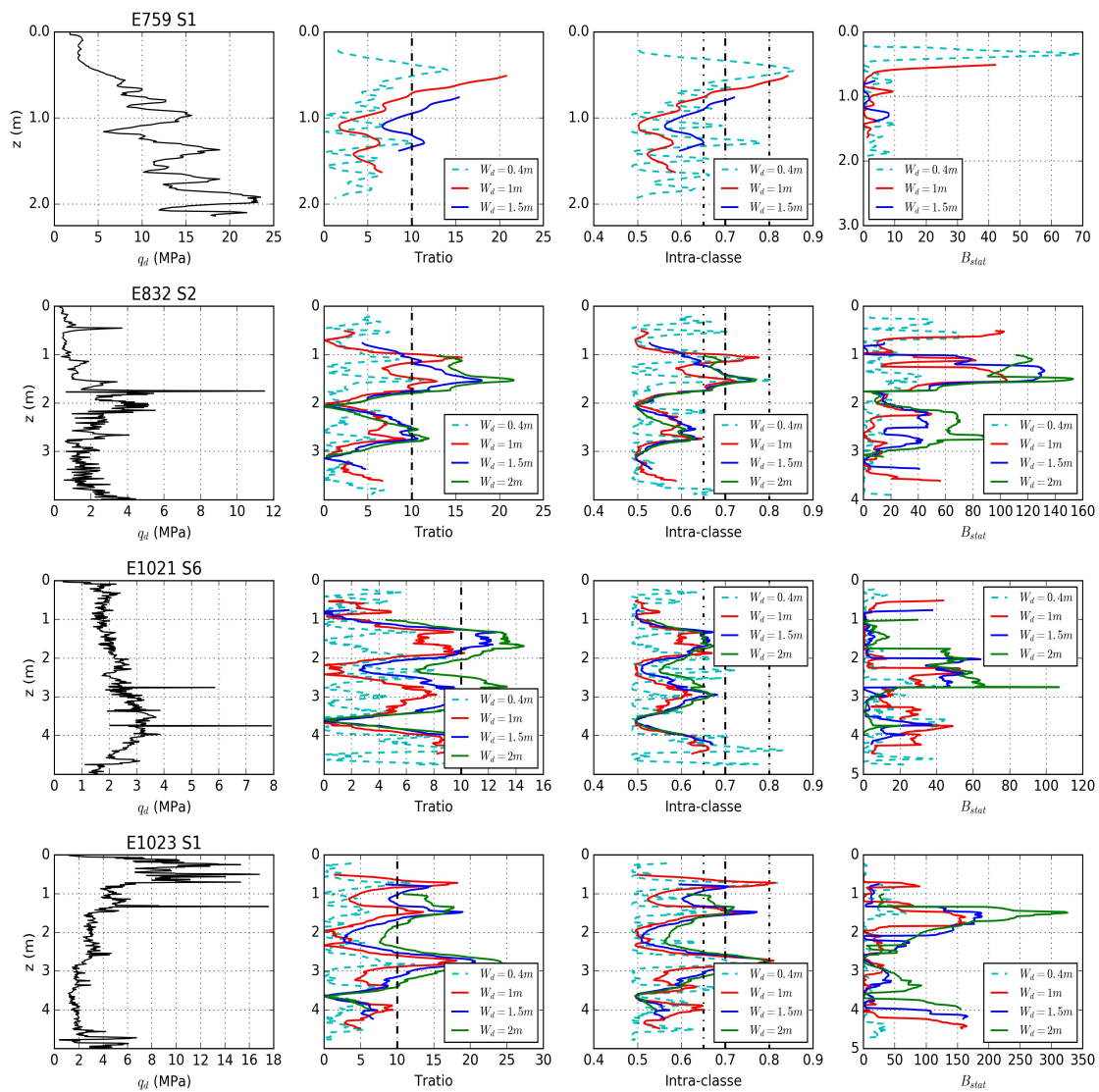


FIGURE A.10 – Courbes des tests statistiques selon la hauteur de la fenêtre utilisée pour les sondages in-situ analysés par les experts

B. Annexes B

B.1 Le krigeage ordinaire (KO)

On suppose que la valeur de la moyenne μ est inconnue. Sous la contrainte de non-biais l'espérance de l'erreur d'estimation s'écrit :

$$E[V^*(X) - V(X)] = \sum_{i=1}^n \lambda_i E[V_i] - E[V(X)] = \mu \left(\sum_{i=1}^n \lambda_i - 1 \right) \quad (\text{B.1})$$

Avec la condition d'universalité ou de non biais :

$$\sum_{i=1}^n \lambda_i = 1 \quad (\text{B.2})$$

D'après la condition d'optimalité, la variance de l'estimation est :

$$\sigma_E^2(X) = \text{Var}[V^*(X) - V(X)] = E\{[V(X) - V^*(X)]^2\} \quad (\text{B.3})$$

elle peut être écrite sous la forme :

$$\sigma_E^2(X) = \text{Var}[V^*(X)] + \text{Var}[V(X)] - 2 \text{cov}[V^*(X), V(X)] \quad (\text{B.4})$$

compte tenu de :

$$\text{Var}[V^*(X)] = \text{Var} \left[\sum_{j=1}^n \lambda_j V_j \right] = \sum_{i,j=1}^n \lambda_i \lambda_j C_V(X_i - X_j) \quad (\text{B.5})$$

et de :

$$\text{cov} [[V^*(X), V(X)] = \text{cov} \left[\sum_{i=1}^n \lambda_i V_i, V(X) \right] = \sum_{i=1}^n \lambda_i C_V(X - X_i) \quad (\text{B.6})$$

On peut réécrire la variance de l'estimation comme ci-suit :

$$\sigma_E^2(X) = \text{Var}[V(X)] + \sum_{i=1}^n \sum_{j=1}^n \lambda_i \lambda_j C_V(X_i - X_j) - 2 \sum_{i=1}^n \lambda_i C_V(X - X_i) \quad (\text{B.7})$$

quantité à minimiser sous la condition d'universalité en appliquant la technique des multiplicateurs de Lagrange. On introduit donc une inconnue auxiliaire v et on obtient ainsi le système de $N + 1$ équations du krigeage ordinaire, correspondant aux N dérivées partielles

qui doivent s'annuler, auxquelles s'ajoute la condition de non biais. D'après la relation entre le variogramme et la covariance, les équation du krigeage deviennent :

$$\begin{cases} \sum_{j=1}^n \lambda_j \gamma(X_i - X_j) + v = \gamma(X_i - X) \quad \forall i = 1 \dots n \\ \sum_{i=1}^n \lambda_i = 1 \end{cases} \quad (\text{B.8})$$

ou soit en notation matricielle :

$$\begin{pmatrix} \gamma(X_1 - X_1) & \dots & \gamma(X_1 - X_n) & 1 \\ \vdots & \ddots & \vdots & \vdots \\ \gamma(X_n - X_1) & \dots & \gamma(X_n - X_n) & 1 \\ 1 & \dots & 1 & 0 \end{pmatrix} \begin{pmatrix} \lambda_1 \\ \vdots \\ \lambda_n \\ v \end{pmatrix} = \begin{pmatrix} \gamma(X_1 - X) \\ \vdots \\ \gamma(X_n - X) \\ 1 \end{pmatrix} \quad (\text{B.9})$$

B.2 Méthodes proposées pour déterminer expérimentalement l'échelle de fluctuation

Ici, on présente succinctement plusieurs approches proposées dans la littérature pour déterminer expérimentalement la valeur de l'échelle de fluctuation.

Méthode simplifiée de Vanmarcke

Une procédure simple pour déterminer l'échelle de fluctuation à partir des distances limitées par les intersections du profil de mesures avec le modèle de tendance ou de dérive choisie est donné par (Vanmarcke (1977a)) :

$$\theta \approx \bar{\Delta} \sqrt{\frac{2}{\pi}} \quad (\text{B.10})$$

où la longueur moyenne des intersections est calculée comme suit :

$$\bar{\Delta} = \frac{1}{n_c - 1} \sum_{i=1}^{n_c} \Delta_i \quad (\text{B.11})$$

avec Δ_i la distance limitée pour chaque intersection et n_c le nombre total d'intersections.

Méthodes basée sur la fonction de variance

L'échelle de fluctuation θ peut être approchée numériquement grâce à l'expression suivante proposée par (Wickremesinghe and Campanella (1993)) :

$$\theta \approx \frac{\sigma_n^2}{\sigma^2} W_n \quad (\text{B.12})$$

avec σ^2 la variance calculée pour la série complète des données et σ_n^2 la variance estimée pour n effectifs à l'aide d'une fenêtre glissante de longueur W_n . La valeur maximale de l'expression (éq.B.12) correspond à l'échelle de fluctuation.

Limite issue du test statistique de *Barlett*

Dans le domaine de l'analyse des séries temporelles, l'approximation de *Barlett* est une technique très populaire pour déterminer la distance de corrélation. Les limites de Barlett l_B peuvent être calculées en fonction du nombre de données disponibles comme ci-après ([Box et al. \(2015\)](#)) :

$$l_B = \frac{2}{\sqrt{n}} \left(1 + 2 \sum_{i=1}^{k-1} r_i^2 \right) \approx \pm \frac{1.96}{\sqrt{n}} \quad (\text{B.13})$$

([Magnan \(1982\)](#)), ([Jaksa \(1995\)](#)), ([Jaksa et al. \(1999\)](#)) montrent que l'échelle de fluctuation peut être déterminée d'une manière relativement simple comme la distance r_B à la quelle l'autocorrélogramme empirique intercepte les limites de Barlett.

C. Annexes C

C.1 Le site expérimentale : Le Solar de Castelló d'Empúries

C.1.1 Localisation du site

Le *Solar de Castelló d'Empúries* se situe au sud de la ville de Castelló d'Empúries (Alt Empordà) dans la province de Gérone dans la province de la Catalogne en Espagne (Fig.C.1(a)). Ce terrain, appartenant à la société IGEOTEST, se trouve dans une zone industrielle à la sortie droite GIV-6216 du rond-point n°7 sur l'autoroute C-260, reliant Figueras à Rosas (Fig.C.1(b)). La surface du terrain étudié, délimité au nord par le Carrer Gironès, au sud par le Carrer Garrotxa et à l'est par le carrer Rec Madral, est de l'ordre de $1600m^2$. Au niveau de la topographie il s'agit d'un terrain plat, avec des écarts d'altitude maximales de 30 cm d'après les mesures d'altimétrie réalisées.



FIGURE C.1 – (a) Situation géographique du terrain. (b) À droite emplacement du Solar de Castello d'Empúries.

C.1.2 Géologie du site

Ce site expérimental est situé dans l'Alt Empordà (Castelló d'Empuries) dans la partie centrale plate de la *Costa Brava*. Celle-ci se situe entre le relief paléozoïque des Pyrénées et le massif du Mésozoïque Montgri. Cette zone est une plaine alluviale alimentée essentiellement par les sédiments des rivières Fluvià et Muga et formant un dépôt de type deltaïque méditer-

ranéen. Les dépôts de l'Holocène atteignent dans cette zone des épaisseurs d'environ 20 à 30 m. En profondeur il existe une alternance d'horizons sableux (i.e. des dunes) principalement et des horizons formés par des dépôts limoneux et argileux (marais côtiers), avec la présence de passages graveleux.

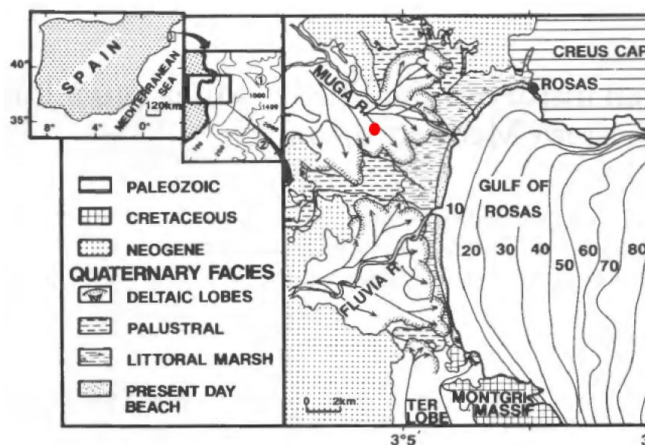


FIGURE C.2 – Carte géologique (source : Díaz and Ercilla (1993)) et situation géographique du site expérimental (point rouge)

C.1.3 Campagne d'essais

La campagne d'essais a été réalisée le long de trois jours de travail sur site, plus précisément du lundi 22 au mercredi 24 Février 2016. On dispose des essais suivants réalisés sur site :

- 8 sondages pénétrométriques Panda 2 ;
- 4 sondages géondoscopiques.

L'implantation des sondages a été réalisée d'après le schéma montré dans la figure (Fig.C.3).

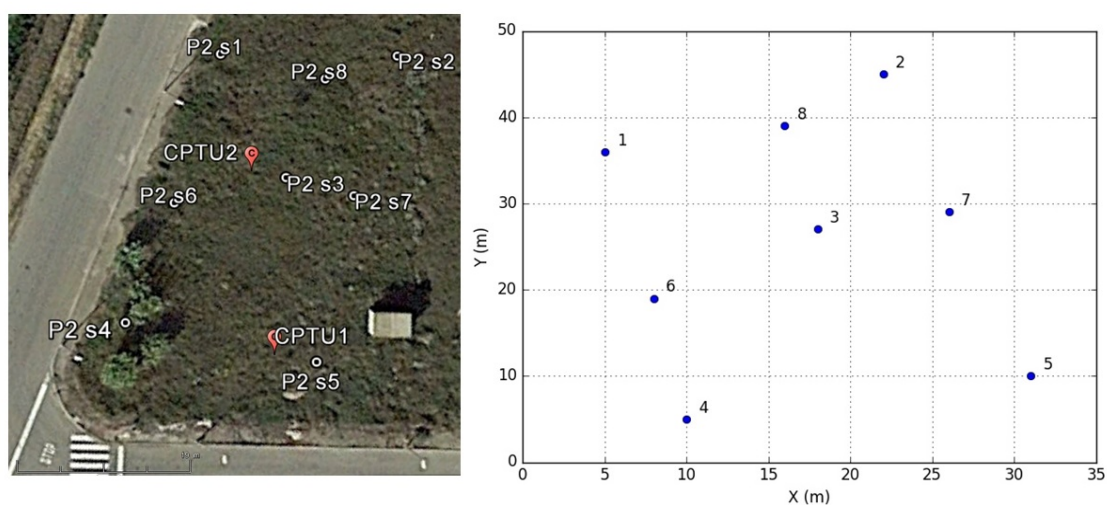


FIGURE C.3 – À gauche, le schéma d'implantations des essais in situ. À droite, la position relative des sondages pénétrométriques

La configuration des sondages suit un maillage assez régulier, et leur positionnement a été géoréférencé à l'aide d'un GPS de précision. Les coordonnées UTM, le nombre de mesures et la profondeur du chaque sondage sont récapitulés dans le tableau (Tab.C.1).

TABLE C.1 – Sondages Panda, nombre de mesures, coordonnées UTM et profondeur du sondage

Sondage	N° mesures	X	Y	Prof (m)
Pda2 S1	951	505815	4677180	7
Pda2 S2	635	505832	4677189	7
Pda2 S3	510	505828	4677171	7
Pda2 S4	602	505820	4677149	7
Pda2 S5	482	505841	4677154	7
Pda2 S6	1950	505818	4677163	13.6
Pda2 S7	1017	505836	4677173	7.8
Pda2 S8	595	505826	4677183	7

C.2 Statistiques descriptives de la campagne de sondages

Les statistiques descriptives de base pour les 8 sondages réalisés sur le site de Castello sont regroupées dans le tableau (Tab.C.2), pour les sondages bruts et régularisés avec un pas de régularisation égal à 10mm. On observe que la régularisation entraîne un affaiblissement des valeurs des moments statistiques par rapport aux analyses des mesures brutes.

TABLE C.2 – Statistiques de base du q_d par sondage

Statistiques	Sondages brutes							
	S1	S2	S3	S4	S5	S6	S7	S8
<i>Taille</i>	744	383	687	450	411	345	658	412
<i>Moyenne</i> (MPa)	9.83	9.19	6.28	6.56	9.75	7.52	5.48	7.59
<i>Variance</i> (MPa ²)	156.91	75.35	49.01	39.86	92.09	53.09	23.85	40.56
<i>Écart-type</i> (MPa)	12.53	8.68	7.00	6.31	9.60	7.29	4.88	6.37
<i>CV</i>	1.27	0.94	1.11	0.96	0.98	0.97	0.89	0.84
Statistiques	Sondages régularisés							
	S1	S2	S3	S4	S5	S6	S7	S8
<i>Taille</i>	499	499	499	499	499	499	499	499
<i>Moyenne</i> (MPa)	6.15	4.91	4.76	4.96	5.67	4.52	4.52	4.67
<i>Variance</i> (MPa ²)	81.61	39.79	32.44	29.01	52.80	31.26	18.79	21.73
<i>Écart-type</i> (MPa)	9.03	6.31	5.70	5.39	7.27	5.59	4.33	4.66
<i>CV</i>	1.47	1.28	1.20	1.09	1.28	1.24	0.96	1.00

C.3 Statistiques de q_d pour chaque unité des modèles de terrain établis

Dans ce paragraphe on fournit les statistiques de base de la résistance de pointe q_d pour chaque unité en fonction du modèle de terrain établi pour le site de Castello.

C.3.1 Modèle M2

TABLE C.3 – Statistiques du modèle M2, unité 1 (U1)

Sondages	Statistiques			
	Moyenne (MPa)	Variance (MPa ²)	Écart-type (MPa)	CV
S1	14.04	38.92	6.24	0.44
S2	11.71	18.20	4.27	0.36
S3	11.76	47.49	6.89	0.59
S4	13.86	54.57	7.39	0.53
S5	10.77	37.08	6.09	0.57
S6	11.18	36.45	6.04	0.54
S7	10.06	11.18	3.34	0.33
S8	10.25	19.30	4.39	0.43
Moyenne	11.70	32.90	5.58	0.47
Écart-type	1.52	15.22	1.42	0.10
CV	0.13	0.46	0.25	0.20

TABLE C.4 – Statistiques du modèle M2, unité 2 (U2)

Sondages	Statistiques			
	Moyenne (MPa)	Variance (MPa ²)	Écart-type (MPa)	CV
S1	1.37	0.96	0.98	0.72
S2	1.68	1.30	1.14	0.68
S3	1.55	0.19	0.43	0.28
S4	1.63	0.31	0.56	0.34
S5	1.48	0.23	0.48	0.32
S6	1.64	0.70	0.84	0.51
S7	1.83	0.31	0.55	0.30
S8	1.95	0.39	0.63	0.32
Moyenne	1.64	0.55	0.70	0.43
Écart-type	0.19	0.40	0.26	0.18
CV	0.11	0.73	0.37	0.41

C.3.2 Modèle M3

TABLE C.5 – Statistiques du modèle M3, unité 1 (U1)

Sondages	Statistiques			
	Moyenne (MPa)	Variance (MPa ²)	Écart-type (MPa)	CV
S1	16.60	50.41	7.10	0.43
S2	12.59	23.56	4.85	0.39
S3	16.59	30.81	5.55	0.33
S4	18.42	43.09	6.56	0.36
S5	14.49	33.09	5.75	0.40
S6	14.68	37.70	6.14	0.42
S7	11.71	9.58	3.10	0.26
S8	12.08	14.35	3.79	0.31
Moyenne	14.65	30.32	5.36	0.36
Écart-type	2.43	13.94	1.37	0.06
CV	0.17	0.46	0.26	0.15

TABLE C.6 – Statistiques du modèle M3, unité 2 (U2)

Sondage	Statistiques			
	Moyenne (MPa)	Variance (MPa ²)	Écart-type (MPa)	CV
S1	10.91	7.39	2.72	0.20
S2	10.63	9.76	3.12	0.25
S3	5.88	4.62	2.15	0.29
S4	8.31	12.42	3.52	0.30
S5	6.23	4.52	2.13	0.34
S6	6.92	1.83	1.35	0.37
S7	8.05	5.83	2.42	0.42
S8	8.01	16.35	4.04	0.51
Moyenne	8.12	7.84	2.68	0.33
Écart-type	1.74	4.45	0.80	0.09
CV	0.21	0.57	0.30	0.27

Les statistiques du q_d de l'unité n°3 du modèle M3 ne sont pas fournies puisqu'elles sont identiques à celles de l'unité n°2 du modèle M2.

C.3.3 Modèle M4

Pour ce modèle, on ne présente que les tableaux avec les moments statistiques des unités n°3 et n°4. En effet, celles relatives aux unités n°1 et n°2 sont les mêmes que celles du modèle M3.

TABLE C.7 – Statistiques du modèle M4, unité 3 (U3)

Sondage	Statistiques			
	Moyenne (MPa)	Variance (MPa ²)	Écart-type (MPa)	CV
S1	1.42	2.22	1.49	1.05
S2	2.19	2.70	1.64	0.75
S3	1.36	0.11	0.33	0.25
S4	1.31	0.09	0.29	0.22
S5	1.47	0.28	0.53	0.36
S6	1.76	1.68	1.30	0.74
S7	1.77	0.32	0.57	0.32
S8	1.87	0.46	0.68	0.36
Moyenne	1.64	0.98	0.85	0.51
Écart-type	0.31	1.05	0.54	0.30
CV	0.19	1.07	0.63	0.59

TABLE C.8 – Statistiques du modèle M4, unité 4 (U4)

Sondages	Statistiques			
	Moyenne (MPa)	Variance (MPa ²)	Écart-type (MPa)	CV
S1	1.34	0.16	0.40	0.30
S2	1.35	0.14	0.37	0.28
S3	1.67	0.20	0.44	0.27
S4	1.83	0.35	0.59	0.32
S5	1.49	0.19	0.44	0.30
S6	1.57	0.07	0.26	0.17
S7	1.87	0.29	0.54	0.29
S8	2.01	0.35	0.59	0.29
Moyenne	1.64	0.22	0.46	0.28
Écart-type	0.25	0.10	0.11	0.05
CV	0.15	0.46	0.25	0.17

C.4 Analyse structurale

C.4.1 Modèles de dérives ajustés pour chacun des sondages (modèle M2)

Les modèles de dérives linéaires ajustés aux profils du logarithme de q_d pour les deux formations lithologiques identifiées sur le site de Castello sont regroupés dans la figure (Fig.C.4).

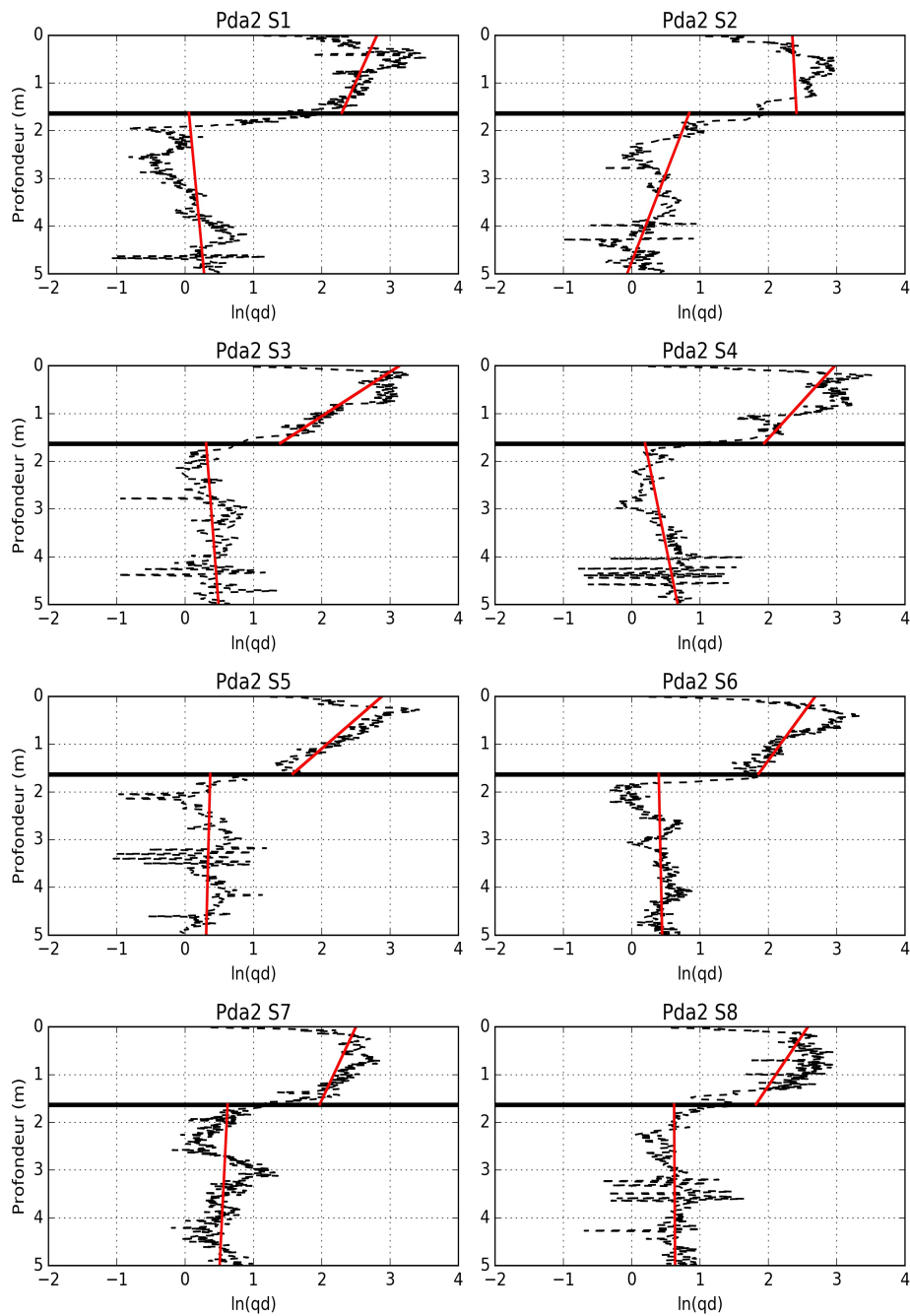


FIGURE C.4 – Calage de dérives linéaires pour $\ln(q_d)$ au sein des deux unités lithologiques identifiés

C.4.2 Estimation des dérives globales pour les modèles (M1, M3, M4)

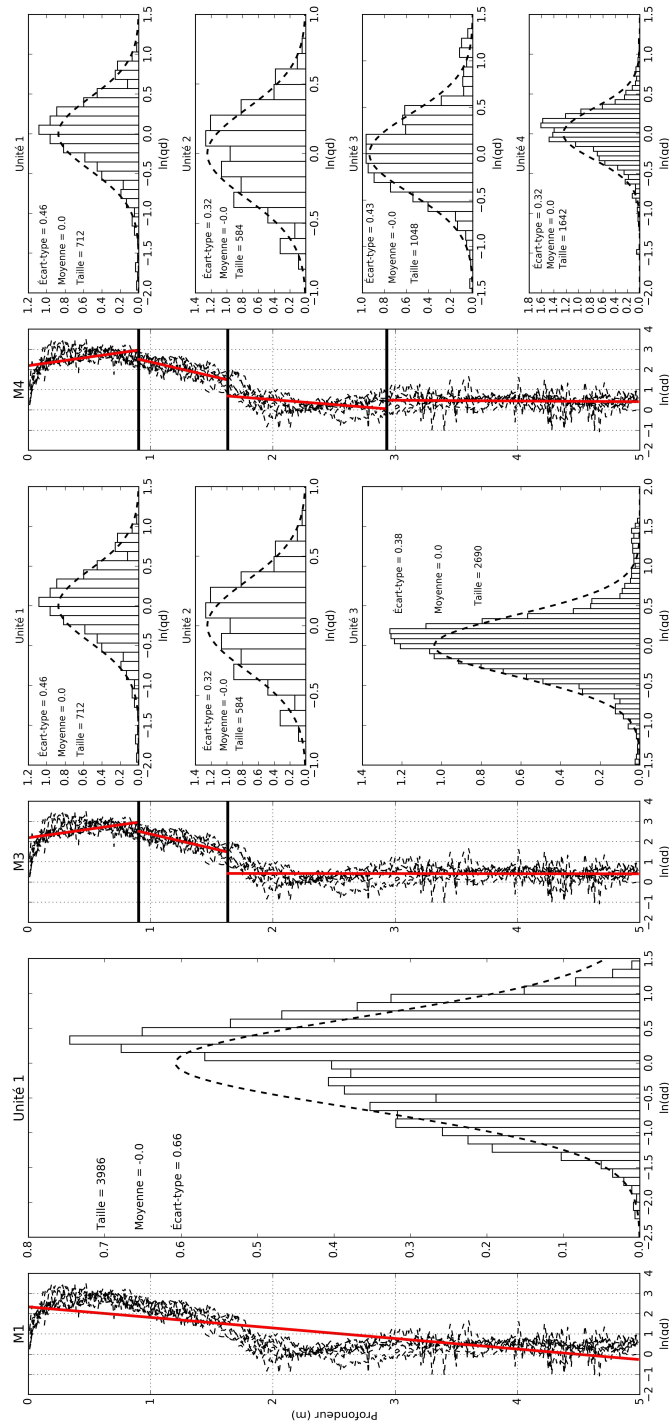


FIGURE C.5 – Dérives linéaires et histogrammes des résidus sur chacune des unités pour les différentes unités de terrain

C.5 Conditionnement du champ par krigeage. Réduction de l'incertitude

Le tableau (Tab.C.9) regroupe les résultats numériques de réduction de l'incertitude atteinte grâce au conditionnement par krigeage sur le site de Castello. La première colonne recueille les valeurs des échelles de fluctuation horizontales considérées pour l'analyse ; la deuxième, la réduction de l'incertitude atteinte quand on conditionne le champ avec 6 sondages ; la troisième, pour les 8 sondages. La dernière colonne notée $\Delta I(\%)$ représente le gain d'incertitude lié au passage de 6 à 8 sondages.

TABLE C.9 – Réduction de l'incertitude en fonction de l'échelle de fluctuation horizontale et du nombre de sondages

$\theta_H(\text{m})$	I 6 Pda	I 8 Pda	$\Delta I(\%)$
2	0.37	1.32	0.96
4	1.85	3.21	1.37
6	4.64	6.70	2.06
8	8.31	11.07	2.76
10	12.51	15.81	3.30
13	19.15	22.90	3.75
15	23.50	27.36	3.86
18	29.63	33.48	3.86
20	33.37	37.16	3.80
25	41.50	45.05	3.55
50	64.70	67.09	2.38
75	74.93	76.66	1.73
100	80.53	81.87	1.33
200	89.54	90.22	0.68
300	92.76	93.22	0.45
400	94.44	94.78	0.34
500	95.49	95.75	0.27
600	96.20	96.42	0.22
700	96.71	96.90	0.19
800	97.10	97.27	0.17
900	97.41	97.56	0.15
1000	97.66	97.79	0.13
1500	98.42	98.51	0.09
2000	98.81	98.87	0.07
2500	99.04	99.09	0.05
5000	99.52	99.54	0.03
10000	99.76	99.77	0.01

D. Annexes D

D.1 Le site expérimental d'Aulnat

D.1.1 Localisation du site

Le terrain expérimental d'étude est situé dans la commune d'Aulnat (63), située dans le département du Puy-de-Dôme en région d'Auvergne-Rhône-Alpes (Fig.D.1). Il fait partie de l'aire urbaine de Clermont-Ferrand. Il s'agit d'une parcelle agricole, de surface triangulaire d'environ $4500 m^2$. Il est délimitée par le croisement entre la D54 et la DR54B sur la route de Malintrat.

D.1.2 Géologie du site

Le site d'étude d'Aulnat fait partie du complexe de la Grande Limagne (Fig. D.2), vaste plaine d'effondrement au relief relativement contrasté. Après les dépôts bariolés, discontinus, de l'Oligocène basal, le bassin de la Limagne est le siège d'une sédimentation fluviatile et laguno-lacustre. Il s'agit presque principalement de formations sablo-argileuses à argileuses sur une profondeur de 4 à 5 mètres appelées « complexe K ». Ces couches surmontent une marne devenant très compacte avec la profondeur.

Les complexes K de Limagne sont alimentés pour l'essentiel par des colluvions marneuses auxquelles s'ajoutent en quantité variable d'alluvions, voire des apports éoliens. La réputation de la Grande Limagne est fondée sur la fertilité des sols à cause d'une première formation

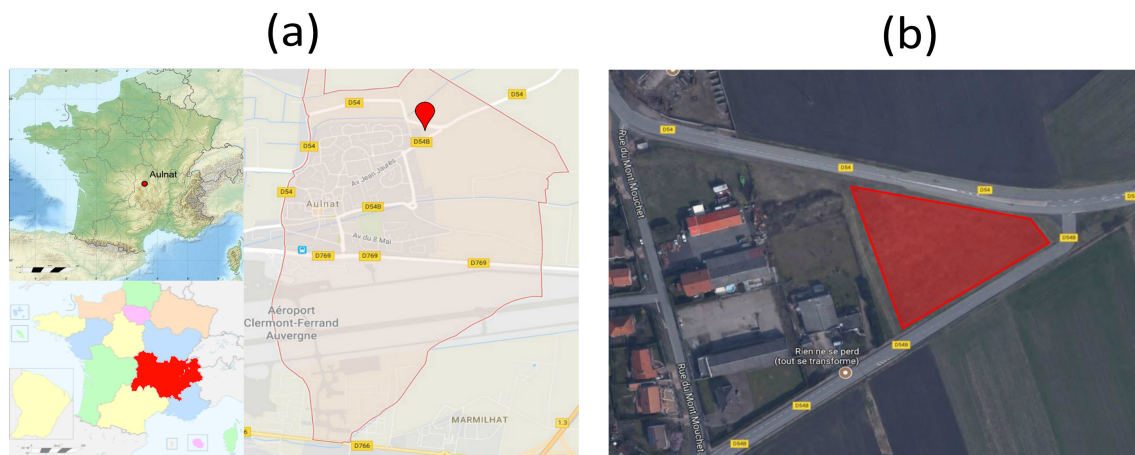


FIGURE D.1 – (a) Situation géographique du terrain. (b) Emplacement de la parcelle d'Aulnat.

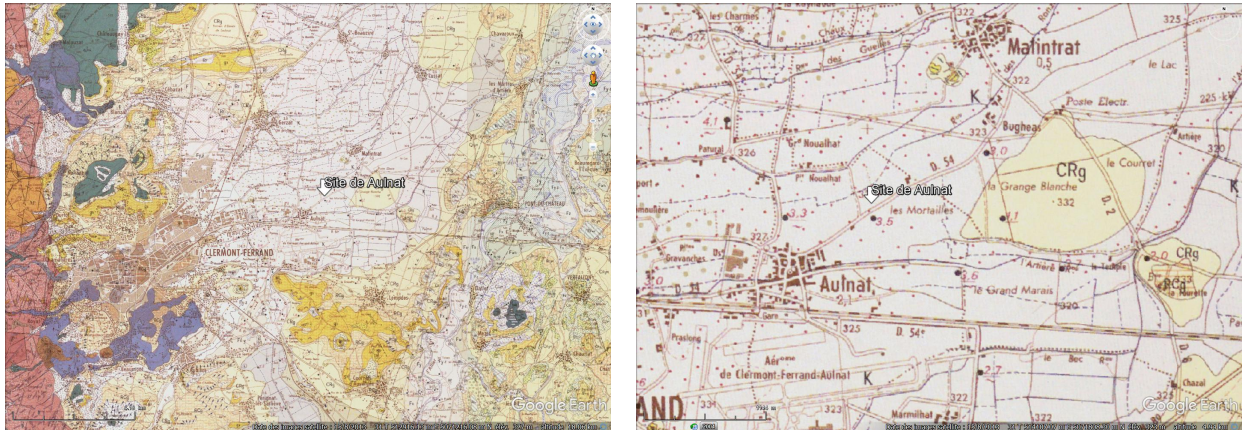


FIGURE D.2 – Zone d'étude. Carte géologique de la région de Clermont-Ferrand 1 : 50000. Source : <http://planet-terre.ens-lyon.fr/article/BRGM-kml.xml>

de remblaiement composée par des sols limoneux souvent regroupés sous le nom de terres noires.

La formation argilo-calcaire, base du complexe K, a été décrite sous le nom de « vase », « marne » ou « limon » (Rudel (1957); Gachon (1963)). Du point de vue granulométrique, la formation est caractérisée par une large prédominance de la fraction inférieure à 20μ : 85 à 95 % (Gachon (1963)) dont 45 à 55 % d'éléments inférieurs à 2μ . La teneur moyenne en calcaire, de l'ordre de 40 à 45 % montre que la formation dérive pour l'essentiel des marnes et calcaires oligocènes.

Une coupe-type du complexe K comprend, de haut en bas :

- 0.6 à 2 mètres des terres noires ;
- 0.1 à 1.5 mètres d'alluvions, sables et graviers où dominant des minéraux d'origine volcanique ;
- 1 à 3 mètres d'une formation argilo-calcaire, meuble mais compacte, à première vue très homogène de couleur beige ou grise s'éclaircissant en profondeur.

Cette formation repose soit sur les marnes grises oligocènes connues à l'affleurement, soit sur des marnes de couleur vert clair. Cette coupe-type est relativement constante en Grande Limagne mais dans le détail, elle est cependant beaucoup plus compliquée. C'est ainsi que la présence d'alluvions multiplie les variantes.

D.1.3 Campagnes d'essai

Première campagne d'essais

Une première campagne d'essais a été réalisée du lundi 3 au lundi 10 Octobre 2016. Pour ce site, on dispose des essais suivantes :

- 1 forage à la tarière
- 9 sondages pénétrométriques Panda2®
- 3 profils de tomographie électrique (ERT)

La position des essais réalisés est indiquée dans la figure (Fig.D.3). Le sondage S9, réalisé deux mois après, est également retenu dans les analyses à réaliser.

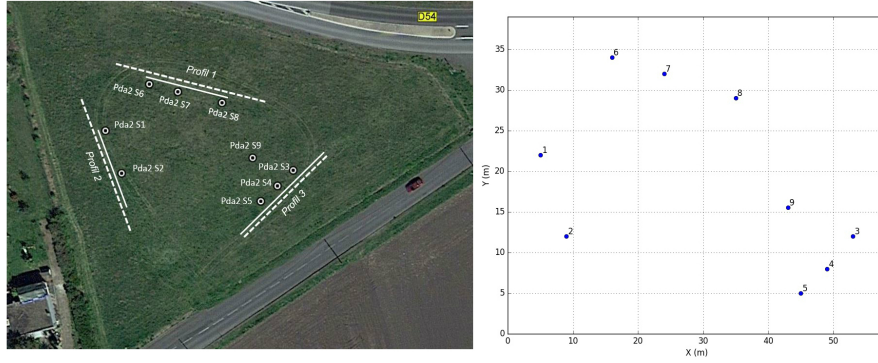


FIGURE D.3 – Première campagne. À gauche, le schéma d’implantations des essais in situ. À droite, la position relative des sondages pénétrométriques

Les sondages Panda ont été effectués en pointe de 4cm^2 et géoréférencés à l’aide d’un GPS de précision. Pour chaque sondage, les coordonnées UTM, le nombre de mesures et la profondeur d’étude sont récapitulés dans le tableau (Tab.D.1).

TABLE D.1 – Sondages Panda, nombre de mesures, coordonnées UTM et profondeur du sondage

Sondage	N° mesures	X	Y	Prof (m)
Pda2 S1	1090	513467	5071936	4
Pda2 S2	931	513471	5071926	4
Pda2 S3	1192	513515	5071926	4
Pda2 S4	1178	513511	5071922	4
Pda2 S5	961	513507	5071919	4
Pda2 S6	1254	513478	5071948	4
Pda2 S7	1206	513486	5071946	3
Pda2 S8	994	513497	5071943	3
Pda2 S9	638	513505	5071929	4

Deuxième campagne d’essais

Pour étudier la variabilité spatiale des propriétés du sol, une pratique courante est d’utiliser un pas d’échantillonnage régulier le long d’un profil longitudinal ou suivant un maillage régulier en forme de grille carrée. Pour ce terrain on a pu profiter de la réalisation d’une série de sondages pénétrométriques plus exhaustive afin de mener un étude plus rigoureuse de la variabilité selon le plan horizontal. Cette deuxième campagne a été menée durant 5 jours de travail du lundi 31 Juillet au vendredi 4 Août 2017, avec un total de 60 sondages Panda réalisés à environ 2 et 3.5 m de profondeur. Les essais ont été repartis le long de deux profils longitudinaux, soit 30 essais dans l’axe nommé X et 30 essais dans l’axe nommé Y, avec un espacement de 0.5 m. L’ensemble des sondages forme ainsi un crois de 15×15 m. Le plan d’implantation de la campagne d’essais est donné dans la figure (Fig.D.4). Les sondages ont été réalisés à proximité du sondage S9 qui fait partie de la première campagne.

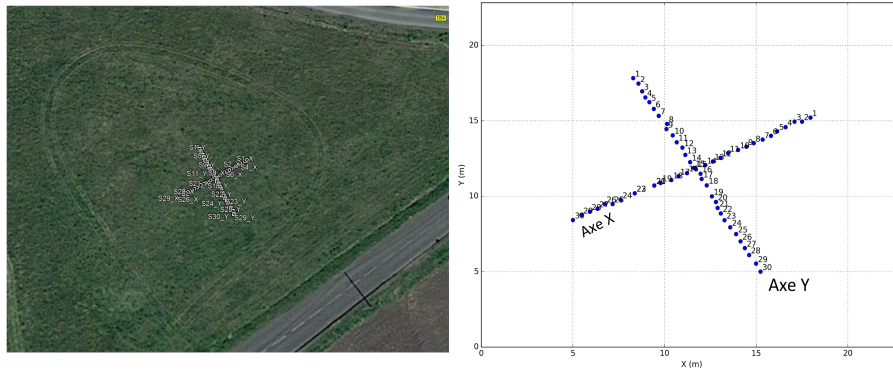


FIGURE D.4 – Deuxième campagne. À gauche, le schéma d’implantation des essais in situ. À droite, la position relative des sondages pénétrométriques

Coupe lithologique

La stratigraphie du site a été déduite à l’aide d’un sondage à la tarière mécanique mené jusqu’à 6 mètres de profondeur (cf. Fig.D.5). Les premiers 30 cm sont composés par la terre végétale. Ensuite une formation limoneuse qui correspond aux terres noires de la Limagne a été identifiée jusqu’à environ 2 m suivie par une couche d’argile bleu grise avant d’atteindre vers 4 m le substratum marneux. Le refus du sondage est atteint à 6 m.

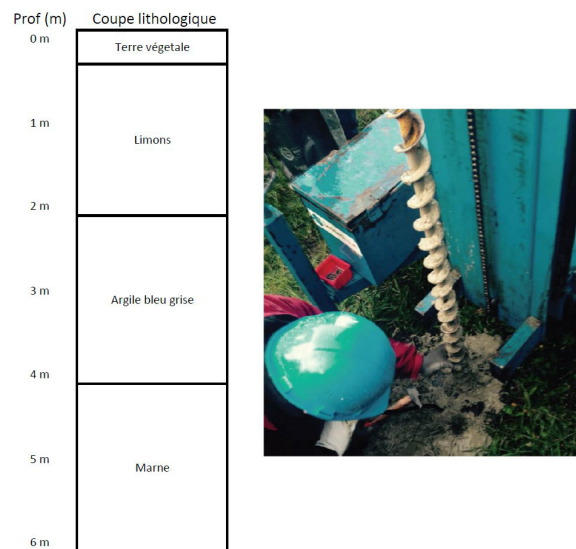


FIGURE D.5 – Coupe lithologique déduite du sondage à la tarière

D.1.4 Application des techniques stratigraphiques pour la définition des unités homogènes

Première campagne d'essais

Les méthodes statistiques par fenêtre glissante ont été appliquées à l'ensemble des sondages Panda effectués sur le site. Trois longueurs de fenêtre : 1, 1.5 et 2 ont été testées. À titre d'exemple on montre les résultats du découpage statistique pour le sondage S4 (Fig.D.6), avec une fenêtre de 1 m. Les limites entre couches identifiées au moyen du test *Tratio* sont identiques à celles détectées grâce au test *Intraclasse*. Le profil statistique issu du test de *Barlett* montre aussi un maximum à 2m, et un deuxième de valeur plus faible à environ 1m de profondeur.

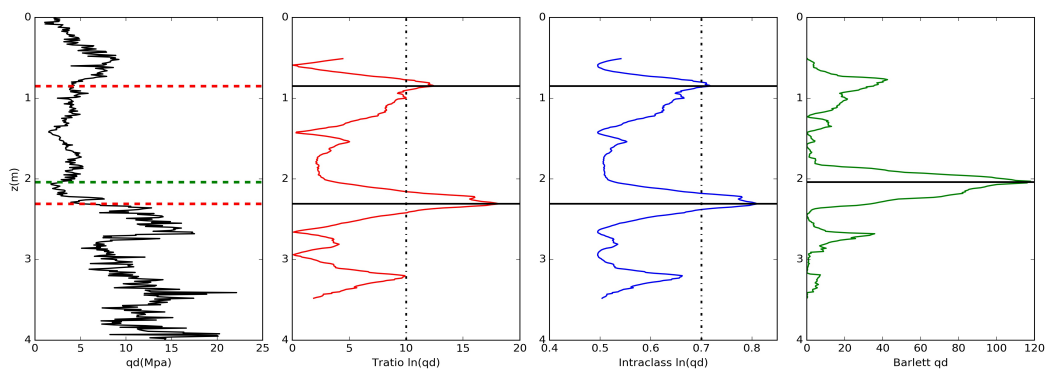


FIGURE D.6 – Application des techniques statistiques du découpage pour le sondage S4. Code couleur : *Tratio*, *Intra-classe* et *Barlett*. Longueur de fenêtre $W_d = 1\text{m}$

Les profondeurs des limites de couches détectées avec les trois méthodes statistiques sont présentées dans la figure (Fig.D.7). Indépendamment de la taille de la fenêtre retenue, une interface est toujours identifiée entre 2 et 2.5 m, cohérente avec l'interface à 2 mètres entre la formation limoneuse et argileuse. Pour les analyses faites avec une fenêtre de 1 mètre, une première couche plus en surface apparaît pour 4 des 7 sondages analysés (S1, S2, S4 et S6), avec une limite à environ 1 m de profondeur.

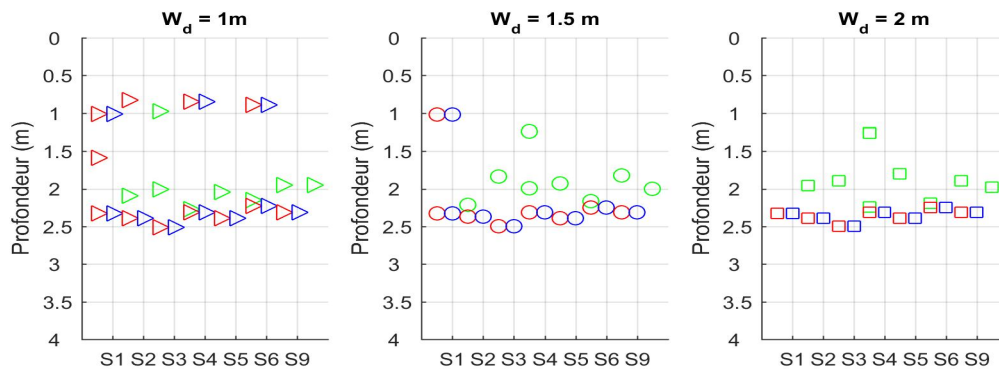


FIGURE D.7 – Limites de couches détectées avec les méthodes statistiques. Code couleur : *Tratio*, *Intra-classe* et *Barlett*. Avec W_d la longueur de la fenêtre testée.

Deuxième campagne d'essais

Les sondages réalisés durant la deuxième campagne ont été analysés par les mêmes méthodes statistiques.

La profondeur atteinte pour la plupart des sondages réalisés dans l'axe X varie entre 3.20 et 3.50m. Une largeur de fenêtre de 2m présente des dimensions trop grandes par rapport à la longueur des sondages à analyser. On a constaté des résultats très similaires à ceux obtenus grâce à l'analyse des essais de la première campagne. La profondeur moyenne et l'écart-type des limites identifiées à l'aide de test *Tratio* sont regroupés dans le tableau (Fig.D.2). À titre d'exemple, on a représenté les interfaces identifiées pour chacun des sondages dans la figure (Fig.D.8). On précise qu'ils sont les mêmes que celles détectées à l'aide de la méthode *Intra-classe*.

TABLE D.2 – Analyse profil X. Résultats *Tratio*

Longueur fenêtre	Limite 1		Limite 2	
	Moyenne (m)	Écart-type (m)	Moyenne (m)	Écart-type (m)
Wd = 1 m	0.75	0.06	2.45	0.14
Wd = 1.5 m	0.79	0.02	2.47	0.14

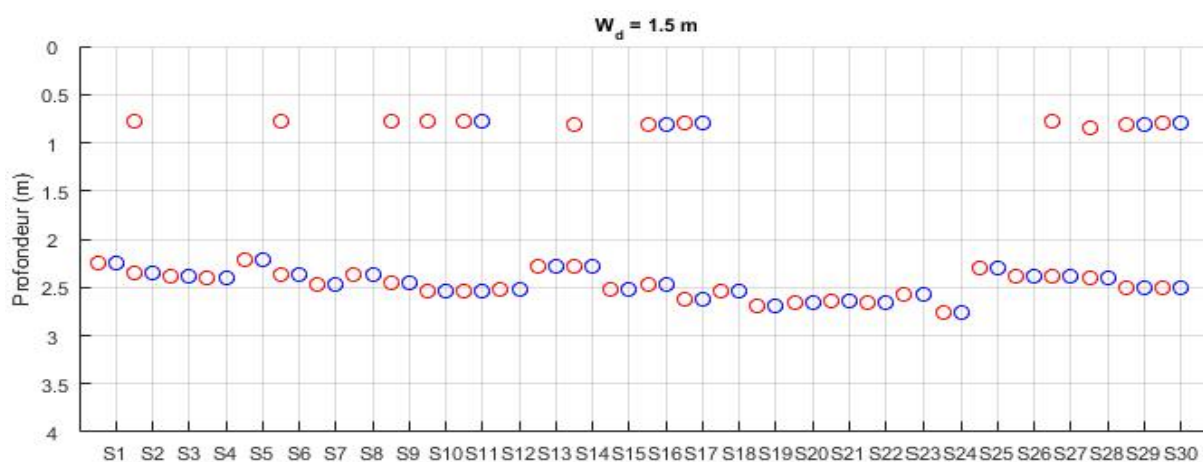


FIGURE D.8 – Analyse profil X. Limites de couches détectées avec les méthodes statistiques pour une longueur de la fenêtre égale à 1.5m. Code couleur : *Tratio*, *Intra-classe*.

Enfin, les essais placés dans l'axe Y ont été analysés avec une fenêtre dont la longueur est égale à 1 m. La profondeur atteinte pour la plupart des sondages est inférieure à 2.5 m. Les tests ont identifié la limite en surface pour certains sondages environ 0.8m avec un écart type de 0.16 m.

D.1.5 Tomographie électrique

Trois profils de tomographie électrique ont été réalisés à Aulnat : un profil de 42 électrodes équidistantes de 0.5m (ERT1) et deux profils de 42 et 24 électrodes respectivement écartés 1m (ERT2 et ERT3). Les pseudo-sections mesurées, calculées et le modèle de sol obtenu après inversion sont donnés dans les figures (Fig.D.9), (Fig.D.10) et (Fig.D.11) pour les trois panneaux électriques réalisés. La configuration des panneaux est du type Wenner-Schlumberger et les cartes de résistivité inversée sont calculés à l'aide du logiciel RES2DINV (Loke (2002)). On peut observer que les valeurs de résistivité mesurées variant de 1 à 70 Ωm sont situées dans la plage de variation typique des sols argileux, de 1 à 100 Ωm (Reynolds (2011)). Le tableau (Tab.D.3) regroupe les caractéristiques d'acquisition pour les trois panneaux réalisés.

TABLE D.3 – Caractéristiques des profils électriques réalisés et paramètres de l'inversion

	ERT1	ERT 2	ERT 3
Itérations	7	3	5
RMS (%)	0.73	1.67	1
Longueur (m)	41 m	41	23
$\Delta x(m)$	0.5	1	1
nb. électrodes	42	42	24
nb. mesures	40-78	19-39	17-21
Prof. (m)	5.5	5.9 m	4

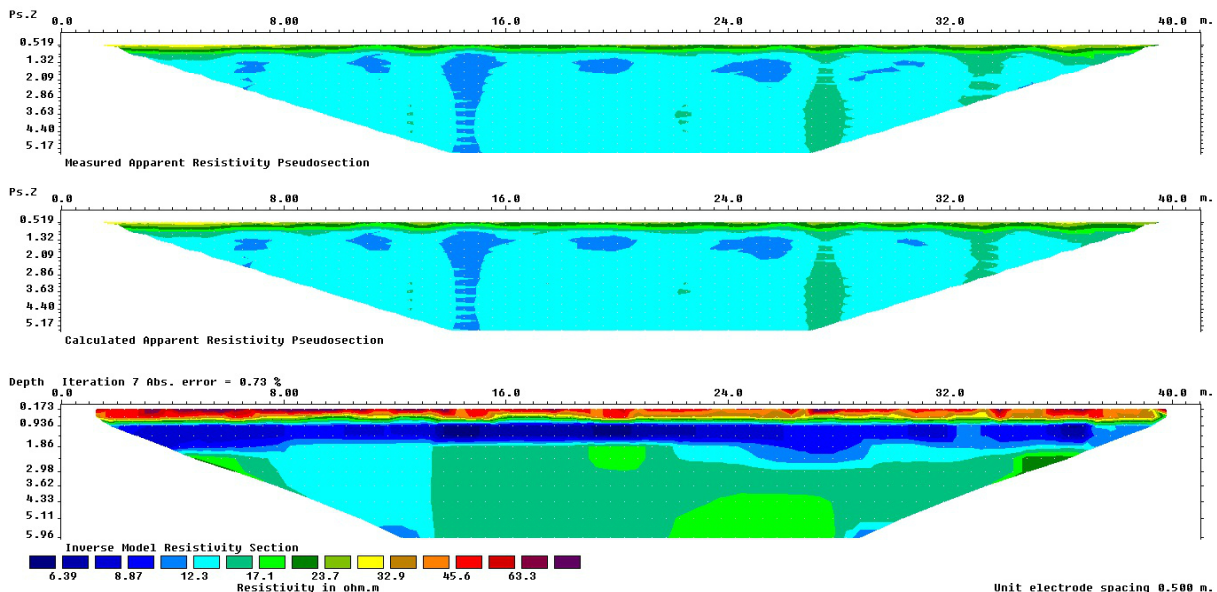


FIGURE D.9 – Mesures de résistivité issues du premier panneau électrique (ERT1).

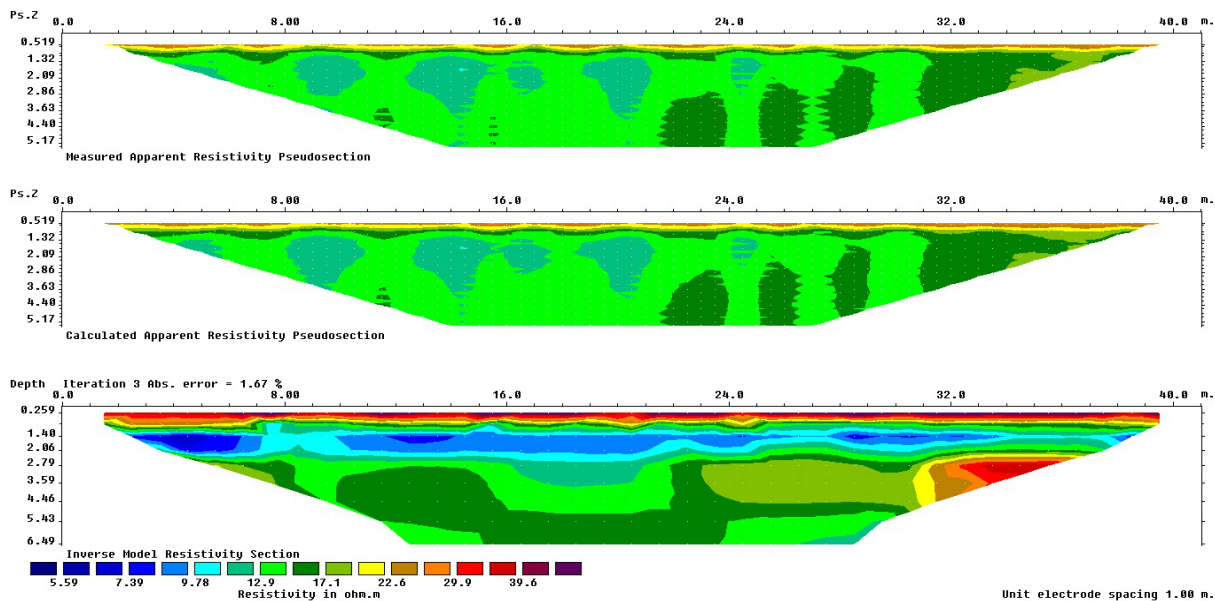


FIGURE D.10 – Mesures de résistivité issus du deuxième panneau électrique (ERT2).

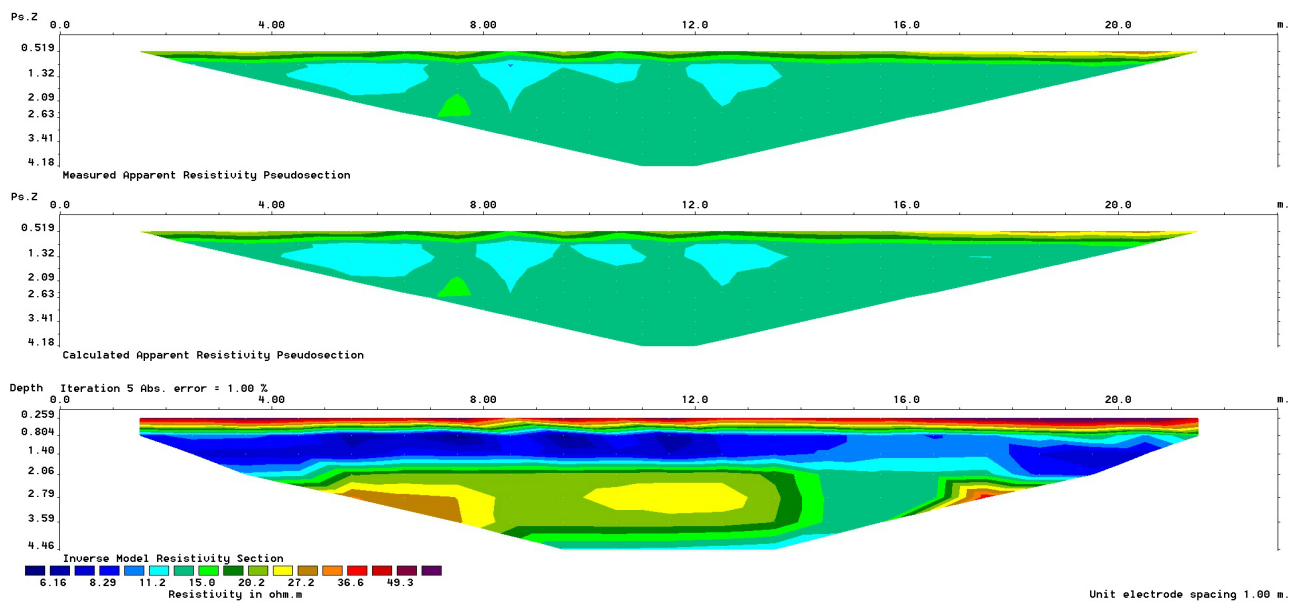


FIGURE D.11 – Mesures de résistivité issus du troisième panneau électrique (ERT3).