

# ESTIMATION DE LA NATURE DU SOL À PARTIR D'UN ESSAI PANDA2<sup>®</sup> AVEC UN RESEAUX DE NEURONES ARTIFICIEL

## AUTOMATIC METHODOLOGY TO PREDICT NATURAL GEOTECHNICAL PROFILE FROM DYNAMIC PENETRATION TEST USING NEURAL NETWORKS

C. Sastre<sup>1,2</sup>, P. Breul<sup>2</sup>, M.A. Benz-Navarrete<sup>1</sup>, C. Bacconnet<sup>2</sup>, R. Gourvès<sup>1</sup>

<sup>1</sup> Sol Solution Géotechnique-Réseaux, Riom, France

<sup>2</sup> Université Blaise Pascal, Laboratoire de Génie Civil Polytech Clermont-Ferrand

**RÉSUMÉ** – Le Panda2<sup>®</sup>, développé par Roland Gourvès en 1991, est un pénétromètre dynamique léger à énergie variable. Malgré tous ses avantages l'essai est dit aveugle puisqu'on ne prélève pas d'échantillon. Dans ce travail on propose une méthodologie de classification automatique de la nature du sol à partir du signal pénétrométrique Panda2<sup>®</sup> basées sur les réseaux de neurones à apprentissage supervisé.

**ABSTRACT** – The Panda 2<sup>®</sup>, developed by Roland Gourvès in 1991, is a lightweight dynamic cone penetrometer. Nevertheless it cannot take soil samples so the penetration test is called 'blind'. The aim of this paper is to propose an automatic methodology to predict the natural geotechnical profile from the cone resistance log using artificial neural networks.

### 1. Essai de pénétration dynamique Panda2<sup>®</sup>

Le pénétromètre dynamique Panda2<sup>®</sup> a été créé pour la reconnaissance de sols à faible profondeur jusqu'à environ cinq mètres (Benz-Navarrete, 2009). Largement répandu en France et dans le monde, il s'agit d'un pénétromètre léger dont l'énergie variable provenant de la frappe d'un marteau est mesurée pour chaque coup. La formule de battage dite « des Hollandais » permet de calculer la résistance de pointe pour chaque enfoncement. Les résultats sont usuellement présentés sous la forme d'un pénétoگرامme représentant la résistance dynamique de pointe  $q_d$  en fonction de la profondeur  $z$ . Un des grands intérêts de l'appareil est sa capacité à enregistrer pratiquement en continu la résistance du sol. Le pénétoگرامme ainsi obtenu est très riche d'informations sur la stratigraphie et sur les différentes propriétés du sol (Shahour et Gourvès, 2005) avec un nombre de données enregistrées important.

Malgré tous les avantages que présente le Panda2<sup>®</sup>, il n'en reste pas moins qu'on ne prélève pas d'échantillons lors de l'essai et on ne dispose donc d'aucune information sur la nature du sol. Cependant, on constate empiriquement que la forme du pénétoگرامme diffère d'un type de sol à un autre. La figure 1 montre à titre d'illustration trois pénétoگرامmes réalisés en chambre de calibration de 80 cm de hauteur au laboratoire, correspondant à trois types de sol de nature différente mais avec une densité et un état hydriques similaires.

Tableau 1. Paramètres d'état des sols représentés en figure 1

Nature	GTR	$\gamma_h$ (kN/m <sup>3</sup> )	$\gamma_d$ (kN/m <sup>3</sup> )	w (%)	$w_{OPN}$ (%)	État hydrique
Limon DGA	A1	17,8	16,2	10,0	18,4	Sec
Argile Laschamps	A2	17,6	15,6	12,9	18,1	Sec
Sable Sayat	B2	17,3	16,0	7,9	11,0	Sec

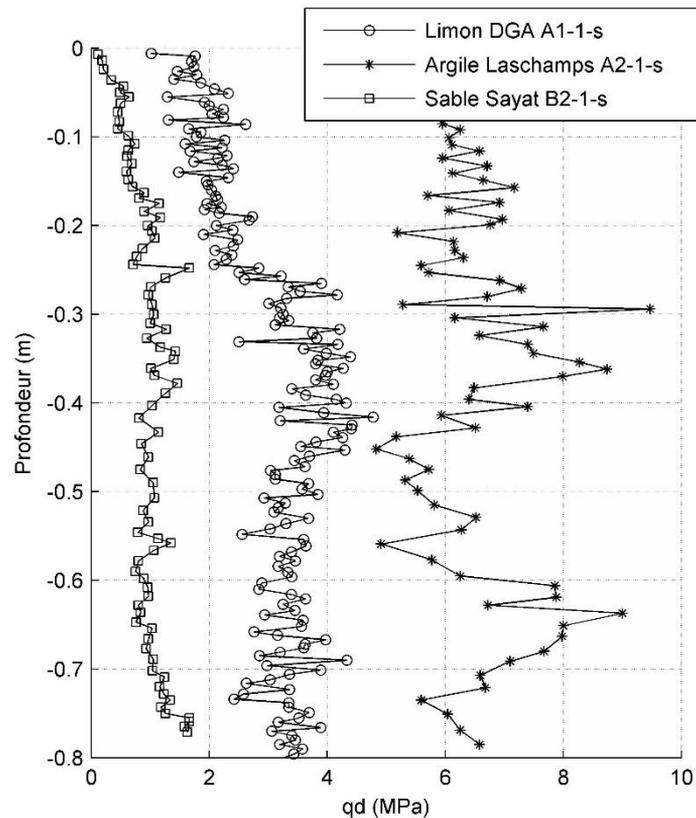


Figure 1. Exemple de pénétrogrammes obtenus dans des sols de nature différente mais avec des paramètres d'état similaires

Dans l'exemple choisi, les paramètres d'état (densité et teneur en eau) sont similaires. Toutefois, on peut remarquer aisément les différences morphologiques entre les trois pénétrogrammes. L'idée de base de la méthodologie est de proposer un outil permettant de différencier automatiquement ces variations qu'un œil expérimenté peut déceler dans la forme du signal en vue d'identifier la nature du sol testé. Bien que ces sols soient de natures différentes, ils peuvent présenter des valeurs de résistance moyenne très similaires tout en exhibant des morphologies de signaux très différentes. Pour mener à bien cette méthodologie, on propose de s'appuyer sur la théorie des Réseaux de Neurones Artificiels (RNA) et la Reconnaissance de Formes.

## 2. Les Réseaux de Neurones Artificiels

### 2.1. Généralités

Les RNA (Bishop, 1995), sont des modèles mathématiques basés sur le fonctionnement des réseaux de neurones biologiques. Ainsi, un réseau de neurones est constitué d'un graphe dont les nœuds sont les neurones, c'est-à-dire, les unités de calcul. Pour cet ensemble de neurones, on établit un flux d'information par le biais d'une topologie d'interconnexions (synapses). La sortie de chaque neurone est filtrée par une fonction d'activation ou fonction de transfert du neurone.

La littérature propose actuellement de nombreux RNA d'architecture différente. Dans cette étude, notre choix s'est porté sur deux classifieurs neuronaux supervisés : Perceptron Multicouche (PMC) et Réseau Probabiliste (PNN) dont

l'architecture est visible (figure 2). Ils sont appelés réseaux de type Perceptron ou en anglais Feed-Forward puisque l'information se propage vers l'avant sans retour en arrière possible.

Le PMC (Rumelhart et al. 1986) est l'architecture neuronale la plus répandue dans le domaine de la géotechnique (Shahin et al. 2008). Il s'agit d'une architecture stratifiée de neurones qui traitent l'information selon le modèle proposé par McCulloch-Pitts (McCulloch et Pitts, 1943). Le choix du PMC pour un problème de classification, comme celui qui nous intéresse, est justifié par le fait que, avec un nombre de neurones et de données d'apprentissage suffisant, ils sont capables de réaliser une approximation juste de n'importe quel type de fonction arbitraire (Poggio et Girosi, 1990).

Le PNN (Spetch, 1990) est un Réseau à Fonction de Base Radiale (RBF) dont la fonction d'activation est une fonction exponentielle gaussienne à la différence du PMC dont la fonction d'activation est souvent la tangente hyperbolique ou la fonction sigmoïde. Le PNN est capable d'estimer les fonctions de densité de probabilité pour chaque classe de sortie du réseau, grâce au théorème de Bayes, dans un problème de classification. Les avantages du PNN sont la fourniture des sorties du réseau en termes de probabilité, la rapidité avec laquelle l'apprentissage est réalisé et la facilité d'implémentation.

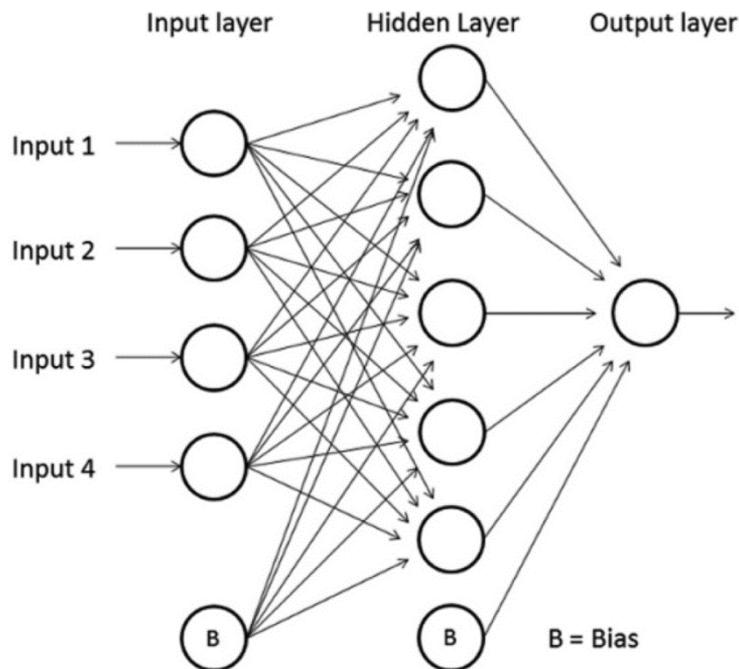


Figure 2. Structure général d'un réseau Feed-Forward

## 2.2. L'apprentissage du RNA. L'apprentissage supervisé

La construction d'un RNA comporte plusieurs étapes (entraînement, validation et test). L'étape d'apprentissage qui est sa caractéristique principale, correspond à la mise à jour des poids des connexions au sein du réseau, afin de réussir la tâche qui lui est demandée. Dans le cadre de l'apprentissage supervisé, on établit une correspondance entre une entrée et une sortie préalablement définie. On connaît donc, en tous points ou seulement en certains points, les valeurs que doit avoir la sortie du réseau en fonction des entrées correspondantes. Le réseau s'adapte par comparaison entre le résultat qu'il a calculé, en fonction des entrées fournies, et la réponse attendue en sortie. Ainsi, le réseau va se modifier jusqu'à ce qu'il trouve la bonne sortie, c'est-à-dire celle attendue, correspondant à une entrée donnée. Pour notre cas, ce type d'apprentissage est idéal

puisqu'on dispose des entrées, les pénétrogrammes, et des sorties, la classe du sol en fonction de sa nature permettant d'établir une sorte de mémoire associative intelligente.

### **2.3. Utilisation des Réseaux de Neurones en Géotechnique**

A l'heure actuelle, l'utilisation des algorithmes de Machine Learning, concrètement les RNA devient de plus en plus répandue dans les différents domaines de la géotechnique et de la mécanique de sol. Pour une vision plus globale de la mise en pratique des RNA dans ce domaine, le lecteur intéressé pourra se référer aux travaux de (Shahin et al. 2001) et (Shahin et al. 2009).

Une approche alternative de l'usage des RNA en géotechnique est présentée dans le travail de Sulewska (Sulewska, 2011). L'article analyse l'application des RNA dans 6 problèmes de géotechnique. Tous les réseaux utilisés ont une typologie PMC, avec une couche cachée. Le nombre de neurones de calcul varie entre 1 et 8 en fonction du problème traité. Les excellents résultats confirment l'intérêt de l'application des RNA dans notre domaine. C'est pour cette raison que le nombre d'utilisations de RNA en problèmes de régression en géotechnique est en augmentation constante. Enfin, l'application des RNA floue et les méthodes d'inférence Bayésienne s'avèrent d'après l'auteur comme un outil très puissant pour l'analyse de problèmes géotechniques dans l'avenir.

Si les RNA sont très bien adaptés pour des problèmes très complexes et pas complètement maîtrisés, la capacité explicative des réseaux supervisés reste limitée. Cette limite est due à l'opacité des RNA qui empêche une analyse pertinente des solutions obtenues. Les RNA peuvent être assimilés à une boîte noire qui donne une réponse quand on lui fournit des données mais qui ne délivre pas toujours de justification simple à analyser. Les liens existant entre les variables du modèle ne sont pas toujours détectés et de ce fait il s'agit d'un des principaux inconvénients pour son application.

L'autre grande contrainte des RNA est liée à la phase d'apprentissage. En effet, pour un fonctionnement correct du réseau, les données d'apprentissage doivent être représentatives de la situation à apprendre et doivent être en quantité suffisante, situation qui, surtout dans notre domaine, n'est pas toujours possible.

## **3. Démarche proposée**

L'idée de ce travail est de développer un RNA permettant d'identifier une classe de nature de sol à partir de la signature de son pénétrogramme obtenu à l'aide du Panda. Les étapes de ce développement sont: choix et établissement d'une base de données servant à l'apprentissage et à la validation du réseau ; définition des entrées du RNA à partir de la codification des pénétrogrammes ; définition des sorties du RNA sous forme de classes de nature de sols.

### **3.1. Choix de la base de données**

Pour que le RNA puisse faire un bon apprentissage et obtenir ainsi de bons résultats, on doit disposer d'une base de données robuste, suffisamment importante et représentative de la diversité des situations auxquelles le Réseau devra répondre. La base de données créée est composée de 218 pénétrogrammes obtenus dans des sols suffisamment homogènes et identifiés. Parmi ces essais, 149 ont été obtenus au laboratoire et 69 in situ.

Pour les essais au laboratoire, il s'agit des pénétrogrammes issus d'essais Panda2<sup>®</sup> réalisés en chambre de calibration de 80 cm de hauteur, pour une trentaine de sols « modèles » et pour lesquels on dispose de l'ensemble des informations de caractérisation géotechnique (Chaigneau, 2000).

Les échantillons restants sont des pénétrogrammes issus d'essais Panda2<sup>®</sup> réalisés in situ lors de campagnes géotechniques menées principalement en région Auvergne. Pour chacun des essais, nous disposons ainsi de l'identification des matériaux traversés à partir des essais de caractérisation au laboratoire sur les échantillons prélevés.

### 3.2. Définition des entrées du RNA

La seconde étape du développement du réseau consiste à identifier les paramètres permettant de caractériser le signal pénétrométrique Panda et qui constitueront le vecteur des entrées du RNA. Une analyse de sensibilité aux 26 paramètres initiaux a été menée par l'approche One-Factor-At-A-Time (OAT) en vue d'exclure les paramètres d'entrée n'étant pas très discriminants au sein du RNA. La méthode consiste à modifier chaque paramètre d'entrée du RNA d'une valeur de 10%, 50% et 100% de son écart type et de sa variance en gardant fixes les paramètres internes du RNA déjà calculés. L'effet de chaque modification est analysé en fonction de la variation de l'erreur finale du modèle (figure 3) afin de choisir les plus pertinents pour la classification (figure 3).

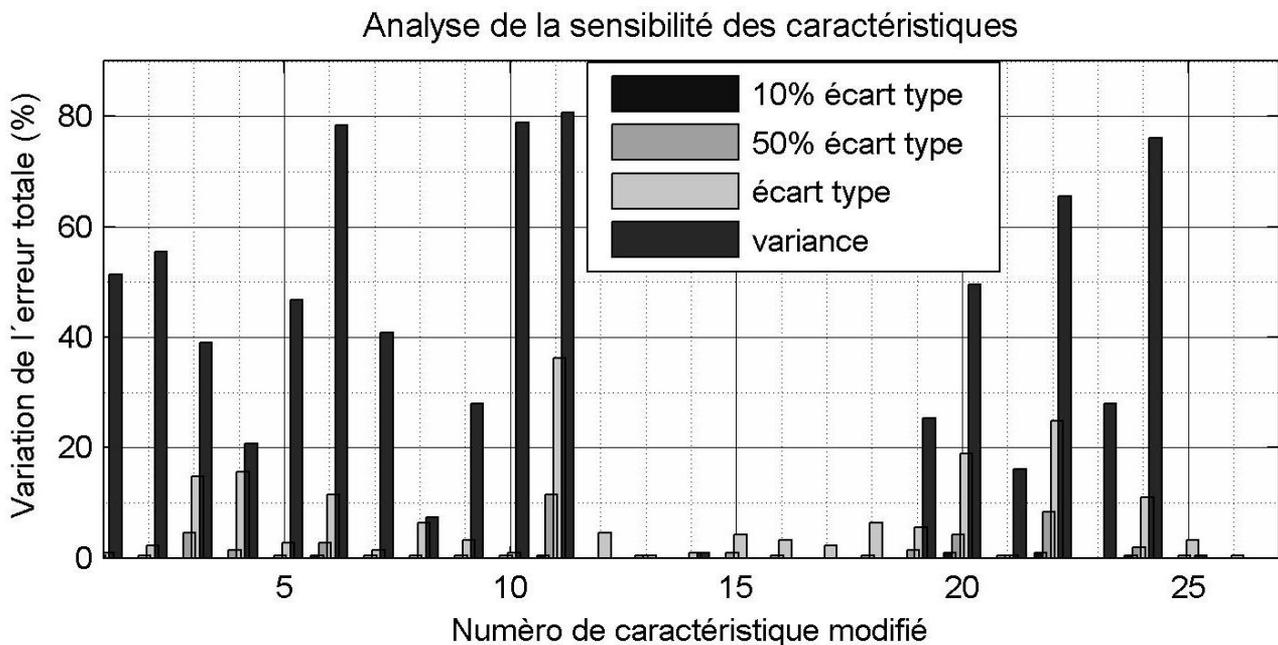


Figure 3. Résultats de la méthode PSA pour les 26 paramètres initiaux

Au final, on a réduit le nombre de paramètres du vecteur d'entrées, définis à partir de l'analyse (statistique, entropique, morphologique et spectrale) du signal pénétrométrique (résistance dynamique de pointe  $q_d$  et enfoncement  $e$ ), de 26 à 17 (tableau 2). On note que d'après l'analyse réalisée, les paramètres relatifs à l'enfoncement n'ont aucune influence dans la classification, déduction logique, compte tenu qu'il s'agit d'un essai à énergie variable.

Tableau 2. Paramétrage final du pénétrogramme

1. $q_d$ moyenne	6. Intervalle $q_d$	11. Entropie logarithmique $q_d$	16. Coefficient 2 droite régression
2. $q_d$ médiane	7. Écart interquartile $q_d$	12. Changement de signe de la pente $q_d$	17. Puissance spectral maximale
3. Écart type $q_d$	8. Asymétrie $q_d$	13. Croissement avec la ligne de tendance	
4. Coefficient de	9. Coefficient $q_d$	14. Forme de l'onde	

variation $q_d$	aplatissement		
5. Variance $q_d$	10. Entropie Shannon $q_d$	15. Coefficient 1 droite régression	

### 3.3. Définition des sorties du RNA

L'étape suivante a consisté à définir le nombre de classes de sortie du RNA fournissant la nature du sol. L'ensemble des sols testés dans notre base de données a été classifié selon la nomenclature GTR (NF P 11-300). Dans un premier temps, nous avons proposé de définir 4 classes de sortie différentes sur la base des paramètres granulométriques des matériaux (tableau 3).

Tableau 3. Classes de sol testées

Classe proposée	Classe GTR				Nature	Codification			
Classe 1	A1	A2	A3	A4	Sols fins	1	0	0	0
Classe 2	B5		B6		Sables/graves avec fines	0	1	0	0
Classe 3	D1	B1		B2	Sables	0	0	1	0
Classe 4	D2		B3	B4	Graves	0	0	0	1

### 3.4. Définition de l'architecture du PMC

Dans le cadre du Machine Learning, l'ensemble des données est normalement divisé en ensemble d'entraînement, validation et test (Bishop 1995, Ripley 1996), technique connue comme sous-échantillonnage. La pratique courante de sous-échantillonnage consiste à diviser la base de données de la manière suivante : 75% des échantillons pour l'ensemble d'entraînement et 15% pour les ensembles de validation et de test (ensembles indépendants et inconnus pour le RNA). Cette division a été retenue pour le calcul du réseau final dont la matrice de confusion (matrice fournissant le résultat final du RNA) est affichée dans la figure 4.

Un des inconvénients présentés par le PMC, à la différence du PNN, est l'absence d'une méthode systématique pour définir la meilleure topologie du réseau et le nombre de neurones cachés. On a donc adopté une stratégie afin de trouver la meilleure configuration du réseau par le biais d'une analyse multicritère, en étudiant plusieurs configurations possibles et l'influence des divers paramètres. Pour chaque neurone de calcul, on a entraîné 50 réseaux différents. Parmi tous les PMC testés, le modèle le plus performant pour le problème traité a les caractéristiques suivantes : une couche cachée de 12 neurones, un algorithme d'entraînement de Levenberg-Marquardt (Levenberg, 1944 ; Marquardt 1963), une normalisation des données dans l'intervalle  $[-1, 1]$  et une fonction de transfert de type tangente hyperbolique.

### 3.5. Performance du modèle et validation

A partir du meilleur modèle PMC, on a comparé la performance du PMC avec le PNN suivant leur taux de réussite sur le même ensemble de validation. Leurs taux de validation sont très similaires et élevés : 89% pour le PNN et 97% pour le PMC. Sur la base de ces résultats nous avons retenu le réseau PMC. La figure 4 fournit les matrices de confusion du réseau PMC final retenu. Un des intérêts de la matrice de confusion est qu'elle montre rapidement si le système parvient à classer correctement. Chaque colonne de la matrice représente le nombre d'occurrences d'une classe de référence, tandis que chaque ligne représente le nombre d'occurrences d'une classe estimée par le réseau. La première file, est celle de la classe dite 1 et la quatrième sera celle de la classe de sol 4. On affiche également les pourcentages d'échantillons correctement classés par ligne, colonne et le total de chaque ensemble.

		Entraînement							Validation				
Classe estimée		75	2	0	0	97%	Classe estimée		17	1	0	0	94 %
		1	18	0	0	94%			0	2	0	0	100 %
		0	0	25	0	100%			0	0	7	0	100 %
		0	0	0	31	100%			0	0	0	6	100%
		98%	90%	100%	100%	98%			100%	67%	100%	100%	97%
		Classe réelle							Classe réelle				
		Test							Total				
Classe estimée		18	0	0	0	100%	Classe estimée		110	3	0	0	97%
		0	3	0	0	100%			1	23	0	0	96%
		0	0	5	0	100%			0	0	37	0	100%
		1	0	1	5	71%			1	0	1	42	96%
		94,7%	100%	83%	100%	94%			98%	89%	97%	100%	97%
		Classe réelle							Classe réelle				

Figure 4. Matrices de confusion du PMC retenu comme classifieur final

Globalement, le PMC calculé a classifié correctement en 4 classes de sol 212 pénétrogrammes Panda, c'est-à-dire 97% du total des pénétrogrammes testés. Dans les phases de validation et de test, les taux de réussite finaux sont respectivement de 97% et 94%, ce qui démontre que la capacité de généralisation est très satisfaisante. Le tableau 4 ci-dessous, fournit un récapitulatif des résultats en termes de nature du sol.

Tableau 4. Bilan et compte de résultats du classifieur neuronal

Classe de sol	Tamisé	N <sup>bre</sup> d'échantillons	Classification correcte
Sol fins	80µm > 35%	113	98 %
Sables/graves avec fines	12% < 80µm ≤ 35%	24	89 %
Sables fins	80 µm ≤ 12% & 2mm > 70%	37	97 %
Graves	80 µm ≤ 12% & 2mm ≤ 70%	44	100 %

#### 4. Conclusions et perspectives

On a mis en place une méthodologie automatisée de prédiction de la nature du sol en fonction de ses paramètres granulométriques à partir de l'essai de pénétration dynamique Panda2<sup>®</sup>. Pour ce faire, le pénétrogramme a été paramétré à partir de quatre analyses de signal : statistique, non linéaire, morphologique et spectrale. Le paramétrage constitue le vecteur d'entrées d'un réseau de neurones artificiels ayant pour but de classer la nature du profil géotechnique sur la base de critères granulométriques. Ces réseaux ont été entraînés grâce à une base de données d'essais Panda2<sup>®</sup> réalisés in situ et au laboratoire. Les deux typologies de RNA choisies (PNN et PMC) ont fourni des résultats très prometteurs. Néanmoins, le PMC a eu des performances légèrement supérieures au PNN et a été donc retenu. Par contre, à long terme lorsqu'on incorporera de nouveaux échantillons, le PNN pourrait être plus approprié parce que c'est un réseau très rapide à construire et à entraîner. De plus, le fait d'exprimer la classe de sol en termes de probabilité est très séduisant puisque les résultats pourraient être accompagnés d'un indice de fiabilité.

La suite de ce travail consistera dans un premier temps à poursuivre la construction de la base de données afin d'optimiser la méthode et à continuer à étudier son applicabilité in

situ. Par la suite, l'incorporation d'autres paramètres mesurés sur les sols, tels que ceux obtenus à partir des images géoendoscopiques (Breul 1999, Haddani 2004) ou ceux du Panda 3 (Benz, 2009 ; Escobar, 2014), permettra d'améliorer encore cette classification.

## 5. Références bibliographiques

- Benz-Navarrete M.A. (2009). *Mesures dynamiques lors du battage du pénétromètre PANDA 2*. Chemical and Process Engineering. Université Blaise Pascal-Clermont-Ferrand II.
- Breul P. (1999). *Caractérisation endoscopique des milieux granulaires couplée à l'essai de pénétration*. Thèse de doctorat : Physique ; Clermont-Ferrand 2.
- Bishop C.M. (1995). *Neural Networks for Pattern Recognition*. Oxford, University Press.
- Chaigneau L., Boissier D., Bacconnet C. (2000). *Un modèle mécanique du complexe Sol/tranchée/réseau*. 1<sup>ères</sup> rencontres du réseau doctoral génie civil, Aussois 31 Janvier – 2 février.
- Escobar E.J. (2015). *Mise au point et exploitation d'une nouvelle technique pour la reconnaissance des sols : le Panda 3*. Thèse de doctorat Génie Civil Clermont-Ferrand.
- Haddani Y. (2004). *Caractérisation et classification des milieux granulaires par géoendoscopie*. Thèse de doctorat Génie civil Clermont-Ferrand 2.
- Levenberg K. (1944). *A Method for the Solution of Certain Problems in Least Squares*. Quart. Appl. Math. 2, 164-168.
- Marquardt D. (1963). *An Algorithm for Least-Squares Estimation of Nonlinear Parameters*. SIAM J. Appl. Math. 11, 431-441.
- McCulloch W.S., Pitts W. (1943). *A logical calculus of the ideas immanent in nervous activity*. Bulletin of Mathematical Biophysics 5:115-133.
- Poggio T., Girosi F. (1990). *Regularization algorithms for learning that are equivalent to multilayer networks*. Science. 247:978 982.
- Ripley, B.D. (1996). *Pattern Recognition and Neural Networks*. Cambridge: Cambridge University Press.
- Rumelhart D.E., Hinton G.E., Williams R.J. (1986). *Learning internal representation by error propagation*. Parallel distributed processing: Explorations in the microstructure of cognition, vol. 1, Pages 318-362.
- Shahin M.A., Jaksa M.B., Maier H.R. (2001). *Artificial Neural Network applications in geotechnical engineering*. Australian Geomechanics-March.
- Shahin M.A., Jaksa M.B., Maier H.R. (2008). *State of the art of artificial neural networks in geotechnical engineering*. Electronic Journal of Geotechnical Engineering.
- Shahin M.A., Jaksa M.B., Maier H.R. (2009). *Recent Advances and Future Challenges for Artificial Neural Systems in Geotechnical Engineering Applications*. Advances in Artificial Neural Systems.
- Shahrour I., Gourvès R. (2005). *Reconnaissance des terrains in-situ*. Hermès-Lavoisier.
- Specht D.F. (1990). *Probabilistic neural networks*. Neural Networks 3: 109-119.
- Sulewska M.J. (2011). *Applying Artificial Neural Networks for analysis of geotechnical problems*. Computer Assisted Mechanics and Engineering Sciences, 18: 231-241.