

Approches statistiques et numériques permettant d'estimer et prédire la densité sèche in situ à partir de données géotechniques

Statistical and numerical approaches for estimating and predicting in situ dry density from geotechnical data

Nayel EZZAIRI^{1,2}, Matthias FERREIRA^{3*},

1 Ecole Nationale Supérieure des Mines de Nancy, Nancy, France

2 Université de Lorraine, Nancy, France

3 Groupe Hydrogéotechnique, Eslettes, France

**m.ferreira@hydrogeotechnique.com*

RESUME : La densité sèche apparente des sols constitue un paramètre central dans la modélisation géotechnique, puisqu'elle influence directement ou indirectement – à travers l'état de contraintes initial – le dimensionnement et la stabilité des infrastructures. Or, sa détermination en laboratoire ne peut pas être systématiquement réalisée en raison de contraintes pratiques sur les échantillons prélevés in situ. Le choix de cette valeur repose donc souvent sur l'expérience de l'ingénieur géotechnicien, ce qui introduit une part de subjectivité et un risque d'incertitude dans la conception. Afin d'apporter un support objectif à la prise de décision, ce travail propose le développement d'un outil permettant d'estimer la densité sèche à partir de paramètres mesurables de manière plus courante. L'approche retenue consiste à exploiter des données issues d'essais pressiométriques, de la valeur au bleu et des analyses granulométriques, spécifiquement appliquées aux sols inorganiques. Pour cela, une base de données a été constituée à partir de campagnes de mesure réalisées sur les sols du Bassin parisien, représentatifs d'une grande diversité de faciès. Une analyse statistique exploratoire a d'abord permis d'identifier les corrélations existantes entre la densité sèche et les paramètres disponibles. Sur cette base, différents modèles d'apprentissage automatique supervisé ont été mis en œuvre afin d'évaluer leur capacité prédictive. L'outil développé vise ainsi à constituer une aide pratique pour les ingénieurs, en intégrant l'apprentissage automatique au service du dimensionnement géotechnique.

ABSTRACT: The dry bulk density of soils is a key parameter in geotechnical modelling, as it directly influences the design and stability of infrastructure. However, although laboratory testing provides accurate results, it cannot be carried out systematically due to feasibility constraints. In practice, the choice of this value is often based on the experience of the geotechnical engineer, which introduces a degree of subjectivity and a risk of uncertainty into the design. In order to provide objective support for decision-making, this work proposes the development of a tool for estimating dry density from more commonly measurable parameters. The approach adopted consists of using data from pressuremeter tests, blue value and grain size analyses, specifically applied to inorganic soils. To this end, a database was compiled from measurement campaigns carried out on soils in the Paris Basin, representative of a wide variety of facies. An exploratory statistical analysis was first used to identify the correlations between dry density and the available parameters. On this basis, various supervised machine learning models were implemented to assess their predictive capacity. The tool developed is therefore intended to be a practical aid for engineers, integrating machine learning into geotechnical design.

Mots-clés : Densité sèche ; essai pressiométrique ; courbe granulométrique ; apprentissage automatique ; analyse en composantes principales

1 INTRODUCTION

La densité sèche est une propriété intrinsèque du sol, dont la connaissance influe directement ou indirectement sur la conception de modèle géotechnique (Spagnoli et al., 2020) mais également en pédologie (Al-Shammary et al., 2018). Sa connaissance nous permet de connaître indirectement

le degré de compaction d'un sol sous tout type de conditions climatiques. Pour autant, plusieurs difficultés techniques et opérationnelles peuvent rendre la détermination de la densité sèche difficilement réalisable. Lorsqu'un carottage est effectué, la bonne tenue de l'échantillon prélevé constitue un facteur crucial pour la fiabilité des

mesures de densité. En effet, des mesures de densités sèches sont en pratique rarement réalisées sur des sols ne présentant pas ou peu de cohésion : sables, graves, vase, etc. Bien que la présence de cohésion dans les échantillons prélevés puisse constituer un premier biais dans la mesure de la densité sèche, une première hypothèse posée dans cet article est l'indépendance de la densité sèche vis-à-vis de la cohésion. Au cours des dernières décennies, de nombreux articles ont tenté d'établir des corrélations entre la densité sèche et d'autres paramètres géotechniques facilement mesurables, dont certaines sont compilées dans le Tableau 1. Néanmoins, peu d'articles mentionnent l'utilisation de méthode d'analyse statistiques exploratoires afin de déterminer les paramètres les plus pertinents pour estimer la densité sèche. Dans la littérature, il est toutefois observé l'usage de plus en plus conséquent de l'apprentissage automatique au travers différents modèles pour la prédiction de densité ou d'autres paramètres géotechniques.

Tableau 1. Quelques prédictions de la densité apparente (sèche ou humide) dans la littérature

Article	Relation / modèle	R ²
Tranter et al., 2007	Régression linéaire	0,41
Sequeira et al., 2013	Arbres décisionnels	0,50-0,71
Hattefard et al., 2023	Régression linéaire	0,06-0,97
<i>Id.</i>	Arbres décisionnels	0,79-0,88
<i>Id.</i>	Réseau de neurones	0-0,76
<i>Id.</i>	Machine à vecteurs supports	0,19-0,59

Cet article aborde la corrélation de la densité sèche vis-à-vis des paramètres obtenus dans le cadre des déterminations de classes de sol au sens du Guide des Terrassements Routiers (GTR) du Cerema (ex SETRA) ainsi que des paramètres obtenus dans le cadre d'essai pressiométrique. Des paramètres indirects ont également été calculé afin d'affiner les travaux statistiques antérieurs aux modèles de prédiction.

2 VARIABLES CALCULÉES

2.1 Passants

En considérant un échantillon de sol, le passant p est défini comme la fonction d'une variable réelle à valeur dans $[0,1]$ qui à un individu $\log(D)$ associe la part relative de la masse des grains dont le diamètre est inférieur à D . En pratique, la mesure de granulométrie est discrète et n'est faite que sur un nombre limité de passants : seules quelques valeurs de p sont connues pour un échantillon donné. Dans le cas d'un essai granulométrique permettant d'établir la classe de sol

au sens du GTR, les valeurs de p sont connues pour 0,08, 0,4, 2, 5, 20 et 50 mm. Le guide actualisé en 2024 introduit les passants pour 0,063, 0,63, 6,3 et 63 mm. Pour simplifier les notations, les variables paramétrées suivantes sont définies pour tout diamètre D exprimé en millimètres (1) :

$$p_D \stackrel{\text{def}}{=} p(\log D) \quad (1)$$

La définition d'une telle fonction aura son utilité pour le calcul de la graduation, définie ci-dessous. Les variables p_D pour les diamètres donnés par la première version du GTR sont donc des valeurs mesurées. p étant croissante, une corrélation positive est observée entre ces différentes variables. Afin de les décorrélérer partiellement, des paramètres sont définis pour tout diamètre $D_1 < D_2$:

$$p_{[D_1;D_2]} \stackrel{\text{def}}{=} p_{D_2} - p_{D_1} \quad (2)$$

Les paramètres $p_{[0,08;0,4]}$, $p_{[0,4;2]}$, $p_{[2;5]}$, $p_{[5;20]}$ et $p_{[20;50]}$ seront donc considérés pour la suite de ces travaux.

2.2 Argilosités synthétiques

L'essai au bleu de méthylène (VBS) est préconisé pour la classification d'un sol faiblement plastique au sens du GTR, Pour des sols en apparence plus cohérents (sols argileux très plastiques), l'essai établissant les limites d'Atterberg (indice de liquidité Wl , indice de plasticité IP ainsi que dans certains cas l'indice de consistance IC) est suggéré. L'alternance de ces deux essais sur les sols pose un problème dans la base de données : si un essai est systématiquement privilégié, cela introduit un biais évitable.

Sur la base de précédents travaux, des paramètres synthétisant les valeurs de ces deux essais sont calculés :

$$VBS_{syn} \stackrel{\text{def}}{=} 0,15IP + 2 \quad (3)$$

Ces paramètres synthétiques seront utilisés dans la base de données en complément des valeurs mesurées lors des classifications des sols au sens du GTR.

2.3 Graduation

Concrètement, la graduation d'un échantillon s'appuie sur le même raisonnement que le coefficient de Gini d'une population en analyse macro-économique. Ce coefficient permet de synthétiser les inégalités de richesses au sein d'une population dans une seule grandeur, comprise entre 0 et 1. Lorsque le coefficient de Gini est nul, une égalité parfaite de richesse est

constatée au sein de la population. Lorsque ce coefficient est égal à 1, un seul individu concentre l'ensemble des richesses de celle-ci.

Pour traduire ce raisonnement, la graduation exprime l'inégalité de répartition massique d'un échantillon de sol suivant le diamètre du grain. De la même manière que pour le coefficient de Gini, une graduation de 1 souligne une distribution parfaite de la masse suivant le diamètre des grains (équivalent anglophone du « *well-graded soil* » : grains hétérogènes) quand une valeur de 0 signifierait que ce sol contiendrait un seul diamètre de grain (équivalent du « *single sized* » ou « *poorly sized soil* » : grains homogènes).

Une première difficulté dans la définition de la graduation provient de l'échelle multiplicative utilisée pour mesurer la granulométrie d'un échantillon. En pratique une échelle semi-logarithmique est souvent utilisée pour représenter graphiquement une granulométrie. Ainsi, il sera considéré qu'une graduation ne pourra se faire que dans un intervalle de mesure compris entre un diamètre de grain minimal arbitraire noté d_{min} , et le diamètre du plus gros grain observé, noté d_{max} .

Définissons $I = [\log(d_{min}); \log(d_{max})]$ la population fictive de notre échantillon dans l'intervalle choisi. La population alors définie ne concerne pas la nombre de grain, il s'agit d'une « population » constituée de diamètres corrigés du logarithme.

On suppose que la fonction p qui à un individu $\log(D)$ associe le passant d'un échantillon inférieur à D est continue et de classe C^1 par morceaux. p étant croissante, la dérivée de p est positive lorsqu'elle peut être définie. Cette dernière peut être interprétée comme étant la densité de grain pour un diamètre donné.

Définissons à présent une fonction de tri \mathcal{T} qui à p' associe une fonction $\mathcal{T}(p')$ définie dans le même espace et vérifiant pour tout $\log(D)$ dans I :

$$\mathcal{T}(p')(\log(D)) \stackrel{\text{def}}{=} \inf\{t \in \mathbb{R} \mid \lambda(\{y \in I : p'(y) \leq t\}) \geq \log(D) - a\} \quad (6)$$

Où λ désigne la mesure de Lebesgue.

Une démonstration simple utilisant l'identité de Fubini permet de déduire la convexité positive de $\mathcal{T}(p')$ ainsi que les égalités suivantes :

$$p_{d_{max}} - p_{d_{min}} = \int_I p'(x) dx = \int_I \mathcal{T}(p')(x) dx \quad (7)$$

Nous pouvons alors définir la graduation g de la façon suivante :

$$g \stackrel{\text{def}}{=} \left(\int_I \int_{\log(d_{min})}^y \mathcal{T}(p')(x) dx dy \right) g_{max}^{-1} \cdot (8)$$

Où g_{max} est définie par :

$$g_{max} \stackrel{\text{def}}{=} \left(\int_I \int_{\log(d_{min})}^y \frac{p_{d_{max}} - p_{d_{min}}}{\log d_{max} - \log d_{min}} dx dy \right) \quad (9)$$

Cette définition mathématique permet de généraliser la notion de graduation pour tous types de sol et de mesures granulométriques. Pour autant, seules six valeurs de p sont mesurées lorsqu'une classification conforme au GTR est réalisée. Une extrapolation est alors réalisée ; en considérant deux à deux des valeurs mesurées consécutives p_{D_i} et $p_{D_{i+1}}$ l'extrapolée \tilde{p} est définie dans l'intervalle $[d_{min}; d_{max}]$. Pour tout x dans $[\log D_i; \log D_{i+1}]$:

$$\tilde{p}(x) \stackrel{\text{def}}{=} p_{D_i} + \frac{p_{D_{i+1}} - p_{D_i}}{\log D_{i+1} - \log D_i} (x - \log D_i) \quad (10)$$

\tilde{p} étant dès lors une fonction vérifiant les mêmes hypothèses que p , il est donc possible de calculer sa graduation sur l'intervalle $[d_{min}; d_{max}]$.

L'intérêt d'une telle variable est grand puisque la distribution des diamètres des grains dans un sol sec impacte directement sa densité sèche. En effet, une distribution, où une très forte homogénéité des grains est observée, aura une densité sèche bien plus faible qu'un sol où les interstices laissés par cette homogénéité serait comblée par des grains plus petits (Roquier, 2022).

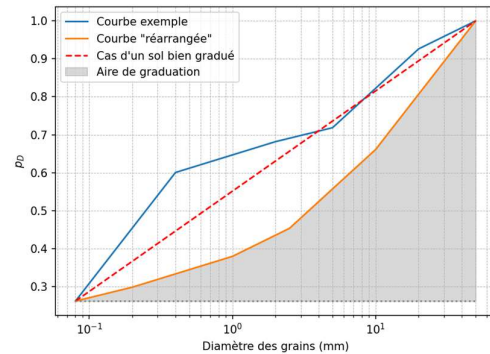


Figure 1. Exemple de calcul de la graduation sur un échantillon suivant la classification du GTR de 1992. La courbe « réarrangée » correspond à une primitive de $\mathcal{T}(p')$. La graduation correspond à l'aire grisée par rapport à l'aire sous la courbe d'un sol bien gradué (en rouge pointillé)

Une corrélation entre la graduation calculée et la densité sèche est bien observée en Figure 2. D'une façon plus générale, définir une telle variable permet d'attribuer une grandeur quantitative à un paramètre jusqu'à présent exprimé à l'aide de termes qualitatifs.

Tableau 2. Caractéristiques des variables des bases de données

Variables	Base d'entraînement				Base de test			
	Moyenne	Ecart-type	Min	Max	Moyenne	Ecart-type	Min	Max
$p_{0,08}$	56,26	28,51	3,20	99,90	53,86	30,08	4,99	99,85
$p_{[50;20]}$	6,10	8,39	0,00	50,00	6,27	10,47	0,00	72,89
$p_{[20;5]}$	7,55	8,67	0,00	40,53	8,25	9,80	0,00	38,76
$p_{[5;2]}$	3,25	3,71	0,00	22,85	3,54	4,32	0,00	21,00
$p_{[2;0,4]}$	6,04	7,66	0,00	70,04	6,36	8,60	0,00	62,00
$p_{[0,4;0,08]}$	18,43	20,64	0,00	93,88	19,22	21,98	0,00	94,31
g	0,39	0,17	0,15	0,82	0,38	0,18	0,16	0,79
$VBS_{p_{20}}syn$	2,29	2,44	0,04	13,25	2,15	2,45	0,00	12,47
ρ_h	1,99	0,16	1,52	2,56	1,99	0,16	1,60	2,52
ρ_d	1,62	0,26	1,03	2,43	1,62	0,25	1,02	2,31
w_n	0,24	0,11	0,01	0,59	0,24	0,11	0,03	0,60
$\log E_m$	2,41	1,268	-1,11	6,32	2,49	1,22	-0,03	6,13
$\log pl^*$	-0,01	1,03	-3,61	1,92	0,02	1,04	-2,39	1,66
$\log pf^*$	-0,52	1,15	-4,02	1,92	-0,51	1,17	-3,32	1,61

3 ANALYSE EXPLORATOIRE DES DONNEES

3.1 Base de données

Le jeu de données est une compilation des sondages et essais pressiométriques entre 2013 et 2025 sur les départements de Seine-Maritime et de l'Eure. La base complète contient 1635 entrées, mais seules 904 d'entre elles présentent également des mesures de densités.

Chaque entrée correspond à une combinaison d'un essai pressiométrique, d'une classification GTR et d'une densité à une profondeur donnée. Étant donné que les sondages carottés et les essais pressiométriques ne sont pas réalisés aux mêmes coordonnées, une distance de 20 mètres au maximum est acceptée, sous réserve que le relief du terrain naturel, la description lithologique restent similaire entre ces deux points. Les profondeurs doivent également ne pas différer de moins d'un mètre. Ce dernier critère s'entend par le fait qu'un essai pressiométrique et qu'une classification GTR restent des mesures statistiques sur des échantillons d'environ 1 mètre de hauteur.

3.2 Corrélations

La visualisation des corrélations de Pearson existantes entre chaque paramètre des modèles testés fait partie intégrante de l'analyse exploratoire de données. Celles-ci sont résumées en Figure 2. Pour la variable qui nous intéresse, à savoir ρ_d , nous observons une forte corrélation négative avec la teneur en eau w_n ainsi que des corrélations moins marquées avec

$\log E_m$, $p_{[5;2]}$ ou bien $p_{0,08}$. Le paramètre ρ_h (densité humide apparente) présente en toute logique une forte corrélation. Celle-ci ne sera pas exploitée étant donné que ρ_h est directement utilisée pour calculer ρ_d .

3.3 Analyse en composantes principales

Une analyse en composantes principales (ACP) est réalisée sur l'ensemble des paramètres du jeu de données. Cette analyse permet de résumer l'information contenue dans la base de données en un nombre réduit de variable, indépendantes entre elles.

En pratique, sur les grandes bases de données, l'ACP est une étape de pré-calcul pour l'apprentissage automatique. Le choix a été fait de conserver les paramètres d'origines lors des calculs de prédictions afin de rendre l'interprétation des résultats plus aisée.

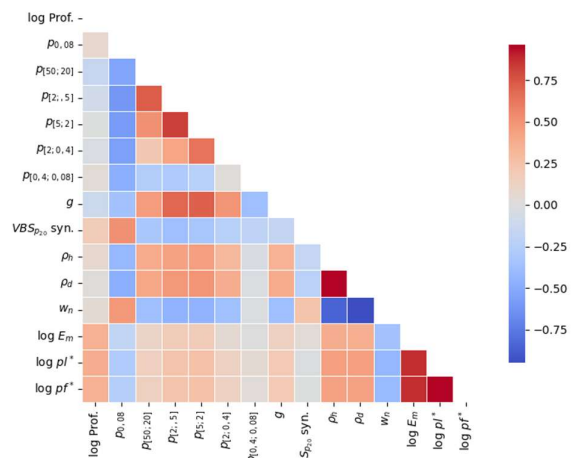


Figure 2. Matrice des corrélations entre chaque variable de la base de données

Dans la Figure 3, la première composante semble être une réduction aux diamètres des grains, où des valeurs négatives correspondent à des argiles et des limons et les valeurs positives des sables et des graves. La deuxième composante est plus difficile à interpréter mais semble être liée à la compaction des sols. Le Tableau 3 indique en effet que les deux principales composantes de la base de données n'expliquent que 56% de la variance observée. Les observations réalisées sur la troisième composante indiquent que celles-ci caractérisent la présence de matériaux sableux fins, principalement corrélées à $p_{[0,4;0,08]}$.

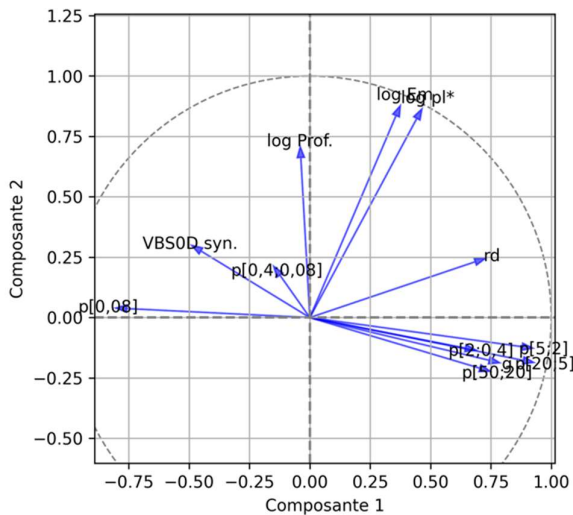


Figure 3. Cercles de corrélations pour les 2 composantes principales. Notes de lecture : Sur le plan de ces composantes, la densité sèche ρ_d (notée rd) est corrélée négativement à $p_{0,08}$. $p_{[0,4;0,08]}$ semble corrélée à $VBS_{p_{20}} syn.$ (notée VBS0D syn.). Pour autant, les deux variables sont mal représentées sur ce plan : elles contribuent davantage à d'autres composantes.

Tableau 3. Part de la variance expliquée par chacune des 5 composantes principales

Dimension	Part de variance expliquée	Cumul
1	0,384	0,384
2	0,176	0,560
3	0,140	0,700
4	0,079	0,779
5	0,059	0,839

L'observation d'une forte corrélation entre la teneur en eau de la base de données et les densités sèches ont souligné l'intérêt de l'estimation de ces dernières à partir de ce seul paramètre. L'outil le plus naturel pour cette estimation est la régression linéaire. Une autre approche est néanmoins développée en partie suivante.

Enfin, l'analyse en composantes principales tend à montrer que les 3 premières composantes de la base de données suffisent à expliquer une bonne part de la variance observée.

4 REGRESSION ET PREDICTION

4.1 Teneur en eau

L'approche proposée consiste à estimer la densité sèche à partir de la teneur en eau :

$$w = \frac{\rho_h}{\rho_d} - 1 \quad (11)$$

Soit :

$$\rho_d = \frac{\rho}{w+1} \quad (12)$$

Généralement, sur les essais réalisés, la valeur de ρ_h est comprise entre 1,8 et 2,1. Ainsi une première estimation de la densité sèche peut être faite à partir de la teneur en eau dans cette fourchette. Pour pousser le raisonnement, une dépendance de la densité humide à la teneur en eau est observée.

La corrélation donnée avec un R^2 de 0,725 est :

$$\rho_h = 2,2963 - 1,2604w \quad (13)$$

La relation que nous proposons pour la densité sèche est avec un R^2 de 0,920 :

$$\rho_d = \frac{-1,26w+2,3}{w+1} \quad (13)$$

Les grands avantages de cette corrélation sont sa fiabilité et la facilité à mettre en œuvre un essai de teneur en eau. Qui plus est, celle-ci semble correspondre à tous les types de sols de la région, comme visible en Figure 4.

4.2 Limites de cette approche

Dans la base de données, la mesure de la teneur en eau est réalisée systématiquement dans le cadre d'une mesure de densité des sols. Cette mesure est réalisée sur un échantillon intact, au même titre que la mesure de densité. Il n'est donc pas certain que la teneur en eau mesurée sur des échantillons déstructurés reproduise la même tendance. Des corrélations entre la teneur en eau sur des échantillons déstructurés (pouvant provenir des classifications suivant le GTR) et la teneur en eau sur des échantillons intacts devront être posées.

Cette approche suppose qu'un matériau naturellement sec présente toujours la même densité sèche apparente, autour de 2,30. De façon implicite, ce constat souligne que la portée de cette relation est limitée aux régions où les minéraux constitutifs des sols sont similaires à ceux du Bassin parisien.

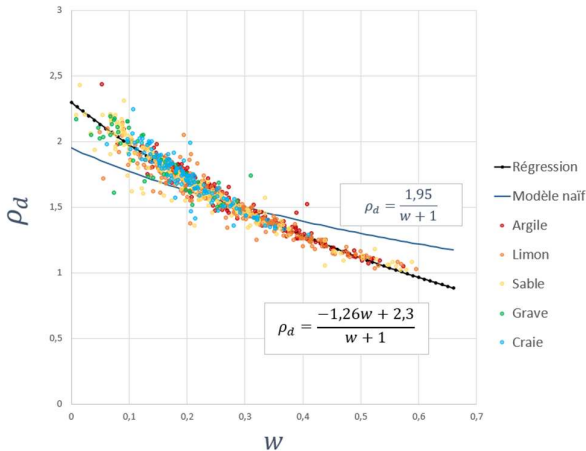


Figure 4. Prédiction de la densité sèche suivant la teneur en eau.

Des paramètres climatiques (pluviométrie, température, altitude) ou géologiques (terrains alluviaux, colluviaux, etc.) pourraient également influencer sur la valeur de la densité sèche apparente pour un matériau sec. Les effets de ces paramètres sur la densité sèche restent encore à déterminer.

Pour autant, cette corrélation confirme la caractérisation d'un sol compressible et évolutif par une faible densité sèche ou une forte teneur en eau.

4.3 Apprentissage automatique

La deuxième approche est plus indirecte et se base sur les diverses données issues d'essais pressiométriques et des classifications au sens de GTR de 1992. Des données pressiométriques, il est possible de déduire :

- La pression limite nette, la pression limite de fluage ainsi que le module pressiométrique de Ménard. De ces trois grandeurs, la pression limite de fluage n'est pas utilisée car sa détermination sur les courbes de pression/volume est délicate. Il nous reste alors le module pressiométrique et la pression limite nette. L'analyse exploratoire de données nous donne d'ailleurs de bonnes corrélations avec la densité sèche.

Tableau 4. Modèles d'apprentissage automatique et résultats associés

Modèle	Références	Hyperparamètres ¹	Nombres de paramètres ²	Résultats		
				RMSE	R^2_{train}	R^2_{test}
Perceptron multicouche (réseau de neurone)	MLP-5		5	0,183	0,562	0,471
	MLP-6	early_stopping, alpha,	6	0,176	0,563	0,511
	MLP-11	hidden_layer_size	11	0,164	0,564	0,574
	MLP-16		16	0,158	0,717	0,608
K plus proches voisins	KNN-5		5	0,204	0,542	0,342
	KNN-6	n_neighbors	6	0,186	0,489	0,452
	KNN-11		11	0,177	0,519	0,505
Machine à vecteurs de support	SVR-5		5	0,179	0,455	0,491
	SVR-6	C, epsilon, kernel, gamma	6	0,171	0,531	0,541
	SVR-11		11	0,169	0,555	0,548
Amplification de gradient	GB-5	n_estimators,	5	0,183	0,627	0,473
	GB-6	learning_rate, subsample	6	0,176	0,600	0,512
Forêts aléatoires	RF-5	n_estimators, max_depth,	5	0,186	0,516	0,459
	RF-6	max_features	6	0,176	0,533	0,511
Régression linéaire	Linear-5		5	0,192	0,360	0,421
	Linear-6		6	0,187	0,375	0,447
	Linear-11	-	11	0,186	0,380	0,455
	Linear-16		16	0,181	0,399	0,485

¹ : Les hyperparamètres permettent d'ajuster le modèle aux caractéristiques de la base de données (taille, nombre de paramètres) afin d'éviter au maximum tout sur- ou sous-apprentissage. Pour chaque modèle, seuls les hyperparamètres les plus importants sont ajustés, les autres ayant des valeurs par défaut suivant la documentation de la librairie `sk_learn`.

² : Sur les modèles à 5 paramètres, les données d'entrée sont : $p_{0,08}$, g , $VBS_{p_{20}}_{syn.}$, $\log pl^*$ et $\log E_m$. Sur les modèles à 6 paramètres, $p_{[5;2]}$ est rajouté. Le type de sol est donné sur les modèles à 11 paramètres (argile, limon, sable, grave et craie). Enfin, $p_{[50;20]}$, $p_{[20;5]}$, $p_{[2;0,4]}$, $p_{[0,4;0,08]}$ et $\log Prof.$ sont donnés en entrée pour MLP-16 et Linear-16.

- Une lithologie supposée, basée sur l'interprétation des paramètres de forages et les essais pressiométriques, permettent de déduire de façon formalisée une variable qualitative dans 5 valeurs possibles : {argile, limon, sable, grave, craie}

Les données de classification selon le GTR ne peuvent être obtenues sans réaliser de sondages carottés. En ce sens, pour obtenir l'ensemble des paramètres d'entrée de chaque modèle, la réalisation de sondages carottés et pressiométriques restent nécessaires.

Avant l'entraînement, les données sont normalisées : elles sont centrées et réduites. Cette étape fait partie intégrante du pré-calcul avant tout entraînement de modèle. Cette étape est essentielle lorsque les paramètres d'entrée n'ont pas les mêmes unités. Elle permet d'adimensionner ces paramètres avant l'analyse statistique mais également lors de l'entraînement des modèles.

En tout, 18 modèles sont entraînés puis validés dans la base de test. Les résultats sont résumés dans le Tableau 4 en Figure 5. Sur l'ensemble des tests réalisés, les meilleurs résultats sont obtenus avec le modèle MLP-16. Il s'agit d'un modèle de perceptron multicouche (*multilayer perceptron* ou *MLP*), un type de réseau de neurones artificiels. Ce modèle est organisé en une couche d'entrée de neurones permettant la lecture des 16 paramètres, en 2 couches dites « cachées », comptant respectivement 50 et 20 neurones, permettant le traitement de ces données et d'une couche de sortie pour le système global.

Les meilleurs résultats obtenus à l'aide du modèle MLP-16 par rapport au modèle Linear-16, ayant le même nombre de paramètres en entrée, soulignent le

caractère non linéaire des relations entre les paramètres géotechniques et la densité sèche.

Avec des coefficients de détermination inférieur à 0,608 et un grand nombre de paramètres utilisés, ces modèles ne dépassent toutefois pas les estimations de la densité sèches à partir de la teneur en eau. Leur intérêt réside dans une approche pratique globale d'estimation des paramètres d'un modèle géotechnique : une classification au sens des GTR et un essai pressiométrique suffisent à calculer dans un premier temps ces 16 paramètres puis dans un deuxième temps à estimer la densité sèche avec une erreur moyenne de 0,16.

Aussi ces résultats démontrent que ces modèles peuvent servir à mettre en cohérence des données en apparence peu corrélées. L'usage de l'apprentissage automatique pour estimer la densité sèche ou d'autres paramètres géotechniques à partir de données pressiométriques ou autres (pénétrömètre statique ou dynamique par exemple) est prometteur en ce sens.

5 CONCLUSIONS

En conclusion de cet article, la mise en œuvre d'approches statistiques et prédictives sur les bases de données constituées lors de précédents forages géotechniques est un axe de développement technique majeur pour la profession.

La teneur en eau sur échantillon intact s'est avérée être un très bon estimateur de la densité sèche. Pour autant, les modèles d'apprentissage automatique donnent de bons résultats pour la prédiction de cette densité à partir de données pressiométriques et de classification au sens du GTR 1992.

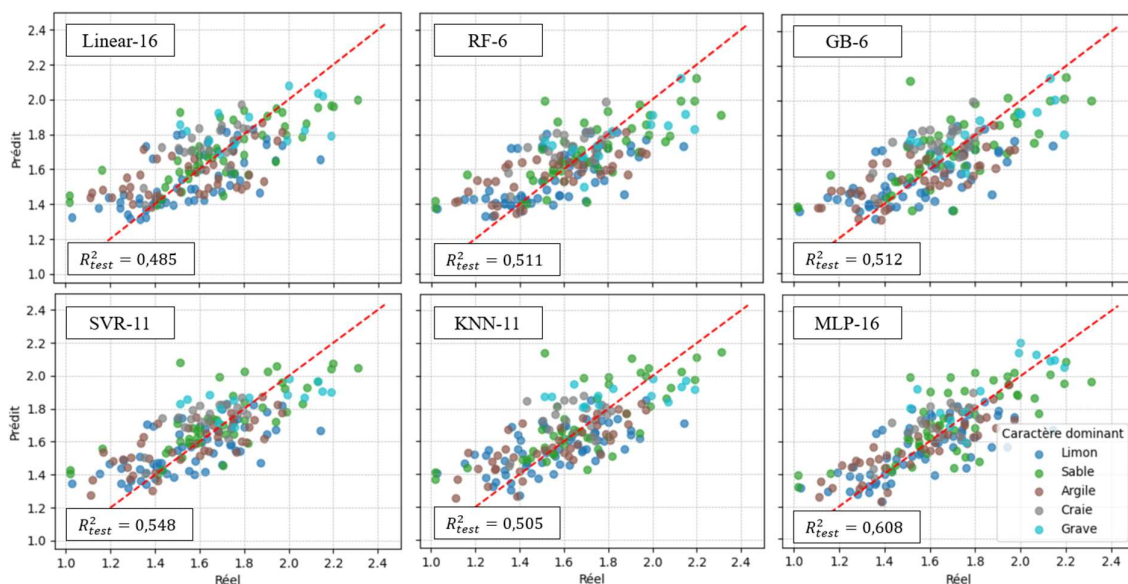


Figure 5. Prédiction de la densité sèche à partir de quelques modèles sur la base d'entraînement.

Plusieurs pistes de recherches émergent des suites des résultats de cet article :

- Les mesures de teneur en eau étant réalisées sur des échantillons intacts, cohérents, il reste à démontrer que ces mesures restent similaires sur un matériau déstructuré afin de pouvoir confirmer la relation (13) pour la mesure de la teneur en eau suivant la norme NF EN ISO 17892-1
- La base de données sur laquelle sont basés les résultats de cet article est relativement petite. Davantage de données permettront d'affiner, voire d'améliorer les modèles d'apprentissage automatique.

Dans un premier temps, pour l'estimation et l'utilisation de la relation (13) la réalisation de la teneur en eau sur échantillon intact est préconisée. Pour l'apprentissage automatique, davantage de données devront être formatées pour l'entraînement de ces modèles afin d'améliorer leurs résultats.

6 REMERCIEMENTS

Les auteurs souhaitent remercier l'ensemble de l'équipe de l'agence d'Eslettes (76) du bureau d'étude HYDROGEOTECHNIQUE pour l'aide technique et la mise à disposition des résultats des essais ayant servi à la constitution de la base de données préalable à la rédaction de cet article.

7 REFERENCES BIBLIOGRAPHIQUES

- Al-Shammary, A. A. G., Kouzani, A. Z., Kaynak, A., Khoo, S. Y., Norton, M., & Gates, W. (2018). Soil bulk density estimation methods: A review. *Pedosphere*, 28(4): 581–596. [https://doi.org/10.1016/S1002-0160\(18\)60034-7](https://doi.org/10.1016/S1002-0160(18)60034-7)
- Hateffard, F., Szatmári, G., & Novák, T. J. (2023). Applicability of machine learning models for predicting soil organic carbon content and bulk density under different soil conditions. *Soil Science Annual*, 74(1): Article 165879. <https://doi.org/10.37501/soilsa/165879>
- Roquier, G. (2022). A versatile model for estimating the minimum void ratio of binary soil mixtures based on packing density theory. *Engineering Geology*, 296: Article 106453. <https://doi.org/10.1016/j.enggeo.2021.106453>
- Sequeira, C. H., Wills, S. A., Seybold, C. A., & West, L. T. (2014). Predicting soil bulk density for incomplete databases. *Geoderma*, 213: 64–73. <https://doi.org/10.1016/j.geoderma.2013.07.013>
- Spagnoli, G., & Shimobe, S. (2020). State-of-the-art overview on compaction characteristics of soils by laboratory tests. *Engineering Geology*, 278: Article

105830.

<https://doi.org/10.1016/j.enggeo.2020.105830>

- Tranter, G., Minasny, B., McBratney, A. B., Murphy, B., McKenzie, N. J., Grundy, M., & Brough, D. (2007). Building and testing conceptual and empirical models for predicting soil bulk density. *Soil Use and Management*, 23(4): 437–443. <https://doi.org/10.1111/j.1475-2743.2007.00092.x>