



Rapport de stage

Page 1/66

Réf. :
Aix-Marseille Université

Date : 22/07/2025

Indice :

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES

Direction des Énergies

Institut des Sciences Appliquées et de la Simulation pour les énergies bas carbone

Département de Modélisation des Systèmes et Structures

Service d'Etudes Mécaniques et Thermiques

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES**BOU SERHAL Elio**

Aix-Marseille Université

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

	Rapport de stage	Page 2/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

MOTS CLES

Interaction sol -structure, apprentissage automatique, interface sol-fondation, modèle non-linéaire, python

RÉSUMÉ/CONCLUSIONS rédigé pour être de niveau DL si le document est de niveau DR, S ou TS

Les fondations basculantes offrent un mécanisme potentiel pour améliorer la performance sismique des systèmes structurels en autorisant des mouvements contrôlés de soulèvement, tassement et de rotation. Cependant, les incertitudes inhérentes à la modélisation de l'interface dynamique sol-fondation limitent leur emploi comme solution de conception parasismique.

Dans ce travail, tout d'abord le comportement sismique du système sol-fondation-structure (SFSI) a été investigué au moyen d'une base de données expérimentales ensuite numérique générée à partir de simulations paramétrées intégrant la variabilité des propriétés du sol, des caractéristiques de la fondation et la superstructure ainsi que du chargement sismique. Les résultats bruts de calculs ont été exploités et post-traités afin de dériver des indicateurs représentatifs tels que : le déplacement maximal (PGD), la vitesse maximale (PGV), l'accélération maximale (PGA), l'intensité d'Arias (Ia), le facteur de sécurité (FS), l'énergie dissipée normalisée (NED), le tassement normalisé (NS) et le rapport d'amplification (AAR).

Une analyse statistique exploratoire a été réalisé sur l'ensemble de données expérimentales et numériques afin d'étudier la distribution des variables, leurs corrélations et l'effet de transformation logarithmique. Malheureusement la base de données ne nous a pas permis d'atteindre nos objectifs et c'est pour cela on a décidé de continuer avec une base de données numériques qui enrichie a servi à l'entraînement et à l'évaluation de plusieurs modèles d'apprentissage automatique notamment le K-Nearest Neighbors (KNN), le Gradient Boosting (GB) et le Extreme Gradient Boosting (XGBoost).

Les résultats obtenus avec les trois modèles KNN, GB et XGB mettent en évidence que la combinaison de la simulation numérique et les méthodes d'apprentissage automatique évaluent l'impact des paramètres matériau et géométriques sur la réponse dynamique des structures basculante. Parmi ces trois modèles appliqués, XGBoost est le plus performant qui a capté la non-linéarité. En se basant sur les résultats de ce modèle, l'impact des variables sur le système est classé par un ordre croissant comme le suivant : PGD, PGA, FS, PGV et Ia.

Ces résultats ouvrent de nouvelles perspectives, en particulier l'intégration de données expérimentales réelles et le recours à des modèles d'apprentissage plus avancés pour optimiser la conception parasismique des structures.

	Rapport de stage	Page 3/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

SOMMAIRE

1 INTRODUCTION	9
2 BIBLIOGRAPHIE	10
2.1 Interaction Sol Fondation structure (SFSI).....	10
2.1.1 Méthodes Numériques	10
2.1.2 Méthodes Expérimentales.....	11
2.1.3 Méthode de Machine Learning.....	13
2.2 Idée générale sur les ML.....	14
2.3 Régression Linéaire (LR) et K-plus proche voisins (KNN)	15
2.3.1 Gradient Boost (GB) et XGBoost (XGB)	17
3 BASE DE DONNEES EXPERIMENTALES	19
3.1 Première base de données : « FoRDy »	19
3.2 Deuxième Base de données « CAMUS IV »	20
3.3 Exploitation et Méthodologie	21
3.4 Conclusion	27
4 BASE DE DONNEES NUMERIQUE	28
4.1 Introduction	28
4.2 Modélisation numérique	28
4.2.1 Structure générale du modèle	30
4.2.2 Noeud et éléments du modèle	31
4.2.3 Sauvegarde des résultats numériques.....	32
4.3 Génération des paramètres.....	33
4.3.1 Paramètres du sol	33
4.3.2 Paramètres du chargement.....	36
4.3.3 Paramètres de la structure	36
4.3.4 Combinaison et Fusion des paramètres	39
5 EXPLOITATION ET PRETRAITEMENT	39
5.1 Introduction	39
5.2 Importance des fichiers « . Outs » générés	39
5.3 Post-traitement des variables	40
5.4 Analyse Univariée	42
5.5 Transformation logarithmique et nettoyage.....	44

	Rapport de stage	Page 4/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

5.5.1	Représentation Graphiques des Features	44
5.5.2	Représentation graphique des outputs	48
5.6	Analyse Multivariée	50
5.6.1	Matrice de corrélation.....	50
5.6.2	Matrice de distribution	52
5.7	Résultats et Conclusion	52
6	PARTITIONNEMENT ET APPLICATION DES ML.....	54
6.1	Introduction	54
6.2	Application du premier modèle « KNN ».....	54
6.2.1	Ajustement des paramètres du modèle KNN	54
6.2.2	Comparaison des mesures métriques du modèle KNN	55
6.2.3	Evaluation d'importance des variables selon le modèle KNN	56
6.3	Application du deuxième modèle « GB ».....	57
6.3.1	Ajustement des paramètres du modèle GB	57
6.3.2	Comparaison des mesures métrique du modèle GB	57
6.3.3	Evaluation d'importance des variables selon le modèle GB	59
6.4	Application di troisième modèle « XGB »	60
6.4.1	Ajustement des paramètres	60
6.4.2	Comparaison des mesures métrique du modèle XGB.....	60
6.4.3	Evaluation d'importance des variables selon le modèle XGB	61
7	CONCLUSION ET PERSPECTIVES	63
8	REFERENCES	65

		Rapport de stage	Page 5/66
		Réf. : Aix-Marseille Université	
		Date : 22/07/2025	Indice :
	RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

Liste des Figures

Figure 1 : Comparaison des deux modèles KNN et MLR	16
Figure 2 : Matrice de corrélation des variables de "FoRDY et CAMUS.....	22
Figure 3 : Evaluation de la performance des variables du "FoRDY et CAMUS	24
Figure 4 : Diagramme en barres d'Importance des variables du "FoRDY et CAMUS"	26
Figure 5 : Matrice d'importance des variables du "FoRDY et CAMUS"	27
Figure 6 : Schéma du modèle BNWF (Raychowdhury & Ugalde, s.d.)	29
Figure 7 : Schéma du modèle CIM (Raychowdhury & Ugalde, s.d.)	30
Figure 8 : Illustration du modèle (Raychowdhury & Ugalde, s.d.).....	32
Figure 9 : Répartition des valeurs de phi selon le type du sol	35
Figure 10 : Répartition des valeurs de phi selon module d'Young.....	35
Figure 11 : Les différentes configuration de la structure.....	38
Figure 12 : Distribution du PGD après prétraitement.....	45
Figure 13 : Distribution du PGV après prétraitement	46
Figure 14 : Distribution du PGA après prétraitement	46
Figure 15 : Distribution de la après prétraitement.....	47
Figure 16 : Distribution du FS après prétraitement	47
Figure 17 : Distribution de NED après prétraitement	48
Figure 18 : Distribution de NS après prétraitement.....	49
Figure 19 : Distribution de AAR après prétraitement	49
Figure 20 : Matrice de corrélation des variables d'intérêt après prétraitement	51
Figure 21 : Matrice de distribution des variables d'intérêt après prétraitement	53
Figure 22 : Evaluation de la performance du modèle KNN sur l'SFSI	55
Figure 23 : Evaluation de l'importance du modèle GB sur l'SFSI normalisées [0-1]	56
Figure 24 : Evaluation de la performance du modèle GB sur l'SFSI	58
Figure 25 : Evaluation de l'importance du modèle GB sur l'SFSI normalisées [0-1]	59
Figure 26 : Evaluation de la performance du modèle XGB sur l'SFSI	60
Figure 27 : Evaluation de la performance du modèle XGB sur l'SFSI	62
Figure 28 : Evaluation de la performance du modèle XGB sur l'SFSI	60

	Rapport de stage	Page 6/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
	RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES	

Liste des Tableaux

Tableau 1 : Tableau de corrélation des Indicateurs sismique selon Gavras (Gavras et al., 2020)	12
Tableau 2 : MAPE des modèles MLR et KNN selon Gajan (Gajan, 2021)	14
Tableau 3 : Objectifs principaux du « FoRDY »	20
Tableau 4 : Tableau statistiques des variables de "FoRDY et CAMUS"	21
Tableau 5 : Ajustement des paramètres "FoRDY et CAMUS"	23
Tableau 6 : Paramètres du sol (« Geomechanics- Bornes des paramètres », s.d.).....	34
Tableau 7 : Paramètres du chargement.....	36
Tableau 8 : Paramètres de la structure (« Geomechanics- Bornes des paramètres », s.d.)	37
Tableau 9 : Contenu et utilité principale des fichiers «. Outs ».....	40
Tableau 10 : Variables d'entrées et de sorties.....	41
Tableau 11 : Tableau statistique des variables avant prétraitement	43
Tableau 12 : tableau statistique des variables après prétraitement	44
Tableau 13 : Ajustement des hyperparamètres du modèle KNN	54
Tableau 14 : Evaluation de la performance du modèle KNN sur l'SFSI	55
Tableau 15 : Ajustement des hyperparamètres du modèle GB	57
Tableau 16 : Evaluation de la performance du modèle GB sur l'SFSI.....	58
Tableau 17 : Ajustement des hyperparamètres du modèle XGB	60
Tableau 18 : Evaluation de la performance du modèle XGB sur l'SFSI	60

		Rapport de stage	Page 7/66
		Réf. : Aix-Marseille Université	
		Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

Glossaire

Terme / sigle	Définition
Opensees	
Fondation oscillante (Rocking foundation)	Système de fondation qui est conçu pour permettre un léger basculement ou pivotement sous l'effet de sollicitations sismiques, afin de limiter les forces transmises à la superstructure.
Interaction sol-fondation-structure (SFSI)	Effet combiné du sol, de la fondation et de la structure sur le comportement global d'un ouvrage soumis à des charges dynamiques, notamment sismiques.
Basculement (Rocking motion)	Mouvement rotatif d'une fondation autour de ses bords sous l'effet de forces latérales, entraînant une perte partielle de contact avec le sol.
Non-linéarité du sol	Comportement du sol qui ne suit pas une relation proportionnelle entre contrainte et déformation, surtout en cas de fortes sollicitations comme un séisme.
Machine Learning ML (Apprentissage automatique)	Ensemble de méthodes statistiques permettant aux ordinateurs d'identifier des modèles dans des données sans être explicitement programmés pour cela.
Feature importance (Importance des variables)	Mesure de l'influence relative de chaque paramètre d'entrée sur les résultats d'un modèle d'apprentissage automatique.
Cdg	Centre de Gravité
Campagne CAMUS IV	Programme expérimental réalisé sur des structures en béton armé pour étudier leur comportement sous sollicitations sismiques.
Méthode de validation croisée (Cross-validation CV)	Technique utilisée pour évaluer la performance d'un modèle de machine Learning en le testant sur différentes sous-parties du jeu de données.
Modèle prédictif	Outil mathématique ou statistique qui vise à anticiper des résultats futurs en fonction d'un ensemble de variables explicatives.
Variable Explicatives	Variables contenant des valeurs importantes pour faire des prédictions
Variables Cibles	Variables intéressantes sur lesquelles la prédition est faite
Features ou Input	Les paramètres utilisés pour s'entraîner afin de faire des prédictions
Output	Les paramètres que l'algorithme cherche à prédire ou calculer

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

		Rapport de stage	Page 8/66
		Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :	
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

Underfitting (Sous apprentissage)	Modèle trop simple qui donne des mauvaises performances sur l'ensemble d'entraînement et celui de test
Overfitting (Surapprentissage)	Modèle trop complexe qui donne une bonne performance sur l'ensemble d'entraînement mais pas sur l'ensemble de test
KNN	K-Nearest Neighbor : modèle d'apprentissage automatique
GB	Gradient Boost : modèle d'apprentissage automatique
XGB	XGBoost : modèle d'apprentissage automatique
IQR	IQR = Q1 – Q3 = c'est la dispersion centrale des données (en utilisant quantile 1 et quantile 3) cette mesure est utilisé pour détecter les valeurs aberrantes avec la règle IQR par défaut k=1.5) tel que : $x < (Q1 - k) \cdot IQR$ ou $x > (Q3 + k) \cdot IQR$
NED	Normalized Energy Dissipation : dissipation d'énergie normalisée.
NS	Normalized Settlement : tassement normalisé
AAR	Amplification Acceleration Ratio : rapport d'amplification des accélérations.
Ia	Instantaneous Arias Intensity : intensité d'Arias.
FS	Facteur de Sécurité (capacité portante / sollicitation).
PGD	Peak Ground Displacement : déplacement horizontal maximal.
PGV	Peak Ground Velocity : vitesse associée au déplacement horizontal.
PGA	Peak Ground Acceleration : accélération maximale mesurée.

	Rapport de stage	Page 9/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

1 INTRODUCTION

En génie parasismique, la rotation des fondations offre un mécanisme potentiel pour améliorer la performance sismique des systèmes structurels en permettant des mouvements contrôlés de soulèvement et de tassement. Plusieurs études expérimentales et numériques ont montré qu'on peut profiter du comportement oscillant pour réduire la demande sismique sur les structures (Deng & Kutter, 2012 ; Gajan & Kutter, 2008).

Cependant, malgré ces avantages, l'intégration de la rotation des fondations dans la conception parasismique des structures reste limitée. Cela s'explique par un certain nombre de difficultés et d'incertitudes, parmi lesquelles :

- La complexité de la modélisation : le comportement du système sol-fondation-structure (SFSI) est fortement non linéaire. Il implique à la fois l'interaction non linéaire à l'interface sol-fondation, définie comme non linéarité géométrique tels que le soulèvement, le tassement ou le glissement, et le développement de la non linéarité des matériaux au cours du chargement sismique. La prise en compte de ces mécanismes nécessite des modèles numériques sophistiqués, nécessitant, pour certains modèles d'interface de contact, une calibration délicate et un cout de calcul en temps et en mémoire considérable.
- Le manque de données expérimentales : Les bases de données disponibles sont limitées et généralement hétérogènes. La variabilité provient des moyens d'essais, des matériaux employés pour le sol et la structure, des configurations considérées (fondation enfoncée, sol hétérogène, sol saturé en eau etc.) et des sollicitations de chargements. Ce défi d'homogénéité données fiables réduit la robustesse des prédictions et alimente les incertitudes.

Ce travail de stage vise à identifier les paramètres physiques influençant la rotation, le tassement et le décollement des fondations superficielles soumises à des charges sismiques. Pour se faire, des approches issues de l'apprentissage automatique sont appliquées tout d'abord à une base de données expérimentale en libre accès à laquelle les données de la campagne expérimentale Camus IV ont été rajoutées. Par la suite, cette base de données a été enrichi par des données simulées numériquement avec le logiciel OpenSees et Python.

Le mémoire est structuré de la manière suivante : tout d'abord, une bibliographie divisée en deux parties est présentée : la première concerne l'introduction sur l'ISFS, et la deuxième porte sur les modèles d'apprentissage automatique. Ensuite, deux bases de données, expérimentales et numériques, sont détaillées. Par la suite, les modèles d'apprentissage automatiques sont explorés, avant de présenter les résultats obtenus à l'application de modèles d'apprentissage automatique non linéaires et terminer par les conclusions et les perspectives de ce travail.

	Rapport de stage	Page 10/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

2 BIBLIOGRAPHIE

2.1 Interaction Sol Fondation structure (SFSI)

Un séisme est le tremblement de terre causé par la libération soudaine d'énergie suite à la rupture d'une faille dans la croûte terrestre. Cette libération d'énergie engendre la propagation des ondes sismiques à travers les couches géologiques profondes modifiant ainsi les conditions dynamiques du milieu.

Une fois arrivée en surface, les ondes sismiques sollicitent les structures et induisent l'SFSI. Cette interaction se traduit par :

- 1- L'interaction cinématique : le sol modifie le mouvement imposé à la base de la structure.
- 2- L'interaction inertielle : les forces générées par la structure rétroagissent sur la fondation et le sol.

Il existe trois parties à modéliser : la première c'est le sol, la deuxième c'est la fondation qui est souvent représenté par des raideurs et amortissements. La troisième partie qui est la structure sur laquelle la réponse du séisme apparaît sous forme de translation, rotation et basculement. Donc, le système SFSI est un phénomène complexe à modéliser, vu qu'il dépend de nombreuses caractéristiques mécaniques du sol et de la structure.

Le Basculement superficiel c'est un mécanisme non linéaire observé lors de secousses sismiques où la fondation pivote autour d'un bord entraînant un renversement qui s'accompagne souvent de tassements et rotations résiduels. La stabilité de la fondation dépend des rapports sans dimension comme : Facteur de sécurité FS et niveau de Moment qui dépend eux aussi de plusieurs paramètres (sol, fondation et structure).

Plusieurs méthodes d'analyse du système sol-fondation-structure ont été adaptées, notamment la première : méthode numérique (Drosos, Gerolymos, & Gazetas, 2012 ; Gajan, Raychowdhury, Hutchinson, Kutter, & Stewart, 2010a ; Lee & Salgado, 2000). La deuxième méthode : les méthodes expérimentales (Deng & Kutter, 2012 ; Gavras et al., 2020). Finalement, la troisième : l'utilisation des modèles d'apprentissage automatique (Gajan, 2021 ; Soundararajan & Gajan, 2020). Alors pour mieux comprendre on va expliquer la différence entre ces 3 méthodes utilisées pour améliorer la compréhension du système.

2.1.1 Méthodes Numériques

Les méthodes numériques visent à modéliser le comportement non linéaire des fondations superficielles et leur interaction avec la structure. Parmi ces approches, on retrouve les modèles Beam-on-Nonlinear Winkler Foundation (BNWF), les macro-éléments ou encore les simulations en éléments finis. Gajan et al. (2010) ont proposé deux outils pratiques (BNWF et CIM) implémentés dans OpenSees, validés par comparaison avec des essais centrifuge (Gajan et al., 2010a). De leur côté, Chatzigogos et al. (2009) ont développé une formulation de macro-élément permettant de représenter

	Rapport de stage	Page 11/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

les non-linéarités matérielles et géométriques dans le calcul de la fondation (Pecker, Paolucci, Chatzigogos, Correia, & Figini, 2014). Ces approches sont plus flexibles que les seules expériences, car elles permettent d'explorer de larges espaces paramétriques, mais elles nécessitent une calibration précise à partir des données expérimentales.

2.1.2 Méthodes Expérimentales

Les méthodes expérimentales constituent la base de l'étude du système sol–fondation–structure (SFSI). Elles reposent principalement sur des essais en centrifugeuse, sur table vibrante ou encore in situ, permettant de reproduire les conditions sismiques réelles à échelle réduite. Ces campagnes fournissent des données fiables sur les mécanismes de basculement, la dissipation d'énergie et les tassements permanents. Par exemple, Gavras et al. (2020) ont compilé la base FoRDy, regroupant plus de 200 cas expérimentaux dynamiques, tandis que Deng et Kutter (2012) ont utilisé la centrifugeuse pour analyser le comportement de ponts sur fondations basculantes. Ces expériences constituent une référence indispensable pour comprendre le comportement réel du SFSI et valider les modèles numériques et prédictifs. (Deng & Kutter, 2012) (Gavras et al., 2020)

En se basant sur l'article de Gavras l'objectif c'est de constituer une base de données expérimentale (FoRDy) de 245 essais normalisés pour évaluer la performance des fondations superficielles basculantes soumises à des charges sismiques en intégrant une variabilité importante de dans les paramètres des matériaux, des géométries et des chargements dynamiques. En plus, Identifier les indicateurs sismiques les plus corrélés à leur déplacement. La problématique réside dans l'assemblage, la structuration et l'analyse des résultats expérimentaux hétérogènes pour mieux comprendre le comportement dynamique des fondations superficielles et améliorer les modèles prédictifs ? La démarche est la suivante :

- Collecte des données : 245 expériences sur des centrifugeuses et des tables vibrantes 1g.
- Sélection / Calcul des variables (Input & Output): PGA (Peak Ground Acceleration), PGV (Peak Ground Velocity), PGD (Peak Ground Displacement), Sd (Standard deviation), Cr (Critical ratio) etc...
- Normalisation et Transformation : valeurs centrées autour de zéro entre -1 et 1.
- Analyse corrélationnelle : Étude de la corrélation entre le déplacement maximal des structures et divers indicateurs d'intensité sismique pour identifier ceux les plus pertinents.
- Comparaison et Conclusion

Après le prétraitement de la base de données (159 de 245 essais expérimentaux sont extraits de la base FoRDy (excluant certains biais initiaux ou cas particuliers), Gavras à calculer les variables d'intérêt comme des indicateurs de performances dans le but d'expliquer la variable cible selon la formule. Après le prétraitement de la base de données (159 de 245 essais expérimentaux sont extraits de la base FoRDy (excluant certains biais initiaux ou cas particuliers), Gavras à calculer les variables d'intérêt comme des indicateurs de performances dans le but d'expliquer la variable cible selon la formule suivante [1]

	Rapport de stage	Page 12/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

$$yipk = \frac{y_{max} - y_{end}}{H}$$

[1]

Où : y_{max} est le déplacement latéral maximum au sommet de la structure pendant le séisme (en valeur absolue), y_{end} est le déplacement résiduel au sommet de la structure à la fin de l'essai et H = hauteur de la structure (ou distance entre la base et le point de mesure).

Ensuite, la performance prédictive de chaque indicateur sismique (IM) est évaluée via : Le coefficient de détermination R² donnée au-dessous [2]:

$$R = \sqrt{\frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2}$$

[2]

Tableau 1 : Tableau de corrélation des Indicateurs sismique selon Gavras (Gavras et al., 2020)

IM (Indicateur Sismique)	R ² (tous)	Commentaires
PGA (Peak Ground Acceleration)	0.46	Corrélation faible – peu représentatif du drift
PGV (Peak Ground Velocity)	0.84	Corrélation forte – meilleur indicateur que PGA
PGD (Peak Ground Displacement)	0.85	Meilleure corrélation globale – indicateur le plus fiable
Spectre de déplacement moyen (Sd,geo[T1∈[0.2,3]]/h)	~0.83–0.86	Très bonne corrélation – sensible à la forme du spectre

Conclusion : L'indicateur PGD ressort comme le meilleur prédicteur du drift maximum incrémental (Yipk), suivi de PGV. En outre, les indicateurs spectraux comme Sd, geo(T1∈[0.2,3])/h sont également efficaces, à condition que la forme du spectre soit bien prise en compte. Alors, l'utilisation de PGA n'est pas recommandée seule pour estimer la réponse mécanique des structures à fondation basculante.

	Rapport de stage	Page 13/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

2.1.3 Méthode de Machine Learning

Les méthodes de Machine Learning se sont imposées comme un outil complémentaire pour l'étude du SFSI. Ces approches permettent de capturer les relations complexes et non linéaires entre les paramètres d'entrée (ex. géométrie, propriétés du sol, intensité sismique) et les performances globales (dissipation d'énergie, tassement, amplification d'accélération). Gajan (2021) a montré que les algorithmes de régression (MLR, k-NN) surpassaient les modèles statistiques simples pour prédire le comportement de fondations basculantes. De même, Soundararajan & Gajan (2020) ont exploité une base de 142 essais dynamiques pour identifier les corrélations clés entre le coefficient de rocking et la performance en utilisant 3 modèles KNN, DTR et SVR. Ces méthodes ouvrent la voie à une meilleure généralisation des comportements observés et à une optimisation des conceptions parasismiques. (Gajan, 2021 ; Soundararajan & Gajan, 2020)

En se basant sur l'article de Gajan, L'objectif c'est de prédire la performance des fondations superficielles basculantes lors des séismes en utilisant l'IA spécifiquement les modèles d'apprentissage automatique. C'est le premier article qui introduit l'IA dans le domaine sismique.

La problématique de cet article c'est : Comment Prendre en considération tous paramètres physiques et les conditions sismiques pour prédire la performance des fondations superficielles et modéliser le système sol-fondation-structure en utilisant l'Apprentissage Automatique.

Les démarches qui ont été suivies :

- Collecte des données : 140 de 245 essais expériences sur des centrifugeuses et des tables stables.
- Sélection / Calcul des paramètres (Input & Output) : la, Cr, NED, NS et AAR
- Normalisation et Transformation en log : valeurs centrées autour de zéro entre -1 et 1.
- Partitionnement de la Base de Données : deux ensembles Train & Test.
- Application d'Apprentissage Automatique : modèle linéaire et non-linéaire : MLR et KNN.
- Comparaison et Conclusion

Conclusion, les résultats que Gajan a obtenu après l'application du ML est présenté sous forme d'un tableau ci-dessous qui montre que la valeur de MAPE du modèle MLR est plus grande que celle du modèle KNN.

	Rapport de stage	Page 14/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

Tableau 2 : MAPE des modèles MLR et KNN selon Gajan (Gajan, 2021)

Paramètre évalué	MAPE	
	MLR	K-NN
NED	0.76	0.42
NS	4.05	0.97
AAR	0.26	0.15

En résumé, les deux articles nous ont indiqués les variables inputs, outputs et quelles sont les variables d'intérêt ensuite l'article de Gajan nous a facilité la tâche de choisir les modèles d'apprentissage automatique non-linéaire. Donc, maintenant ma étude va suivre les mêmes démarches mais avec des améliorations : création de ma propre base de données, ajout de nouvelles variables explicatives, application d'autres algorithmes de ML plus avancés dans le but d'obtenir une analyse plus robuste, comparative et mieux comprendre l'interaction sol-fondation-structure (SFSI).

2.2 Idée générale sur les ML

Les Machine Learning ou l'Apprentissage Automatique est un algorithme qui permet aux ordinateurs d'apprendre à partir de données sans être explicitement programmés pour chaque tache. L'objectif principal de ces algorithmes c'est de construire des modèles capables de détecter des relations complexes entre variables d'entrée (Features) et variable de sortie (cibles), afin de prédire ou classer de nouvelles observations.

Dans ce travail, le recours au ML se justifie par plusieurs raisons : L'étude du système sol-structure sous chargement de séisme présent souvent des relations non-linéaires et difficiles à modéliser par des simples équations analytiques. En plus, les modèles d'apprentissage automatique prennent en compte un grand nombre de variables explicatives et offrent une meilleure robustesse face aux données bruitées ou hétérogènes. Ainsi que ces modèles facilitent la comparaison entre eux selon des critères statistiques choisis pour identifier quel modèle fournit la meilleure précision prédictive.

	Rapport de stage	Page 15/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

En résumé, l'utilisation des modèles ML dans ce travail permet d'améliorer la qualité des prédictions et de mieux comprendre l'influence des paramètres sur les variables cible autrement dit, c'est de mieux comprendre le comportement de ces paramètres et variables sur l'interaction du système sol-fondation-structure. En général, chaque modèle admet une équation, fonction d'optimisation et paramètres à ajuster. Dans ce travail on va utiliser 3 modèles non-linéaire KNN, GradientBoost et XGBoost. Ce choix des modèles vient des résultats des articles qui montrent que la relation de ce système est non-linéaire., (Gajan, Raychowdhury, Hutchinson, Kutter, & Stewart, 2010b), (Anastasopoulos & Kontoroupi, 2014), (Soundararajan & Gajan, 2020) et (Gajan, 2021) etc ...

2.3 Régression Linéaire (LR) et K-plus proche voisins (KNN)

En se basant sur l'article (Gajan, 2021) ou ils ont utilisé deux modèles un linéaire et un autre non linéaire pour faire la comparaison et montrer que les modèles non linéaire décrivent mieux le système, on va montrer dans un tableau la différence entre ces deux modèles ensuite expliquer l'algorithme de chacun: Comme c'est mentionné dans le tableau ci-dessus, le MLR a une équation du problème linéaire (combinaison linéaire) alors que celle de KNN dépend du poids (pondération). En plus, le KNN n'admet pas une fonction de cout c'est un algorithme non paramétrique basé sur la similarité entre observation seulement ce qui nous intéresse pour optimiser c'est de trouver à chaque fois la distance minimale entre le nouveau point x_{test} et les K points de données x_{train} dans le but de faire des prédictions en fonction de ces K voisins : cette prédition se fait En régression (la prédition est la moyenne des valeurs cibles des voisins) ou en classification (la prédition correspond à la classe majoritaire parmi les voisins), cela dépend sur la variable cible si elle est quantitative ou qualitative, mais ici dans ce travail les variables cible qu'on a sont quantitatives alors on va travailler avec des classifications pour calculer à chaque fois l'équation du problème, la fonction de cout qui donne l'erreur MAPE minimale. Alors que le modèle de la Régression linéaire est un modèle statistique classique qui suppose une relation linéaire entre la variable Cible et plusieurs variables explicatives. Ce modèle dépend des paramètres Beta qui prennent des valeurs selon l'algorithme de descente de gradient et calcule l'équation du problème ensuite la fonction de cout à chaque fois pour comparer à la fin ces fonctions et choisir la meilleure qui prédit avec une erreur MAPE minimale (voir Figure 1).

		Rapport de stage	Page 16/66
		Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :	
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

	Multiple Linear Regression (MLR)	K- Nearest Neighbor (KNN)
Type du modèle	Linéaire	Non-Linéaire
Equation du Problem	$\hat{y} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p + \xi$	$\hat{y} = \frac{\sum_{i=1}^k w_i \cdot y_i}{\sum_{i=1}^k w_i}$
Mesure Metrique	$MAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{\hat{y}_i - y_i}{y_i} \right ^1 \times 100$	$MAPE = \frac{1}{n} \sum_{i=1}^n \left \frac{\hat{y}_i - y_i}{y_i} \right ^1 \times 100$
Fonction du Cout	$L(\beta) = \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2$	Pas de fonction globale, mais optimisation locale des distances
Explication des parametres	<p>\hat{y} = valeur predite (NED, NS et AAR)</p> <p>$\beta_0, \beta_1, \dots, \beta_p$ = Coeff du modèle ajustés avec l'algorithme de Gradient Stochastique</p> <p>x_1, \dots, x_p = Variables Explicatives</p>	$w = e^{\left(\frac{-1}{2} \cdot d(x, x_n) \right)}$ $d(x, x_n) = \sqrt{(x_1 - x_{n1})^2 + (x_2 - x_{n2})^2 + \dots + (x_p - x_{np})^2}$ x_1, \dots, x_p = La nouvelle donnée à classer (x_{Test}) x_{n1}, \dots, x_{np} = Les données d'entraînement (x_{Train}) w = Poids ponderé et $d(x, x_n)$ = Distance Euclidienne
Etape de prédiction	<p>Equation du modèle :</p> $\hat{y} = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$ <p>Comment minimiser $\xi = \hat{y} - y$?</p> <ol style="list-style-type: none"> Initialiser : β_0, β_1, \dots et β_p Calculer la dérivée : $\frac{\partial L(\beta)}{\partial \beta_j} = \frac{\partial}{\partial \beta_j} \frac{1}{n} \sum_{i=1}^n (\hat{y}_i - y_i)^2 = \frac{2}{n} \sum_{i=1}^n (\hat{y}_i - y_i) x_{ij}$ <ol style="list-style-type: none"> Mettre à jour β_j avec la DG : $\beta_j = \beta_j - \alpha \frac{\partial L(\beta)}{\partial \beta_j}$ Répéter jusqu'à convergence (lorsque les β ne changent plus significativement) et comme ça on obtiendra les β_j de l'équation \hat{y} 	<p>Equation du modèle : $\hat{y} = \frac{\sum_{i=1}^k w_i \cdot y_i}{\sum_{i=1}^k w_i}$</p> <p>Comment minimiser $\xi = \hat{y} - y$?</p> <ol style="list-style-type: none"> Calculer la distance entre x_{Test} et x_{Train} $d=?$ Augmenter les poids w des k plus proches points de x_{Test} car si d est petite, w est grande et si d est grande, w est petite. Comme ça on pondère les poids pour obtenir \hat{y} et on trouve MAPE

Figure 1 : Comparaison des deux modèles KNN et MLR

	Rapport de stage	Page 17/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

2.3.1 Gradient Boost (GB) et XGBoost (XGB)

Les deux modèles GradientBoost (GB) et XGBoost (XGB) appartiennent de la même familles « modèles ensembliste » qui admettent une équation du problème et une fonction du cout à optimiser. Ces deux modèles sont séquentiels (pas parallèle) c'est-à-dire a chaque fois il corrige l'erreur de la prédiction précédente mais chacun a un algorithme et démarches différentes de l'autre :

Le Gradient Boost commence par un modèle de base (un modèle pas compliqué comme arbre de décision) et à chaque étape, on ajuste un nouvel arbre pour corriger les erreurs commises par le modèle précédent c'est-à-dire on combine plusieurs modèles faibles pour obtenir un modèle final puissant et précis :

- Initialisation : On commence avec une prédiction simple (si variable cible quantitative → on prédit par la moyenne de cette variable, si elle est qualitative → on prédit par la classe majoritaire) alors on a la première fonction du cout selon la formule [3]

$$F_0(x) = \text{Arg Min}_C \sum L(y_i, c) \quad [3]$$

- Calcule des pseudo-résidus : à chaque itération m, on calcule les résidus en utilisant le modèle descente de gradient décrit dans l'équation [4] qui nous donne la direction dans lesquelles le Modèle doit s'améliorer.

$$r_i^{(m)} (\text{grad}) = -\frac{\partial [L(y_i, F_{m-1}(x_i))]}{\partial f(x_i)} \quad [4]$$

- Ajustement d'un nouvel arbre : On entraîne un petit arbre $h_m(x)$ pour prédire ces résidus.
- Calcul d'un coefficient optimal : On trouve le meilleur coefficient d'ajustement γ_m multiplicateur (longueur des pas à faire) qui minimise la perte [5]

$$\gamma_m = \text{Arg Min}_\gamma \sum L(y_i, F_{m-1}(x_i)) + \gamma \cdot h_m(x_i) \quad [5]$$

- Mettre à jour le modèle : en ajoutant ce nouvel arbre pondéré par un taux d'apprentissage n (contrôle la vitesse d'apprentissage) selon la formule suivante [6]

$$F_m(x) = F_{m-1}(x) + \eta \gamma_m \cdot h_m(x) \quad [6]$$

- Répétition : On répète les étapes jusqu'au nombre d'itérations fixé ou jusqu'à ce que l'erreur ne diminue plus.

	Rapport de stage	Page 18/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

Le modèle XGBoost est une version optimisée et améliorée du gradient Boosting développée pour être plus rapide, plus performante et mieux régularisée. Il garde le même principe de base mais introduit des améliorations au niveau de la fonction objective, de la régularisation et de l'optimisation computationnelle :

- Initialisation : c'est comme dans le GB mais ici on ajoute une fonction de régularisation de la complexité des arbres. Selon la formule [7]

$$L = \sum_{i=1}^n l(y_i; \hat{y}_i) + \sum_{k=1}^K \Omega(f_k) \quad \text{Avec } \Omega(f) = \gamma \cdot T + \frac{1}{2} \lambda \|w\|^2 \quad [7]$$

Où : $l(y_i; \hat{y}_i)$ = la fonction de perte qui mesure l'écart entre la valeur réelle y_i et la prédiction \hat{y}_i et $\Omega(f_k)$ = le terme de régularisation qui contrôle la complexité des arbres, T = nombre de feuilles dans l'arbre, w = poids des feuilles et γ, λ = paramètres de pénalisation ce qui permet d'éviter les arbres trop complexes et donc le sur-apprentissage.

- Calcule des pseudo-résidus : à chaque itération m , on calcule les résidus en utilisant la descente de gradient (dérivé premier de la perte [8]) et Hessien (dérivé second de la perte [9]) ce qui permet une mise à jour plus précise.

$$g_i = \frac{\partial l(y_i; \hat{y}_i)}{\partial \hat{y}_i} \quad (\text{gradient, 1^{re} dérivée}) \quad [8]$$

$$h_i = \frac{\partial^2 l(y_i; \hat{y}_i)}{\partial \hat{y}_i^2} \quad (\text{Hessien, 2^e dérivée}) \quad [9]$$

- Construction des arbres avec régularisation : pour contrôler la taille des arbres selon la formule suivante [10]

$$L = \sum_{i=1}^n \left[g_i f_t(x_i) + \frac{1}{2} h_i f_t(x_i)^2 \right] + \Omega(f_t) \quad [10]$$

- Poids optimal de chaque feuille : pour chaque feuille j , la valeur optimale w_j^* qui est le score attribué aux échantillons de la feuille est donnée par la formule [11]

$$w_j^* = -\frac{\sum g_i}{\sum f_i + \lambda} \quad [11]$$

	Rapport de stage	Page 19/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

- Choisir le meilleur arbre en maximisant le gain : Quand on veut découper une feuille en deux sous-feuilles gauche (L) et droite (R), XGBoost calcule le **gain** de ce split, c'est-à-dire l'amélioration apportée à l'objectif selon la formule [12]

$$Gain = \frac{1}{2} \left[\frac{G_L^2}{H_L + \lambda} + \frac{G_R^2}{H_R + \lambda} - \frac{(G_R + G_L)^2}{H_R + H_L + \lambda} \right] - \gamma$$

[12]

Alors, après qu'on a vu les étapes de ces quatre modèles (MLK, KNN, GB et XGB), nous allons utiliser dans ce travail seulement les modèles non-linéaire (KNN, GB et XGB) et pour entrer en détails nous allons exploiter les articles sur lesquels on a basé cette étude.

3 BASE DE DONNEES EXPERIMENTALES

3.1 Première base de données : « FoRDY »

Avant d'entrer en détail de ce travail, il faut savoir l'existence d'une base de données FoRDY, on a téléchargé la base de données en ligne du platform DEED ([FoRDY – datacenterhub](#)) : **Foundation Rocking Dynamic** qui contient des essais dynamiques réalisés avec des centrifugeuses. De même le folder zip contient plusieurs folder pour expliquer et détailler les expériences réalisé mais, malheureusement la base de donnée FoRDY ne contient pas toutes les variables d'intérêt et en plus, elle contient beaucoup des valeurs manquantes ce qui a abouti à éliminer quelque série de tests alors, j'ai suivi les mêmes démarches que la base de donnée foRCY mais ici la différent qu'après prétraitement et post-traitement j'ai obtenu une base de donnée formées de 4 séries de tests : **MAH01**, **PWRI_S**, **S2011SQF1** et **UCSD** telle que chacune contiennent les variables d'intérêt suivantes : **PGD**, **PGV**, **PGA** pour la structure et pour le sol en plus une variable **Ia** et **Cr** (8 variables d'entrées) et j'ai calculer les outputs NED, NS et AAR. En revanche, après le nettoyage et filtrage les essais cette base de données FoRDY ont devenu **100 expériences** seulement alors j'ai décidé d'ajoute les 5 expériences réalisée à CEA Paris-Saclay. Voici un tableau qui montre toutes les expériences de la base de données FoRDY en indiquant pour chaque série de test les objectifs principaux avec leurs références.

		Rapport de stage	Page 20/66
		Réf. : Aix-Marseille Université	
		Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

Tableau 3 : Objectifs principaux du « FoRDY »

Labo / Institution	Type d'essai	Objectifs principaux	Références
CEA – Tamaris (France)	Grande échelle 1g	Étudier le rocking et renversement de structures sur semelles superficielles rigides	(Sollogoub, Combescure, Queval, & Chaudat, s.d.) (Millen & Hare, 2025)
UC Davis – CGM (USA)	Centrifuge	Capacité portante, tassement, dissipation d'énergie de semelles soumises à rocking	(Gajan & Kutter, 2008)
PWRI (Japon)	Table vibrante laminaire 1g	Comportement non linéaire de piles de ponts sur semelles superficielles	(Shirato et al., 2008)
NTUA (Grèce)	Centrifuge	Réponse dynamique de piles de ponts sur sable	(Drosos et al., 2012)
UC Davis – CGM (USA)	Centrifuge	Modélisation sismique de fondations basées sur rocking	(Deng & Kutter, 2012)
UC Davis – CGM (USA)	Centrifuge	Chargement combiné sur semelles superficielles	(Hakhamaneshi et al., 2020)
NTUA (Grèce)	Petite échelle 1g	Comportement au rocking de piles élancées sur sable	(Tsatsis & Anastopoulos, 2015)
LHPOST – UC San Diego (USA)	Table vibrante	Tests sur piles de ponts supportées par semelles basculantes	(Antonellis et al., 2015)
UC Davis – CGM (USA)	Centrifuge	Semelles sur argile molle renforcée par grilles de ciment	(Khosravi et al., 2017)
MFST – Chine	Table vibrante laminaire 1g	Comportement sismique de structures nucléaires sur sols argileux	(Zhao, Chen, & Zhou, 2023)

3.2 Deuxième Base de données « CAMUS IV »

Le programme CAMUS : Comportement et Analyse des Murs en Sollicitation Sismique a été mené au CEA (Cosmiseraï d'Energie Atomique) – laboratoire TAMARIS à Paris Saclay en France, sur la grande table vibrante Azalée (Sollogoub et al., s.d.). Les essais de CAMUS IV ont été réalisé en 1999, sur un mur porteur en béton armé à l'échelle 1\3 reposant sur une fondation superficielle. Cette structure a été soumise à des séismes synthétiques produits sur la table vibrante, avec intensités croissantes. L'objectif principal était d'étudier les phénomènes d'SFSI en conditions réalistes, en incluant les mouvements (basculement, soulèvement, tassement et rotation). Plusieurs chercheurs ont utilisés de nouveau cette base de donnée comme cas de validation (Pecker et al., 2014) (Millen & Hare, 2025).

	Rapport de stage	Page 21/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
	RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES	

Alors vu que la deuxième base de données FoRDY contient que 100 essais et la troisième CAMUS IV contient 9 essais avec 8 variables d'entrées (PGD_S, PGV_S, PGA_S, PGD_B, PGV_B, PGA_B, Ia et Cr) et 3 variables de sorties (NED, NS et AAR). On a décidé de combiner ces deux bases de données pour étudier la performance du système sol-structure selon l'impact des variables calculées dans le post traitement et comme ça on peut classer par ordre croissant l'importance des variables (du plus significative au moins significative) sur le système sol-structure.

3.3 Exploitation et Méthodologie

Tout d'abord, on a combiné les deux bases de données FoRDY et CAMUS IV même si que le nombre d'expérience totale de ces deux bases de données 109 essais n'est pas suffisant pour entraîner des modèles mais on a décidé de continuer ce travail car c'est la seule base de données qui contient des valeurs réelles résultant des expériences. Ensuite, on a suivi les démarches suivantes :

1. Vérifier une deuxième fois après combinaison des deux bases de données filtrées que la base de données finale ne contient pas des valeurs aberrantes ou des valeurs manquantes (n/a, n/r, n/m). Mais malheureusement on a détecté la présence des valeurs aberrantes ce qui nous a donné une base de données filtrée contenant 88 essais.
2. Explorer et Analyser la base de données final et c'est ici où on a appliquée des transformations logarithmiques car les distributions n'étaient pas symétriques Voir Tableau 4 : Tableau statistiques des variables de "FoRDY et CAMUS". Après on a trouvé la matrice de corrélation de ces variables d'intérêt et on a conclu qu'il existe deux variables (PGV_S et PGA_S) à éliminer car ils sont fortement corrélés avec les autres (voir Figure 1).

Tableau 4 : Tableau statistiques des variables de "FoRDY et CAMUS"

	mean	Std	Min	Max	Q1	Médian	Q3	IQR	IQR%	Skewness	Kurtosis
Cr	0.374	0.201	0.103	0.982	0.18	0.347	0.565	0.39	112.392	0.357	-0.88
Ia	12.78	19.522	0.035	70.722	1.14	4.707	13.386	12.246	260.176	1.855	2.047
AAR	1.366	0.356	0.323	2.505	1	1.46	1.561	0.558	38.243	0.022	0.683
NED	74.32	138.04	-0	781.55	0	0.001	119.15	119.15	10501085.2	2.798	9.436
NS	0.018	0.029	0	0.178	0	0.008	0.017	0.013	160.885	3.195	11.234
PGD_B	0.549	1.267	0.021	7.744	0.11	0.169	0.343	0.233	137.87	4.215	18.853
PGV_B	3.562	5.609	0	23.91	0.14	1.27	5	4.859	382.598	2.012	3.194
PGA_B	21.85	24.797	0.059	104.6	4.81	11.1	30	25.195	226.982	1.704	2.623
PGD_S	1.195	5.041	0.053	43.335	0.11	0.139	0.24	0.13	93.525	6.822	50.037
PGV_S	4.022	6.207	0.028	32.38	0.22	1.57	5.3	5.082	323.662	2.464	7.008
PGA_S	26.43	25.516	0.137	102.6	7.13	17.1	46	38.875	227.339	1.263	1.032

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

												Rapport de stage	Page 22/66
												Réf. : Aix-Marseille Université	
												Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES													

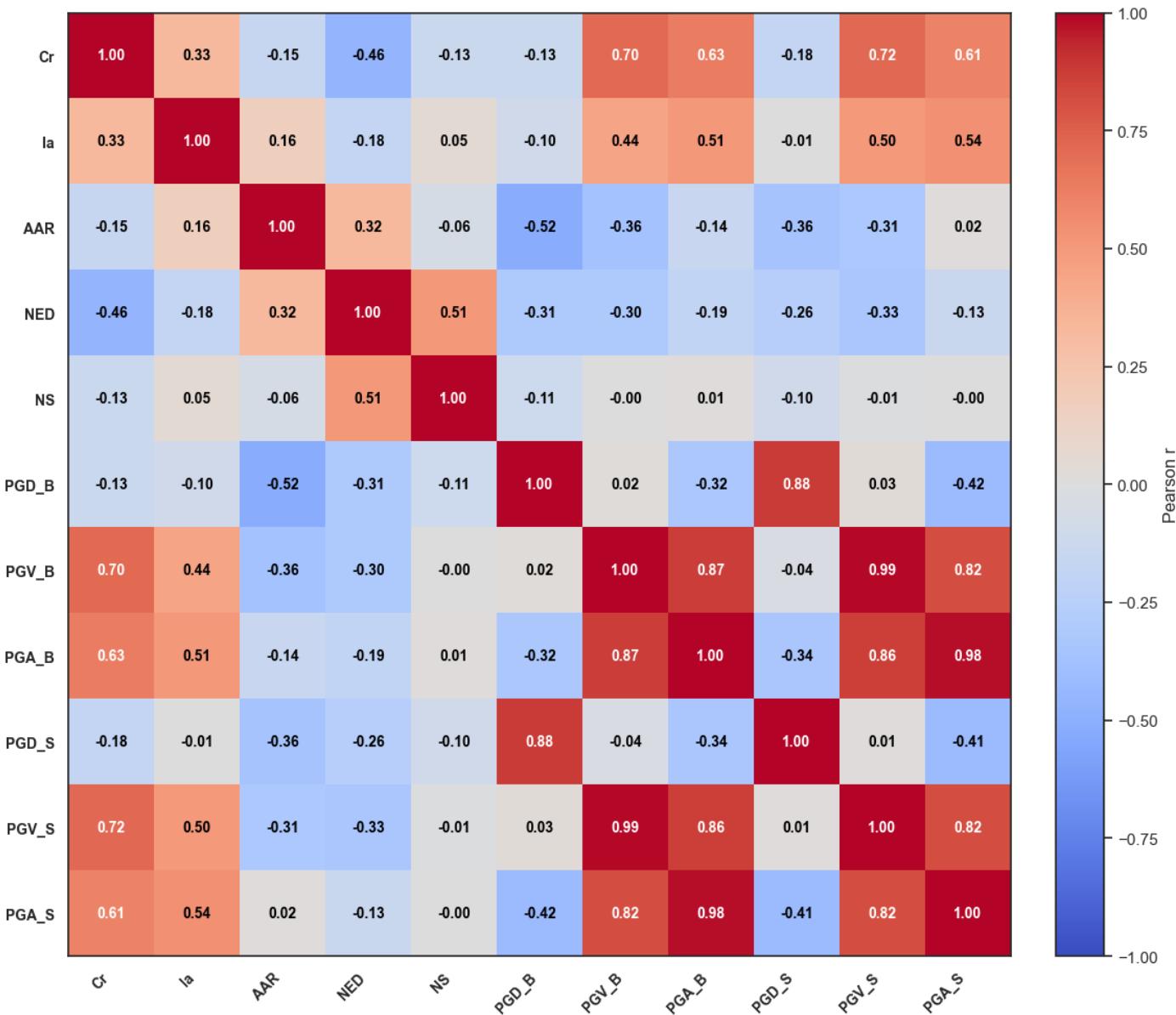


Figure 2 : Matrice de corrélation des variables de "FoRDY et CAMUS

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

		Rapport de stage	Page 23/66
	Réf. : Aix-Marseille Université		
	Date : 22/07/2025	Indice :	
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

3. Application des modèles : Ici on a choisi des modèles Non-linéaire KNN, GradientBoost, XGBoost et SVM en se basant sur les résultats de prédiction de Gajan ou il a utilisé un modèle linéaire (MLR) et un autre non linéaire (KNN) (Gajan, 2021). Tout d'abord, on a partitionné notre base de données filtrée qui contient 88 essais avec 6 variables d'entrées et 3 variables de sorties. Après on a entraîner les modèles et pour chacun on a ajusté les paramètres : (Voir tableau 5)

Tableau 5 : Ajustement des paramètres "FoRDY et CAMUS"

Variable	Modèle	Best Params	RMSE	R ² / Accuracy
NED	KNN	n_neighbors=11	2.357	0.233
	GradientBoost	lr=0.05, max_depth=2, n_estimators=100	2.654	0.233
	XGBoost	lr=0.1, max_depth=2, n_estimators=100	2.809	0.233
NS	KNN	n_neighbors=20	2.628	-0.099
	GradientBoost	lr=0.01, max_depth=2, n_estimators=200	3.043	-0.099
	XGBoost	lr=0.1, max_depth=1, n_estimators=100	3.258	-0.099
AAR	KNN	n_neighbors=3	0.11	0.536
	GradientBoost	lr=0.1, max_depth=3, n_estimators=300	0.059	0.536
	XGBoost	lr=0.1, max_depth=7, n_estimators=50	0.043	0.536

	Rapport de stage	Page 24/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

4. Comparaison des mesures métriques : Avant d'interpréter les résultats ce qu'il faut savoir que plus le RMSE est élevé plus l'erreur de prédiction est grande, et plus R^2 est négatif plus le modèle fait pire que la prédiction par la moyenne de la variable cible. En se basant sur les résultats obtenus on remarque que la prédiction de la variable AAR admet les meilleures valeurs de RMSE et R^2 avec pour les 3 modèles. En plus, pour la variable NED la meilleure valeur de RMSE c'était avec le modèle XGBoost suivi par GB et KNN mais ces modèles n'ont pas trop compris la relation entre ces variables car R^2 ne sont pas élevés. Alors que la prédiction de la variable NS n'a aucune structure claire apprise de ces trois modèles elle n'est pas directement comparable avec les autres (voir Figure 3).

Donc, la prédiction de AAR c'est meilleur en utilisant le modèle XGB et celle de la variable NED est meilleur en utilisant KNN alors que celle de la variable NS n'était pas efficace et cela peut être du à la petite taille de la base de données ou bien au choix des modèles entraînés.

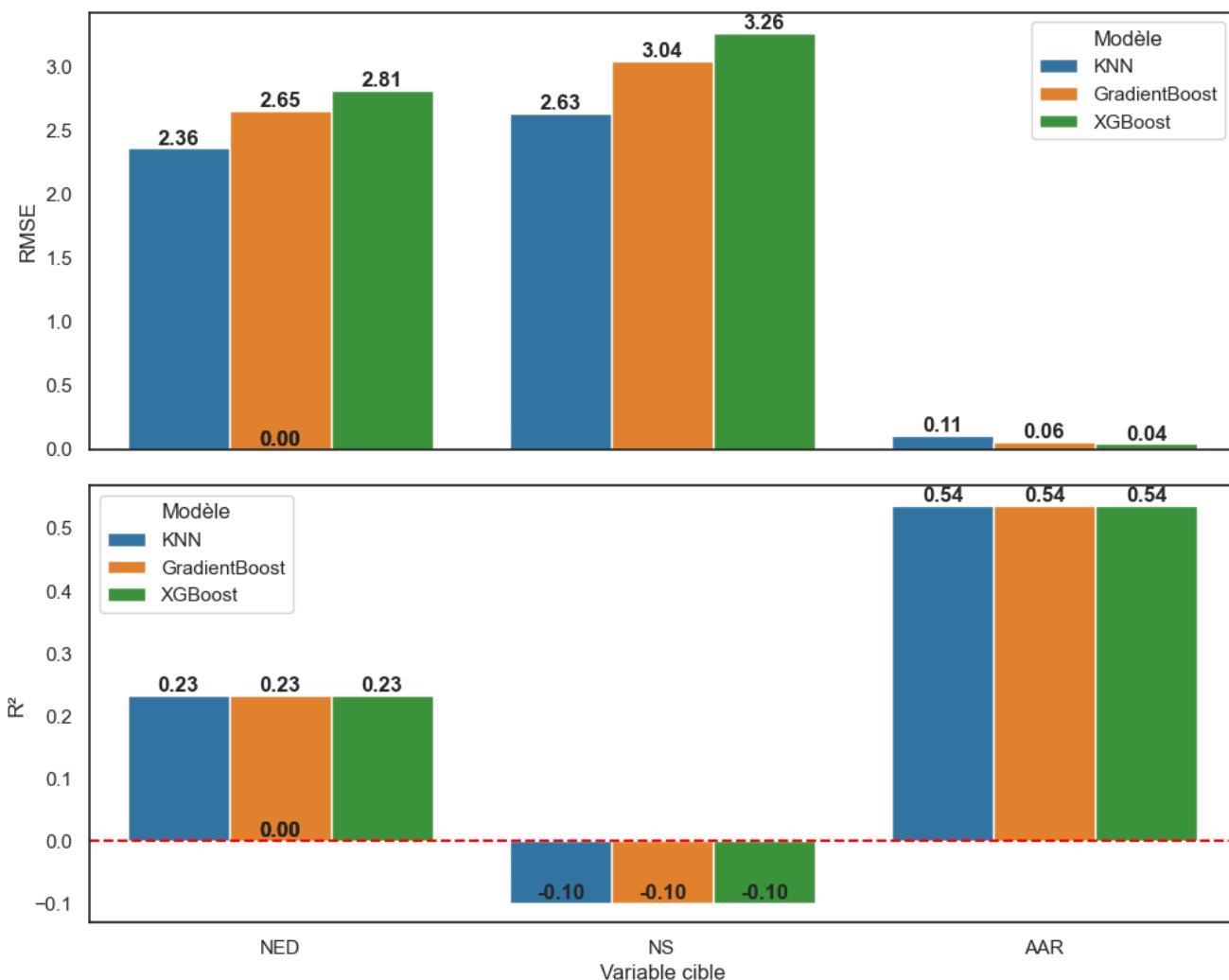


Figure 3 : Evaluation de la performance des variables du "FoRDY et CAMUS"

	Rapport de stage	Page 25/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

5. Classification de l'impact des variables sur la performance du système sol-structure : Après application des ML on peut classer l'efficacités des variables inputs sur celles des outputs :

- NED : la variable Cr est dominant pour tous les modèles, PGD, PGV suivent de près surtout avec GB/XGB alors que la et PGA_B jouent un rôle secondaire. Donc, NED est surtout contrôlé par Cr et les indicateurs de déplacement / vitesses de la Base.
- NS : Il existe une grande variabilité des importances selon les modèles. la et Cr ressortent souvent comme importants pour (GB et XGB) alors que pour KNN on a un poids élevé à Cr uniquement. Les autres variables (PGD, PGV, PGA) ont un impact faible et négatif. Donc, même si Cr et la semblent liés à NS, la relation reste instable et faible ce qui explique les mauvais R^2 (voir Figure 3).
- AAR : la variable PGD_B est toujours la plus importante pour tous les modèles suivis de PGV_B et PGD_S alors que Cr a quasiment zéro importance. Donc, AAR dépend fortement des mesures de déplacement/vitesse et beaucoup moins des autres indicateurs (voir Figure 3).

Conclusion : NED a comme meilleur modèle KNN et comme variable clé c'est **Cr** ce qui est logique car NED est lié à l'efficacité énergétique et donc au critical ratio. NS aucun modèle nous donne des résultats convaincants, les variables sont trop bruitées ou dépendante d'autre facteurs absents de la base mais c qu'on peut constater que la et Cr sont les plus liés à NS. Alors que, AAR a PGD_B et PGV_B comme variables dominantes pour tous les modèles ce qui est logique car elle dépend des accélération / déplacements. Ces résultats confirment les mécanismes physiques attendus et orientent les futures analyses vers un enrichissement des variables explicatives pour mieux capturer NS (voir Figure 4).

6. Classification de l'impact des variables sur la performance du système sol-structure : Après application des ML on peut classer l'efficacités des variables inputs sur celles des outputs :

- NED : la variable Cr est dominant pour tous les modèles, PGD, PGV suivent de près surtout avec GB/XGB alors que la et PGA_B jouent un rôle secondaire. Donc, NED est surtout contrôlé par Cr et les indicateurs de déplacement / vitesses de la Base.
- NS : Il existe une grande variabilité des importances selon les modèles. la et Cr ressortent souvent comme importants pour (GB et XGB) alors que pour KNN on a un poids élevé à Cr uniquement. Les autres variables (PGD, PGV, PGA) ont un impact faible et négatif. Donc, même si Cr et la semblent liés à NS, la relation reste instable et faible ce qui explique les mauvais R^2 (voir Figure 3).

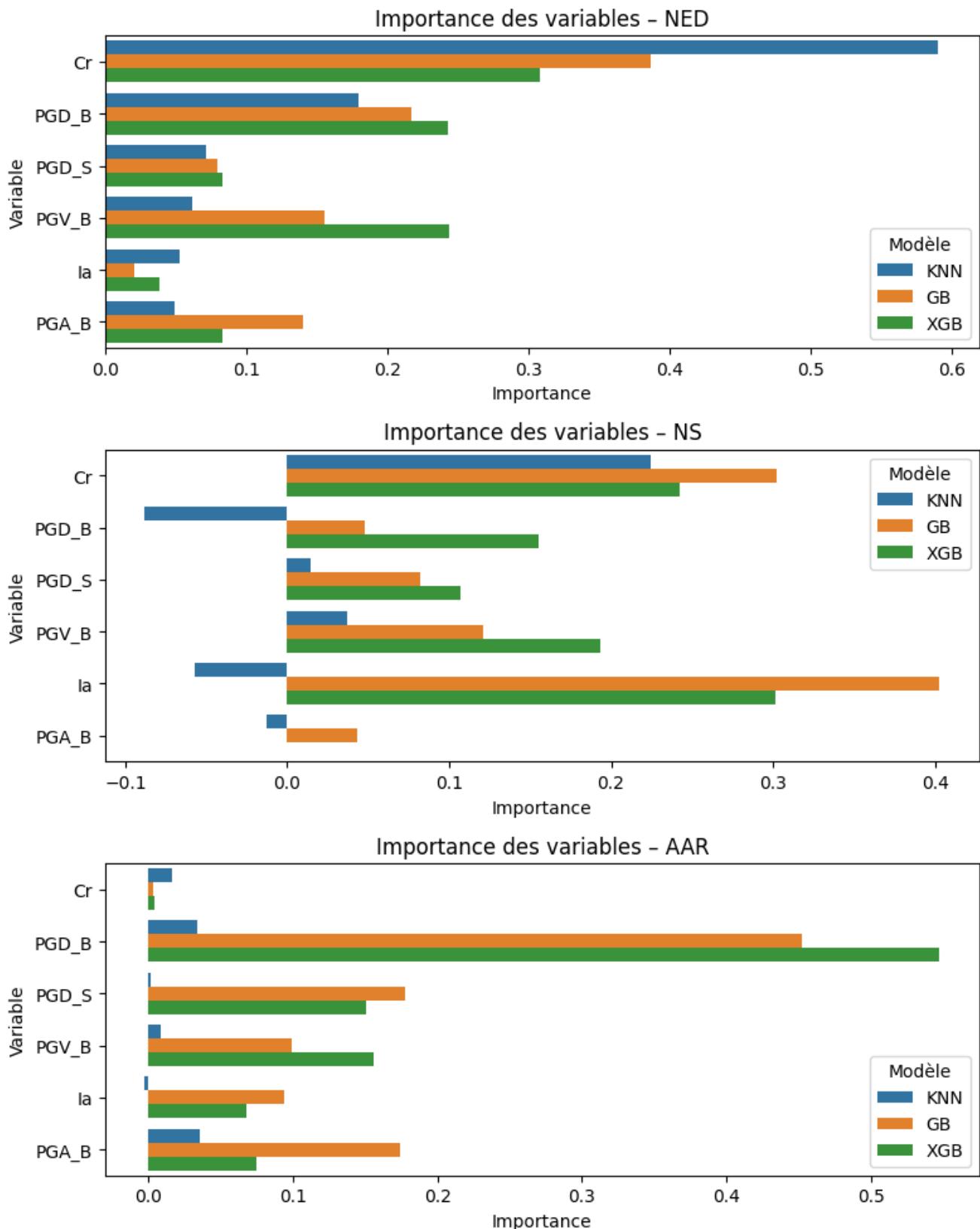

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES


Figure 4 : Diagramme en barres d'importance des variables du "FoRDY et CAMUS"

		Rapport de stage	Page 27/66
		Réf. : Aix-Marseille Université	
		Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

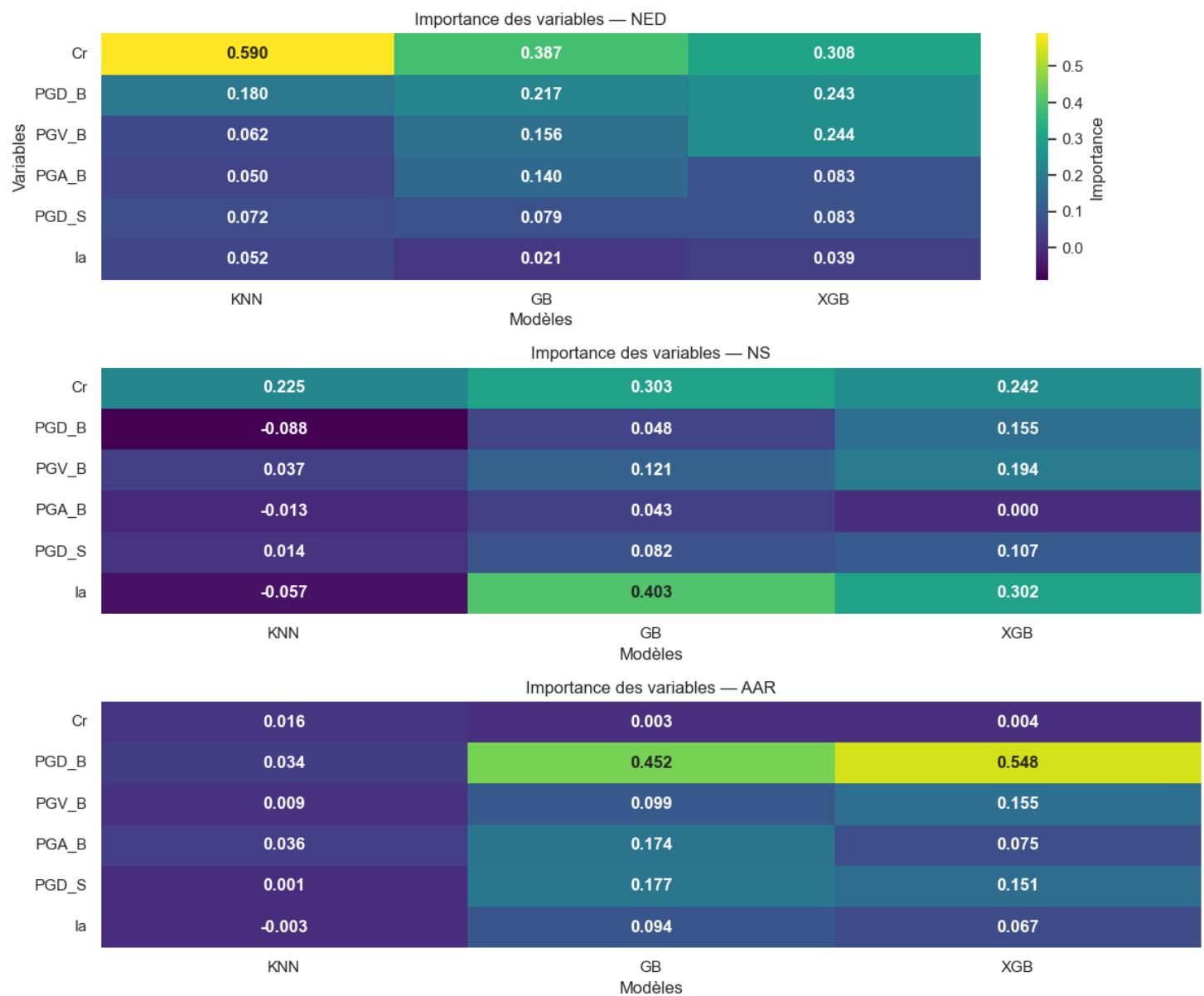


Figure 5 : Matrice d'importance des variables du "FoRDY et CAMUS"

3.4 Conclusion

Donc, Après beaucoup de travail avec les 3 bases de données on n'a pas obtenu des résultats qui répond à nos objectifs, c'est pour cela une nouvelle base de données simulé a été développé à partir d'hypothèses contrôlées dans le cadre de ce travail pour explorer le comportement du système selon un grand nombre d'essais et une grande variété de paramètres.

	Rapport de stage	Page 28/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

4 BASE DE DONNEES NUMERIQUE

4.1 Introduction

La première étape de ce travail c'est de créer une base de données qui présente le SFSI soumis à des charges sismiques. Cette création de la base de données est nécessaire car les essais expérimentaux sont souvent limités, coûteux à obtenir et ne couvre pas une variété suffisante de configurations. Alors que la simulation numérique permet de générer un grand nombre d'essais en modifiant les paramètres du sol, de la fondation, de la structure et du chargement.

Pour créer cette base de données numérique, deux environnements ont été utilisés de manière complémentaire : le premier c'est **OpenSees** via des scripts **TCL** qui nous a permis de modéliser le système et de lancer les calculs dynamiques. Ensuite le deuxième c'est python en utilisant des bibliothèques spécialisées pour générer des paramètres en précisant les bornes, organiser les scénarios de calcul sur chaque nœud du SFSI ensuite trouver les variables qui nous intéressent en appliquant des ML sur les fichiers de résultats obtenus après post traitement.

Autrement-dit, la démarche adoptée suit une chaîne de traitement claire :

1. Génération des paramètres par python.
2. Simulation numérique des réponses sur SFSI dans OpenSees.
3. Sauvegarde et post-traitements des résultats.
4. Application des modèles d'apprentissage automatique sur la base finale pour prédire la performance et étudier la significativité des variables du système sol-structure.

4.2 Modélisation numérique

Le modèle a été implémenté dans un script principal appelé « **main_template.tcl** » qui s'appuie sur deux autres sous-fichiers « **FScale.tcl** » pour calculer le facteur de sécurité et « **GazetasStiff_1.0.tcl** » pour trouver les raideurs équivalentes du sol. Dans ce travail, le modèle s'appuie sur l'approche **BNWF (Beam-on-Nonlinear Winkler Foundation)** pour présenter l'interaction du système sol-structure. Ce modèle considère la fondation comme une poutre élastique discrétisée en éléments poutre-colonne (beam-column) reliés au sol par des ressorts de type zéro-longueur tel que chacun représente un mode particulier de comportement du sol :

- QzSimple1 : comportement vertical charge-déplacement,
- PySimple1 : comportement horizontal passif sur les flancs de la fondation,
- TzSimple1 : comportement de cisaillement horizontal à la base.


RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES

L'ensemble permet de reproduire de façon réaliste les principaux phénomènes observés : **rocking**, **glissements**, **tassements** permanents, dissipation d'énergie par hystérésis et amortissement radiatif. Par ailleurs, la répartition variable des ressorts le long de la semelle assure une bonne représentation de la rigidité en rotation.

Ce modèle a été initialement calibré sur des essais de fondations superficielles (Raychowdhury & Hutchinson, 2009) et adapté dans OpenSees pour des analyses bidimensionnelles (« Boulanger, 2000-BNWF - Modélisation du code », s.d.).

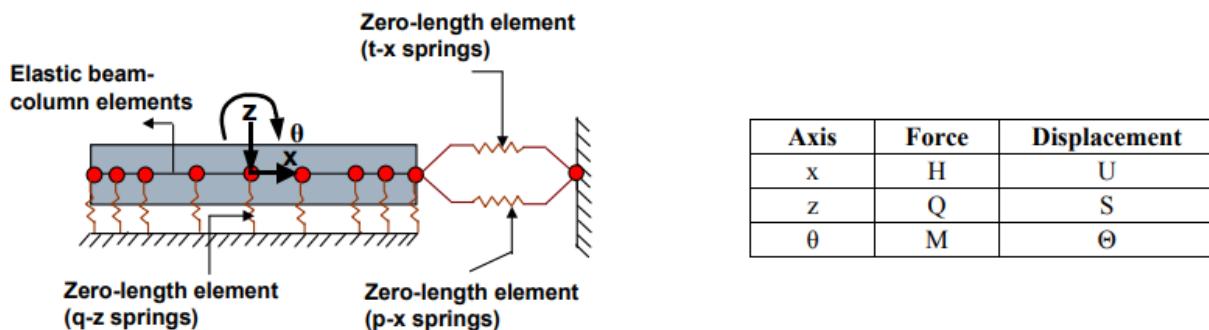


Figure 6 : Schéma du modèle BNWF (Raychowdhury & Ugalde, s.d.)

En complément du modèle BNWF, une autre approche fréquemment utilisée est le **Contact Interface Model** (CIM), proposé par Gajan et Kutter (Gajan & Kutter, 2008). Le CIM est formulé comme un macro-élément placé à l'interface semelle–sol, qui remplace la fondation rigide et la zone de sol influencée. Ce modèle reçoit en entrée des déplacements incrémentaux et fournit en sortie les charges correspondantes, permettant de représenter de manière compacte le comportement global de l'interface. Contrairement au BNWF, où les ressorts sont indépendants par direction, le CIM prend explicitement en compte le **couplage entre glissement, rotation et tassement** grâce à des surfaces de charge normalisées (M–H diagrammes). Cela permet de représenter les interactions complexes observées expérimentalement sur semelles superficielles soumises à chargements cycliques. (Raychowdhury & Ugalde, s.d.)

	Rapport de stage	Page 30/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES

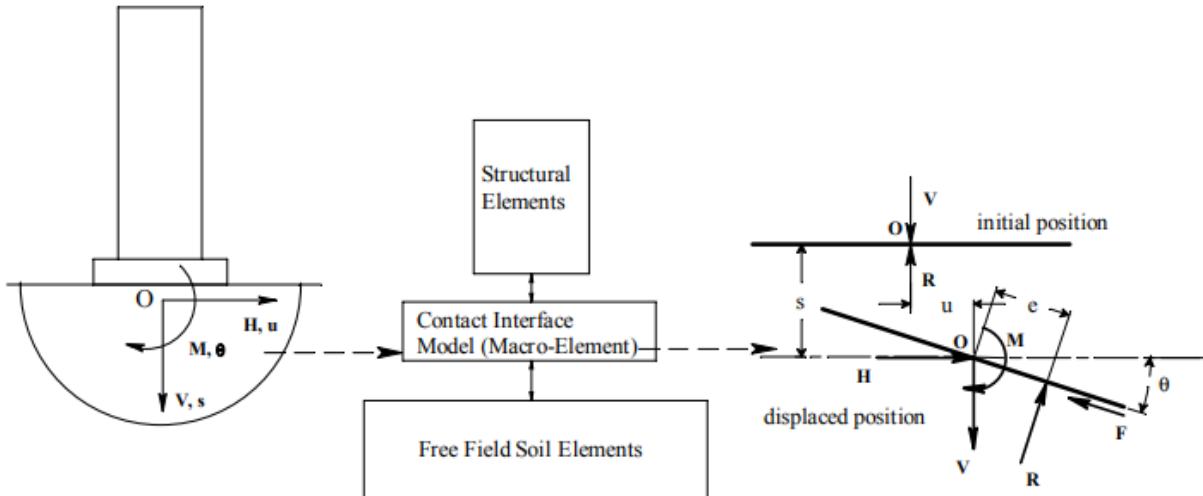


Figure 7 : Schéma du modèle CIM (Raychowdhury & Ugalde, s.d.)

Dans ce travail, le modèle BNWF a été retenu car il offre une description plus pédagogique et modulaire des différents modes de déformation (rocking, glissement, tassement), ce qui est essentiel pour générer une base de données riche et exploitable par la suite en Machine Learning. De plus, son implémentation native dans OpenSees et sa flexibilité pour des analyses paramétriques en font un choix pratique et robuste pour l'étude menée. (Gajan et al., 2010a)

4.2.1 Structure générale du modèle

Le système est modélisé en deux dimension 2D avec 3 degrés de liberté pour chaque nœud. C'est-à-dire sur chaque nœud on mesure la translation horizontale, verticale et la rotation. Ce système modélisé comprend :

- Fondation superficielle qui est représentée par un ressort équivalent (en utilisant Macro-Elément-Modeling) qui a un rôle de capter les raideurs verticales et horizontales du sol.
- Colonne qui est modélisée comme un élément poutre élastique en liant la fondation superficielle avec la superstructure. Cette colonne a des différentes longueurs et largeurs.
- Tablier qui est représenté par une masse concentrée et un moment d'inertie au centre de gravité.

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

	Rapport de stage	Page 31/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

Cette modélisation qui est formée de trois parties différentes permet de reproduire les principaux mécanismes d'interaction d'SFSI tout en limitant le temps de calcul (Raychowdhury & Ugalde, s.d.)

4.2.2 Nœud et éléments du modèle

Vu que le système est modélisé en 3 parties différentes, pour les lier ensemble on définit plusieurs nœuds superposés verticalement pour étudier le mouvement et l'interaction entre sol-fondation et entre fondation-structure. C'est pour cela on a construit 6 nœuds différents :

- Nœuds 1 et 2 : Pour le premier nœud c'est la base du système sol-fondation-structure qui est enfoncé alors que la deuxième c'est la base de la fondation enfoncée dans le sol telle que les deux nœuds 1 et 2 sont liés par un ressort (un élément zeroLengthSection concentrant la rigidité du massif de sol) représentant le comportement global du sol autrement-dit l'interaction sol-Fondation.
- Nœud 3 : c'est le centre de gravité de la fondation qui porte la masse et l'inertie de la semelle (fondation).
- Nœud 4 : c'est le sommet de la fondation qui nous permette d'étudier et d'analyser l'interaction entre la fondation et la structure (pied de la colonne).
- Nœud 5 : C'est la tête de la colonne reliée avec la fondation par des poutres élastiques (elasticBeamColumn) dans le but de modéliser la colonne et les liaisons géométriques.
- Nœud 6 : C'est le centre de gravité du tablier, où est appliquée la masse et l'inertie de la superstructure.

Après l'explication générale de la structure du modèle et le placement des nœuds tout ce qui nous reste c'est de donner des charges sismiques pour étudier le SFSI. Ce qu'il faut savoir que le nœud 1 est fixé pour présenter l'ancre à la base or que les autres nœuds sont libres, afin de permettre les déplacements dynamiques. Les charges de gravité sont appliquées comme forces verticales équivalentes au poids de la fondation et de la superstructure. En outre, le chargement sismique est introduit à travers d'une excitation appliquée au nœud 1 dans la direction horizontale en utilisant un accélérogramme réel lu depuis un fichier externe (GM_run.txt) qui varient selon le signal donné (accélération et fréquence). La figure ci-dessous représente le modèle créé dans l'article (Raychowdhury & Ugalde, s.d.) contenant 7 nœuds or que la modélisation de notre travail est la même mais avec 6 nœud en considérant nœud 5 et 6 comme un seul nœud.

	Rapport de stage	Page 32/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES

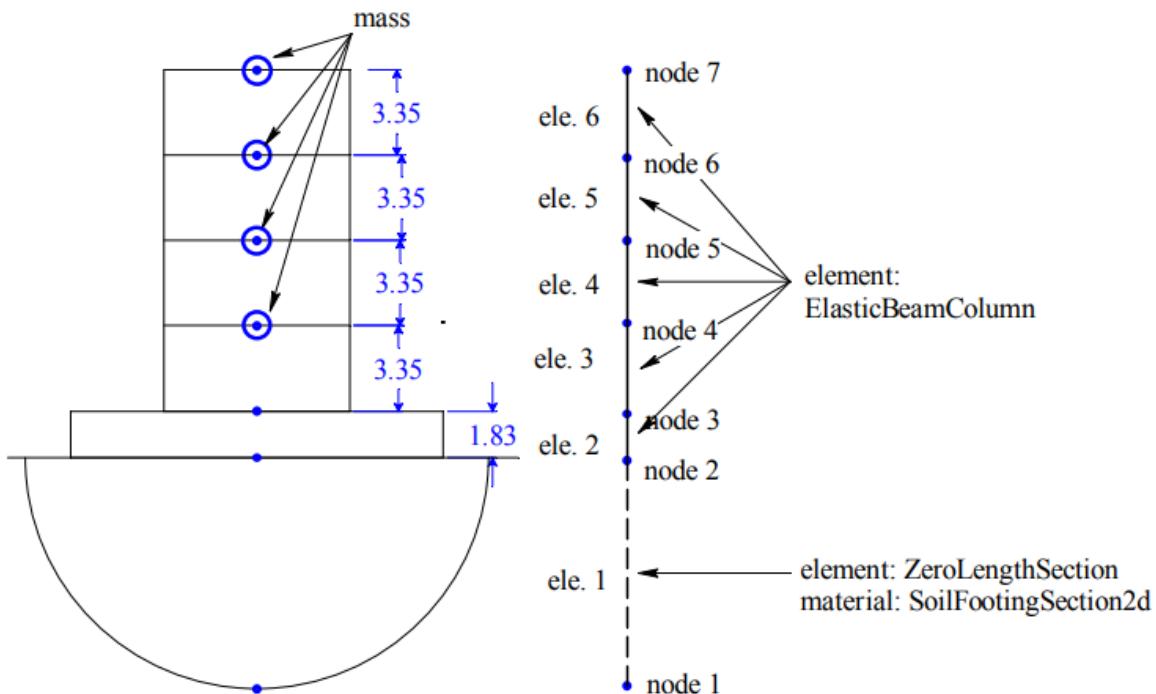


Figure 8 : Illustration du modèle (Raychowdhury & Ugalde, s.d.)

4.2.3 Sauvegarde des résultats numériques

Les résultats numériques obtenus sur chaque nœud sont différents de l'autre par exemple : Pour le nœud 1 et 2 on s'intéresse à mesurer les forces dans l'élément représentant le sol après, pour le nœud 4 et 6 on s'intéresse à étudier le déplacement de la fondation et de la structure suivant les axes x, y, z. en plus, pour le nœud 6 on s'intéresse à mesurer l'accélération après la propagation des ondes du sol en bas jusqu'en haut. Donc ces fichier « .out » constituent la matière brute pour le post-traitement et le calcul des variables d'intérêt.

	Rapport de stage	Page 33/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

4.3 Génération des paramètres

Dans le génie parasismique et mécanique il existe plusieurs hypothèses ou suppositions, alors c'est pareil pour générer des paramètres il faut préciser les bornes de chaque paramètre selon des critères phtisiques, mécanique et parasismiques. Dans ce travail on a 3 catégories de paramètres d'entrée au modèle à générer. Les choix des paramètres et les hypothèses mis selon l'articles.(« Geomechanics-Bornes des paramètres », s.d.)

4.3.1 Paramètres du sol

Dans cette étude les propriétés du sol ont été définis en fonction de trois bandes : Loose ; Medium dense et High dense. Pour chaque bande existe des intervalles de valeurs pour le module d'Young E, le coefficient de poisson nu et l'angle de frottement phi. Alors pour échantillonner efficacement ces intervalles, une méthode de Latin Hypercube Sampling (LHS) avec distribution gaussienne tronquée a été utilisée ce qui permet d'obtenir des échantillons variés mais statistiquement ces valeurs ne dépassent pas les bornes supposées. Maintenant, à partir de ces valeurs générés (E, nu et phi) des grandeurs dérivées sont calculées comme le Module de cisaillement G, la densité rho et le poids volumique gamma, la vitesse des ondes de cisaillement V_s et finalement l'indice normalisé N_{16} lié à la compacité. Pour mieux comprendre ces paramètres voici un tableau qui montre les bornes, l'unité, formule et symbole.

		Rapport de stage	Page 34/66
		Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :	
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

Tableau 6 : Paramètres du sol (« Geomechanics- Bornes des paramètres », s.d.)

Paramètre	Symbol	Unité	Borne	Formule / règle
Module d'Young	E	GPa	[1,39 - 11,11]	[1,39 - 4,17] Loose [4,17- 6,94] Medium Dense [6,94 - 11,11] High Dense
Rapport de Poisson	v	–	[0.20 - 0.40]	[0,20 - 0,36] Loose [0,25- 0,40] Medium Dense [0,30 - 0,40] High Dense
Angle de frottement interne	φ	°	[30 - 40]	[30 - 45] Loose [32 - 37] Medium Dense [35 - 40] High Dense
Densité apparente	ρ	t/m³	[1.76 - 1.80]	1.76 Loose 1.78 Medium Dense 1.80 High Dense
Poids volumique	γ	N/m³	Calculé	$\gamma = \rho \cdot g$
Module de cisaillement	G	Pa	Calculé	$G = E / 2(1+v)$
Vitesse onde cisaillement	V _s	m/s	Calculé	$\sqrt{G/\rho}$
Indice N ₁₆₀	N16	–	Calculé	(Emax/24975) ^0.75
Rebond vertical	R _v	–	0,15	Fixe
Amortissement ξ	ξ	–	0	Fixe
Cohésion non drainée	c	kPa	0	Fixe

Le nombre de possibilité de la bande de sol est 3 alors que celui de l'échantillonnage selon LHS est de 8 ce qui nous donne 3 x 8 = 24 possibilités d'essais pour les paramètres du sol. Nous ne montrant pas les histogrammes de gamma, rho, G_Pa, N16, c, Rv et xi car ces paramètres sont constants ou déduites des variables principales qui sont : E, nu et rho. Alors, voici des représentations graphiques pour mieux comprendre la distribution des paramètres principaux du sol.

La première figure ci-dessous montre la répartition des valeurs de phi selon les trois types de sol ce qui valide l'hypothèse des bornes choisis pour l'angle de frottement selon le type du sol (voir Figure 8). En plus, la deuxième figure qui montre les points colorés selon le type du sol occupent uniformément le domaine spécifié par les bornes de chaque classe et cela est dû grâce à échantillonner selon LHS Gaussien (voir Figure 9).

	Rapport de stage	Page 35/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES

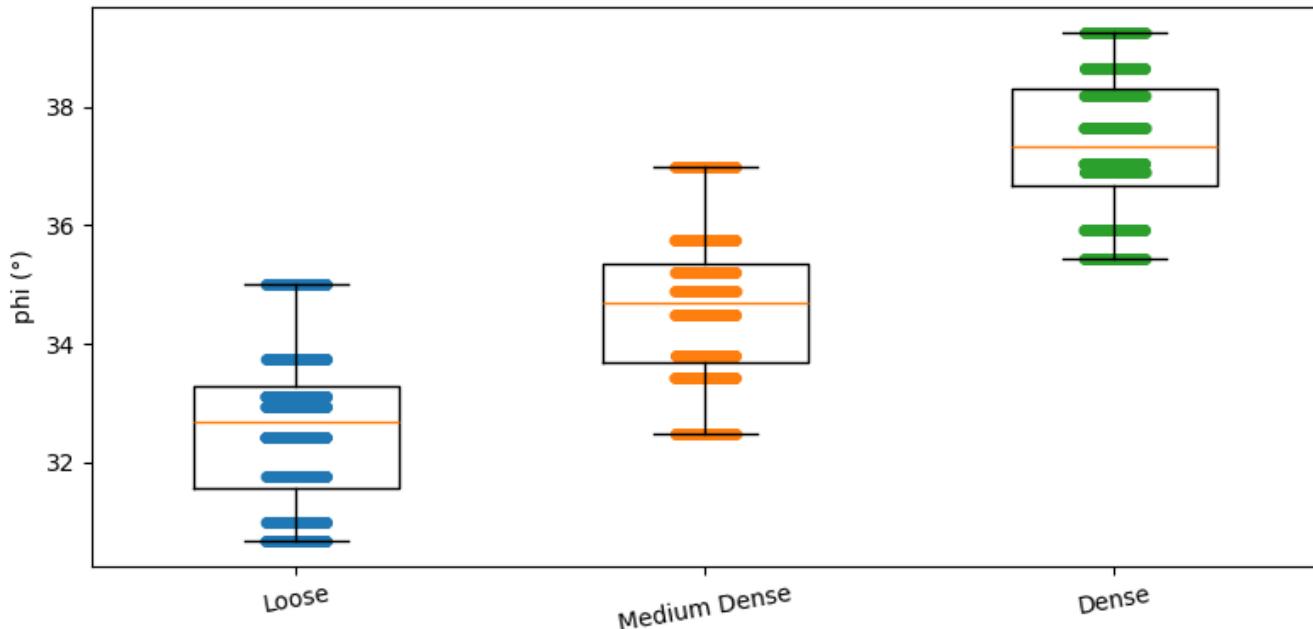


Figure 9 : Répartition des valeurs de ϕ selon le type du sol

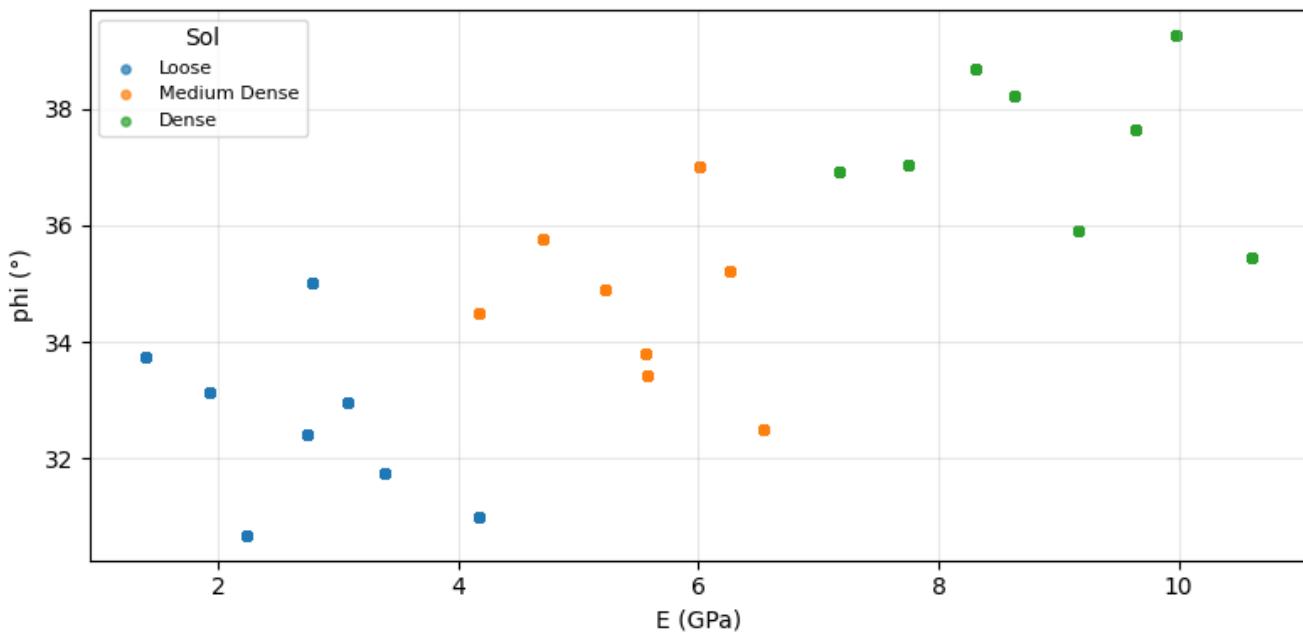


Figure 10 : Répartition des valeurs de ϕ selon module d'Young

		Rapport de stage	Page 36/66
		Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :	
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

4.3.2 Paramètres du chargement

Les paramètres de chargement jouent un rôle important sur la performance du système alors il faut définir les bornes scientifiquement avec des valeurs qu'on peut les appliquées ou les réalisés dans un laboratoire. Dans ce travail, le chargement sismique est défini par deux paramètres principales : la **fréquence** dominante du signal et l'**accélération** maximale. Pour chaque expérience on définit les paramètres du chargement différemment et cela grâce au fichier d'accélérogramme sur lequel on change le signal en variant la fréquence de 4 à 8.5 Hz avec un pas de 0.5 Hz et en variant aussi l'accélération de 0.50 à 0.85 g avec un pas de 0.05 g ce qui nous donne des différents signaux et différentes valeurs d'accélération donc différentes fichiers « **GM_runs** » pour explorer plusieurs possibilités de paramètres dynamique.

Tableau 7 : Paramètres du chargement

Paramètre	Symbol	Unité	Borne	Formule / règle
Fréquence	f	Hz	[4 - 8.5]	Avec un pas de 0.5 Hz
Accélération	PGA_in	g	[0.50 - 0.85]	Avec un pas de 0.05 g

Le nombre de possibilité de la fréquence est 9 alors que celui de l'accélération est 8 ce qui nous donne $9 \times 8 = 72$ possibilités d'essais pour les paramètres de chargement.

4.3.3 Paramètres de la structure

Les paramètres de la structure se distinguent selon les **dimensions** de la fondation, de la colonne et du tablier. Supposons qu'il existe dix typologies structurelles différents (S1 à S10) alors pour chaque typologie, les caractéristiques suivantes sont calculées : Masse et moment d'inertie du tablier, de la fondation et la position des centres de gravité ensuite propriétés de la colonne. Dans cette étude nous allons travailler comme si la position est toujours sur la surface du sol c'est-à-dire prenons seulement le cas où le centre de gravité de la structure n'est pas enfoncé dans le sol. Voici un tableau qui montre la dépendance des dimensions et les différentes valeurs de masse que la structure peut prendre.(« Geomechanics- Bornes des paramètres », s.d.)

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

		Rapport de stage	Page 37/66
		Réf. : Aix-Marseille Université	
		Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

Tableau 8 : Paramètres de la structure (« Geomechanics- Bornes des paramètres », s.d.)

Paramètre	Symbol	Unité	Borne	Formule / règle
Diamètre de la colonne circulaire	dCol	m	[0.5 - 1]	Avec un pas de 0.1
Hauteur de la colonne	Hcol	m	[8 - 11]	Avec un pas de 1
Longueur du tablier	deckL	m	[8 - 14]	Avec un pas de 2
Largeur du tablier	deckW	m	[10 - 14]	Avec un pas de 2
Épaisseur du tablier	deckT	m	{0.6 ; 0.8 ; 0.9 ; 1 ; 1.2}	Dépend de deck L et W
Longueur de la semelle	L	m	{2.7 ; 3.6 ; 5.4}	Standard
Largeur de la semelle	B	m	{2.7 ; 3.6 ; 5.4}	Standard
Hauteur de la semelle	Hfoot	m	{1 ; 1.5 ; 2}	Standard
Aire de la colonne	A	m^2	Calculé	$A = \pi d^2/4$
Inertie de la colonne	I	m^4	Calculé	$I = \pi d^4/64$
Masse du tablier	Mdeck	kg	{120 ; 180 ; 240 ; 315 ; 378 ; 420 ; 588}	Standard
Masse de la semelle	Mfoot	Kg	{18225 ; 48600 ; 145800}	Standard
Moment d'inertie du tablier	Ideck	$Kg \cdot m^2$	Calculé	$Mdeck \times (deckW/2)^2$
Moment d'inertie de la semelle	Ifoot	$Kg \cdot m^2$	Calculé	$Mfoot \times (B/2)^2$
Position du cdg de la semelle	Hfootcg	m	Calculé	Hfoot / 2
Position du cdg du tablier	Hdeckcg	m	Calculé	Hfoot + Hcol + deckT / 2
Profondeur d'encastrement de la semelle	Df	m	0	Fixée de niveaux 0

Le nombre de combinaison de la structure est 10 possibilités avec une seule position d'encastrement même si on a différentes valeurs de longueur, largeur et hauteur mais ici ces dimensions sont dépendantes. La plupart des paramètres de la structure sont discrète mais les primaires sont : dCol, Hcol, deckL, deckW, deckT, Hfoot, B, et L alors qu'Es et Df sont des constantes. On a l'aire A et l'inertie I sont déjà 100% corrélées a dCol, or que hfootcg et Hdeckcg sont des combinaisons linéaires des hauteurs. Mdeck, Mfoot, Ideck et Ifoot sont des dérivées de la géométrie. Donc les histogrammes des paramètres dérivés (A, I, Hfootcg, Hdeckcg) et des constantes (Es, Df) ne sont pas reproduits : ils n'apportent pas d'information nouvelle. Nous concentrons la présentation sur les **paramètres primaires** et sur des **indicateurs synthétiques** (EI, λ , B/L, ratios de masses) qui résument mieux la diversité structurale des 10 modèles S1–S10.

	Rapport de stage	Page 38/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES

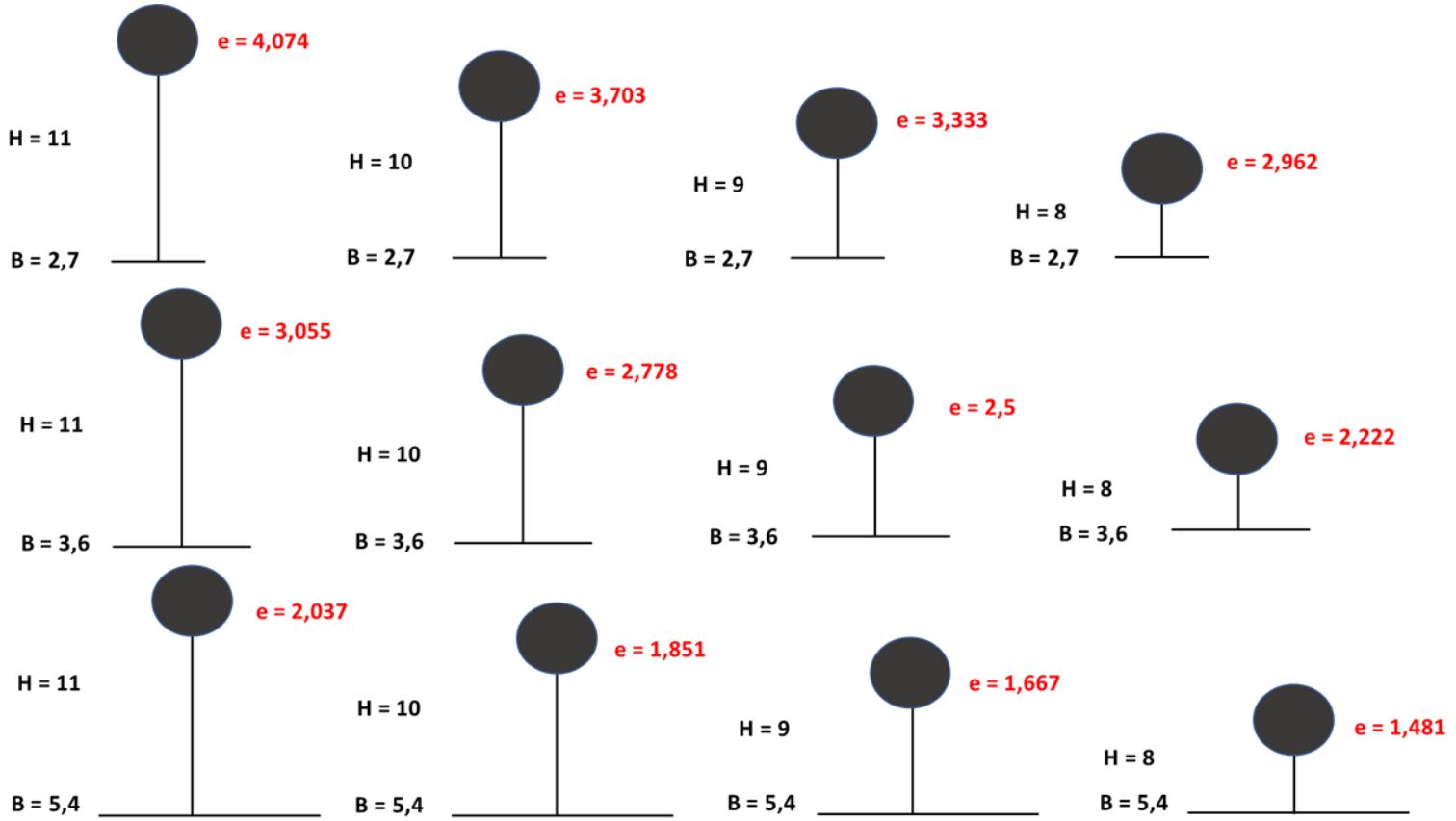


Figure 11 : Les différentes configurations de la structure

La figure positionne les différentes configurations de la structure en fonction de la dimension de base B et de la hauteur de colonne H_{col} alors pour chaque cas on peut trouver l'élancement $e = H_{col}/B$ qui varie entre 1.5 et 4 En plus, ce qu'il faut savoir que plus augmente \rightarrow la raideur latérale est plus faible \rightarrow les déplacements/dérivé sont plus élevés et potentiellement plus d'énergie dissipée si des non-linéarité s'activent (sol-fondation-structure). A l'inverse, des bases plus larges (B élevé) tendent à réduire NS et filtrer l'accélération de sortie (AAR plus faible) (voir Figure 10).

	Rapport de stage	Page 39/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

4.3.4 Combinaison et Fusion des paramètres

Enfin, les trois ensembles des paramètres sol, chargement, structure sont combinés de manière cartésienne pour obtenir toutes les possibilités dans un ensemble complet des scénarios de simulations ou chaque ligne de cet ensemble complet correspond à une combinaison unique de propriétés de sol, configuration structurelle et paramètre de chargement. Le fichier final complet contenant toute la combinaison vaut 17 280 essais qui seront chacun simulées dans OpenSees afin d'alimenter la base de données.

5 EXPLOITATION ET PRETRAITEMENT

5.1 Introduction

Une fois la génération des paramètres terminé, la deuxième étape consiste à exploiter toutes ces combinaisons obtenues pour simuler des valeurs numériques qui nous aideront dans la partie du post traitement. Voici les démarches qu'on va suivre pour découvrir et mieux comprendre la base de données :

- Simulation des cas de charges pour chaque combinaison générée en sauvegardant ces valeurs.
- Post-traitement de ces résultats pour calculer les variables d'intérêt (Inputs et Outputs).
- Exploitation de la base de données avec des analyses univariées et multivariées pour décrire les distributions de chaque variable et identifier les relations entre elles.
- Nettoyage de la base de données en éliminant les valeurs aberrantes et/ou manquantes.
- Conclusion qui résume la nouvelle base de données propre, cohérente et prête à être exploitée par des modèles prédictifs.

5.2 Importance des fichiers « .Outs » générés

Chaque simulation sur OpenSees donne un ensemble de fichiers de sorties « .out » décrivant des enregistrements temporels de déplacements, accélérations, rotations et efforts internes. Ces fichiers contenant les valeurs numériques simulées sont très essentielles car ils constituent les données brutes à partir desquelles sont calculées les variables d'intérêt (Inputs et Outputs). Voici un tableau résumant le contenu et l'utilité des fichiers outs. Alors, comme c'est mentionné, pour étudier l'SFSI c'est très important d'étudier le déplacement de tous les nœuds mais spécifiquement nœud 4 et 6 car pour calculer les variables d'intérêt on est besoin de savoir comment les ondes sismiques se propagent dans le sol ensuite, frappent la fondation pour arriver à la structure.

		Rapport de stage	Page 40/66
		Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :	
	RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

Tableau 9 : Contenu et utilité principale des fichiers «. Outs »

Fichier .out	Contenu principal	Utilité principale
node6_disp.out	Temps, déplacements (ux, uy), rotation (rz) au nœud 6	Calcul de PGD, PGV, NS
node6_accel_x.out	Accélération horizontale au nœud 6	Calcul de PGA_out et AAR
node4_rot.out	Rotation en base de colonne (nœud 4)	Calcul de NED
col4_force.out	Efforts internes (moment en pied de colonne)	Calcul de NED
FS_value.out	Facteur de sécurité FS	Variable de sortie directe
GM_run.txt	Signal sismique généré (entrée)	Calcul de PGA_in et Ia

5.3 Post-traitement des variables

Le post traitement consiste à transformer les séries temporelles enregistrées selon les nœuds avec OpenSees sous des fichiers «. Outs » en grandeurs synthétiques caractérisant la réponse du système sous forme des variables divisées en inputs et outputs. Les variables d'intérêt proviennent directement du fichier « params.tcl » tel que chaque combinaison est décrite sous forme de ce fichier contenant les valeurs des paramètres du sol, chargement et de la structure. Pour mieux comprendre le post traitement voici un tableau qui explique ces variables en précisant la formule, unité, symbole et source. On a choisi ces variables en se basant sur les variables qui ont été choisi dans les différentes article pour étudier l'interaction du système sans utilisé le ML (Soundararajan & Gajan, 2020) (Gavras et al., 2020) et avec le ML (Gajan, 2021)

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

		Rapport de stage	Page 41/66
		Réf. : Aix-Marseille Université	
		Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

Tableau 10 : Variables d'entrées et de sorties

Variable	Type	Symbol	Définition / Formule	Unité	Source
Déplacement max	Input	PGD	PGD = Max U	m/s	En tête de tablier node6_disp.out
Vitesse max	Input	PGV	PGV = Max $\left \frac{\partial u}{\partial t} \right $	m/s	Dérivé du déplacement node6_disp.out
Accélération max sortie	Input	PGA	PGA = Max $\left \frac{\partial^2 u}{\partial^2 t} \right $	m/s ²	Dérivé Second du déplacement node6_disp.out
Intensité d'Arias	Input	Ia	Ia = $\frac{\pi}{2g} \int_{t_0}^{t_f} a(t)^2 dt$	m/s	GM_run.txt
Facteur de sécurité	Input	FS	FS = $\frac{q_{ult}}{q_{applied}}$	–	FS_value.out
Settlement normalisé	Output	NS	NS = Max $\left \frac{PGD}{B} \right $	–	node6_disp.out et largeur B
Dissipation d'énergie normalisée	Output	NED	NED = $\frac{1}{B \cdot V_{tot}} \int M d\theta$	–	col4_force_out (moments M) node4_rot.out (rotations θ) θ = interp(t_M, t_rot, θ_rot)
Amplification d'accélération	Output	AAR	AAR = $\frac{a_{max, out}}{a_{max, in}}$	–	node6_accel_x.out GM_run.txt

Les variables d'intérêts calculées à partir des fichiers de sortie des simulations sont résumées dans le tableau ci-dessous ou chaque formule repose sur des grandeurs extraites directement ou indirectement des fichiers « Outs » produits par OpenSees et traduit un aspect physique particulier du système. Détaillant chaque variable pour mieux comprendre :

- PGD : Correspond au déplacement horizontal maximal du nœud 6 qui présente la tête du tablier. Les valeurs de PGD sont obtenues partit du fichier node6_disp.out qui dépend directement des conditions de chargement et de la rigidité de la structure.
- PGV : Correspond au dérivé de la vitesse associée au déplacement horizontal du nœud 6. Pour trouver les valeurs de PGV il suffit de calculer le dérivé premier de la série de déplacement dans le fichier node6_disp.out.
- PGA : Correspond à l'accélération maximale mesurée au nœud 6, sa valeur est définie à partir du fichier node6_accel_x.out qui nous donne les valeurs d'accélération en g en traduisant l'intensité des efforts inertIELS subis par la superstructure.

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

		Rapport de stage	Page 42/66
	Réf. : Aix-Marseille Université		
	Date : 22/07/2025	Indice :	
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

- Ia : c'est un indicateur énergétique, calculé par intégration de l'accélération d'entrée au carré sur la durée totale du signal en utilisant les données du fichier « GM_run.txt ». Cette variable représente la quantité d'énergie sismique transmise au système.
- FS : C'est le facteur de sécurité qui nous informe sur la stabilité globale vis-à-vis de la portance. Il est calculé comme le rapport entre la charge limite du sol (q_{ult}) et la charge appliquée ($q_{applied}$). Telle que, q_{ult} est calculé dans le script « FScale.tcl » à partir des formules classiques de portance (terzaghi/Hanss) en fonction de la Largeur B, longueur L de la fondation et de la profondeur d'encastrement Df en plus, les propriétés mécaniques du sol : angle de frottement phi, cohésion c et poids volumique gamma alors que le $q_{applied} = V_{tot}/A$ avec $V_{tot} = (M_{deck} + M_{foot}).g$ qui correspond au poids total de la superstructure et de la fondation et $A = B \cdot L$ qui correspond à l'Aire de la semelle. Donc FS dépend à la fois des propriétés du sol (c, phi, gamma) et des charges appliquées (M_{deck} , M_{foot}), ainsi que de la **géométrie de la fondation** (B, L, Df).
- NS : c'est défini comme le déplacement horizontal maximal (PGD) rapporté à la largeur de semelle B. Ce rapport rend le tassement comparable entre différentes configurations géométriques.
- NED : cette variable mesure la capacité du système à dissiper de l'énergie au travers des cycles hystéritiques. Elle est obtenue en intégrant le produit du moment en pied de colonne en utilisant les valeurs du fichier « col4_force.out » par la rotation correspondante du fichier « node4_rot.out » telle que cette intégration mesure l'énergie perdue lors des cycles sismiques. De plus, il faut multiplier par B. V_{tot} le facteur de normalisation pour comparer différentes configurations avec : B largeur de la semelle et V_{tot} le poids total appliqué au système. Ainsi, NED dépend à la réponse dynamique (moments et rotations), à la géométrie de la fondation (B) et au poids appliqué (M_{deck} , M_{foot}).
- AAR : c'est le rapport entre l'accélération maximale en sortie « node6_accel_x.out » et celle en entrée « GM_run.txt », cette variable quantifie l'effet d'amplification ou d'atténuation induit par le système sol-fondation-structure.

5.4 Analyse Univariée

Après le post traitement, c'est le temps d'explorer la distribution de chaque variable séparément pour mieux comprendre le comportement de la performance du système étudier en évaluant la forme de la distribution (symétrie ou asymétrie), la présence d'extrêmes et la dispersion.

		Rapport de stage	Page 43/66
		Réf. : Aix-Marseille Université	
		Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

Tableau 11 : Tableau statistique des variables avant prétraitement

	Min	Q25	Médian	Q75	Max	Mean	Std	Skew	Kurt	iqr	IQR_%
AAR	0.9542	1.0127	1.0234	1.0438	7872.0548	2.7691	92.1996	79.5632	6677.5265	0.0311	9.2625
FS	1.2	2.1166	3.7739	5	5	3.5077	1.46	-0.3529	-1.4505	2.8834	0
Ia	1.0819	3.0235	4.2115	5.713	12.1683	4.496	1.8339	0.6719	0.0421	2.6895	0.6375
NED	5.57E-07	7.94E-06	1.80E-05	4.37E-05	10109.5969	1.7577	119.1483	79.0846	6553.2717	3.58E-05	11.15
NS	0.0003	0.0012	0.002	0.0036	798.7963	0.1313	9.0265	86.7952	7666.4124	0.0023	8.6625
PGA	0.4872	0.6045	0.708	0.8116	5510.4383	1.9223	64.8881	78.6799	6547.9861	0.2072	3.0875
PGD	0.0017	0.0045	0.007	0.0125	2156.75	0.3716	24.4783	85.8033	7535.2495	0.008	8.2875
PGV	0.084	0.1518	0.1881	0.2442	652.0375	0.3397	7.5536	81.562	6954.6841	0.0924	4.025

Le tableau ci-dessus présente N= 8000 essais pour chaque variable, aucune valeur manquante NA=0 et des indicateurs classiques pour des critères de décision :

Les colonnes Minimum et Maximum on peut prendre une idée générale sur les bornes de chaque variable calculée et pour les quantiles on peut analyser le pourcentage de la distribution.

Les colonnes q25, médiane, q75 racontent le cœur de la distribution de chaque variable et pour mieux comprendre la dispersion des points on regarde la colonne iqr = q75 – q25 si IQR grand → la variable est hétérogène bien dispersée et si IQR petit → la variable est stable peu dispersée. En se basant sur nos résultats on peut dire que FS et la sont bien dispersées contraire des autres variables.

La colonne Skew nous donnent des informations sur la distribution : concernant la queue : skew < 0 → queue à gauche, skew > 0 → queue à droite et concernant la normalité : 0.5 < skew < 1 → Asymétrie légère et 1 < skew < 2 → Asymétrie modérée et skew > 2 → très Asymétrie. Donc en se basant sur notre résultat on constate que les variables AAR, NED, NS, PGD, PGV, PGA présentent des queues a droite et des distributions trop asymétrique ce qui montre l'existence des points aberrantes et la preuve c'est les pourcentages dans la colonne d'interquartiles IQR %. La colonne Kurt nous donne des informations sur l'aplatissement selon les coefficients d'aplatissement on a : si Kurt est presque nul 0 → distribution proche de normale (mesokurtique), si 0-1 → légèrement leptokurtique ; si 1-2 → modérément leptokurtique ; si 2-5 → fortement leptokurtique (queues lourdes) ; si > 5 → extrême (queues très lourdes, outliers nombreux) et si <0 → platikurtique (plus "plate" que normale). Alors, en se basant sur les résultats du tableau on conclut qu'aussi les variables AAR, NED, NS, PGD, PGV, PGA ont un coefficient d'aplatissement extrêmes ce qui montre une queue très lourde alors que les autres coefficients des deux variables FS et la présentent des distributions plus plates que normale.

		Rapport de stage	Page 44/66
		Réf. : Aix-Marseille Université	
		Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

5.5 Transformation logarithmique et nettoyage

D'après les résultats de l'analyse univarié, il faut appliquer des transformations logarithmiques sur les variables qui présentent des queues lourdes et une forte asymétrie c'est pour cela on va visualiser la distribution de ces variables après transformation sous forme des histogrammes et présenter un tableau statistique pour comparer avant et après la transformation pour logarithmique et l'élimination des points aberrantes.

Tableau 12 : tableau statistique des variables après prétraitement

	Skew_1	Skew_2	Kurt_1	Kurt_2	IQR_%_1	IQR_%_2
FS	-0.3529	-0.3529	-1.4505	-1.4505	0	0
Ia	0.6719	0.5486	0.0421	-0.3946	0.6375	0.1384
Log AAR	11.1842	1.1033	170.5493	0.7547	9	3.4478
Log NED	2.3965	0.1687	18.3471	-0.3107	1.6	0.1905
Log NS	2.3075	0.4074	16.8155	-0.3047	1.5125	0.1904
Log PGA	8.0787	-0.0785	111.1769	-0.8594	2.6875	0.0514
Log PGD	2.7566	0.5439	20.6003	-0.3315	1.6375	0.1652
Log PGV	3.7125	0.295	44.5096	-0.3685	1.175	0.1518

Voici le tableau ci-dessus qui nous montre la différence des valeurs de skew, Kurt et IQR_% avant et après la transformation : Vous pouvez remarquer les valeurs de skew_1, Kurt_1 et IQR_%_1 comment ils ont tous diminué sauf la variable FS car elle ne contient pas des valeurs aberrantes. Ce qui nous aboutit à une base de données plus cohérent, robuste et bien nettoyée.

Par ailleurs, les variables sont mesurées sur des échelles hétérogènes (m, m/s, g, ratios), ce qui justifie une standardisation avant l'analyse multivariée et les modèles sensibles à l'échelle comme SVM et KNN qu'on va les utiliser, au contraire des autres arborescents qui ne l'exigent pas.

5.5.1 Représentation Graphiques des Features

La distribution de la variable PGD est asymétrique à droite, cohérente avec des déplacements majoritairement faibles et quelques épisodes forts. En plus, les points aberrants présents en queue c'est pour cela il faut appliquer la transformation logarithmique pour réduire l'impact des cas extrêmes. Voici une figure après le filtrage et la transformation logarithmique qui montre une distribution unimodale, plus symétrique et l'IQR capture la variabilité utile sans domination des extrêmes (voir Figure 10).

	Rapport de stage	Page 45/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

La distribution de la variable PGV est asymétrique à droite, avec des vitesses surtout modérées et de rares valeurs élevées. En plus les points aberrants sont visibles ce qui nécessite une application logarithmique pour homogénéiser l'échelle et limiter l'effet de queue. Voici une figure après le filtrage et la transformation logarithmique qui montre une distribution plus équilibrée avec un faible taux d'IQR (voir Figure 11).

La distribution de la variable PGA est asymétrique à droite avec une majorité d'accélérations modérées et quelques valeurs fortes. En plus, les points aberrants sont visibles en queue c'est pour cela il faut appliquer une transformation logarithmique pour compresser l'échelle et fiabiliser les statistiques. Voici une figure après transformation qui montre la distribution nettement plus symétrique avec un IQR qui reflète mieux la variabilité centrale et les extrêmes qui ont été écartés (voir Figure 12).

La distribution de la variable FS est compact et légèrement asymétrique à gauche car le Skew est négatif, mais on constate qu'elle est en cohérence avec des valeurs souvent au-dessus de 1. En plus, cette distribution ne présente pas des outliers (quasi pas d'outliers). Donc, voici une figure qui représente la distribution de FS sans transformation nécessaire car l'échelle actuelle est déjà adéquate (voir Figure 13).

La distribution de la variable la est modérément à droite avec quelques épisodes d'énergie élevée, mais IQR raisonnable et faible pourcentage d'aberrantes. Donc, voici une figure présentant la distribution d'la sans modification car le comportement reste lisible et stable (voir Figure 14).

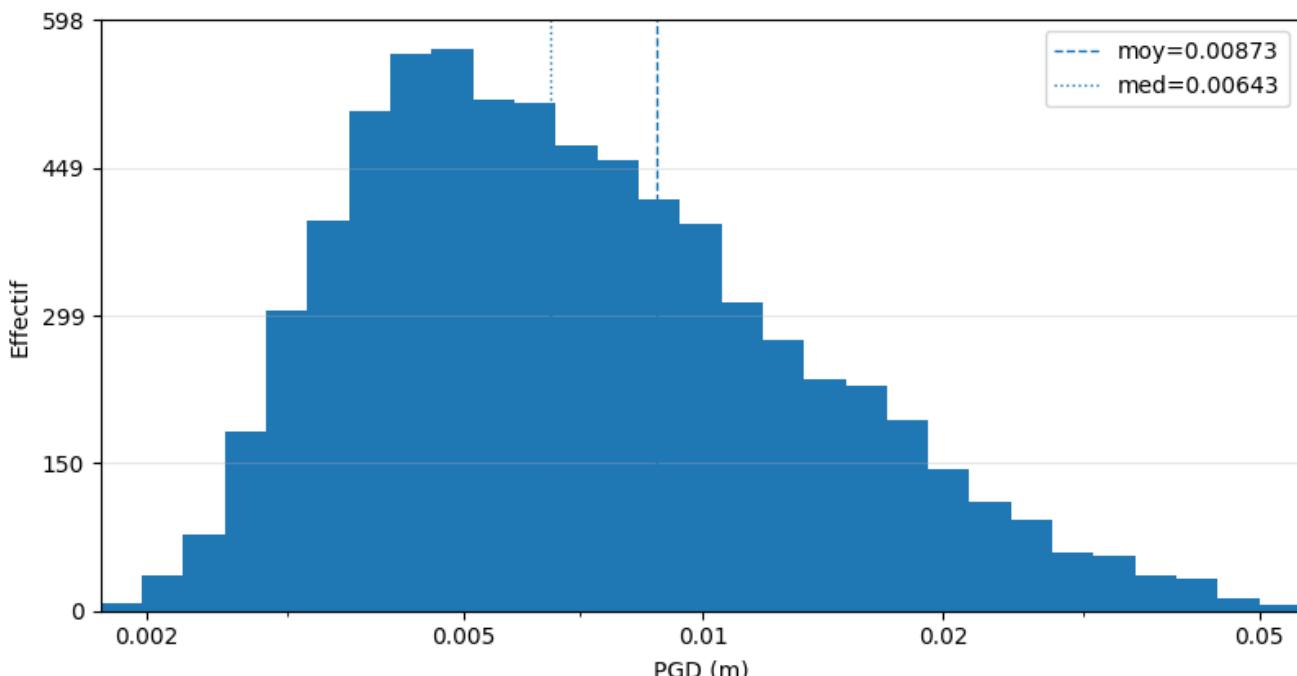


Figure 12 : Distribution du PGD après prétraitement

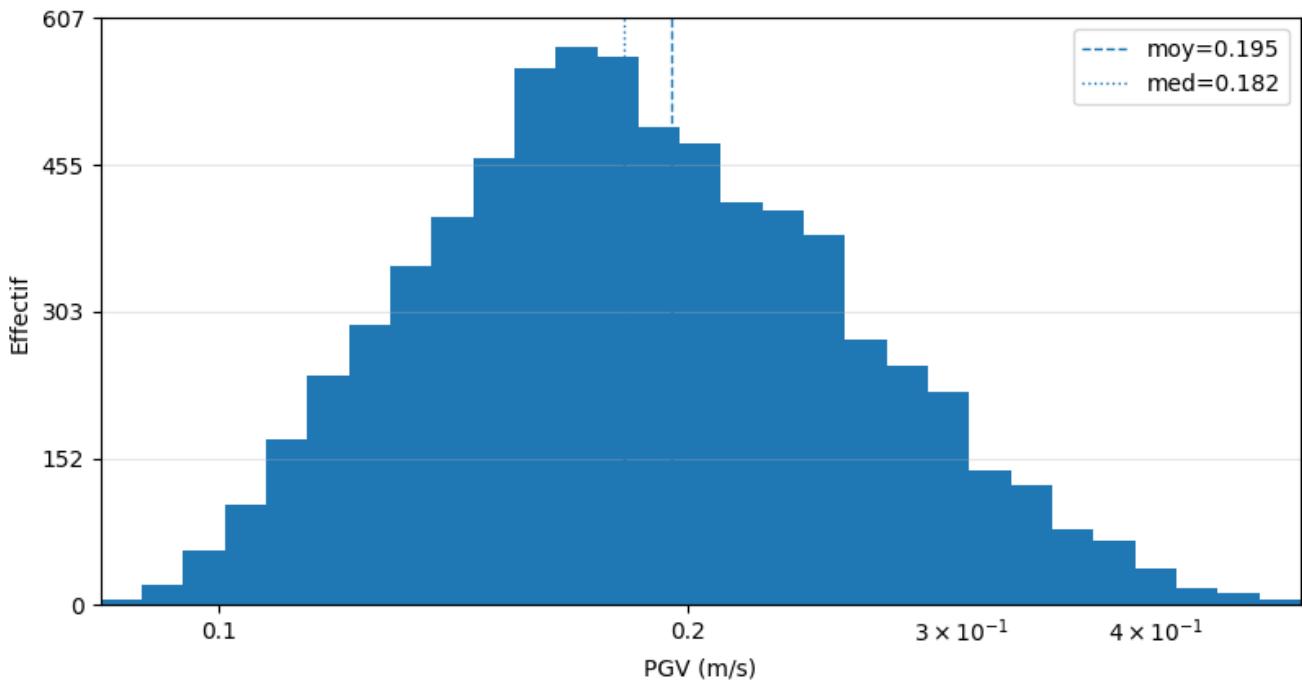

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES


Figure 13 : Distribution du PGV après prétraitement

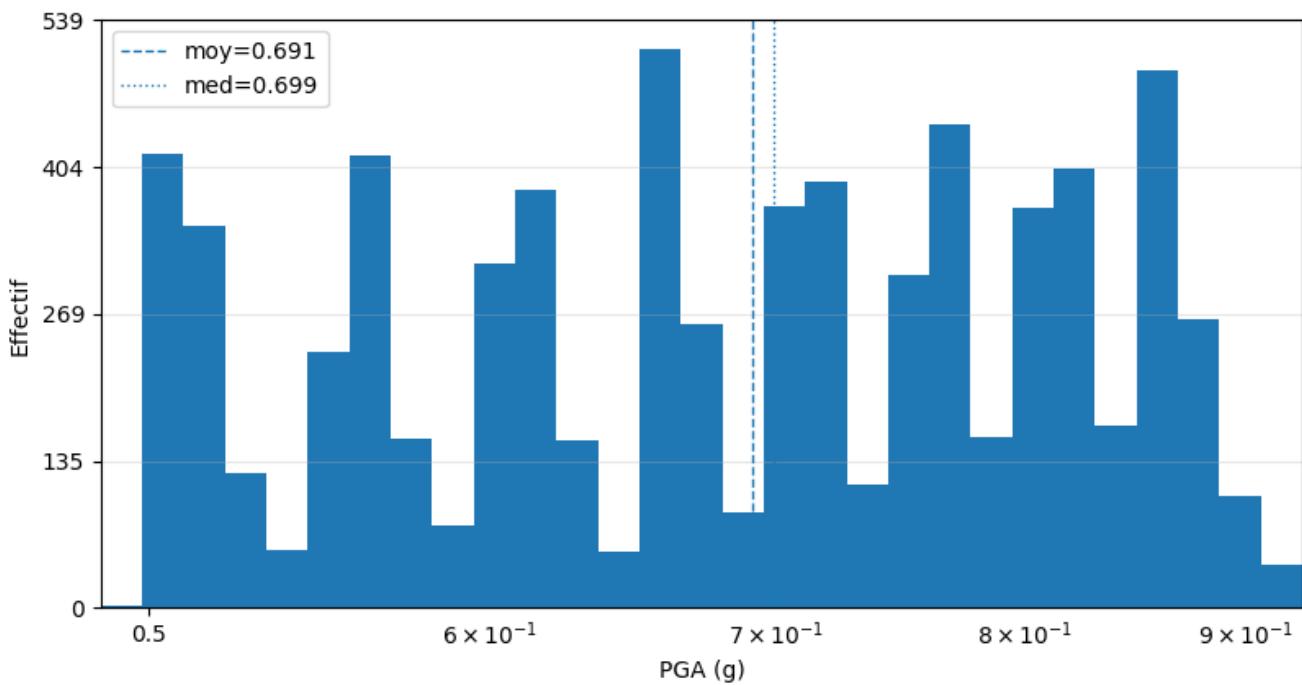


Figure 14 : Distribution du PGA après prétraitement

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur



Rapport de stage

Page 47/66

Réf. :
Aix-Marseille Université

Date : 22/07/2025 Indice :

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES

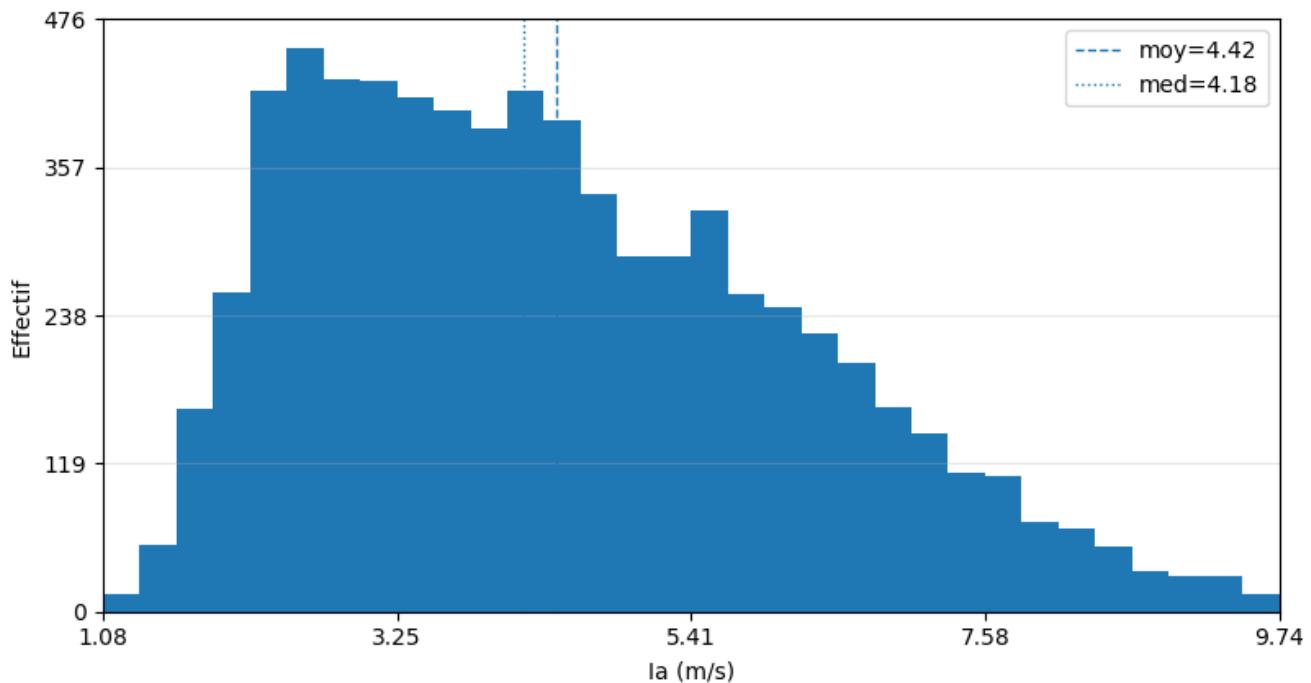
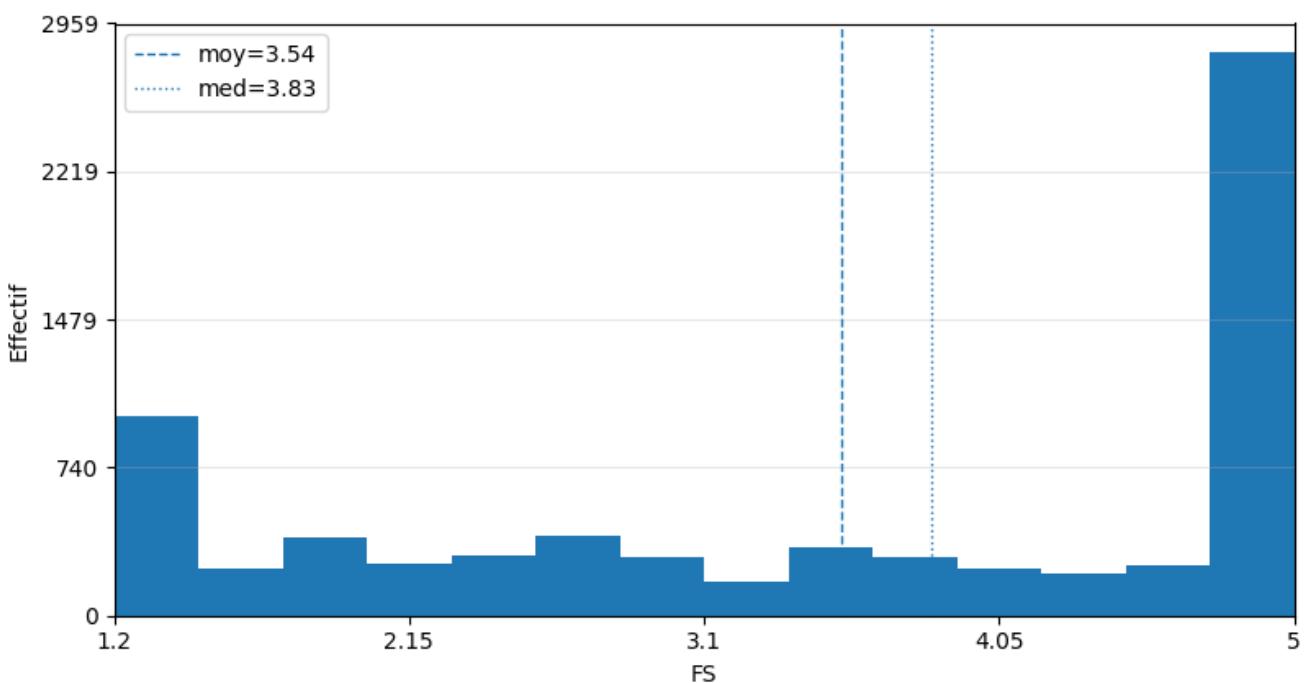


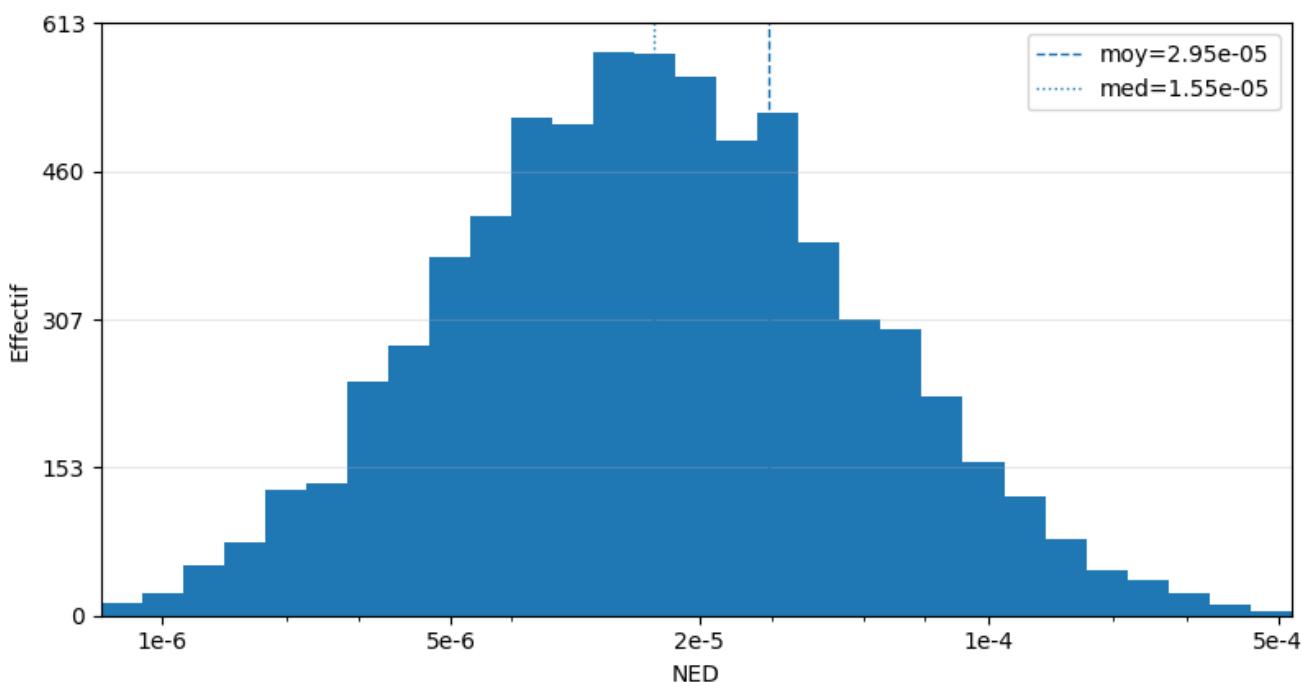
Figure 15 : Distribution de la après prétraitement



	Rapport de stage	Page 48/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
	RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES	

5.5.2 Représentation graphique des outputs

- La distribution de la variable NED est fortement asymétrique à droite (queue lourde), avec un noyau de valeurs faibles et quelques pics très élevés. En plus, l'IQR est étroit mais le % des points aberrantes est notable (Voir tableau 14) indiquant une dispersion extrême dans la queue. Alors, une transformation log10 s'impose pour compacter l'échelle et réduire l'influence des valeurs extrêmes. Donc voici une figure qui après transformation La distribution qui a devenu quasi symétrique autour de sa médiane, avec un IQR resserré et un taux de points d'aberrantes faible c'est-à-dire le log10 a stabilisé la variance et rendu NED exploitable pour l'analyse multivariée (voir Figure 15).
- La distribution de la variable NS est très asymétrique à droite, dominée par de petites valeurs avec quelques cas élevés. En plus, l'IQR couvre une plage restreinte mais la queue droite reste lourde (outliers présents). Alors, on applique une transformation logarithmique pour réduire la dissymétrie et la sensibilité aux extrêmes. Donc, voici après les modifications, une distribution beaucoup plus équilibrée, avec un IQR réduit et peu d'outliers autrement-dit, Le log10 a amélioré la lisibilité et la comparabilité de NS avec les autres indicateurs (voir Figure 16).
- La distribution de la variable AAR est centrée près de 1 avec un IQR très faible, mais une queue droite sporadique liée à quelques amplifications ponctuelles (outliers). Alors, le log10 est pertinent pour écraser ces cas extrêmes et clarifier le comportement "typique". Donc, voici après transformation, une distribution resserrée et quasi symétrique autour de la médiane tel que, les pics d'amplification atypiques ont été neutralisés par le filtrage. Ce qui nous q donné un AAR plus stable pour la corrélation (voir Figure 17).



DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur



Rapport de stage

Page 49/66

Réf. :
Aix-Marseille Université

Date : 22/07/2025 Indice :

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES

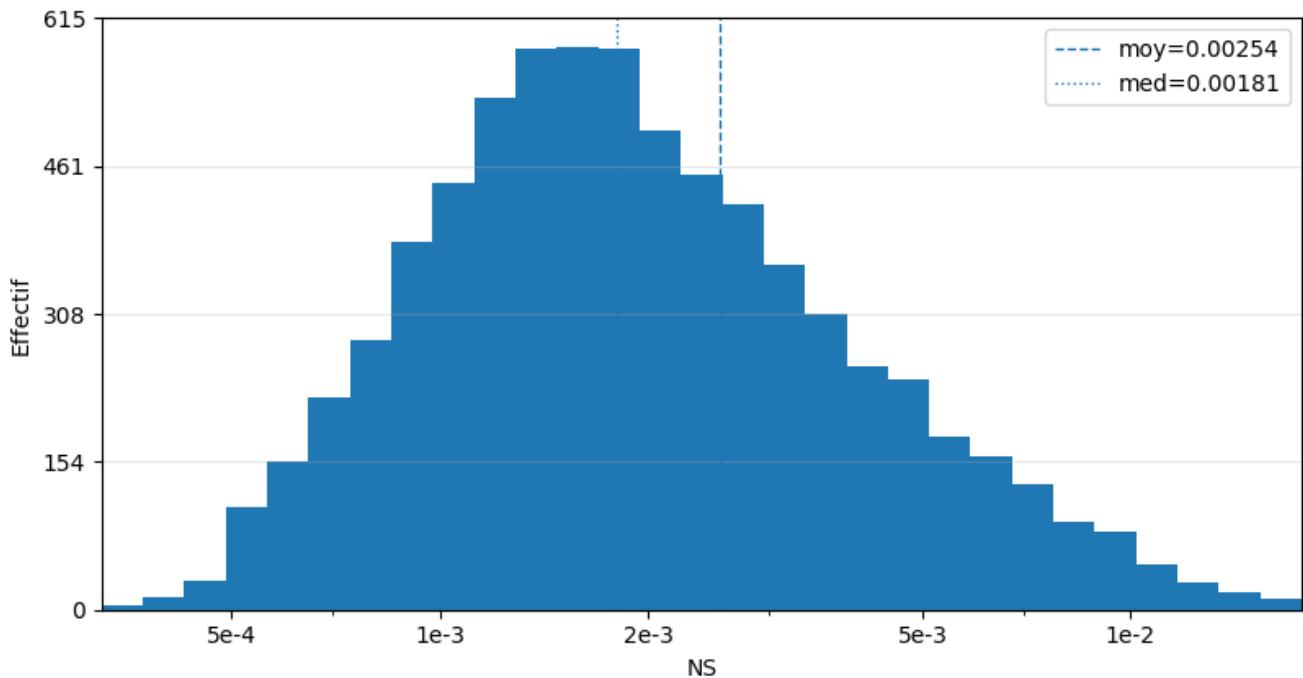


Figure 18 : Distribution de NS après prétraitement

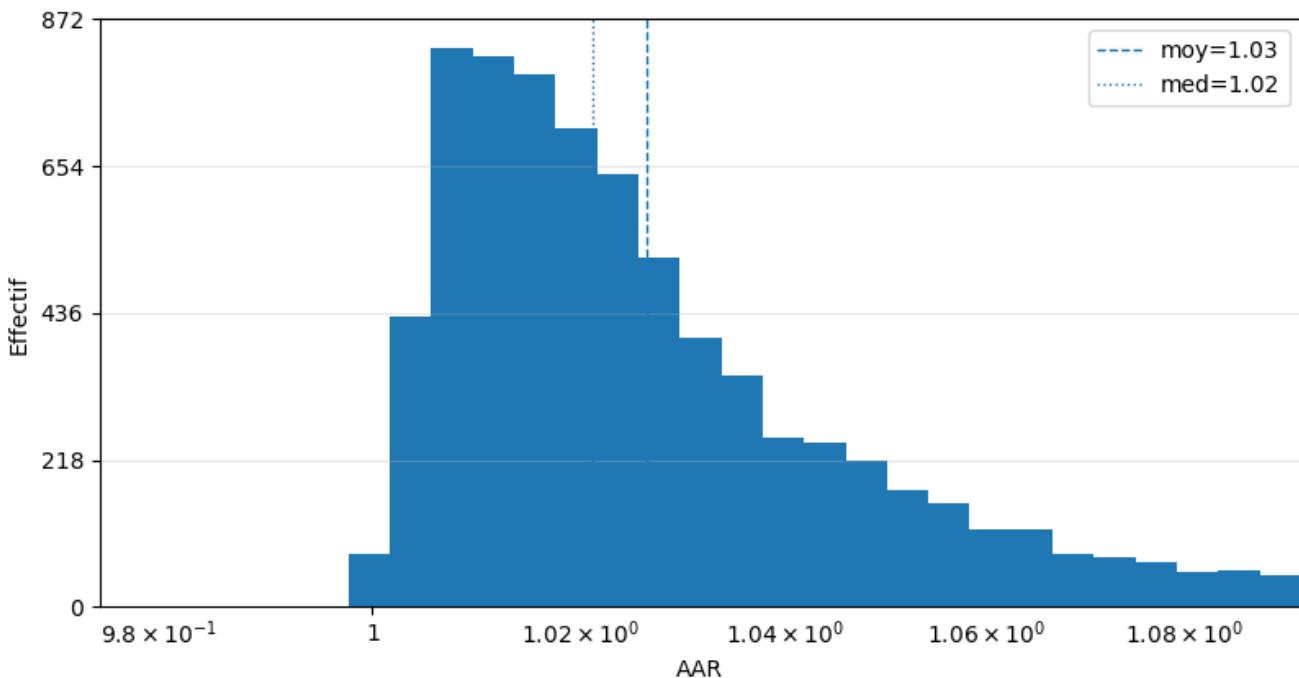


Figure 19 : Distribution de AAR après prétraitement

	Rapport de stage	Page 50/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

5.6 Analyse Multivariée

5.6.1 Matrice de corrélation

Les résultats de la matrice de corrélation présentent 3 niveaux de corrélations : faible ; moderne et forte. En se basant sur la formule et la modélisation du système sol-structure on constate que :

- Les variables qui sont fortement corrélées sont : PGD et NS avec $r= 0.915 \rightarrow$ ce qui est normale car en calculant NS on a PGD comme numérateur PGD. Apres, on a PGD et PGV avec $r=0.875 \rightarrow$ ce qui est normal car la vitesse est le dérivé du déplacement, de même pour PGA et la avec $r=0.817 \rightarrow$ ce qui est logique car en calculant la on intègre l'accélération au carré. Ensuite on a PGV et PGA avec $r=0.602 \rightarrow$ tout simplement c'est que les deux variables montent ensemble. Donc cette forte corrélation de ces variables est due a des cohérence cinématique et énergétique.
- Les variables qui sont modérément corrélées sont : NED et PGV avec $r=0.696$, NED et PGD avec $r=0.668$ et NED avec AAR avec $r=0.632 \rightarrow$ ce qui montre que l'énergie dissipe augmente quand l'excitation est riche en vitesses et amplification.
- Les variables qui sont faiblement corrélées (légèrement négatif) sont : FS avec NED et NS \rightarrow ce qui est logique car plus l'énergie dissipé et le tassement augmente plus le facteur de sécurité stress et diminue.

Parfois la matrice de corrélation ne montre pas la relation directement entre deux variables comme AAR et PGA (avec $r=0.10$). L'amplification AAR est définie comme un rapport entre PGA et accélération du sol telle que le PGA c'est le résultat sur le nœud 6 donné par des valeurs numériques simulée avec OpenSees et qui sont contrôlée par la proximité fréquentielle, l'amortissement et les non-linéarités sol/structure alors vu que le modèle de notre système est quasi linéaire l'effet d'amplitude s'annule dans le ratio, pour mieux comprendre les corrélations modérées d'AAR avec PGV ($r=0.49$) et PGD ($r=0.47$) confirment l'influence du contenu spectral et des mécanismes de déformation plutôt que l'amplitude d'entrée seule. Les démarches suivis c'étaient les mêmes que dans l'article (Gavras et al., 2020) dans le but d'étudier l'SFSI en utilisant la matrice de corrélation.

Voici une matrice de corrélation qui rend l'idée plus claire et facile pour comprendre quelles sont les variables les plus et les moins corrélées (voir Figure 18).

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

		Rapport de stage	Page 51/66
		Réf. : Aix-Marseille Université	
		Date : 22/07/2025	Indice :
		RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES	

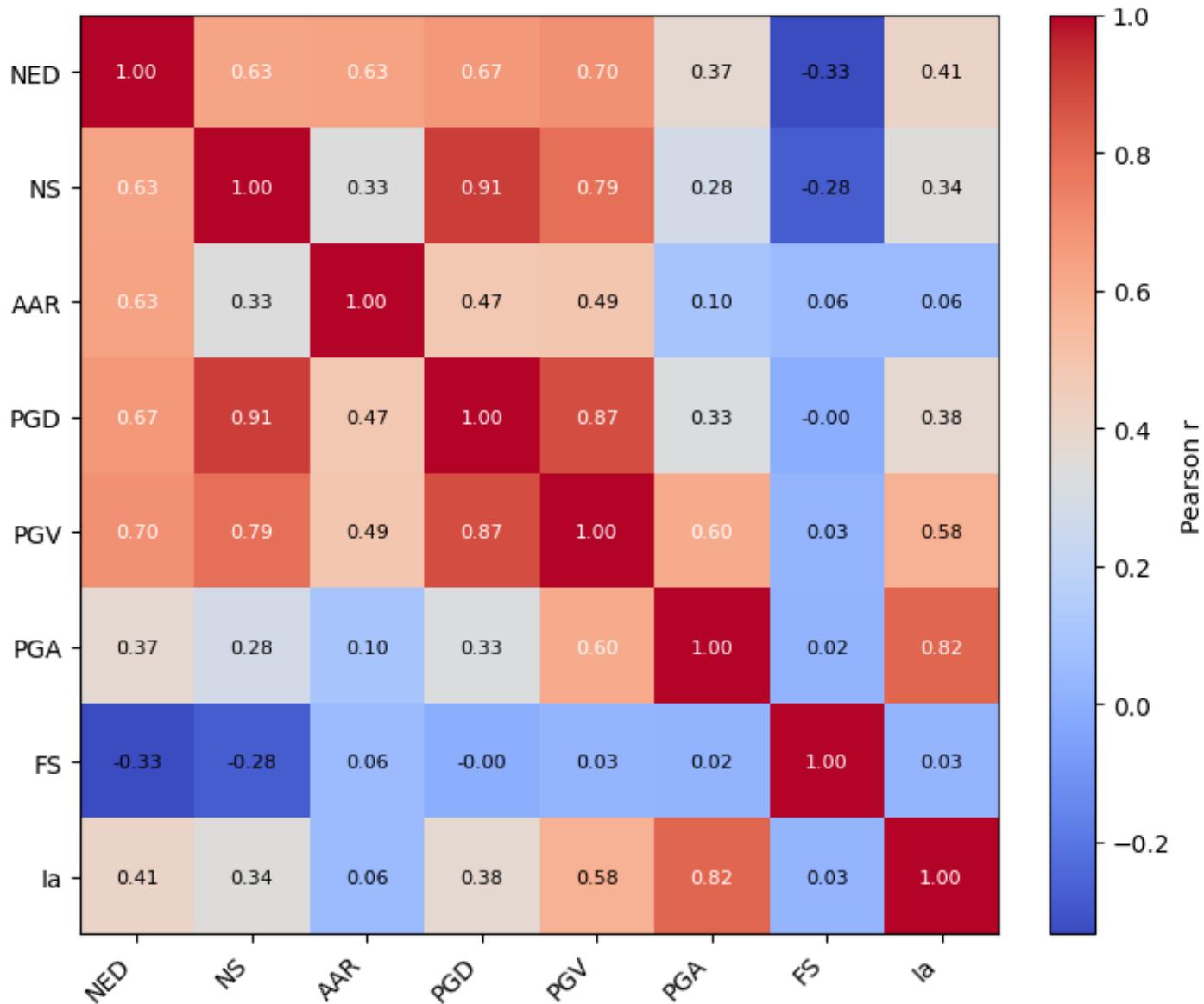


Figure 20 : Matrice de corrélation des variables d'intérêt après prétraitement

	Rapport de stage	Page 52/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

5.6.2 Matrice de distribution

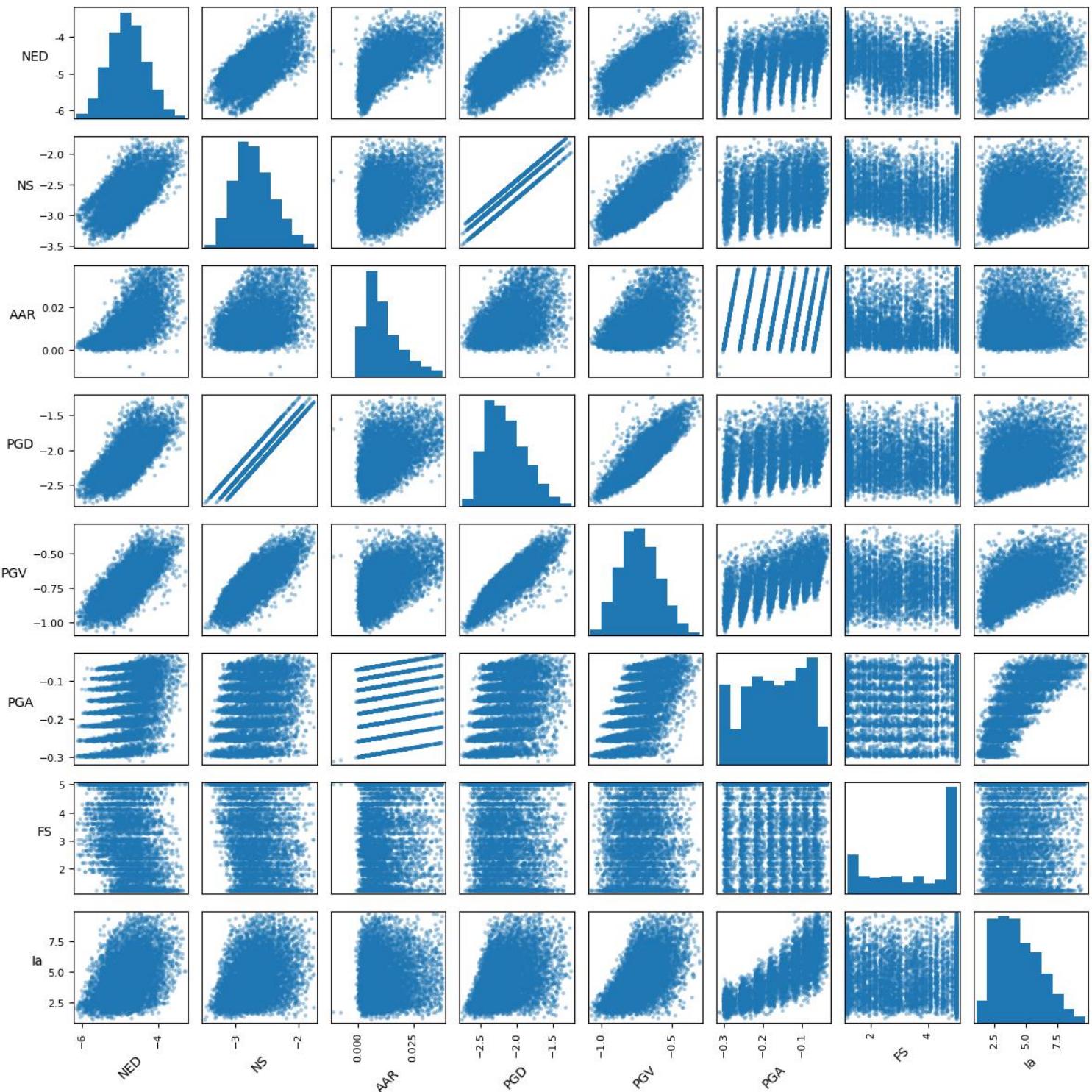
Voici une matrice qui représente toutes les distributions des variables Features et Outputs après transformation logarithmique et filtrage. Tout d'abord, vous pouvez constater que les distributions sur la diagonale sont unimodales et globalement plus symétriques sauf celle de la variable AAR (légère queue droite) et FS (légère queue gauche) restent un peu dissymétriques. Ensuite cette matrice représente des relations bivariées en regardant les termes diagonaux de la matrice on conclut qu'il existe des relations presque linéaires, modérées et des autres faibles (voir Figure 1).

- Relation quasi linéaire : Première relation c'est NED vs NS, Deuxième relation c'est PGD et PGV ; Troisième relation c'est NS avec AAR, PGA et la ; Quatrième relation c'est AAR avec PGA ; Cinquième relation c'est PGD avec PGV
- Relation modérée : Première relation c'est NED vs AAR, PGA et la ; Première relation c'est NS avec PGD ; Troisième relation c'est AAR avec PGD, PGV et la ; Quatrième relation c'est PGD avec PGA et la ; Cinquième relation c'est PGV avec PGA et la ; Sixième relation c'est PGA avec la
- Relation faible : Première relation c'est NED vs FS ; Deuxième relation c'est NS avec FS ; Troisième relation c'est AAR avec FS ; Quatrième relation c'est PGD avec FS ; Cinquième relation c'est PGV avec FS ; Sixième relation c'est PGA avec FS ; Septième relation c'est FS avec la.

Alors, qu'on déduit que NS, PGV, PGD et NED covariant alors que AAR reste relativement indépendante de PGA, et FS aussi reste indépendante indirectement de ces variables. Conclusion La représentation en **log** a clarifié la structure des relations et prépare bien la phase de ML (voir Figure 19).

5.7 Résultats et Conclusion

Pour résumer, nous avons travaillé avec une base de données initiale contenant **8000 essais** et **42 colonnes** (variable + paramètres), Tout d'abord on a appliqué une **transformation logarithmique** des variables NED, NS, AAR, PGD, PGV, PGA et après on a **nettoyé les valeurs aberrantes** par la règle IQR. Ensuite, on a obtenu une base de données filtrée qui comporte **7 203 essais** (90% des essais ont été conservés) c'est-à-dire le retrait des points aberrantes représente typiquement 1.6% des observations par variable. Malgré qu'on ait remarqué des corrélations entre certaines mesures d'exigence, aucune variable n'est redondante vis-à-vis de toutes les autres. Donc, la base de données filtrée est prête pour entraîner et étudier les différents modèles d'apprentissage automatique afin d'atteindre les objectifs de ce travail.


RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES


		Rapport de stage	Page 54/66
		Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :	
	RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

6 PARTITIONNEMENT ET APPLICATION DES ML

6.1 Introduction

Nous cherchons à prédire trois sorties sismiques — NED, NS et AAR — à partir de cinq variables d'entrée (PGD, PGV, PGA, Ia, FS). Avant modélisation, les variables fortement asymétriques ont été log-transformées et les valeurs aberrantes retirées via IQR, ce qui a amélioré la symétrie (skewness) et aplati les queues (kurtosis). Pour évaluer les modèles, nous séparons les données en train/test et nous utilisons la validation croisée pour l'ajustement des hyperparamètres. L'importance des variables est estimée par Permutation Importance (baisse de R² quand on permute une feature sur l'ensemble de test), ce qui offre une mesure robuste et comparable entre modèles.

6.2 Application du premier modèle « KNN »

6.2.1 Ajustement des paramètres du modèle KNN

Le modèle KNN n'admet pas une fonction globale à paramètres comme les autres, ici la prédiction c'est la moyenne des K voisins les plus proches en mesurant ça par la distance de Minkowski. Pour appliquer le meilleur modèle de KNN il faut choisir les meilleurs hyperparamètres c'est pour cela on a optimisé 3 par GridSearchCV (5-fold) :

- K = c'est le nombre de points voisins, si k petit → modèle plus local (pas besoin de prendre en considération beaucoup de points pour classer une nouvelle observation ce qui montre que la distribution n'est pas très dispersée), si k grand → modèle plus lisse, plus biais (dans ce cas-là il existe une dispersion et on a besoin de prendre en considération plusieurs points pour classer la nouvelle observation).
- Weights = il existe deux types de pondération, le première « Uniform » c'est que tous les voisins sont pondérés pareil, et le deuxième « Distance » c'est que les voisins roches pèsent plus. Ici dans cette étude bien sûr qu'on a pris la deuxième pondération : distance.
- P = C'est la norme de Minkowski, il existe deux normes : le premier p=1 → Manhattan et le deuxième p=2 → Euclidienne. Alors dans notre cas on a pris en considération p=1 car elle est plus robuste aux valeurs extrêmes quand les variables sont standardisées.

En appliquant le modèle KNN, on a exploité l'espace suivant : n_neighbors ∈ {3,5,7,9,11,15,25}, weights ∈ {uniform, distance} et p ∈ {1,2}. Après que le modèle a essayé toutes les combinaisons on a trouvé les meilleurs réglages qui sont affichés dans le tableau ci-dessus

Tableau 13 : Ajustement des hyperparamètres du modèle KNN

Cible	K (n_neighbors)	Weights	P
NED	25	Distance	Manhattan
NS	15	Distance	Manhattan
AAR	15	Distance	Manhattan

	Rapport de stage	Page 55/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

Donc, K est plus grand pour NED (relation plus lisse) et plus petit pour NS/AAR (relations plus locales). Avec Weights=distance et p=1 améliorent la sensibilité aux vrais voisins proches.

6.2.2 Comparaison des mesures métriques du modèle KNN

Après l'ajustement des hyperparamètres du modèle KNN, maintenant c'est le temps de faire des prédictions des trois variables cibles NED, NS et AAR en utilisant les cinq variables d'entrées PGD, PGV, PGA, la et FS. En se basant sur les résultats on remarque que :

- NS est très bien prédit par KNN ($R^2 \approx 0.91$), signe d'une structure de relation fortement locale et régulière entre entrées et NS. En plus, ce qui valide que ce modèle a bien prédit c'est les valeurs de RMSE ≈ 0.09 et MAE ≈ 0.07
- NED atteint un niveau correct ($R^2 \approx 0.66$) avec un k plus élevé, cohérent avec une relation plus lisse/agrégée. Alors le modèle KNN n'est pas le meilleur pour prédire NED malgré le niveau correcte R^2 , il a présenté des grandes erreurs en comparant avec NS et AAR
- AAR reste difficile à prédire avec KNN ($R^2 \approx 0.43$). De plus, le MAPE très élevé (142 %) vient du fait qu'il est calculé dans l'espace log (sorties très petites en valeur absolue \rightarrow pourcentages instables). Alors pour AAR, il est important de tester autre modèle de prédiction.

Tableau 14 : Evaluation de la performance du modèle KNN sur l'SFSI

Model	Target	R2_test	RMSE_test	MAE_test	MAPE_%
KNN	NED	0.6603	0.2765	0.223	4.6955
KNN	NS	0.908	0.0922	0.0742	2.7783
KNN	AAR	0.4298	0.006	0.0045	142.083

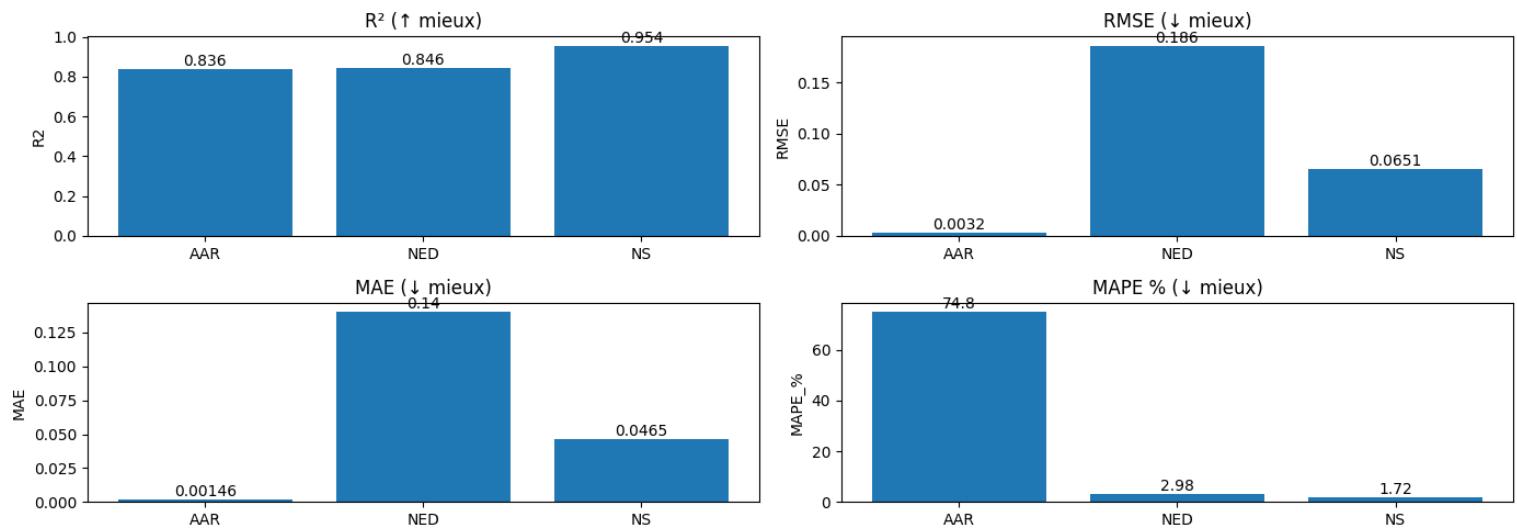


Figure 22 : Evaluation de la performance du modèle KNN sur l'SFSI

	Rapport de stage	Page 56/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

6.2.3 Evaluation d'importance des variables selon le modèle KNN

Pour comprendre l'interaction du système sol-fondation-structure il faut étudier l'importance des variables d'entrées sur les variables de sorties :

- NED : piloté par la résistance/robustesse structurelle et l'amplitude des mouvements avec un rang des entrées : FS (0.275) > PGD (0.217) ≈ PGV (0.212) ≫ la (0.040) ≈ PGA (0.039). Ce qui montre la capacité portante (FS) et les niveaux de déplacement/vitesse de pic (PGD/PGV) expliquent l'essentiel de la variabilité de NED ; l'énergie intégrée (la) et l'accélération de pic (PGA) jouent un rôle marginal.
- NS : quasi entièrement expliqué par le déplacement de pic avec un Rang : PGD (0.934) ≫ FS (0.178) > PGV (0.102) ≫ PGA (0.014) ≈ la (0.003). Ce qui montre que PGD domine très largement, ce qui confirme une relation locale, monotone et très forte entre le déplacement de pic et NS ; le reste n'apporte que des corrections secondaires.
- AAR : dépend surtout de la cinématique et du contenu énergétique du signal avec un Rang : PGV (0.329) > la (0.175) ≈ PGA (0.167) > PGD (0.126) ≫ FS (0.033). Ce qui montre le rapport d'amplification d'accélération est surtout sensible à la vitesse de pic (PGV), au contenu énergétique (la) et à PGA ; la contribution structurelle FS est faible.

Donc, la qualité de prédiction selon R^2 est classé comme le suivant : NS >> NED > AAR telle que NS suit PGD de façon quasi monotone, NED cohérente avec la structure physique **FS + (PGD/PGV)** alors que AAR est un rapport plus sensible et bruité qui dépend du cinématique (PGV, PGA), d'intensité la et indirectement des autres paramètres.

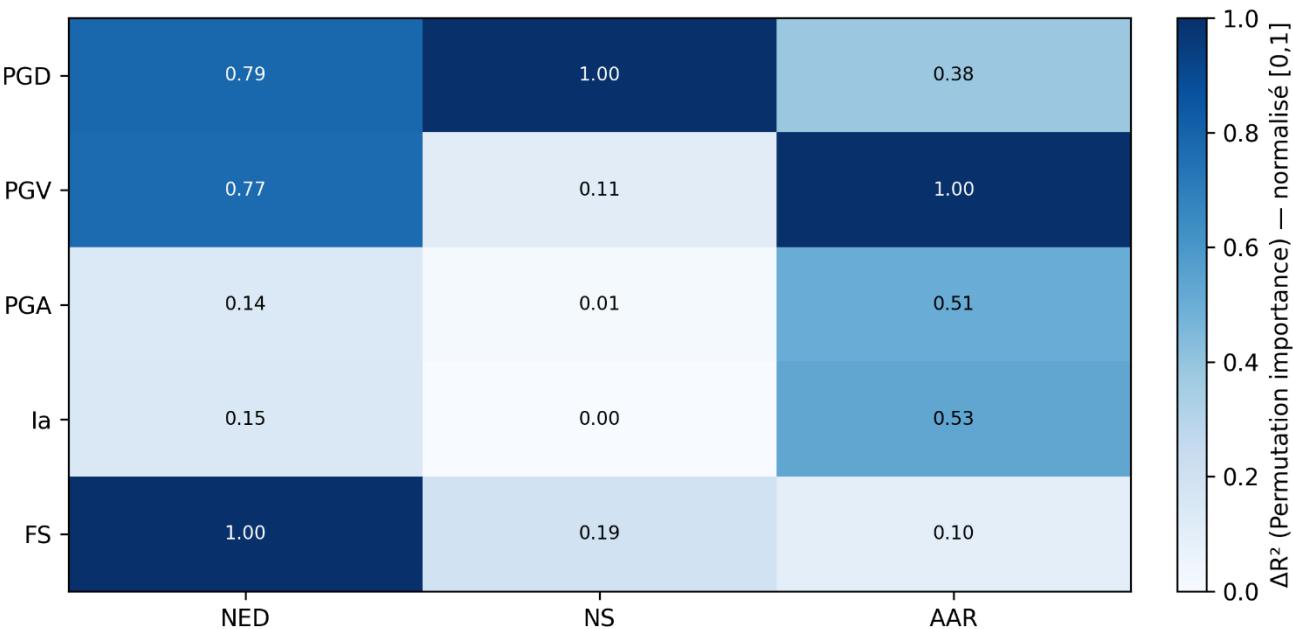


Figure 23 : Evaluation de l'importance du modèle GB sur l'SFSI normalisées [0-1]

	Rapport de stage	Page 57/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

6.3 Application du deuxième modèle « GB »

6.3.1 Ajustement des paramètres du modèle GB

Le Gradient Boosting combine séquentiellement des arbres faibles pour corriger les erreurs du modèle précédent. Alors, pour appliquer le meilleur modèle GB il faut choisir les meilleurs hyperparamètres c'est pour cela par validation croisé on va tester : learning_rate, n_estimators, max_depth et subsample. Tels que chacun a un rôle :

- Learning_rate : c'est le pas de gradient (plus il est faible, plus il faut davantage d'arbres) autrement dit ce paramètre améliore souvent la généralisation.
- N_estimators : c'est le nombre d'arbres qui contrôle la complexité globale du modèle.
- Max_depth : c'est la profondeur d'un arbre (contrôle le niveau d'interactions captées)
- Subsample : fraction de l'échantillon tirée à chaque itération

Après la combinaison de plusieurs valeurs de ces paramètres on a obtenu à la fin les meilleurs pour chaque prédiction mentionnée dans le tableau ci-dessous :

Tableau 15 : Ajustement des hyperparamètres du modèle GB

Cible	Learning_rate	N_estimators	Max_depth	Subsample
NED	0.1	500	4	0.8
NS	0.2	300	4	1
AAR	0.05	500	4	1

Dans les modèles ensemblistes il faut prendre en considération la liaison des hyperparamètres : plus le max_depth augmente → plus la variance augmente → ce qui aboutit à un surapprentissage. En outre plus le subsample est < 1 → plus la variance diminue → ce qui aboutit à un sous-apprentissage. En plus, les Learning_rate augmente → plus le nombre d'arbre N_estimator diminue (et l'inverse) de même avec Max_depth et N_estimator

6.3.2 Comparaison des mesures métrique du modèle GB

Après l'ajustement des hyperparamètres du modèle GB, maintenant c'est le temps de faire des prédictions des trois variables cibles NED, NS et AAR en utilisant les cinq variables d'entrées PGD, PGV, PGA, la et FS. En se basant sur les résultats on remarque que :

	Rapport de stage	Page 58/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

- NS est très bien modélisé ($R^2 \approx 0.95$), mieux encore qu'avec KNN et la preuve c'est les valeurs de RMSE et MAE sont plus basses qu'avec le modèle KNN.
- NED progresse fortement ($R^2 \approx 0.84$), signe que GB capte des interactions non linéaires que le modèle KNN n'a pas pu les capturer.
- AAR s'améliore nettement ($R^2 \approx 0.82$). Le MAPE reste élevé (cible en log et valeurs proches de 0 ⇒ pourcentages instables) ; Mais on constate que la prédiction d'AAR avec des modèles ensemblistes correspond mieux que KNN

Tableau 16 : Evaluation de la performance du modèle GB sur l'SFSI

Model	Target	R2_test	RMSE_test	MAE_test	MAPE_%
GB	NED	0.8441	0.1873	0.1415	3.0063
GB	NS	0.95	0.068	0.0487	1.8089
GB	AAR	0.8213	0.0033	0.0016	79.4369

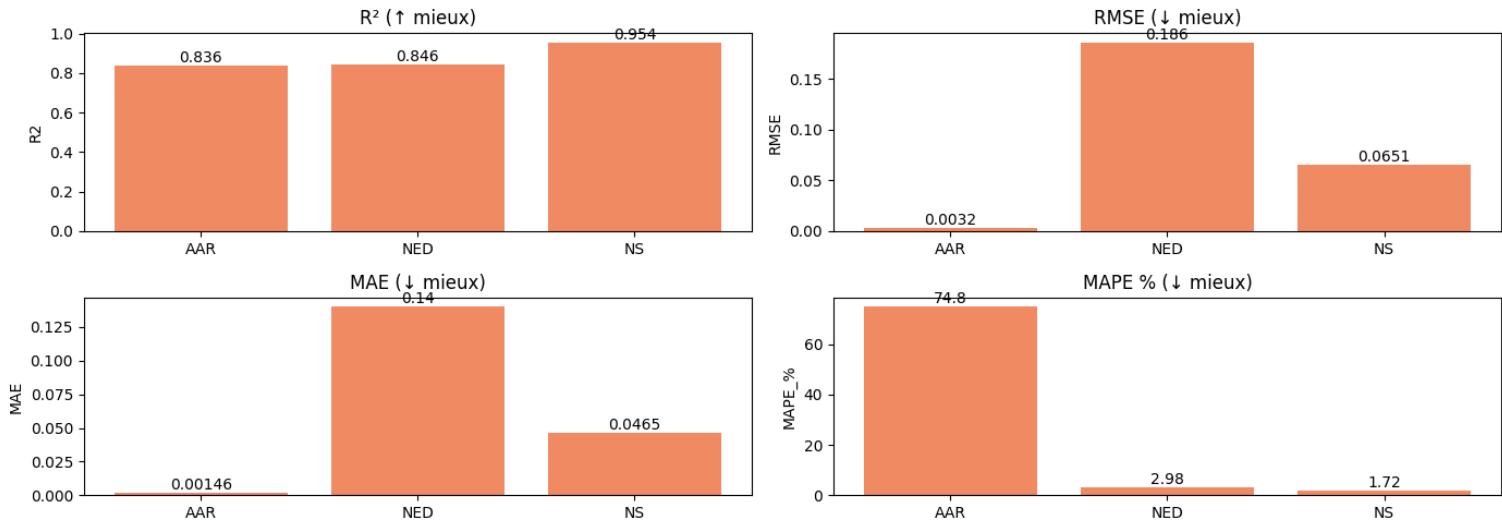


Figure 24 : Evaluation de la performance du modèle GB sur l'SFSI

	Rapport de stage	Page 59/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

6.3.3 Evaluation d'importance des variables selon le modèle GB

Pour comprendre l'interaction du système sol-fondation-structure il faut étudier l'importance des variables d'entrées sur les variables de sorties. Voici un tableau au-dessus qui montre les valeurs d'importance selon chaque prédiction du modèle GB

- NED : on a FS (0.408) → PGD (0.261) ≈ PGA (0.234) → PGV (0.162) → la (0.128). ce qui montre que la robustesse structurelle (FS) demeure le facteur dominant et les amplitudes PGD/PGA complètent l'explication.
- NS : on a PGD (1.452) → FS → (0.209) → PGV (0.052) → PGA (0.021) → la (0.008). ce qui montre que NS est quasi entièrement piloté par PGD, résultat très cohérent physiquement ; FS n'apporte qu'une correction secondaire.
- AAR : on a PGA (1.259) → PGV (0.357) → PGD (0.085) → la (0.049) → FS (0.013). ce qui montre l'accélération de pic (PGA) et la vitesse de pic (PGV) dominent le rapport d'amplification alors que la et PGD jouent un rôle second ; FS est marginal.

Donc, la qualité de prédiction selon R^2 est classé comme le suivant : AAR >> NED > NS telle que AAR suit PGA de façon quasi monotone, NED bruité qui dépend de la structure physique FS + (PGD/PGV/PGA) alors que NS est un rapport plus sensible qui dépend fortement du cinématique PGD.

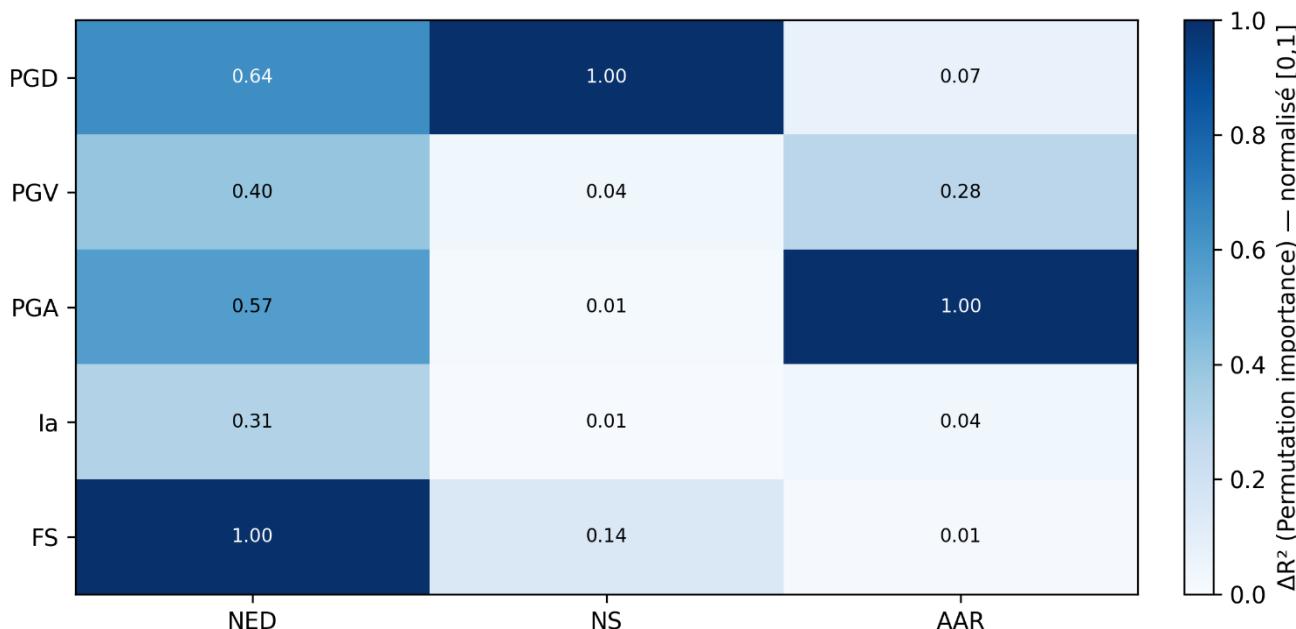


Figure 25 : Evaluation de l'importance du modèle GB sur l'SFSI normalisées [0-1]

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

		Rapport de stage	Page 60/66
	Réf. : Aix-Marseille Université		
	Date : 22/07/2025	Indice :	
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES			

6.4 Application di troisième modèle « XGB »

6.4.1 Ajustement des paramètres

Le modèle XGBoost est une combinaison **d'arbres** régularisé, avec un contrôle fin de la complexité. Alors pour entraîner ce modèle et avoir des meilleures prédictions il faut ajuster les paramètres par GridSearchCV. Vu que c'est le même modèle que GB mais avec des régularisations et des contrainte alors on a les mêmes paramètres que GB (Learning rate, N_estimator, Subsample et Max_depth) en plus de ça, on a les paramètres de régularisation

Colsample_bytree : fraction de colonnes par arbre utilisé pour réduire les corrélations entre arbres.

Min_child_weight : poids minimal dans un nœud (empêche la création de feuilles trop petites ce qui empêche aussi le surapprentissage).

Reg_lambda : pénalisation **L2** des feuilles (régularisation) aussi pour diminuer la corrélation entre les variables d'entrées.

Tableau 17 : Ajustement des hyperparamètres du modèle XGB

Cible	Learning_rate	N_estimators	Max_depth	Subsample	Colsample_bytree	Min_child_weight	reg lambda
NED	0.1	800	4	1	0.7	5	1
NS	0.1	800	3	1	1	5	1
AAR	0.02	800	6	0.9	0.9	1	1

6.4.2 Comparaison des mesures métrique du modèle XGB

Après l'ajustement des hyperparamètres du modèle XGB, maintenant c'est le temps de faire des prédictions des trois variables cibles NED, NS et AAR en utilisant les cinq variables d'entrées PGD, PGV, PGA, Ia et FS. En se basant sur les résultats on remarque que :

- NS : meilleur score ($R^2 \approx 0.954$), parmi les deux modèles KNN et GB.
- NED : très bon ($R^2 \approx 0.846$), légèrement supérieur au GB.
- AAR : nettement meilleur que KNN et supérieur à GB ($R^2 \approx 0.836$). Le MAPE reste élevé, car la cible est en log avec des valeurs proches de zéro

Tableau 18 : Evaluation de la performance du modèle XGB sur l'SFSI

Model	Cible	R ²	RMSE	MAE	MAPE %
XGB	NED	0.846	0.186	0.14	2.98
XGB	NS	0.954	0.065	0.047	1.72
XGB	AAR	0.836	0.0032	0.0015	74.85

	Rapport de stage	Page 61/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES

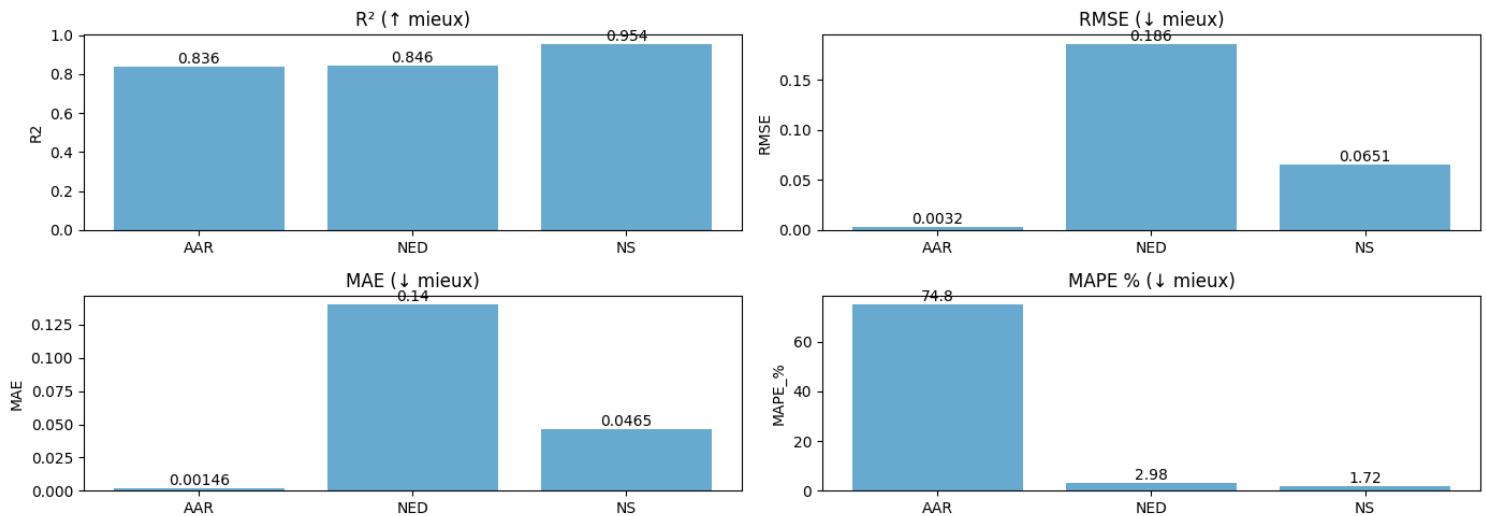


Figure 26 : Evaluation de la performance du modèle XGB sur l'SFSI

6.4.3 Evaluation d'importance des variables selon le modèle XGB

Pour comprendre l'interaction du système sol-fondation-structure il faut étudier l'importance des variables d'entrées sur les variables de sorties. Voici un tableau au-dessus qui montre les valeurs d'importance selon chaque prédiction du modèle XGB

- NED: on a FS (0.436) \rightarrow PGD (0.247) \rightarrow PGA (0.202) \rightarrow PGV (0.158) \rightarrow la (0.118). Ce qui montre la capacité structurelle (FS) domine, puis l'amplitude (PGD/PGA/PGV). Cohérent avec GB.
- NS: on a PGD (1.666) \rightarrow FS (0.212) \rightarrow PGV (0.025) \rightarrow PGA (0.016) \rightarrow la (0.007). Ce qui montre NS est quasi exclusivement expliqué par PGD ; arbres peu profonds (d=3) suffisent.
- AAR: on a PGA (1.056) \rightarrow PGV (0.325) \rightarrow la (0.072) \rightarrow PGD (0.040) \rightarrow FS (0.012). Ce qui montre l'accélération et la vitesse de pic gouvernent le rapport d'amplification ; la apporte un complément, PGD/FS sont secondaires.

XGBoost délivre les meilleures performances globales (R^2) sur les trois cibles, devant GB et KNN. La hiérarchie des variables est stable et physiquement cohérente : NS \leftarrow PGD (très dominant) ; NED \leftarrow FS + amplitudes (PGD/PGA/PGV) ; AAR \leftarrow PGA + PGV (+ la). Ces conclusions guident la priorisation des capteurs et des paramètres de conception : mesurer avec précision PGD pour NS, caractériser FS et les amplitudes pour NED, et assurer la qualité de PGA/PGV (et la) pour AAR.

	Rapport de stage	Page 62/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

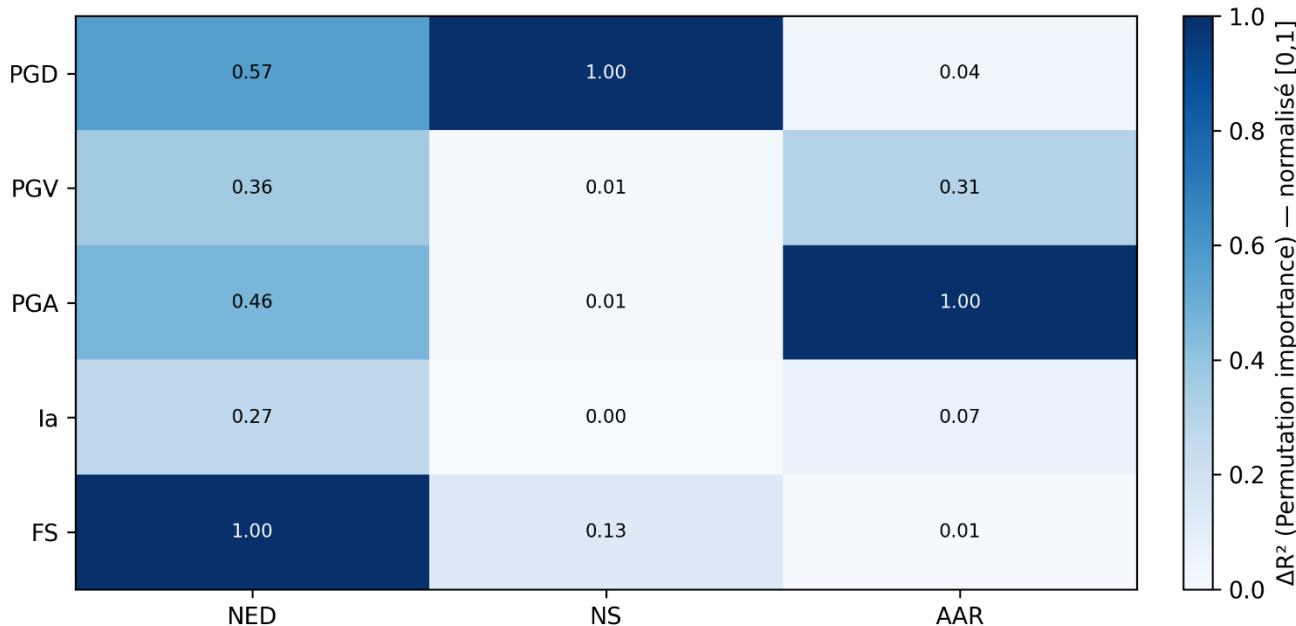


Figure 27 : Evaluation de l'importance du modèle XGB sur l'SFSI normalisées [0-1]

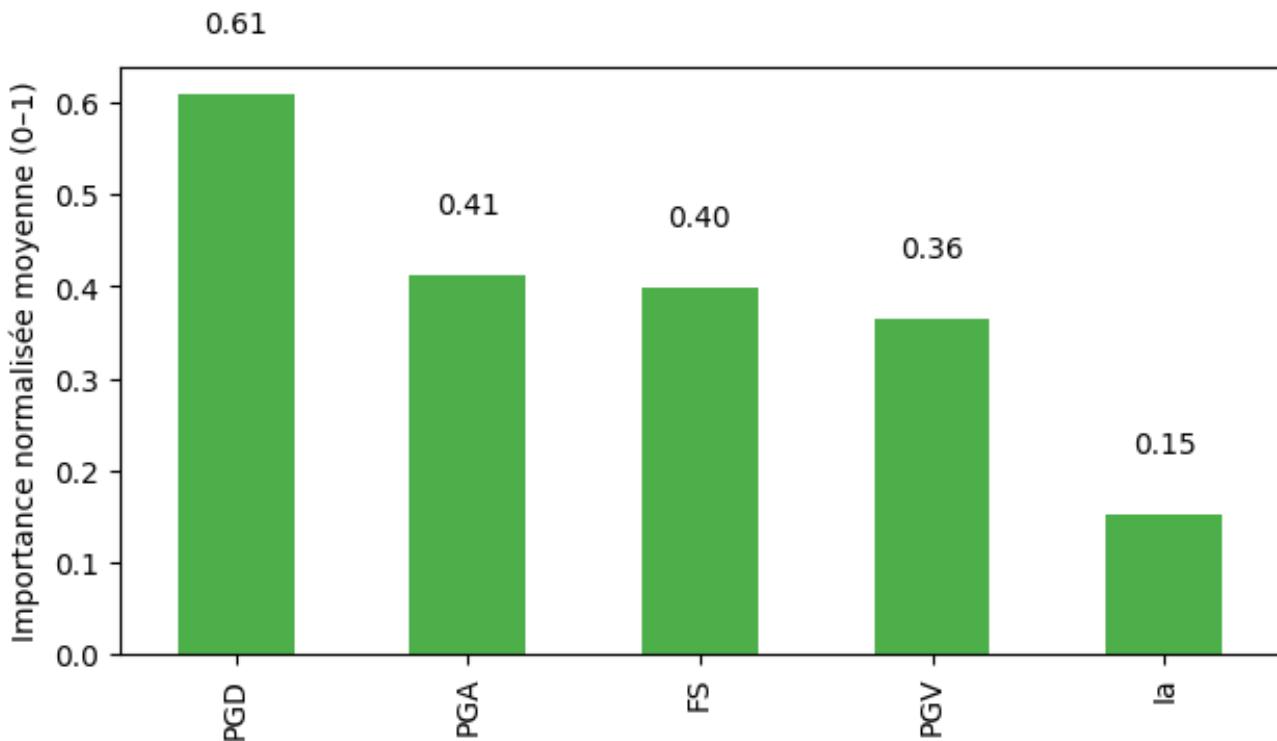


Figure 28 : Classement globale des features selon la moyenne d'importance normalisée

	Rapport de stage	Page 63/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

7 CONCLUSION ET PERSPECTIVES

Ce travail a permis d'investiguer le comportement sismique du système sol–fondation–structure (SFSI) en s'appuyant sur une base de données numérique issue de simulations paramétriques intégrant la variabilité des propriétés du sol, des caractéristiques de la fondation, de la superstructure et du chargement sismique. Un prétraitement incluant la transformation logarithmique des variables asymétriques et l'élimination des valeurs aberrantes via la méthode d'écart interquartile (IQR), a permis de stabiliser les distributions et d'améliorer la généralisation du jeu de données. La base finale, composée de 7200 observations filtrées, a servi à l'entraînement de trois familles de modèles d'apprentissage automatique présentant des niveaux de performance distincte.

Les résultats montrent que XGBoost constitue le meilleur compromis entre biais, variance, robustesse et capacité d'interprétation, par permutation importance. Le GradientBoost présente des performances proches avec une mise en œuvre relativement simple, tandis que KNN s'avère utile comme modèle de référence pour diagnostiquer des structures locales et pas complexes et non-linéaires. Concernant les cibles, le tassement normalisé (NS) est plus prédictible ($R^2 \approx 0.95$ avec GB et XGB), suivie par l'énergie dissipée normalisée (NED) ($R^2 \approx 0.84\text{--}0.85$), tandis que le rapport d'amplification en accélération (AAR), plus complexe, reste bien modélisé par GB et XGB ($R^2 \approx 0.82\text{--}0.84$).

La hiérarchie des facteurs déterminants apparaît stable entre les modèles et cohérente physiquement :

- NS est gouverné quasi exclusivement par le déplacement maximal (PGD), le facteur de sécurité (FS) jouant un rôle secondaire, tandis que le PGV et le PGA restent marginaux.
- NED dépend principalement de FS et les facteurs de PGD, PGV et PGA jouent un rôle mineur.
- AAR est conditionnée par le PGA et le PGV et la contribution des facteurs PGD et FS sont secondaires.

Ces résultats soulignent la nécessité : (i) de renforcer la qualité de l'estimation de PGD pour la prédiction de NS, (ii) bien caractériser le FS et les facteurs PGD, PGV et PGA pour NED, et (iii) de fiabiliser les facteurs PGA et PGV pour la prédiction de AAR.

Le classement global des variables selon leur importance normalisée moyenne est le suivant PGA (0.61), PGV (0.41), FS (0.4), PGD (0.36) et la (0.15). Ces enseignements orientent à la fois les stratégies de mesure et le choix de conception de nouvelles campagnes expérimentaux pour enrichir la base de données expérimentale (voir Figure 28).

Pour aller au-delà des résultats de ce travail, quelques perspectives peuvent être envisagées :

- Application d'autres modèles d'apprentissage automatique notamment SVM (Support Vector Machine) ou ANN (Artificial Neural Network). Ces modèles plus complexes peuvent offrir un moyen pour modéliser la complexité du phénomène de l'ISFS.

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

	Rapport de stage	Page 64/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES

- Enrichir la base de données expérimentale. Ce travail offre à la fois un moyen pour planifier des essais en variant les paramètres physiques du système. Ainsi, permettant d'observer les phénomène souhaités (i.e. tassemement, glissement, soulèvement, rotation). Par ailleurs, met en évidence le besoin de compléter la base de données expérimentale pour aboutir à un modèle fiable et prédictif.

	Rapport de stage	Page 65/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :

RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES

8 REFERENCES

- Anastasopoulos, I., & Kontoroupi, Th. (2014). Simplified approximate method for analysis of rocking systems accounting for soil inelasticity and foundation uplifting. *Soil Dynamics and Earthquake Engineering*, 56, 28-43. <https://doi.org/10.1016/j.soildyn.2013.10.001>
- Antonellis, G., Gavras, A. G., Panagiotou, M., Kutter, B. L., Guerrini, G., Sander, A. C., & Fox, P. J. (2015). Shake Table Test of Large-Scale Bridge Columns Supported on Rocking Shallow Foundations. *Journal of Geotechnical and Geoenvironmental Engineering*, 141(5), 04015009. [https://doi.org/10.1061/\(ASCE\)GT.1943-5606.0001284](https://doi.org/10.1061/(ASCE)GT.1943-5606.0001284)
- Boulanger, 2000- BNWF - Modélisation du code. (s.d.).
- Deng, L., & Kutter, B. L. (2012). Characterization of rocking shallow foundations using centrifuge model tests. *Earthquake Engineering & Structural Dynamics*, 41(5), 1043-1060. <https://doi.org/10.1002/eqe.1181>
- Drosos, V. A., Gerolymos, N., & Gazetas, G. (2012). Constitutive model for soil amplification of ground shaking: Parameter calibration, comparisons, validation. *Soil Dynamics and Earthquake Engineering*, 42, 255-274. <https://doi.org/10.1016/j.soildyn.2012.06.003>
- Gajan, S. (2021). Application of machine learning algorithms to performance prediction of rocking shallow foundations during earthquake loading. *Soil Dynamics and Earthquake Engineering*, 151, 106965. <https://doi.org/10.1016/j.soildyn.2021.106965>
- Gajan, S., & Kutter, B. L. (2008). Capacity, Settlement, and Energy Dissipation of Shallow Footings Subjected to Rocking. *Journal of Geotechnical and Geoenvironmental Engineering*, 134(8), 1129-1141. [https://doi.org/10.1061/\(ASCE\)1090-0241\(2008\)134:8\(1129\)](https://doi.org/10.1061/(ASCE)1090-0241(2008)134:8(1129))
- Gajan, S., Raychowdhury, P., Hutchinson, T. C., Kutter, B. L., & Stewart, J. P. (2010a). Application and Validation of Practical Tools for Nonlinear Soil-Foundation Interaction Analysis. *Earthquake Spectra*, 26(1), 111-129. <https://doi.org/10.1193/1.3263242>
- Gajan, S., Raychowdhury, P., Hutchinson, T. C., Kutter, B. L., & Stewart, J. P. (2010b). Application and Validation of Practical Tools for Nonlinear Soil-Foundation Interaction Analysis. *Earthquake Spectra*, 26(1), 111-129. <https://doi.org/10.1193/1.3263242>
- Gavras, A. G., Kutter, B. L., Hakhamaneshi, M., Gajan, S., Tsatsis, A., Sharma, K., ... Gazetas, G. (2020). Database of rocking shallow foundation performance: Dynamic shaking. *Earthquake Spectra*, 36(2), 960-982. <https://doi.org/10.1177/8755293019891727>
- Geomechanics- Bornes des paramètres. (s.d.).
- Hakhamaneshi, M., Kutter, B. L., Gavras, A. G., Gajan, S., Tsatsis, A., Liu, W., ... Gazetas, G. (2020). Database of rocking shallow foundation performance: Slow-cyclic and monotonic loading. *Earthquake Spectra*, 36(3), 1585-1606. <https://doi.org/10.1177/8755293020906564>
- Khosravi, M., Boulanger, R. W., Wilson, D. W., Olgun, C. G., Tamura, S., & Wang, Y. (2017). Dynamic centrifuge tests of structures with shallow foundations on soft clay reinforced by soil-cement grids. *Soils and Foundations*, 57(4), 501-513. <https://doi.org/10.1016/j.sandf.2017.06.002>
- Lee, J., & Salgado, R. (2000). Analysis of calibration chamber plate load tests.
- Millen, M., & Hare, J. (2025). Development of provisions for simplified design of rocking foundations. *Bulletin of the New Zealand Society for Earthquake Engineering*, 58(1), 40-51. <https://doi.org/10.5459/bnzsee.1659>

DIFFUSION LIMITÉE

Document propriété du CEA – Reproduction et diffusion externes au CEA soumises à l'autorisation de l'émetteur

	Rapport de stage	Page 66/66
	Réf. : Aix-Marseille Université	
	Date : 22/07/2025	Indice :
RAPPORT DE STAGE – EVALUATION DE L'IMPACT DES PARAMETRES PHYSIQUES DANS LA REPONSE DE BASCULEMENT DES STRUCTURES PAR APPRENTISSAGE AUTOMATIQUES		

- Pecker, A., Paolucci, R., Chatzigogos, C., Correia, A. A., & Figini, R. (2014). The role of non-linear dynamic soil-foundation interaction on the seismic response of structures. *Bulletin of Earthquake Engineering*, 12(3), 1157-1176. <https://doi.org/10.1007/s10518-013-9457-0>
- Raychowdhury, P., & Hutchinson, T. C. (2009). Performance evaluation of a nonlinear Winkler-based shallow foundation model using centrifuge test results. *Earthquake Engineering & Structural Dynamics*, 38(5), 679-698. <https://doi.org/10.1002/eqe.902>
- Raychowdhury, P., & Ugalde, J. (s.d.). 242 PUBLICATIONS 5,677 CITATIONS SEE PROFILE.
- Shirato, M., Kouno, T., Asai, R., Nakatani, S., Fukui, J., & Paolucci, R. (2008). Large-Scale Experiments on Nonlinear Behavior of Shallow Foundations Subjected to Strong Earthquakes. *Soils and Foundations*, 48(5), 673-692. <https://doi.org/10.3208/sandf.48.673>
- Sollogoub, P., Combescure, D., Queval, J.-C., & Chaudat, T. (s.d.). IN PLANE SEISMIC BEHAVIOUR OF SEVERAL 1/3RD SCALED R/C BEARING WALLS - TESTING AND INTERPRETATION USING NON-LINEAR NUMERICAL MODELLING.
- Soundararajan, S., & Gajan, S. (2020). Effects of Rocking Coefficient on Seismic Energy Dissipation, Permanent Settlement, and Self-Centering Characteristics of Rocking Shallow Foundations. *Dans Geo-Congress 2020* (pp. 123-131). Minneapolis, Minnesota: American Society of Civil Engineers. <https://doi.org/10.1061/9780784482810.014>
- Tsatsis, A., & Anastasopoulos, I. (2015). Performance of Rocking Systems on Shallow Improved Sand: Shaking Table Testing. *Frontiers in Built Environment*, 1. <https://doi.org/10.3389/fbuil.2015.00009>
- Zhao, J., Chen, H., & Zhou, Z. (2023). Shaking table test study of nuclear power plant model considering soil-structure interaction effect. *Soil Dynamics and Earthquake Engineering*, 168, 107859. <https://doi.org/10.1016/j.soildyn.2023.107859>