

# PREVISION EN TEMPS REEL DES TASSEMENTS INDUITS PAR LE CREUSEMENT AU TUNNELIER

## REAL-TIME PREDICTION OF TUNNELING INDUCED SETTLEMENTS

Tatiana RICHA<sup>1,2</sup>, Timothée LAMARE<sup>1</sup>, Jean-Michel PEREIRA<sup>2</sup>, Lina-María GUAYACAN-CARRILLO<sup>2</sup>, Gilles CHAPRON<sup>1</sup>, Francis LANQUETTE<sup>3</sup>

<sup>1</sup> Setec Terrasol, Paris, France

<sup>2</sup> Navier, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France

<sup>3</sup> Setec Tpi, Paris, France

**RÉSUMÉ** – Cet article présente deux approches novatrices pour la prévision en temps réel des tassements lors du creusement de tunnels. La première utilise la méthode Random Forest pour prédire le tassement maximal dans les zones non excavées, tandis que la seconde ajuste la prédiction de la courbe de tassement avec le creusement en utilisant des réseaux de neurones LSTM, adaptés aux séries temporelles. Les données proviennent de deux lignes du Grand Paris Express. Une analyse des avantages et limitations de chaque méthode est fournie.

**ABSTRACT** – This article presents two innovative approaches for real-time prediction of settlements induced by tunnel excavation. The first approach uses Random Forest to predict the maximum settlement in unexcavated zones, while the second continuously adjusts the prediction of settlement progression with excavation using LSTM, an algorithm suited for time series. Data is sourced from two lines of the Grand Paris Express. An analysis of the advantages and limitations of each method is provided.

### 1. Introduction

Le creusement de tunnels dans des zones urbaines denses et des contextes géologiques complexes expose les constructions existantes à des risques significatifs, en raison notamment des tassements en surface pouvant survenir tout au long de l'avancement du front du tunnel. L'objectif de cet article est de présenter deux approches novatrices pour la prévision en temps réel de ces tassements.

La première approche vise à prévoir le tassement maximal observé à la fin du creusement, noté  $s_{max}$  (Figure 1), (Peck, 1969)). L'objectif est de prévoir  $s_{max}$  dans les zones encore non excavées, permettant ainsi de détecter à l'avance si ces déformations dépassent un seuil critique. La seconde approche vise à ajuster en continu la prévision de la courbe de progression du tassement observée tout au long du processus de creusement (Figure 1, (Attewell et al., 1986)).

Cette étude se base sur les données provenant de deux lignes du projet de métro du Grand Paris Express (GPE), dont la constitution de la base de données est détaillée dans (Richa, 2023 ; Richa et al., 2022). Il est important de noter que ces deux lignes partagent un contexte géologique similaire et sont toutes les deux excavées à l'aide d'un tunnelier à pression de terre.

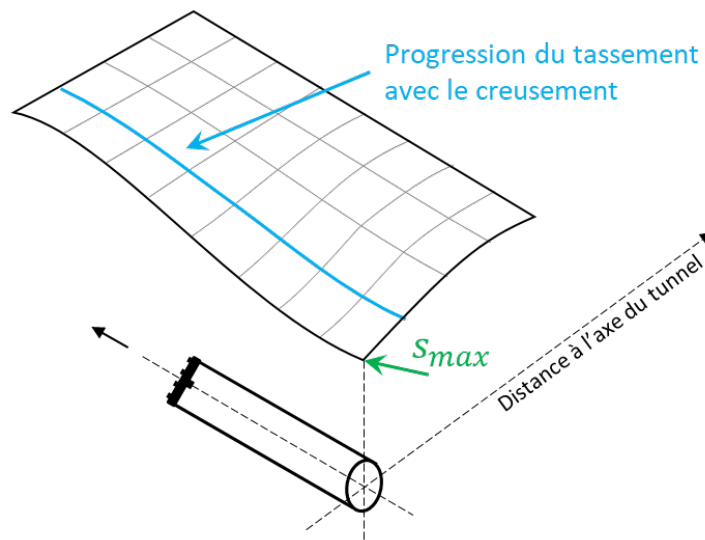


Figure 1. Déformations induites par le creusement au tunnelier

## 2. Prédiction du tassement maximal

### 2.1. Principe

Les tassements induits par le creusement des tunnels se stabilisent à une valeur maximale ( $s_{max}$ , Figure 1) après une centaine de mètres (Richa, 2023). Pour prévoir  $s_{max}$  en temps réel sur les zones non excavées, il est essentiel d'adopter une division non aléatoire des données d'entraînement et de test. Cette approche simule fidèlement une situation réelle : la position du tunnelier est fixée, les données à l'arrière du front du tunnel sont utilisées pour l'entraînement, tandis que les données à l'avant du front sont réservées au test (i.e. prévision en temps réel). En outre, cette méthode permet de déterminer la distance à l'avant du front (zone théoriquement non excavée) sur laquelle la prévision reste fiable. Cette distance est désignée par la suite comme « distance d'extrapolation spatiale ». Pour cette étude, nous avons opté pour l'utilisation d'un algorithme ensembliste, Random Forest (RF), reconnu pour sa performance dans la prédiction de données tabulaires.

### 2.2. Random Forest

Random Forest est un algorithme d'apprentissage automatique supervisé introduit en 2001 (Breiman, 2001). Cet algorithme appartient aux algorithmes de la famille des méthodes ensemblistes (ensemble methods) puisqu'il combine les prédictions de plusieurs arbres de décision (Decision Trees) pour obtenir des résultats plus précis et plus robustes. Chaque arbre de décision individuel est entraîné sur des sous-ensembles aléatoires des données d'entraînement, grâce à la technique du Bagging. Cette approche réduit le risque de surapprentissage et améliore la capacité de généralisation du modèle.

### 2.3. Méthodologie et Résultats

La méthodologie adoptée vise à répondre à trois problématiques majeures. La première consiste à déterminer la distance de creusement nécessaire pour obtenir les informations nécessaires pour un modèle performant. La seconde est d'évaluer la distance d'extrapolation spatiale jusqu'à laquelle la prédiction reste fiable. La troisième vise à évaluer l'impact de l'enrichissement de la base de données d'entraînement par des données provenant d'un tronçon différent de celui sur lequel le modèle est utilisé.

Pour évaluer la qualité des prédictions, nous comparons les valeurs prédites aux valeurs réelles. Si la différence entre ces valeurs est autour de plus ou moins 1 mm, seuil considéré comme acceptable par les experts métiers, les prédictions sont considérées comme réussies.

Le premier test consiste à prendre en compte pour l'entraînement 1000 m de creusement et à faire la prédiction du  $s_{max}$  sur les 2000 m en avant du front (zone à creuser). On obtient ainsi 173 observations pour l'apprentissage et 403 pour le test. La Figure 2 (a) montre que les modèles ont prédit avec succès sur une distance d'environ 300 m (points en rouge et orange), à partir de laquelle les valeurs de  $s_{max}$  commencent à être sur-estimées jusqu'à environ 1000 m. Au-delà de 1000 m, le modèle devient inadéquat et sous-estime les tassements de 2 à 4 mm. Ainsi, pour ce jeu des données, 1000 m de données de creusement pour l'apprentissage sont suffisants pour prévoir  $s_{max}$  correctement sur une distance d'environ 300 m.

Dans le deuxième test, l'algorithme est entraîné sur toutes les mesures obtenues sur 5 km de la ligne 14 Sud du GPE (L14S), soit 795 observations sur 5051 m, ainsi que sur les 500 premiers mètres d'un tronçon de la ligne 15 Sud-Ouest du GPE (TR1), soit 62 observations. Ensuite, le modèle obtenu est utilisé pour prédire  $s_{max}$  sur 500 m du TR1, soit 111 observations. Les résultats présentés dans la Figure 2 (b) montrent que l'entraînement sur les données de la L14S permet d'obtenir de bons résultats jusqu'à 300 m du début de la zone de test, mais détériore les résultats pour les points à partir de 350 m.

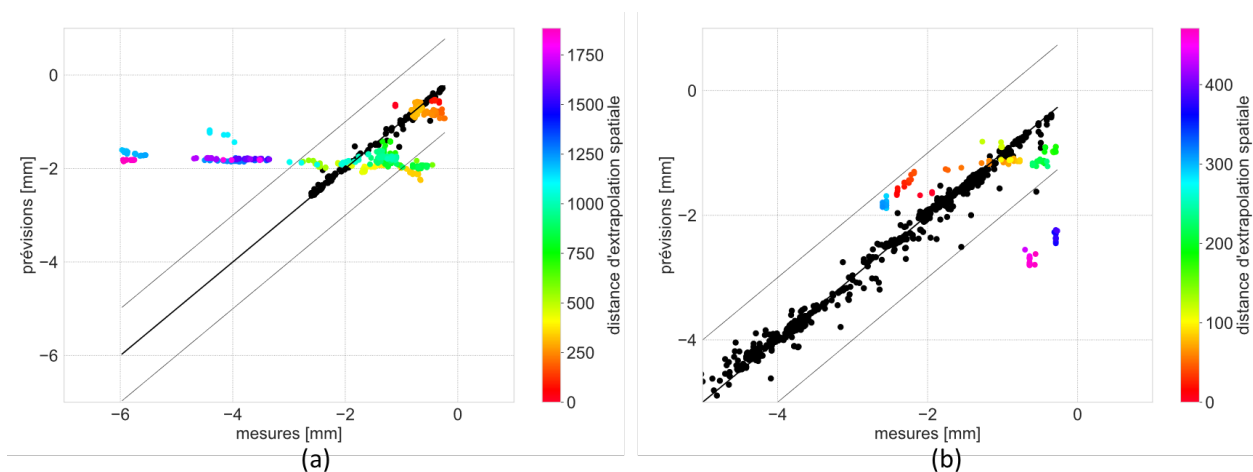


Figure 2. Entraînement sur 1000m d'un tronçon et prévision sur les 2000m suivants (à gauche, (a)) et entraînement sur une autre ligne ainsi que les 500m d'un tronçon et prévision sur les 500 m suivants (à droite, (b))

En conclusion de ces deux exercices, on estime que l'entraînement d'un algorithme de Machine Learning, spécifiquement les Random Forest, à partir de 1000 m de creusement est possible pour prédire  $s_{max}$  sur les 300 m suivants, sur ce cas d'étude. Afin de réduire le temps d'attente, par exemple en n'entraînant le modèle qu'à partir de 500 m de creusement, il est possible d'enrichir la base de données avec des données provenant d'une autre ligne présentant des contextes géologiques et des paramètres de pilotage de tunnelier similaires.

### 3. Prédiction du tassement longitudinal

#### 3.1. Principe

En un point fixe de l'espace, le tassement en surface évolue en fonction de la progression de l'excavation. Pour décrire ces déformations, on prend comme référentiel un point fixe de l'espace et on observe l'évolution des tassements en fonction de la position du front du tunnel (Figure 1). La seconde approche vise donc à ajuster en continu la prévision de la courbe de progression du tassement observé tout au long du processus de creusement. Cette courbe peut être assimilée à une série temporelle, ce qui nécessite l'utilisation d'algorithmes spécifiques, notamment les LSTM (Long Short-Term Memory), appartenant à la famille des réseaux de neurones.

#### 3.2. Long Short-Term Memory Network

Les LSTM (Long-Short Term Memory) sont une variante des réseaux de neurones récurrents (RNN), spécifiquement conçus pour traiter des données temporelles ou séquentielles. À la différence des réseaux de neurones classiques, les RNN intègrent des données d'entrée conventionnelles, des informations des éléments antérieurs dans la séquence, ainsi que les prédictions générées pour ces éléments (Figure 3). Cette caractéristique distinctive confère aux RNN des performances accrues dans le traitement de données séquentielles ou temporelles.

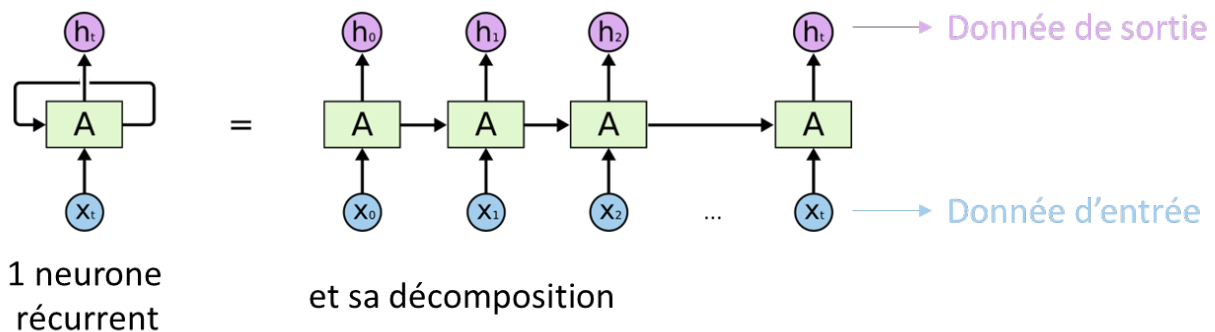


Figure 3. Composition d'un RNN (adapté de Olah, 2015)

#### 3.3. Méthodologie et Résultats

Les spécificités d'une approche basée sur les séries temporelles nécessitent des ajustements des données disponibles. En effet, les RNN requièrent que les données d'entrée soient structurées selon un format particulier, présenté sous forme d'une matrice tri-dimensionnelle (Figure 4). Dans le cadre de cette étude, la première dimension de cette matrice correspond aux capteurs, représentant ainsi chaque position spatiale traversée par le tunnelier au fil du temps. La deuxième dimension capture la séquence de mesures effectuées par chaque capteur, tandis que la troisième dimension englobe les différentes caractéristiques (*features*, i.e. donnée d'entrée).

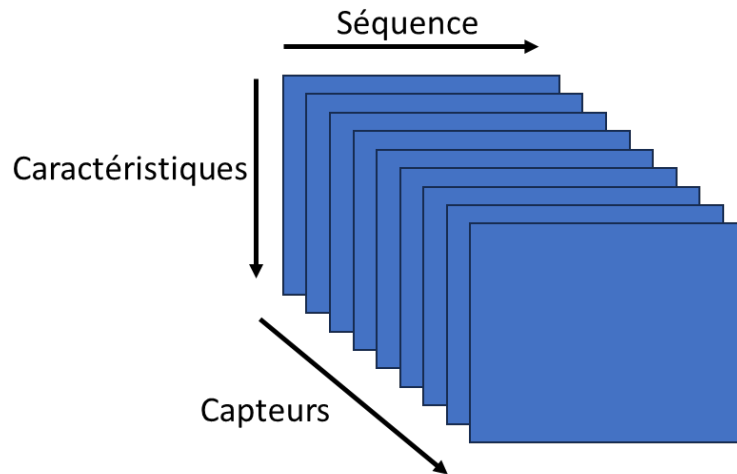


Figure 4. Format des données d'entrée pour un réseau de neurones LSTM

Pour évaluer la qualité des prédictions, on utilise les métriques d'erreur  $R^2$  (coefficient de détermination) et le RMSE (Racine carrée de l'erreur quadratique moyenne). Un RMSE proche de 0 est considéré comme optimal, tandis que des valeurs de  $R^2$  proches de 1 indiquent un meilleur ajustement du modèle aux données. Ainsi, un modèle de prédiction est jugé excellent lorsque le RMSE est proche de 0 et le  $R^2$  de 1.

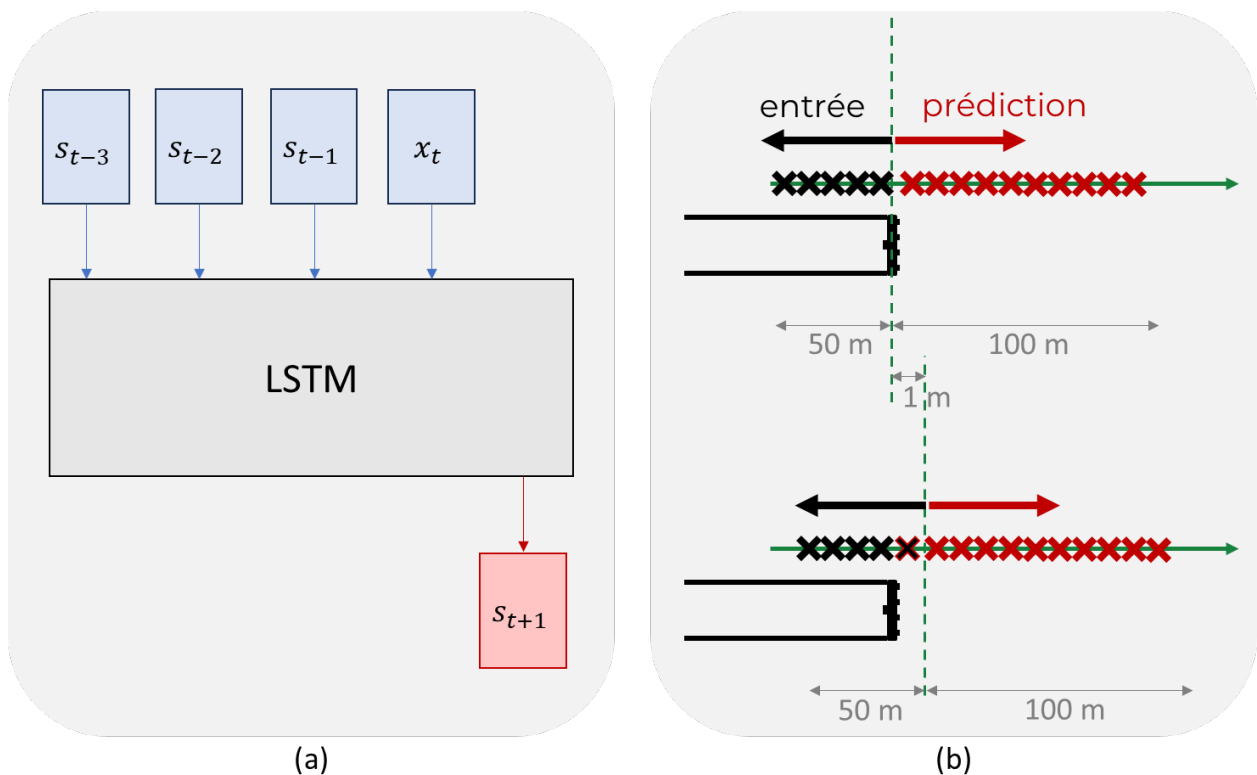


Figure 5. Schématisation des deux approches utilisant les LSTM

Deux approches sont adoptées pour cette partie. Dans la première approche, les séquences de valeurs pour chaque caractéristique sont fournies en entrée, et en sortie, nous obtenons la séquence prévue des tassements pour chaque capteur. Les tassements des éléments précédents,  $s_{t-1}$ ,  $s_{t-2}$  et  $s_{t-3}$  sont également ajoutés en tant que

caractéristiques (Figure 5(a)). En moyenne, on obtient un  $R^2$  de 0.97 et un RMSE de 0.28. Ces résultats particulièrement performants peuvent cacher un phénomène de surajustement du modèle sur les mesures réelles (*overfitting*), comme le montre un exemple dans la Figure 6 où les prédictions sont très proches des mesures réelles, confirmant ainsi la présence de sur-ajustement.

Cette approche semble peu réaliste, car elle se limite à prédire uniquement le prochain tassement. Notre objectif est plutôt de développer un modèle capable de prévoir les tassements avec une certaine anticipation. C'est pourquoi, au lieu d'améliorer le premier modèle, nous optons pour une autre approche que nous appelons « Préviction par tranche à partir des prévisions précédentes ».

L'objectif de cette deuxième approche est de prédire, au moment où le tunnelier passe sous un capteur ( $d_{front} = 0$ ), tous les tassements jusqu'à 250 m en amont du front ( $d_{front}$  entre 0 et 250 m). Cette prédiction est représentée par la courbe en rouge sur la Figure 7. Pour atteindre cet objectif, nous effectuons des prédictions à chaque mètre (que nous appelons « tranche », représentées par les courbes en orange sur la Figure 7), en utilisant à chaque fois les données 50 m en arrière pour prédire les tassements des 100 m suivants. Ainsi, les prédictions d'une tranche servent de données d'entrée pour les tranches suivantes. Enfin, nous calculons la moyenne de toutes les tranches pour obtenir la prédiction finale (courbe rouge).

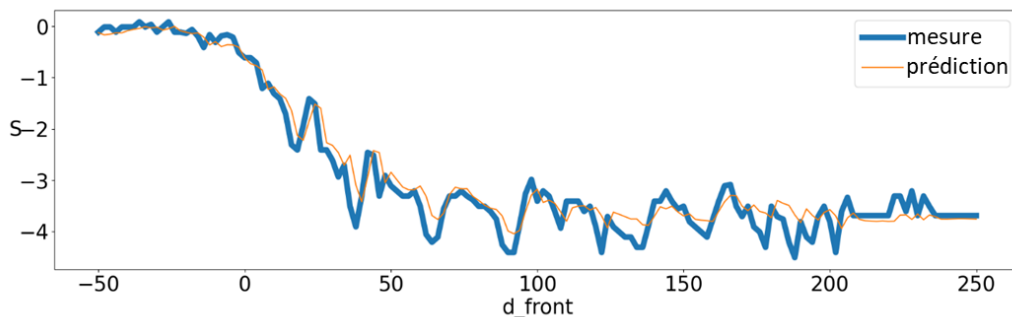


Figure 6. Exemple de résultat de prévision des tassements avec utilisation de  $s_{t-1}, s_{t-2}, s_{t-3}$

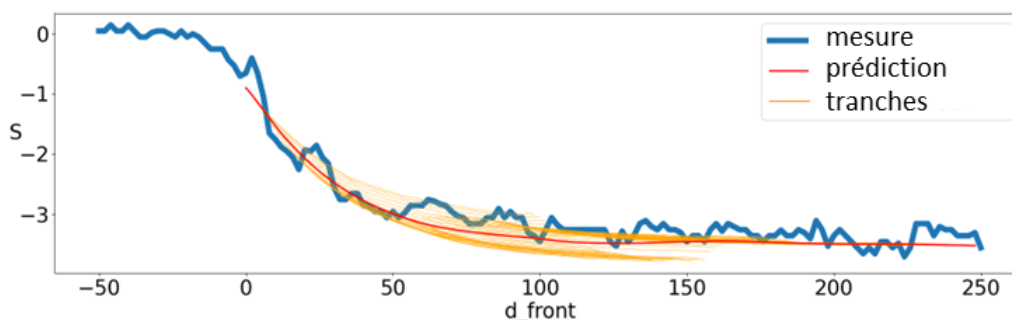


Figure 7. Exemple de résultat de prévision des tassements par tranche à partir de prévision

#### 4. Discussion et Conclusions

Cet article compare deux approches novatrices de prévision en temps réel des tassements induits par le creusement au tunnelier. La prévision du tassement maximal offre une solution « simple », interprétable et rapide, qui permet à l'ingénieur d'avoir une estimation du tassement maximal et d'agir en temps réel si besoin. En effet, l'entraînement des Random Forest ne demande que quelques minutes. La solution à l'aide de LSTM, qui cherche à estimer l'évolution des tassements avec le temps en fonction de la distance au front, est plus complexe à mettre en œuvre et peut nécessiter plusieurs heures de calcul. Il est donc important de bien définir à l'avance l'objectif du projet afin de faire le choix d'utilisation, tout en priorisant le choix du modèle le plus simple répondant aux besoins de renforcement de la sécurité et de l'efficacité des opérations de creusement de tunnels.

Actuellement, dans le cadre d'un projet réel, la prédiction du  $s_{max}$  dans les zones non creusées à l'avant du front de taille suffit pour répondre aux besoins du chantier, à savoir anticiper les risques pour intervenir à temps. Cependant, la prévision du tassement au fur et à mesure du creusement du tunnel offre un niveau de détail supplémentaire qui pourrait être nécessaire, en fonction des enjeux de chaque projet. Dans cette perspective, ces deux approches peuvent être combinées. Dans les travaux à venir, la prédiction de  $s_{max}$  à l'aide des Random Forest, sera utilisée comme donnée d'entrée pour entraîner les LSTM. Cette approche combinée permettrait d'obtenir des résultats plus précis.

Il est important de noter que la performance des algorithmes de Machine Learning est limitée aux zones où le contexte géologique et les paramètres de pilotage du tunnelier sont comparables. En effet, à ce jour, ces algorithmes ne sont pas capables d'extrapolation. Ainsi, ces outils ont pour vocation d'être un outil complémentaire pour l'aide à la décision des ingénieurs, et non pas de les remplacer.

#### 5. Références bibliographiques

- Attewell, P. B., Yeates, J., & Selby, A. R. (1986). Soil Movements Induced by Tunnelling and their Effects on Pipelines and Structures. *Blackie, Glasgow*.
- Breiman, L. (2001). Random Forests. *Machine Learning*, 45, 5–32. <https://doi.org/10.1023/A:1010933404324>
- Olah, C. (2015). *Understanding LSTM Networks*. <http://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- Peck, R. B. (1969). Deep excavations and tunneling in soft ground (State of the art report). *7th International Conference on Soil Mechanics and Foundation Engineering*, 225–290. <https://www.issmge.org/publications/online-library>
- Richa, T. (2023). *Prévision des tassements induits par le creusement au tunnelier : construction d'une base de données et apprentissage automatique* [Ecole des Ponts ParisTech]. <https://pastel.hal.science/tel-04221306>
- Richa, T., Pereira, J.-M., Chapron, G., & Guayacán-Carrillo, L. M. (2022). Constitution d'une base de données des mesures obtenues lors du creusement de deux tunnels du Grand Paris Express. *JNGG*, 1093–1100.