

CONSTITUTION D'UNE BASE DE DONNÉES DES MESURES OBTENUES LORS DU CREUSEMENT DE DEUX TUNNELS DU GRAND PARIS EXPRESS

CONSTITUTION OF A DATABASE OF MEASUREMENTS OBTAINED DURING THE EXCAVATION OF TWO TUNNELS OF THE GRAND PARIS EXPRESS

Tatiana RICHA¹, Jean-Michel PEREIRA², Gilles CHAPRON¹, Lina-María GUAYACAN-CARRILLO²

¹ Terrasol Setec, Paris, France

² Navier, Ecole des Ponts, Univ Gustave Eiffel, CNRS, Marne-la-Vallée, France

RÉSUMÉ – L'instrumentation des zones de creusement des tunnels permet de surveiller les tassements induits en surface. A l'aide d'approches statistiques ou de méthodes d'apprentissage machine, le flux des données collectées pourrait servir à la prédiction des déformations à l'avant du front du tunnel. Cependant, le recours à ces méthodes n'est possible que si les données sont propres et d'une qualité acceptable. Cet article propose un retour d'expérience sur la préparation de ces données pour deux lignes du Grand Paris Express.

ABSTRACT – The instrumentation of tunneling areas allows to monitor the settlements induced on the surface. Using statistical approaches or machine learning methods, the flow of collected data can be used to predict deformations before the tunnel face. However, these methods are only reliable if the data are clean and of acceptable quality. This study provides feedback on the processing of data collected from 2 lines of the Grand Paris Express.

1. Introduction

Le suivi de l'impact sur les avoisinants est un enjeu majeur du creusement des tunnels au tunnelier en environnement urbain. L'instrumentation des Zones d'Influence Géotechnique est essentielle pour surveiller les tassements en surface et valider leur conformité avec les valeurs issues de la conception. L'instrumentation du tunnelier permet par ailleurs d'avoir accès aux paramètres de pilotage de la machine qui exercent une influence sur l'amplitude des tassements observés. Un chantier de tunnel instrumenté génère donc une quantité de données considérable. Aujourd'hui, ces mesures servent à donner l'alerte lorsque des anomalies sont observées, afin de prendre des actions correctives en un délai acceptable, et éventuellement effectuer des recalages ponctuels par méthode numérique. Il serait bien plus utile que ce flux de données important puisse servir à recalculer efficacement les prédictions de déformations en surface à l'avant du passage du front. L'intelligence artificielle, en particulier l'apprentissage machine, apporte des approches intéressantes qui méritent d'être développées (Zhang et al., 2020). Cependant, les méthodes d'analyse de données basées sur l'apprentissage machine sont sensibles à la qualité des données. Ainsi, la présence de bruit et de valeurs aberrantes dans le jeu de données peut conduire à des interprétations incorrectes. Il est donc impératif d'effectuer un travail de nettoyage et d'organisation des données préalablement à leur analyse.

Cette communication s'inscrit dans le cadre d'un projet de thèse portant sur la réduction des incertitudes sur le comportement des sols excavés au tunnelier par l'analyse en continu des données collectées. Celle-ci présente un retour d'expérience sur le traitement des données obtenues lors du creusement de deux lignes du Grand Paris Express. Le travail concerne l'extraction des données depuis diverses plateformes, leur nettoyage et la constitution d'une base de données robuste. Les difficultés rencontrées sont exprimées

dans le but d'inciter l'industrie à suivre certaines recommandations qui facilitent et accélèrent le processus de traitement des données.

2. Contexte

2.1. Description du projet

Le projet du Grand Paris Express (GPE) est le plus grand projet urbain d'Europe avec 200 km de lignes automatiques de métro et 68 nouvelles gares. Le GPE est un projet essentiellement souterrain qui relie les principaux lieux de vie et d'activité en banlieue sans passer par le centre de Paris.

Dans cette étude, deux lignes du GPE sont étudiées : la ligne 15 sud-ouest (L15SO) et le lot GC02 du prolongement du côté sud de la ligne 14 (L14SGC02) (Figure 1). La longueur des tracés est de 7.785 km et 4.578 km respectivement. Ces tracés forment environ 12 km de lignes très bien auscultées et excavées au tunnelier à pression de terre.

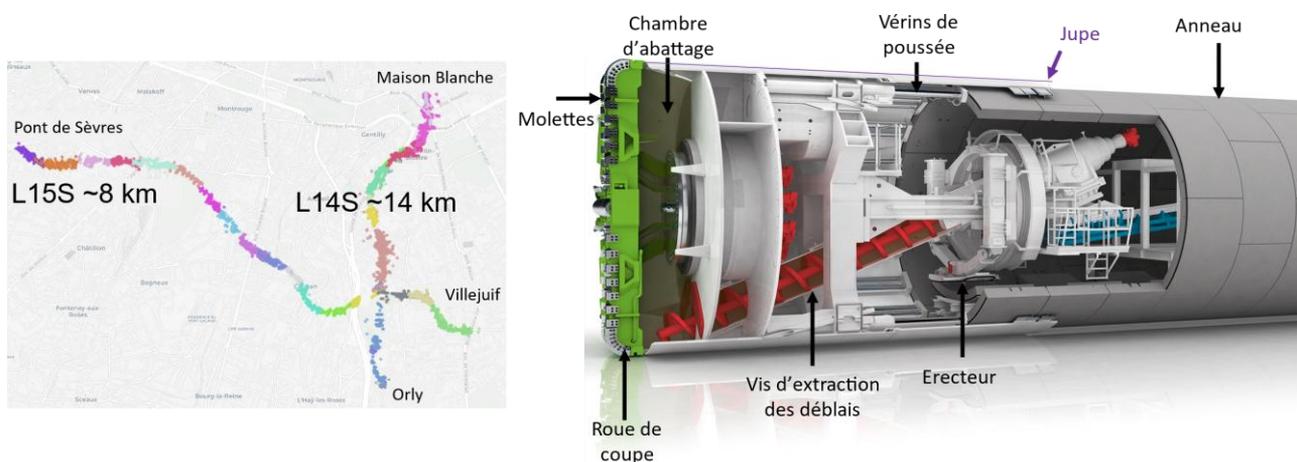


Figure 1 - Tracés des lignes L14SGC02 et L15SO (à gauche) et vue écorchée d'un tunnelier à pression de terre (à droite)

2.2. Tunnelier à pression de terre

Le tunnelier ou TBM (pour Tunnel Boring Machine en anglais) a pour rôles l'excavation du sol, le maintien de la stabilité du front, la pose du soutènement et l'injection de mortier dans les vides annulaires.

La L14SG02 a été excavée par un unique tunnelier alors que la L15SO a été excavée par deux tunneliers. Ces trois TBM sont des tunneliers à pression de terre (Figure 1), type privilégié pour les terrains comportant des argiles, des limons, des craies tendres, des marnes, etc. (AFTES, 2019). Les caractéristiques des machines sont données dans le **Erreur ! Source du renvoi introuvable.**

Tableau 1 - Paramètres des tunneliers utilisés sur les lignes L14SGC02 et L15SO

	Diamètre d'excavation [m]	Diamètre fin jupe [m]	Longueur bouclier [m]	Diamètre intrados du revêtement [m]	Epaisseur du revêtement [m]	Nombre de voussoirs par anneau	Longueur anneau [m]
L15SO	9.87	9.8	13.5	8.7	0.4	7	2
L14S-GC02	8.83	8.76	13.74	7.75	0.35	7	2 - 1.5

2.3. Contexte géologique

Les formations géotechniques majoritairement rencontrées au front le long du tracé (**Erreur ! Source du renvoi introuvable.**) de la L14SGC02 sont du Calcaire Grossier (CG), du Marno-Calcaire de Pantin (MP), des Marnes d'Argenteuil (MA), des Argiles Vertes (AV) et des Argiles Plastiques (AP) (Coblard et al., 2021). Pour la L15SO, le front du tunnel traverse des Argiles Plastiques, Calcaires Grossiers, Masses et Marnes du Gypse (MG), Marnes Supra-gypseuses d'Argenteuil (MSGa) et Marnes de Meudon (MM) (Le Bissonnais et al., 2017). L'épaisseur de la couverture au-dessus des tunnels varie, selon la ligne, entre 15 m et 50 m.

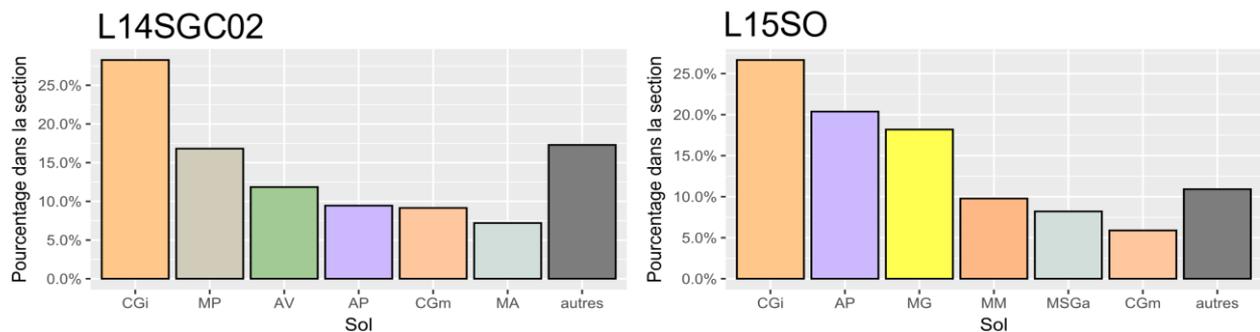


Figure 2 - Principales formations rencontrées au front

2.4. Auscultations et Suivi du creusement

Le développement des méthodes d'auscultation et la capacité actuelle de stockage permettent de collecter une quantité importante de données, qui sont alors mises à disposition des différents acteurs du projet. Aujourd'hui, les auscultations servent essentiellement à visualiser les paramètres mesurés et à contrôler que tout se passe comme prévu. Notre ambition est de profiter de ces auscultations pour prévenir les éventuels dommages à l'avant du front, en recalant les prédictions rapidement et facilement.

Les paramètres de pilotage du tunnelier pendant le creusement, notamment la pression au front, la pression d'injection du mortier, etc., sont mesurés par des capteurs et inspectés par les pilotes de la machine pour assurer son bon fonctionnement et éviter les blocages ou les casses de matériel.

Les déformations en surface sont également mesurées par l'intermédiaires de prismes placés sur les voiries ou les bâtiments (à différents niveaux), et ciblés par des théodolites toutes les 30 minutes (en règle générale).

Nous présentons dans la suite la méthodologie suivie depuis l'extraction de ces données jusqu'à leur organisation dans une base de données performante.

3. Méthodologie et Résultats

Les données disponibles sur un projet de tunnel peuvent être divisées en 3 catégories : les mesures d'auscultations en surface, les paramètres de pilotage du tunnelier, et les données géologiques et géotechniques. L'information brute à laquelle nous avons accès n'est pas toujours facile à extraire, assez disparate et comprend des erreurs. Elle nécessite donc un travail important de nettoyage et d'organisation.

3.1. Extraction des données

Les données relatives au GPE sont mises à disposition sur diverses plateformes pendant la durée de vie du projet. Préalablement à toute étude sur ces données, un travail d'extraction des informations brutes est nécessaire.

Le premier défi est l'acquisition des données dans les délais puisqu'une fois l'exécution du projet achevée, les plateformes ne sont plus disponibles. Il est donc impératif d'extraire les informations et d'en faire une sauvegarde. Le second défi est la diversité des plateformes fournissant les 3 catégories de paramètres nécessaires à nos études. L'accès à ces plateformes se fait par différentes méthodes (application, site web, API), et l'extraction de masse n'est souvent pas évidente (outils non adaptés). Le défi à l'heure actuelle est donc de faire avec, de s'adapter à ces disparités, et d'arriver à créer des outils permettant cet export de façon fiable et automatisée.



Figure 3 - Cabine de pilotage du tunnelier (à gauche) et un exemple d'image de rapport informant sur les paramètres de pilotage du tunnelier (à droite)

Les paramètres relatifs au tunnelier sont affichés sous forme d'images dans la cabine de pilotage de la machine (Figure 3). L'entreprise n'a mis à disposition de la maîtrise d'œuvre pour ces projets qu'une petite partie des flux de données brutes. Pour avoir accès aux paramètres manquants qui nous intéressaient, il a fallu exploiter des captures d'écrans du poste de pilotage, effectuées toutes les 30 secondes et déposées sur une plateforme web dédiée. Nous avons créé des codes (en langage R) pour extraire massivement toutes les images, en améliorer la résolution des caractères et pour finalement les transformer en texte (Figure 4). Pour notre étude, nous avons sélectionné les paramètres les plus utilisés dans la littérature pour les méthodes d'apprentissage machine, à savoir la vitesse d'avancement, le moment de la roue de coupe, la pression au front, la pression et la quantité de mortier injecté et bien d'autres (Chen et al., 2019; Suwansawat & Einstein, 2006).



Figure 4 - Méthode d'extraction des paramètres de pilotage de tunnelier à partir d'images

L'interrogation de la base de données des auscultations était possible via une application ou un site web. L'export massif des données n'étant pas possible via l'application, nous avons eu recours à une technique de type *web scraping* pour extraire plus de 144 millions de mesures associées à près de 16 000 capteurs.

Les données de sol sont de deux types : les données stratigraphiques, qui sont « numérisées » à partir des profils en long dessinés, mètre par mètre (transformation du dessin en table) (Figure 5), et des tableaux de paramètres de calcul, par plage de linéaires.

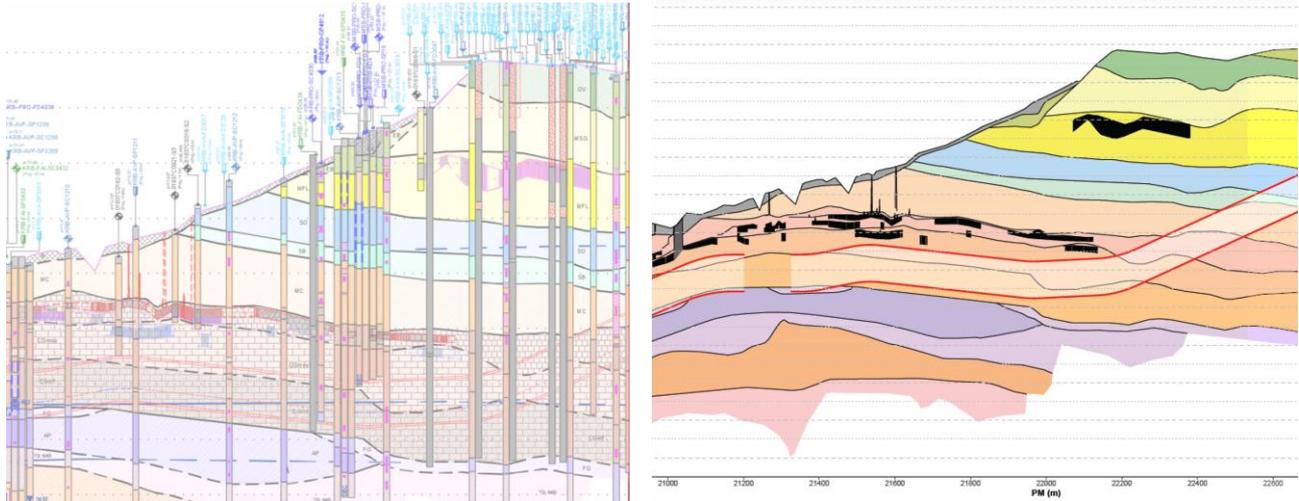


Figure 5 - Numérisation du profil stratigraphique

3.2. Nettoyage des données

Après la sauvegarde, la deuxième étape concerne le nettoyage des données. L'analyse de données basée sur l'apprentissage machine ou des approches statistiques est sensible à la qualité des données. Par conséquent, la présence de bruit et de valeurs aberrantes dans les ensembles de données peut conduire à des interprétations incorrectes (Chu et al., 2016).

Le nettoyage des données consiste à (i) gérer les valeurs manquantes, (ii) supprimer les données dupliquées pour diminuer le volume des données et (iii) détecter les données aberrantes (Krishnan et al., 2016). Le nettoyage des données améliore la qualité des données et apporte ainsi la complétude, la précision et la confiance dans l'ensemble des données ainsi que la fiabilité des résultats d'analyse (Klein & Lehner, 2009).

Concernant les paramètres de pilotage des tunneliers, des corrections ponctuelles sont nécessaires, comme l'ajout de dates importantes manquantes ou la correction des valeurs qui ont été mal interprétées par la reconnaissance de texte. Par exemple, le caractère « 5 » est assez souvent interprété en tant que « 8 » ou « 9 ». Vu la quantité de données et la qualité des images, ces corrections sont faites de façon semi-automatisée.

S'agissant des paramètres d'auscultation, seuls les déplacements en surface nous intéressent dans la présente étude. Nous avons donc sélectionné les données issues du suivi des cibles placées en voirie ou sur les bâtiments. Par ailleurs, plusieurs nettoyages ont été effectués :

(a) Nettoyage du nom des types de cibles pour avoir une nomenclature unifiée. En effet, il y a plusieurs types de capteurs : cibles topographiques, cibles virtuelles, prismes etc. La présence de majuscules, d'espaces, du pluriel et des abréviations sont sources d'erreur et parfois de duplication des données.

(b) Elimination des cibles sans coordonnées spatiales puisque les enregistrements sont alors inutiles. Les cibles concernées par ce défaut représentent moins de 0.01% des cibles considérées dans la présente étude.

(c) Elimination des données dupliquées. Plusieurs sources de duplication coexistent : la plateforme source, l'extraction massive, les types de cibles, les enregistrements dupliqués d'une même cible, etc.

(d) Eliminations des cibles qui n'ont pas de mesures (cibles défectueuses).

Une fois le nettoyage terminé, le jeu de données passe de 15 859 capteurs à 8 689 cibles mesurant les déplacements en surface (sur voirie ou bâtiment).

3.3. Base de données

A la suite de la collecte et du nettoyage des données, on se retrouve avec des fichiers de différents types (csv, json, etc.), sans relation les uns avec les autres, stockés dans différents dossiers et trop lourds pour être chargés en mémoire pour l'exploitation. Travailler avec des données stockées de cette façon pour appliquer des algorithmes d'apprentissage machine n'est pas possible à notre échelle. Il a donc été décidé de créer une base de données relationnelle. Même si la quantité des données à notre disposition est encore loin de l'échelle du Big Data, l'intégration des informations disponibles dans une base de données et le respect des 5V (Volume, Vitesse, Variété, Valeur et Véracité) est un prérequis indispensable pour assurer l'efficacité de la chaîne de valeur.

La mise à disposition des données dans une base unique garantit à des utilisateurs spécifiques une disponibilité et un accès facilité, rapide et sécurisé à un grand volume de données variées (auscultations, paramètres de sol, paramètre de pilotage du tunnelier).

L'intégration ou la fiabilité des données est garantie par le choix de PostgreSQL qui est un système de base de données relationnelle libre de droit (*open access*) et gérant les données spatialisées. Son architecture relationnelle permet de mettre en place les relations entre les différentes entités (tables). Les données sont ainsi bien organisées et sans redondance. Le stockage est donc minimisé et son interrogation par requêtes SQL est rapide et économe en mémoire vive.

Pour partager et sécuriser les données, la base est stockée sur des services *Cloud*.

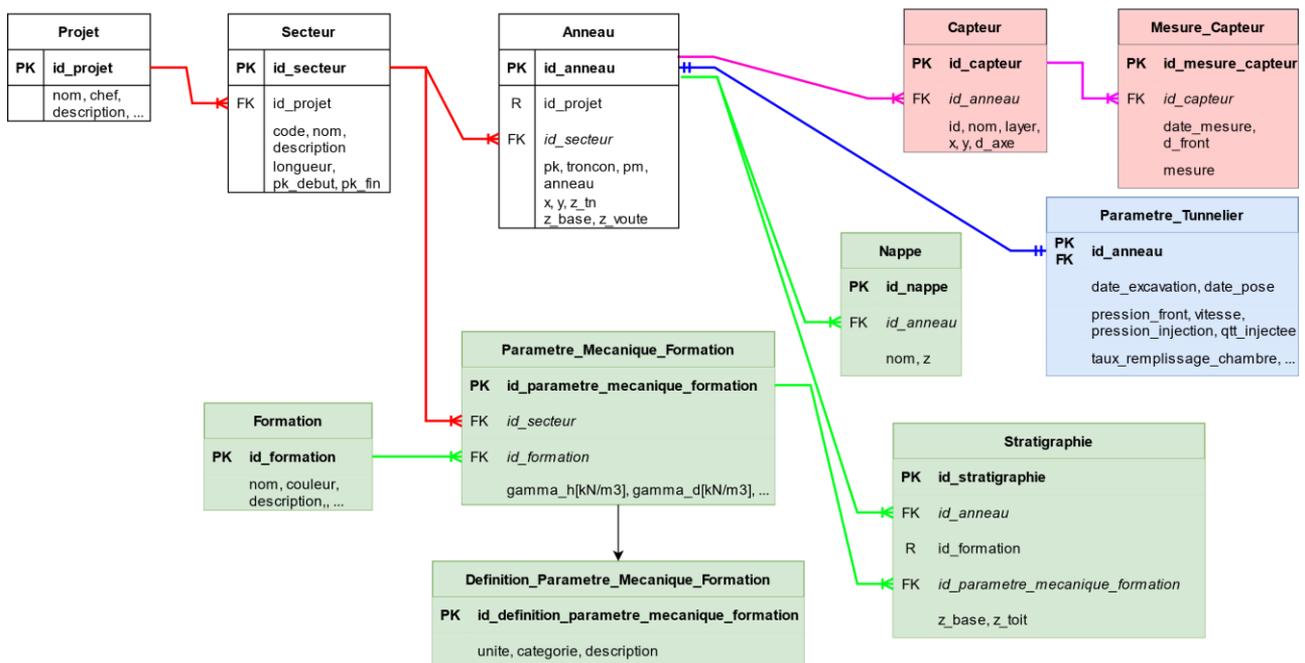


Figure 6 - Diagramme d'entité-relation de la base de données

La création de bases de données d'excavation de tunnels a déjà été abordée par quelques auteurs, notamment Marinos et al. (2013) qui ont regroupé les données de 62 tunnels de l'Egnatia Highway (Grèce). La structure de la base de données choisie dans notre étude est similaire à celle présentée par Marinos et al. (2013) : la base est formée par 3 catégories principales et 1 catégorie supplémentaire, à savoir (a) les mesures d'auscultations en surface, (b) les paramètres de creusement du tunnelier, (c) les données géologiques et géotechniques et enfin (d) les tables d'assistance qui contiennent les abréviations, symboles, unités, etc. La Figure 6 présente la structure de notre base de données (seuls les paramètres principaux de chaque table sont indiqués). La table « parametre_tunnelier » contient par exemple en réalité environ 40 paramètres.

L'entité « projet » décrit le projet de tunnel étudié, le responsable, la localisation, etc. Un projet est divisé en plusieurs « secteurs » qui définissent la géologie et la géotechnique du terrain (« parametre_mecanique_formation »).

Au centre se trouve l'entité « anneau ». Cette dernière représente la position tous les mètres, informant ainsi sur les coordonnées en surface et celles en profondeur pour le tunnelier. L'entité « anneau » est l'entité « parent » à 4 autres entités qui sont les 3 catégories principales : les auscultations (« capteur » et « mesure_capteur »), les « parametre_tunnelier » et les données géologiques et géotechniques (« stratigraphie » et « nappe »).

A noter que les entités d'assistance sont « formation » (symbole, couleur, ère de chaque formation géologique) et « definition_parametre_mecanique_formation » (symbole, unité et catégorie de chacun des paramètres mécaniques de sol).

Notre base est maintenant constituée et prête à être interrogée pour analyser les données collectées et appliquer des algorithmes visant à prédire les déplacements en amont du front du tunnelier. Une première étape dans l'analyse consistera en la qualification de la qualité des mesures issues de chaque capteur. La Figure 7 illustre cet aspect en montrant le tassement mesuré par deux capteurs, l'un qualifié de "bon" et l'autre de mauvais.

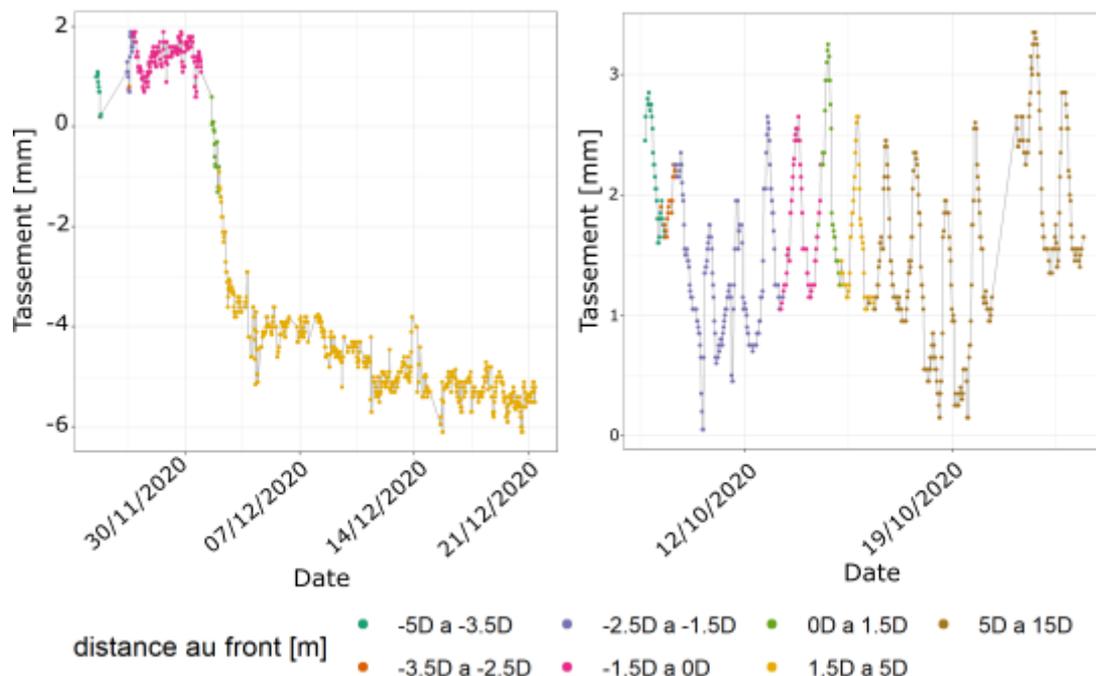


Figure 7 - Illustration de la qualité des données collectées par deux capteurs différents le long du tracé de la L15SO.

4. Conclusions et perspectives

Cette étude décrit le processus d'extraction, de nettoyage et de création d'une base de données rassemblant les données issues du creusement de deux lignes de métro du Grand Paris Express. Les données disponibles sur un projet de tunnel sont divisées en 3 catégories : les mesures d'auscultations en surface, les paramètres de pilotage du tunnelier, et les données géologiques et géotechniques. L'information brute n'est pas toujours facile à extraire, assez disparate et comprend des erreurs.

Notre retour d'expérience est que ce travail prend un temps considérable majoritairement à cause de la multiplicité des plateformes qui hébergent les données, la disponibilité des informations dans des formats non adaptés (e.g. des images), le manque de nettoyage à la source, une nomenclature non unifiée des capteurs, etc. Ce travail est cependant indispensable si l'on veut faire parler et donner de la valeur à ces données.

L'instrumentation des chantiers de tunnel constitue un investissement majeur de la part du maître d'ouvrage. Pour en profiter au maximum, il faudrait que, dès le début du projet, les entreprises s'organisent pour pouvoir fournir aux différents acteurs du projet le flux de données de façon continue et complète dans des bases de données propres et interopérables. Ainsi, des algorithmes d'apprentissage machine pourraient être interfacés avec ces bases dans le but de mettre à jour les prédictions des déformations en surface et ainsi prévenir plutôt que constater les éventuels dégâts aux avoisinants provoqués par les creusements.

La suite de ce projet portera sur l'application de méthodes d'intelligence artificielle à ces données, et la formalisation du processus de leur intégration dans ces algorithmes.

5. Références bibliographiques

- AFTES. (2019). *État de l'art concernant les évolutions des tunneliers et de leurs capacités de 2000 à 2019 (Groupe de Travail n°4 - GT4R6F1)*.
- Chen, R. P., Zhang, P., Kang, X., Zhong, Z. Q., Liu, Y., & Wu, H. N. (2019). Prediction of maximum surface settlement caused by earth pressure balance (EPB) shield tunneling with ANN methods. *Soils and Foundations*, 59(2), 284–295.
- Chu, X., Ilyas, I. F., Krishnan, S., & Wang, J. (2016). Data cleaning: Overview and emerging challenges. *Proceedings of the ACM SIGMOD International Conference on Management of Data, 26-June-20*, 2201–2206. <https://doi.org/10.1145/2882903.2912574>
- Coblard, M., Delattre Levis Leyser, S., Bergère, A., & Nguyen The Dung, L. (2021). Excavation au tunnelier sous les carrières souterraines du Calcaire Grossier de Paris – retour d'expérience de la ligne 14 Sud. *Revue Travaux*, n°974.
- Klein, A., & Lehner, W. (2009). Representing data quality in sensor data streaming environments. *Journal of Data and Information Quality*, 1(2). <https://doi.org/10.1145/1577840.1577845>
- Krishnan, S., Haas, D., Franklin, M. J., & Wu, E. (2016). Towards reliable interactive data cleaning: A user survey and recommendations. *HILDA 2016 - Proceedings of the Workshop on Human-In-the-Loop Data Analytics*, 1, 1–5. <https://doi.org/10.1145/2939502.2939511>
- Le Bissonnais, H., Chapron, G., Veyron, P. L., Pons, G., & Fluteaux, V. (2017). La ligne 15 Sud-Ouest du Réseau de Transport du Grand Paris : Conception et enjeux géotechniques du tronçon T3C Line 15 South West of « Grand Paris Express » : design and geotechnical challenges of T3C section La ligne 15 Sud-Ouest du Réseau de Transport. *Congrès International de l'AFTES*, 1–9.
- Marinos, V., Proutzopoulos, G., Fortsakis, P., Koumoutsakos, D., Korkaris, K., & Papouli, D. (2013). "Tunnel information and analysis system": A geotechnical database for tunnels. *Geotechnical and Geological Engineering*, 31(3), 891–910.
- Suwansawat, S., & Einstein, H. H. (2006). Artificial neural networks for predicting the maximum surface settlement caused by EPB shield tunneling. *Tunnelling and Underground Space Technology*, 21(2), 133–150. <https://doi.org/10.1016/j.tust.2005.06.007>
- Zhang, P., Wu, H. N., Chen, R. P., & Chan, T. H. T. (2020). Hybrid meta-heuristic and machine learning algorithms for tunneling-induced settlement prediction: A comparative study. *Tunnelling and Underground Space Technology*, 99(May 2019), 103383.