

Planification de maintenance en technicentres SNCF: Modélisation et création d'un benchmark

Proposition de stage de 6 mois

Laboratoire d'Informatique de l'X (LIX) - ORAILIX team
Palaiseau, France

En collaboration avec la SNCF

Contexte

Les technicentres industriels de la SNCF assurent la maintenance de l'ensemble du matériel roulant. Leur fonctionnement représente chaque année plusieurs centaines de millions d'euros : optimiser leur activité est donc un enjeu économique, opérationnel et stratégique majeur pour la SNCF.

La planification de la maintenance y constitue un problème de forte complexité combinatoire : il faut coordonner les mouvements des rames sur site, respecter les pas de maintenance pour des raisons de sécurité, et prendre en compte la disponibilité et les compétences des équipes ainsi que les capacités des installations.

De nombreuses approches d'optimisation existent déjà (stages, thèses, etc.), mais elles reposent souvent sur des données partielles ou non publiées, ce qui empêche toute comparaison objective des méthodes. Il devient donc nécessaire de proposer une *modélisation explicite du problème de planification* et un *benchmark d'instances et de solutions de référence* afin d'évaluer de manière transparente l'efficacité réelle des algorithmes sur l'activité des technicentres de la SNCF.

Objectifs

L'objectif principal de ce stage est de partir d'un modèle de données existant (instances et solutions) et d'un ensemble de règles/contraintes métier pour proposer une **modélisation efficace et complète du problème** (PPC, ILP) et l'implémenter dans un solveur dédié. L'objectif est de produire des plannings de maintenance de bonne qualité sur un large ensemble d'instances, et de disposer de premières bases de comparaison.

En parallèle, le/la stagiaire contribuera à la **création d'un jeu de données de référence** à partir de données artificielles générées automatiquement. La génération de ces instances, bien qu'important pour la diffusion du benchmark, ne constituera pas l'unique focus du stage : elle vient en support de la partie modélisation, expérimentation et comparaison d'algorithmes.

Quelles données souhaite-t-on produire ?

L'objectif est de constituer un jeu de plusieurs centaines d'instances variées, validées par des experts ferroviaires, afin de disposer d'un benchmark représentatif.

Chaque instance est décrite par trois composantes principales :

- une infrastructure de site (plan de voie, installations, mouvements autorisés, disponibilités RH, etc.) ;
- une charge de maintenance (opérations, durées, fenêtres de planification, voies admissibles, rame associée, etc.) ;

- un roulement (dates et heures d'entrée/sortie des rames).

En faisant varier ces composantes (disponibilités RH, nombre de rames, niveau de saturation, etc.), on couvre un large éventail de scénarios réalistes. Chaque instance sera associée à une solution de référence, obtenue par la méthode développée lors du stage, éventuellement sous-optimale, servant de baseline pour mesurer les futures améliorations. Le générateur de données et la liste de contraintes génériques seront rendus publics afin de faciliter la réutilisation et la comparaison des approches.

Positionnement du travail dans la littérature scientifique

La recherche en optimisation s'appuie largement sur des jeux d'instances publics devenus des références, comme TSPLIB pour le TSP [3], JSPLIB pour le Job-Shop [8] ou PSPLIB pour le RCPSP [1]. Ces bibliothèques proposent des formats standardisés, des outils de validation et des référentiels de performances (meilleures solutions connues), ce qui facilite la comparaison des approches et garantit la reproductibilité des algorithmes.

Pour des problèmes combinatoires plus complexes et proches des besoins industriels, de tels benchmarks open-source restent toutefois rares. Ces problèmes combinent souvent plusieurs sous-problèmes NP-complets (routage, ordonnancement, allocation de ressources, etc.) et impliquent des données sensibles, ce qui freine la mise à disposition de corpus publics. Quelques initiatives récentes montrent néanmoins la faisabilité d'une telle démarche : Amazon (Last-Mile Routing Research Challenge [5]), Zalando Research [2] ou NVIDIA (OLIST-VRP [6]) ont ainsi proposé des benchmarks industriels ouverts et reproductibles.

Dans ce contexte, le stage s'inscrit dans l'introduction d'un benchmark industriel autour du Train Unit Shunting Problem (TUSP) et de ses extensions, et en particulier du TUSPwSS (Train Unit Shunting Problem with Services Scheduling). Ce problème intègre des aspects de routage (mouvements et conflits d'occupation des voies), d'ordonnancement (fenêtres de maintenance) et de gestion de ressources matérielles et humaines. La littérature existante (formulations MILP et heuristiques) [9, 4, 7] repose essentiellement sur des données propriétaires, sans code ni corpus d'instances standardisé, ce qui rend difficile toute comparaison reproductible des méthodes proposées.

Le projet réalisé dans le cadre de ce stage vise à combler ce manque en introduisant formellement le Train Unit Shunting and Servicing with Scheduling (TUSPwSS) comme benchmark industriel ouvert pour la communauté de recherche. Concrètement, le stage s'articulera autour de trois volets principaux :

1. **Création d'un jeu d'instances multi-technicentres**: conception d'un corpus d'instances stratifié selon le niveau de difficulté, accompagné d'un protocole d'évaluation (programme de vérification des solutions, métriques de performance) afin de permettre des comparaisons reproductibles.
2. **Modélisation et résolution du problème** : formalisation standardisée du problème et implémentation dans un solveur générique (programmation par contraintes, programmation linéaire), de manière à fournir une baseline de référence.
3. **Exploration de méthodes avancées de résolution** : étude et mise en œuvre de méta-heuristiques et/ou d'hybrides (PPC + heuristiques) afin d'améliorer la qualité des solutions obtenues ou de réduire les temps de calcul sur le benchmark proposé.

L'ensemble (jeu d'instances, protocole d'évaluation, modèles et premières méthodes de résolution) a vocation à être rendu public via un dépôt GitHub dédié, constituant un **benchmark de référence** réutilisable par la communauté. Le travail a également pour objectif une **publication scientifique** en Recherche Opérationnelle, particulièrement valorisée car elle ouvre un nouveau problème industriel complexe à la communauté. Le ou la stagiaire sera impliqué.e en tant qu'auteur.e à part entière dans cette publication.

Profil recherché

- Etudiant.e en école d'ingénieurs ou en Master
- Bonnes connaissances en algorithmie
- Bonnes connaissances de programmation en Python et/ou en Java
- Connaissances en Recherche Opérationnelle

Contact

Le stage sera encadré conjointement par l'équipe de recherche ORAILIX du Laboratoire d'Informatique de l'X (LIX) à Palaiseau et par l'équipe OPTIMA pour la SNCF. Les candidats intéressés sont invités à envoyer un CV et un court texte expliquant leurs motivations à l'une des personnes suivantes.

- Mathis LE BAIL, Doctorant (mathis.le-bail@polytechnique.edu)
- Clement ELLIKER, Doctorant (clement.elliker@polytechnique.edu)
- Sonia VANIER, HDR (sonia.vanier@polytechnique.edu)
- Arthur GODET, PhD (arthur.godet@sncf.fr)

References

- [1] PSPLIB: Project scheduling problem library (rcpsp, rcpsp/max, mrcpsp). <https://www.om-db.wi.tum.de/psplib/library.html>.
- [2] Giorgio Abelli, Maximilian Katzmann, Imran Khan, Olaf Maurer, Julius Pätzold, Paweł Pszona, and Jan-David Salchow. Joint order selection, allocation, batching and picking for large scale warehouses, 2025.
- [3] Robert Grant. Tsplib 95: A python library for working with tsplib files. <https://tsplib95.readthedocs.io/en/stable/>, 2020. Documentation, version 0.7.1. Apache-2.0 license.
- [4] Franck Kamenga, Paola Pellegrini, Joaquin Rodriguez, and Boubek Merabet. Solution algorithms for the generalized train unit shunting problem. *EURO Journal on Transportation and Logistics*, 10:100042, 2021.
- [5] Daniel Merchan, Jatin Arora, Julian Pachon, Karthik Konduri, Matthias Winkenbach, Steven Parks, and Joseph Noszek. 2021 amazon last mile routing research challenge: Data set. *Transportation Science*, 2022.

- [6] NVlabs. NVlabs olist vrp benchmark. <https://github.com/NVlabs/olist-vrp-benchmark>, 2025. GitHub repository for the OList VRP benchmark suite.
- [7] Jintang Shi, Haodong Li, and Pieter Vansteenwegen. Joint optimization of train platforming and shunting with service scheduling at a railway hub. *Computers Industrial Engineering*, 197:110544, 2024.
- [8] Tamy0612 (GitHub user). JSPLIB: Benchmark instances for the job shop scheduling problem. <https://github.com/tamy0612/JSPLIB>. GitHub repository of JSSP benchmark instances.
- [9] Roel van den Broek, Han Hoogeveen, Marjan van den Akker, and Bob Huisman. A local search algorithm for train unit shunting with service scheduling. *Transportation Science*, 56(1):141–161, January 2022.