

Ingénierie Mathématique
Master mention Mathématiques et applications

SORBONNE UNIVERSITÉ



Responsables : Cindy Guichard et Marie Postel

Laboratoire Jacques-Louis Lions

Barre 15-25 bureau 313

Téléphone : 01 44 27 54 08

Email : cindy.guichard@sorbonne-universite.fr,marie.postel@sorbonne-universite.fr

Résumés des stages en entreprise et des missions en alternance

effectués par les étudiants du Master Ingénierie Mathématique

Majeure Ingénierie Mathématique Pour l'Entreprise (IMPE)

année 2020-2021

<https://www.lpsm.paris/M2IngMath/impe/>

Secrétariat : Barre 15-25 bureau 107, 4 place Jussieu, 75005 Paris.

Sorbonne Université, BC 187, 4 place Jussieu, 75252 Paris cedex 05

Cette brochure constitue la dix-septième édition de la publication des résumés des stages en entreprise effectués par les étudiants de M2 du parcours Ingénierie Mathématique Pour l'Entreprise de Sorbonne Université.

Elle fait suite à la publication pendant 16 années consécutives d'une brochure des résumés des stages en entreprise effectués par les étudiants du DESS de Mathématiques Appliquées de l'Université Pierre et Marie Curie. Cette formation a changé de nom depuis la rentrée universitaire 2004 à la faveur de la réforme LMD (Licence-Master-Doctorat), puis en 2018 lors de l'ouverture à l'apprentissage, en partenariat avec le CFA des Sciences. Elle est maintenant, sous le nom Ingénierie mathématique pour l'Entreprise (IMPE), une des trois majeures du parcours de M2 "Ingénierie mathématique" de la mention de master Mathématiques et applications, de Sorbonne Université.

Ce document est destiné aux anciens ou futurs étudiants du Master Ingénierie mathématique, aux responsables de stages, aux maîtres d'apprentissage et aux employeurs. Nous espérons qu'il illustrera, par la diversité et la qualité des sujets abordés, le dynamisme de notre formation et la variété des carrières auxquelles elle conduit. Pour cette troisième année de la nouvelle formule où nous accueillons dans le même cursus des étudiants "classiques" et des apprentis, deux étudiants ont suivi la formation en alternance. Cette brochure rassemble les résumés de deux missions d'alternance et dix-sept stages de fin d'études terminés en 2021.

Pour conclure, signalons le groupe LinkedIn Ingénierie Mathématique -MPE (Mathématiques Pour l'Entreprise) et l'association des anciens étudiants A2MAIM.

Responsables pédagogiques dans les différentes spécialisations en 2020-2021

Probabilités - statistiques : Lokmane Abbas Turki, Vincent Lemaire

Laboratoire de probabilités et modèles aléatoires, Campus Pierre et Marie Curie, couloir 16-26 1^{er} étage

Téléphone : 01 44 27 70 47, Télécopie : 01 44 27 72 23

Autres intervenants : M. Abdallah, J.-P. Baudry, I. Kharroubi, Z. Shi, M. Thieullen (SU)

O. Bardou (GDF Suez), A. K. Fermin Rodriguez (Univ Paris Nanterre), T. Guillon et S. Tazi (RTE), F. Pons (CRI4DATA)

Analyse numérique - calcul scientifique : C. Guichard

Laboratoire Jacques-Louis Lions, Campus Pierre et Marie Curie, couloir 15-25, 3^e étage

Téléphone : 01 44 27 42 99, Télécopie : 01 44 27 72 00

Autres intervenants : L. Boudin, M. Boulakia, G. Delay, P. Frey, R. Iakymchuk, S. Labbé, M. Postel (SU)

M. Cerf (Ariane Group), P. Havé (haveneer), X. Juvigny et F.-X. Roux (ONERA)

Mécanique : P. Druault (SU)

Institut Jean Le Rond d'Alembert, Campus Pierre et Marie Curie, couloir 55-65

Téléphone : 01 44 27 37 90

Autres intervenants : M. Abbas et T. Yu (EDF), R. Chakir (IFSTTAR), J. Waeytens (IFSTTAR)

Responsable de l'apprentissage au CFA des Sciences

Nathalie Obert-Ben Taieb

nobert@cfa-sciences.fr

Téléphone : 01 44 27 75 76

Responsables de la majeure

Cindy Guichard, Marie Postel

cindy.guichard@sorbonne-universite.fr, marie.postel@sorbonne-universite.fr

Campus Pierre et Marie Curie, couloir 15-25, 3^e étage, 313

Téléphone : 01 44 27 54 08

Secrétariat du Master Ingénierie Mathématique

Francelise Hardoyal

francelise.hardoyal@sorbonne-universite.fr

Campus Pierre et Marie Curie, couloir 15-25, 1^{er} étage, 107

Téléphone : 01 44 27 51 14

Responsable du Master Ingénierie mathématique

Marie Postel

Axens

TDI (Technologie Développement et Innovation) *

Maître d'apprentissage : Monsieur John CHARBIT

Tuteur académique : Monsieur Frédéric NATAF

Étudiante : Jovana KRSTEVSKA

Mission en alternance : Application des modèles d'apprentissage automatique pour la prédiction du Delta Coke pour les unités de régénération continue des catalyseurs dans les raffineries

Résumé

En vue d'une transformation digitale rapide des industries, les entreprises ont menées des changements dans la façon dont elles surveillent et optimisent leurs opérations. Dans l'industrie pétrochimique et le domaine du raffinage, adopter les nouvelles technologies devient indispensable pour maintenir sa place dans un marché concurrentiel. Le coût des analyses des laboratoires et la faible précision des modèles physico-chimiques ont fait apparaître le besoin de nouvelles approches de modélisation. Ici, on présente l'application des modèles d'apprentissage automatique pour la prédiction d'un indicateur clé de performance KPI des unités de raffinage à régénération continue des catalyseurs (CCR) : la quantité de **coke** sur les catalyseurs à la sortie du régénérateur.

Afin de pouvoir construire des modèles robustes et performants, plusieurs techniques de traitement de données et sélection des variables importantes sont utilisées sur les données clients. Il s'agit d'un problème de régression, ainsi plusieurs modèles linéaires et non-linéaires de régression sont testés et évalués. Des résultats sur six unités différentes sont présentés, ainsi que des comparaisons avec le modèle classique.

Une fois les modèles validés d'un point de vue mathématique, ils sont déployés sur la plateforme *Connect'In*TM et mis à disposition aux experts et au métier, afin d'être validés pour la mise en production. Le cas échéant, nous devons retravailler avec les métiers afin d'améliorer les algorithmes, mais cependant la préparation pour la mise en production est déjà réalisée : le ré-entraînement et le traitement des nouvelles données sont automatisés.

Tout le développement des modules, fonctions, routines, modèles et automatisation des tâches est fait en Python.

*. Adresse de l'entreprise : 91 bd Franklin Roosevelt 92500 Rueil-Malmaison

AXENS

TDI (Technologie Développement et Innovation) *

Responsable : Monsieur John CHARBIT

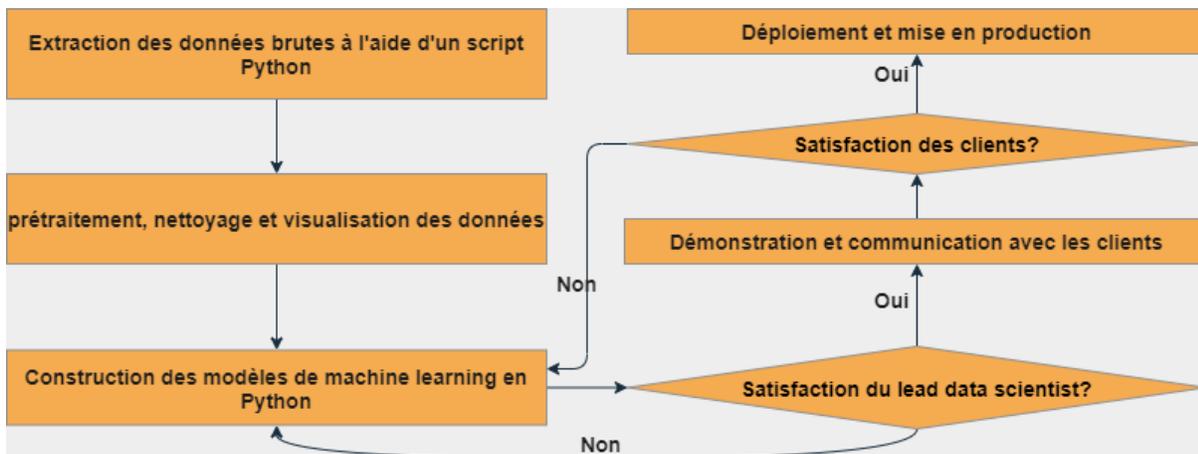
Étudiant : Jijun TANG

Sujet du stage : Data Science

Résumé

Le stage s'est déroulé suivant quatre axes

1. Extraction, nettoyage, visualisation des données sous forme de séries chronologiques issues de procédés industriels de raffinage avec un grand nombre de paramètres
2. Création et optimisation d'algorithmes de machine learning et deep learning en Python
3. Analyse online de la qualité de la prédiction d'algorithmes et échanger avec les experts
4. Déploiement des algorithmes via une API Flask



*. Adresse de l'entreprise : 89, Bd Franklin Roosevelt 92500 Rueil Malmaison

BPCE

Inspection générale *

Responsable : Monsieur Jérémy MERCON

Étudiant : Paul HERVE

Sujet du stage : Stage Data Scientist

Résumé

Lors de ce stage au sein de l'équipe Data de la direction de l'Inspection Générale du Groupe BPCE, plusieurs types de missions et projets m'ont été confiés. J'ai principalement travaillé sur un projet de diagnostic de la qualité de la donnée. J'ai également contribué sur un projet portant sur les écritures manuelles en comptabilité et sur des missions annexes.

Dans le cadre de l'évaluation de la qualité des données, le projet s'est concentré sur la qualité de la donnée brute. L'objectif du projet était de mettre en place un code Python complet qui permet de tester la qualité des bases de données chargées dans le code. Pour cela, nous avons travaillé en collaboration avec les inspecteurs pour déterminer les différents tests à effectuer en fonction des bases, des thématiques et des données. Le code s'est fait sur Jupyter Notebook et Spyder avec la création d'un package correspondant aux différentes fonctions de test et l'utilisation de Git pour rendre le projet accessible au reste de l'équipe. Les résultats sont présentés visuellement grâce à des outils de Data Visualisation (logiciel Spotfire). Ce tableau de bord peut être mis à jour automatiquement une fois les données du code Python chargées. Ce projet sur la qualité de donnée a été mené à son terme et les résultats obtenus sur les bases testées sont encourageants. Il a permis notamment de mettre en lumière l'importance de la qualité de la donnée et de simplifier les investigations des inspecteurs. Par ailleurs, le code est évolutif et pourra être constamment agrémenté de nouveaux tests et de nouvelles fonctions pour répondre aux demandes des inspecteurs.

Lors d'une mission d'audit sur la comptabilité, l'Inspection contrôle les écritures manuelles de comptabilité en s'appuyant sur la norme internationale d'audit. L'objectif du projet sur les écritures manuelles était de simplifier les investigations des inspecteurs en automatisant en partie les analyses qu'ils mènent et ainsi gagner en efficacité. Dans un premier temps, nous avons mis en place un tableau de bord Spotfire qui détecte automatiquement des scénarios d'écritures manuelles à caractères suspects. Dans un second temps, un échantillon d'écritures manuelles, contenant potentiellement des anomalies, est restitué automatiquement dans ce tableau de bord.

Par ailleurs, j'ai participé aux supports techniques auprès de missions et à des chantiers méthodologiques pour faire progresser les processus d'audit tout en travaillant en collaboration avec les inspecteurs.

*. Adresse de l'entreprise : Site Odyssey, 110 avenue de France, 75013 Paris

Brico-Dépôt

Direction Marketing, Digital et Clients *

Maître d'apprentissage : Madame Alimata MARTIN

Tuteur académique : Monsieur Bruno DESPRES

Étudiant : P G

Mission en alternance : Traitement Automatique du Langage et fouille de texte

Résumé

Mon alternance s'est effectuée au sein du pôle Data de Brico-Dépôt qui fait partie de la Direction Marketing, Digital et Clients.

Notre pôle a pour but d'améliorer la connaissance client. En ce sens, l'entreprise nous a demandé de développer un outil afin de pouvoir extraire des informations via des verbatims.

Ces verbatims contenaient des commentaires que les clients laissaient après leurs passages en caisse. Devant le grand nombre de verbatims qui eux même contenaient un grand nombre de commentaires, l'outil que nous avons développé nous a permis un gain de temps énorme.

Afin d'extraire les points positifs et négatifs, nous avons tout d'abord utilisé des modèles de classification de machine learning pour classer ces commentaires en deux catégories : soit positif soit négatif.

La donnée étant textuelle, cela a demandé d'abord une analyse classique de Traitement Automatique du Langage, notamment la constitution d'un dictionnaire pour divers traitements du texte.

Les résultats étaient plutôt bons mais pas excellents sur certains verbatims.

Dans un second temps, afin d'améliorer la précision de l'algorithme, l'utilisation des modèles **Transformers** qui reposent sur les réseaux de neurones nous ont permis d'obtenir des résultats bien supérieurs aux modèles de Machine Learning. Les modèles de types BERT (Bidirectional Encoder Representations from Transformers), pré-entraînés sur un grand corpus de textes (CamemBERT pour le français) nous ont permis d'avoir des résultats plus que satisfaisants.

Ces modèles étant très gourmands en ressources, il nous a fallu utiliser toute la puissance du GPU et ainsi développer un code **CUDA** pour industrialiser notre outil.

*. Adresse de l'entreprise : 30-32 Rue de la Tourelle, 91310 Longpont-sur-Orge

Centre Technique des Industries de la Fonderie (CTIF)

Pôle de Développement Métallurgie et Procédé *

Responsables : Messieurs Patrick HAIRY et Michel STUCKY

Étudiant : Chemsedine BCHIR

Sujet du stage : Développement d'un algorithme génétique en Python ayant pour objectif d'optimiser la composition des alliages métalliques (Alloy Design)

Résumé

Le stage a été réalisé au sein du Pole de Développement Métallurgie et Procédé du CTIF (Centre Technique des Industries de la Fonderie) à Sèvres (92). L'objectif de ce stage était de développer un algorithme génétique en Python servant à faire de l'Alloy Design (création d'un nouvel alliage ou optimisation d'un alliage déjà existant) en extrayant les informations (Fitness Function) à partir de la base de données métallurgique Thermo-Calc via TC-Python (environnement PyCharm).

Thermo-Calc est un logiciel de thermodynamique qui permet, à partir de la composition chimique d'un alliage de prédire les phases métallurgiques présentes et certaines propriétés (masse volumique, Rp0.2, Température de Liquidus, ...). Ce logiciel est utilisé par les Experts de CTIF, mais ne permettait pas de faire une consultation avec une stratégie optimisée comme l'autorisent les algorithmes génétiques.

Dans un premier temps, après avoir réalisé une synthèse documentaire sur les algorithmes génétiques et pris en main l'outil Thermo-Calc, un algorithme « from scratch » intégrant les opérateurs d'évolution de la population d'alliages (sélection, cross-over, mutation) a été développé et testé sur des cas simples et uniquement avec de l'optimisation monocritère. La deuxième partie du stage a consisté à coder le programme en s'appuyant sur des bibliothèques existantes (DEAP) et en particulier à intégrer l'optimisation multicritères (NSGA-III et SPEA-II). Nous avons ensuite testé l'outil (nommé α -screen) sur des cas plus complexes (base fer, nickel, aluminium) fournis par les experts métallurgiques.

Enfin, une interface graphique a été développée (bibliothèque Tkinter) permettant de s'affranchir de l'utilisation de Python pour lancer les calculs. Un développement spécifique a permis en particulier de filtrer les résultats de la dernière génération d'alliages pour faire ressortir les « candidats solutions » les plus pertinents.

Ce stage m'a permis de développer mes compétences en langage Python, de maîtriser de nombreuses bibliothèques, de gérer l'interfaçage (via TC-Python) avec un code du commerce, de mettre en œuvre des stratégies d'optimisation multicritères et enfin de développer de bout en bout un code à usage interne.

Mots clefs : algorithme génétique, optimisation multicritères, Thermo-Calc, Alloy Design, métallurgie, fonderie, Python

Ref : Alpha screen, algorithme génétique pour l'Alloy Design, P. Hairy

*. Adresse de l'entreprise : 44 Av. de la Division Leclerc, 92310 Sèvres

Dassault Systèmes

Département de Recherche *

Responsable : Monsieur LUCAS BRIFAULT

Étudiant : SALAH-EDDINE MAHI

Sujet du stage : Calcul et Optimisation des arbres de construction d'un modèle CAO

Résumé

Ce rapport présente le sujet de mon stage de fin d'études qui s'est déroulé au sein de l'équipe de recherche à Dassault Systèmes sur le site de Vélizy-Villacoublay. Cette équipe a pour mission d'inventer les futures technologies de modélisation 3D, qui serviront à concevoir des expériences de design 3D plus intuitives et automatisées.

Pour concevoir un objet 3D, un designer utilise aujourd'hui un logiciel de Conception Assistée par Ordinateur (CAO), comme CATIA, où il dispose d'un certain nombre d'opérateurs de base qu'il peut enchaîner et combiner de façon à modéliser la forme qu'il souhaite

Le projet dans lequel s'insère mon stage vise à trouver de manière automatique, pour une approximation possiblement bruitée d'un modèle 3D (maillage, nuage de points), la manière optimale de construire ce modèle avec un tel logiciel

L'objectif de mon stage consiste donc, en se basant sur les recherches déjà menées par l'équipe, à :

- Détecter les potentiels opérateurs de fillet (arrondis d'arrêtes) et calculer leurs paramètres.
- Détecter les potentiels opérateurs de sweep (balayage d'un profil par une courbe guide) et calculer leurs paramètres.
- Tester les différents algorithmes retenus et les implémenter en C++

*. Adresse de l'entreprise : 10, Rue Marcel Dassault, 78140 Vélizy-Villacoublay

Dassault Systèmes
Département Techno *

Responsables : Messieurs Jean François LAGUE et Aurélien ALLEAUME

Étudiant : Abel ZEGHIDOUR

Sujet du stage : Qualité des maillages pour les simulations de mécanique des fluides

Résumé

La conception 3D pour les simulations numériques est devenu un outil fondamental pour le développement industriel et technologique de même que pour la recherche scientifique. Les logiciels de maillages se doivent d'être de plus en plus performants afin de répondre à une demande croissante des besoins propres à la virtualisation du monde. Le but de ce stage a donc été de développer des outils et méthodes visant à améliorer la qualité de maillage utilisé en mécanique des fluides.

La méthode des volumes finis est une des principales méthodes utilisées dans la simulation de phénomènes physiques tels que les phénomènes de turbulences et l'écoulement d'air sur des ailes d'avion. Cette méthode se base sur la discrétisation spatiale du domaine d'étude (maillage). En général, la qualité d'une solution est corrélée avec celle du maillage sur lequel elle est implémentée.

Ce stage de 6 mois a été réalisé au sein du département R&D en maillage volumique de Dassault Systèmes du campus Paris 3DS à Vélizy-Villacoublay. Au cours de cette période, nous avons réalisé des recherches bibliographiques afin de déterminer des critères de qualité a priori relatifs à la méthode des volumes finis. Nous avons ensuite implémenté des opérateurs de qualité dans le but d'extraire des données sur les maillages produits par notre équipe. Nous avons en particulier cherché à étendre ces analyses à la structure duale des maillages dont la qualité des éléments intervient dans les calculs de volume finis. Cela nous a conduits à implémenter un code de construction du dual d'un maillage tétraédrique. Enfin, nous avons étudié les données générées pour évaluer la qualité de nos maillages et identifier les configurations faisant défaut à l'aide d'analyses statistiques.

Ce stage aura été l'occasion pour moi d'évoluer dans un environnement de travail agréable et motivant sur un domaine industriel à fort enjeu. J'ai pu travailler en collaboration avec des ingénieurs à l'écoute, partageant leur expérience ainsi que leur expertise. J'y ai beaucoup appris, tant sur le plan théorique que sur le plan pratique, notamment en programmation objet, en architecture informatique ainsi que dans l'organisation du travail en équipe et le métier d'ingénieur logiciel.

*. Adresse de l'entreprise : 10, Rue Marcel Dassault, 78140 Vélizy-Villacoublay

Dorval Asset Management

Gestion Internationale *

Responsable : Monsieur François-Xavier CHAUCHAT

Étudiant : Marin MILLEQUANT

Sujet du stage : Stage en Data Science

Résumé

Dans le cadre du Master 2 – Mathématiques et applications, parcours Ingénierie Mathématique pour l'Entreprise – j'ai réalisé un stage de six mois au sein de Dorval Asset Management.

Dorval AM est une société de gestion affiliée à Natixis Investment Managers. Le poste de stagiaire data scientist était rattaché à l'équipe de gestion internationale qui gère trois fonds. Dorval AM n'ayant aucun data scientist senior dans ses rangs, il était nécessaire de savoir travailler en autonomie. L'objectif était d'assister l'équipe en analysant des données de marché en lien avec la recherche macro-économique et en traitant des données liées aux fonds sous gestion. Pour cela, un accès à l'API Bloomberg a permis d'obtenir le nécessaire pour tester les codes conçus en *Python*. Une partie des projets a été développée directement dans BQuant (un Jupyter Lab embarqué dans l'API).

Concrètement, les missions ont consisté en :

- De nombreuses analyses statistiques concernant des données de marchés.
- Le développement d'un outil automatisant le renouvellement des valeurs en portefeuille.
- L'exploration de cours d'actions en lien avec les évolutions macro-économiques via des méthodes de clustering.
- Le développement d'un outil de filtrage d'articles de presse.

Les trois dernières missions citées représentent les trois grands projets autour desquels le stage s'est articulé ; les analyses statistiques venant ponctuer le quotidien au gré des besoins.

L'outil de renouvellement de portefeuille permet aux gérants d'obtenir les titres qui correspondent le plus fidèlement à un ensemble de critères financiers et extra-financiers et de visualiser la diversification de la composition ainsi obtenue.

Dans le cadre du projet 3, il a été nécessaire d'appliquer des algorithmes de clustering sur des données temporelles. Le traitement de telles données est compliqué par la prise en compte de la temporalité. Il faut donc utiliser des méthodes de comparaison adéquates comme le "dynamic time warping". En effet, les méthodes "classiques" obtiennent des résultats mitigés en ce qui concerne la comparaison et l'agrégation de séries temporelles.

L'outil de filtrage d'article de presse s'inscrit dans le suivi d'actualités concernant les thématiques économiques et financières. Dans un premier temps, un automate de scraping pour récupérer les articles a été conçu. Ensuite, les données textuelles ont été préparées avant d'appliquer les algorithmes de machine learning.

Ainsi, pour mener à bien les différentes missions de ce stage, les compétences développées furent : la programmation en *Python*, la compréhension mathématique des algorithmes de machine learning et la restitution des études à un public non averti.

*. Adresse de l'entreprise : 1 rue de Grammont, 75002 Paris

DreamRs

Data science/Expertise R *

Responsable : Monsieur Victor PERRIER

Étudiante : Samra GOUNRI

Sujet du stage : Stage en Data Science

Résumé

DreamRs est une société de conseil, d'expertise et de formation en data science, spécialisée dans le logiciel R dont l'esprit est en parfaite adéquation avec son slogan : « La data science avec un grand R ». Ses travaux de R&D permettent d'améliorer constamment la qualité de ses services grâce à une veille technologique et stratégique active. La pluralité des types de projets d'un data scientist m'a amenée à participer à diverses missions au sein de l'activité R&D de l'entreprise DreamRs. La compréhension de chaque problématique est importante puisque l'objectif est d'y répondre en analysant les données de façon pertinente.

La première mission consiste à réaliser une classification et une segmentation clients en s'appuyant sur l'étude d'une ACM (Analyse des Correspondances Multiples) et d'une CAH (Classification Ascendante Hiérarchique). L'ACM est réalisée comme prétraitement de la classification. L'application de la CAH permettra d'obtenir une partition et une description de ces classes.

D'autres missions ont pour objectif de créer plusieurs indicateurs statistiques pertinents pour étudier certaines informations sur les observations de plusieurs bases de données. L'étude des données se fait en trois étapes : elle va de la collecte des données, à l'analyse puis à la restitution des données. La restitution des résultats des études est mise en valeur principalement sous forme d'infographie et visuels interactifs et attractifs (réalisation de tableaux résultats de comptage, différentes représentations graphiques statistiques, création d'une nouvelle table de données répertoriant tous les indicateurs, etc). Cette étape est indispensable pour faire passer le bon message de manière simple à ses interlocuteurs.

Par ailleurs, l'entreprise DreamRs développe aussi des outils interactifs sous forme d'add-in RStudio, d'application Shiny ou de packages R, adaptés à la demande et aux besoins des clients. Ainsi, une dernière mission a pour but de construire un module avec la librairie R Shiny que l'on intégrera ensuite à une librairie R de DreamRs déjà existante, nommée « datamods ». La librairie datamods est un package R qui fournit des modules Shiny pour importer des données à partir de sources diverses. Elle permet aussi de manipuler des données en pouvant sélectionner, renommer et convertir des variables d'un tableau de données. Les modules peuvent être utilisés dans n'importe quelle application Shiny standard ou dans un add-in RStudio. L'objectif de ce module est de pouvoir échantillonner un tableau de données fourni par l'utilisateur. D'un point de vue conceptuel, l'interface de ce module doit présenter deux parties distinctes : une première partie permettant d'échantillonner selon un certain pourcentage de lignes choisi par l'utilisateur et une seconde partie permettant d'échantillonner selon un certain nombre de lignes choisi par l'utilisateur.

*. Adresse de l'entreprise : 14 rue Sergent Bobillot, 92400 Courbevoie

EDF R&D

Département ERMES *

Responsable : Monsieur Guillaume DROUET

Étudiant : Eloi GUILBERT

Sujet du stage : Modélisation massivement parallèle des conditions aux limites de type liaison

Résumé

Code_aster est un code open source de simulation numérique en mécanique des structures, développé par EDF R&D. Depuis 2020, une version parallèle massive du code est disponible en version de développement. En 2021, un des objectifs est de rendre disponible comme nouvelles fonctionnalités les conditions limites de type liaison dans cette version. Une liaison permet de définir une relation (généralement linéaire) entre une grandeur physique (degré de liberté, torseur d'efforts,...) d'un ensemble d'éléments. Le stage s'inscrit dans ce travail.

Un calcul de structure mené avec code_aster consiste en l'enchaînement d'un certain nombre de commandes. Chaque commande produit un « concept résultat », qui définit une structure de données que l'utilisateur peut manipuler et réutiliser dans les commandes ultérieures du calcul. La version massivement parallèle du code commence par partitionner le maillage, chaque sous domaine appartient à un seul processus. Chaque processus enchaîne les commandes sur sa partie du maillage indépendamment des autres. Au niveau des conditions limites, un problème est rencontré : que se passe-t-il si un chargement crée une relation entre des éléments appartenant à des sous domaines différents et donc des processus différents ? Pour y remédier, un maillage temporaire est créé, contenant tous les éléments nécessaires au chargement, le traite et communique le résultat aux autres processus, ensuite le maillage temporaire est supprimé.

La première activité du stage a été la mise en place de cas tests : des calculs simples utilisant différents types de liaisons. Ils ont permis de vérifier la validité des résultats du code parallèle sur ces liaisons, en les comparant au code séquentiel. Mais aussi, de détecter des bugs et des anomalies subsistant dans le code, et donc de les corriger.

La seconde partie du stage a été une analyse de faisabilité des liaisons vues dans la première partie. Une partie du temps de chaque calcul est "incompressible", *id* est indépendant du nombre de processus utilisés. Généralement plus un calcul est gros, plus la partie incompressible est petite relativement au reste. Pour cette raison les calculs ont été réalisés sur les super calculateurs d'EDF. Les calculs trop lourds en mémoire avec le code séquentiel sont faisables en parallèle : la mémoire est divisée sur plusieurs noeuds de calcul (un seul pour le séquentiel).

Enfin la dernière partie a été une analyse de scalabilité, toujours portant sur les mêmes liaisons, en temps et en mémoire du code parallèle, c'est à dire regarder la mémoire occupée par processus/le temps pris pour un calcul en fonction du nombre de processus utilisés. Dans les bonnes conditions, en doublant le nombre de processus on divise le temps de calcul/la mémoire par processus par 1.8. Cette partie du stage a aussi permis de détecter et de recenser d'autres bugs qui ne se voyaient pas forcément sur des plus petits calculs.

*. Adresse de l'entreprise : 7 Boulevard Gaspard Monge 91120 Palaiseau

EDF R&D

Département PERICLES, Groupe Systèmes d'Information et Calcul Scientifique*

Responsable : Monsieur Guillaume BOULANT

Étudiante : Sara OUARAB

Sujet du stage : Vérification et Transformation de maquettes BIM pour la simulation numérique

Résumé

Le stage s'inscrit dans le projet SIMBA (Système d'Information et Modélisation des données du BAtiment) 2021.

La surveillance des ouvrages de production d'EDF à grande échelle comprend la vérification de l'intégrité des fonctions structurelles liées aux exigences de sûreté et de puissance. La mise en œuvre de cette démarche nécessite la combinaison de plusieurs informations (plans, maquette numérique, propriétés physiques des composants ...).

La simulation numérique est utilisée à EDF pour analyser le fonctionnement des équipements industriels et proposer des remédiations en cas de dysfonctionnement.

La démarche BIM (Building Information Modeling) introduit une dimension de Système d'Information (SI) dans les projets de construction, exploitation et démantèlement des ouvrages. Le support du modèle BIM d'un ouvrage s'appelle une maquette numérique BIM et prend techniquement la forme de fichiers qui peuvent être dans des formats à spécification ouverte, adaptés à l'échange entre les parties prenantes du projet. C'est le cas du format IFC (Industry Foundation Classes) auquel nous nous sommes intéressés.

Le but de ce stage était de répondre à deux objectifs par la mise en place d'un cadre technique d'analyse de données BIM à travers le format IFC, en utilisant le langage python et la bibliothèque *IfcOpenShell*.

Ainsi, une première mission a été de mettre en place des outils de vérifications respectant les exigences fixées (exemple : avoir les bonnes propriétés physiques renseignées).

Enfin, nous avons pour mission de construire une maquette de calcul idéalisée (dans notre cas une maquette faites d'éléments de structure 2D) sur Salome-Meca à partir des données IFC. Pour cela, nous avons simulé un calcul mécanique afin de pouvoir vérifier la conformité de la maquette (géométrie et maillage).

A ce stade, il n'y a pas de solution générale pour transformer une maquette BIM en une maquette de calcul. En revanche, nous pouvons proposer tout un ensemble de méthodes, en particulier, vérifier dans un premier temps que les informations sont conformes à des exigences prédéfinies à l'aide des outils de vérification développés.

*. Adresse de l'entreprise : 7 Boulevard Gaspard Monge, 91120 Palaiseau

Fotonower

Recherche et Développement *

Responsable : Monsieur Stéphane POIRIER**Étudiante : Kexin MA****Sujet du stage : Analyse des données, apprentissages, évaluations des résultats des différents systèmes algorithmiques****Résumé**

Le but de ce stage est d'explorer l'application de l'apprentissage automatique dans la résolution du problème du traitement des déchets. La contribution de ce stage est d'une part de connaître les différentes méthodes à la base d'apprentissage automatique pour améliorer la résolution du problème et d'autre part de tester la possibilité d'application avec les données de Fotonower.

Pendant le stage, nous avons appliqué des méthodes populaires d'apprentissage automatique sur les données de Fotonower. Le système mask R-CNN nous permet d'avoir une détection en mask pour tous les objets dans une photo. Le thème de classification nous permet de corriger des résultats insatisfaisants de la détection. Par ailleurs, nous avons aussi fait une partie de développement des fonctions en python pour réaliser différentes méthodes de traitement des données.

*. Adresse de l'entreprise : 48 Rue René Clair, 75018 Paris

Framatome

DTIPDM-F : Codes et Méthodes Neutroniques *

Responsables : Messieurs Léandre BRAULT et Xavier HUBERT

Étudiante : Dany UY

Sujet du stage : Inférence statistique de la puissance 3D du coeur à partir de mesures incore partielles

Résumé

Framatome est une entreprise du secteur nucléaire civil. Mon stage se déroule dans le département Codes et Méthodes Neutroniques sous la direction de Léandre Brault. L'équipe travaille principalement sur le développement et la maintenance des codes scientifiques de modélisation (neutronique du coeur, avec couplage thermo-hydraulique).

L'inférence de la puissance 3D du coeur est utilisée pour vérifier la conformité du coeur chargé (et indirectement la validité des modèles physiques) et contribue à la calibration des systèmes de protection des réacteurs.

Ce procédé est aussi appelé reconstruction de la puissance, il s'agit de l'estimation d'un champs 3D à partir de mesures partielles. Autrement dit, nous souhaitons estimer la puissance sur tout le volume du coeur à partir d'un nombre limité de données (variable selon le modèle de réacteur).

Le département souhaite améliorer la méthode actuelle basée sur la correction des estimations théoriques du modèle physique. Pour réaliser cette correction, on utilise une fonction Δ^* obtenue en ajustant un modèle statistique aux données réelles issues des réacteurs. La difficulté est de choisir un nouveau modèle statistique plus performant que le modèle actuel. Donc ma mission principale était l'étude des outils de validation et benchmarking de modèles statistiques.

Mes travaux ont commencé par une étude de la bibliographie en statistiques spatiales et l'analyse descriptive des données des réacteurs présents en France et en Chine (modèles 1300 MWe et EPR *). Puis, nous avons transposé (en partie) des méthodes de validation-croisée spatiale proposées par Roberts, Bahn et al. † pour disposer d'estimateurs d'erreurs des modèles statistiques. L'objectif est d'utiliser un de ces estimateurs pour comparer de potentiels nouveaux modèles (benchmarking). L'étude de ces méthodes a été réalisée avec du calcul parallèle et des simulations d'observations préparées par L. Brault. Les résultats obtenus sont très satisfaisants et seront certainement réutilisés pour des travaux ultérieurs.

En parallèle, une étude de la méthode actuelle a été réalisée pour les données EPR, celle-ci portait sur l'influence des observations disponibles selon leurs positions dans le coeur.

Mes missions ont été très enrichissantes et variées. En particulier, ces travaux ont été l'occasion d'appliquer mes connaissances issues de mes études en M1 et M2, notamment : la programmation en Python et R, les statistiques inférentielles et le calcul parallèle sur CPU.

*. Adresse de l'entreprise : La Défense Tour Areva, 1 Place Jean Miller, 92084 Paris La Défense Cedex

*. Evolutionary Power Reactor

†. D. R. Roberts, V. Bahn et al., Cross-Validation strategies for data with temporal, spatial, hierarchical, or phylogenetic structure, *Ecography*, 2017

GRTgaz

RICE - Research and Innovation Center For Energy *

Responsable : Monsieur Florent BRISSAUD

Étudiant : Théo COUSINO

Sujet du stage : Recherche et Innovation en Méthodes statistiques pour la maintenance

Résumé

Durant son cycle de vie, un système industriel est communément sujet à des défaillances et actions de maintenance. L'enjeu est de garantir des conditions opérationnelles répondant à des conditions de sécurité, de disponibilité et de coût. Après chaque défaillance (ayant lieu à une date imprévisible et donc considérée comme aléatoire), le système est réparé, c'est-à-dire remis dans un état où il peut assurer ses fonctions. Il s'agit de maintenances correctives : le système étant réparable, et les dates de défaillance étant qualifiées d'événements récurrents. Il est aussi possible de retarder les défaillances en intervenant sur le système sans qu'il soit en panne, il s'agit alors d'une maintenance préventive.

Les dates successives de maintenances préventives et correctives constituent des données intéressantes pour définir une gestion optimale du système tout au long de son utilisation. En effet, si par exemple des défaillances ont lieu de manière trop rapprochée, le gestionnaire du système pourra modifier les dates des futures maintenances préventives, afin de ralentir la dégradation du système et de repousser les prochaines défaillances. La planification des instants de maintenances préventives peut se faire de diverses manières. Le plus simple est de fixer à l'avance toutes les dates de maintenance, par exemple périodiquement. Ces dates ont alors un caractère déterministe et ne dépendent pas du fonctionnement du système.

Par nature, les défaillances sont des événements aléatoires. Les outils probabilistes et notamment les processus stochastiques, constituent donc un arsenal privilégié pour les modéliser. Plusieurs modèles existent dans la littérature scientifique pour répondre à cette problématique de modélisation. Ceux-ci peuvent n'inclure que les maintenances correctives, ou bien intégrer les maintenances préventives ainsi que les effets de ces maintenances sur les probabilités de défaillance.

L'objectif de mon stage était de se familiariser avec les méthodes et outils d'estimations paramétriques pour les études de fiabilité et de maintenance utilisés à GRTgaz. Puis de développer un code R qui avait pour objectif d'estimer les modèles de maintenances qui expliquent les défaillances issues des données collectées à GRTgaz en incluant des caractéristiques propres aux systèmes industriels, notamment le traitement du manque de données et la prise en compte des facteurs extérieurs...

*. Adresse de l'entreprise : 1 Rue du Commandant d'Estienne d'Orves, 92390 Villeneuve-la-Garenne

IRT SystemX

Calcul Scientifique et Optimisation *

Responsables : Madame Chiara NARDONI et Monsieur David DANAN

Étudiant : Amine SEKOURANE

Sujet du stage : Raffinement de maillage adaptatif en optimisation topologique

Résumé

Lors de mon stage, j'ai mis en place une méthode de raffinement adaptatif de maillage. L'idée est que ce type d'approche d'adaptation a vocation à être utilisée en optimisation topologique. L'optimisation topologique consiste à résoudre des problèmes de conception optimale de structure. Plus précisément, elle consiste à trouver la répartition optimale de matière dans un volume de conception donné tout en respectant un cahier de charges. On cherche à résoudre un problème de la forme suivante :

$$\min_{\Omega \subset U_{ad}} J(\Omega), \quad \text{tel que } C(\Omega) \leq 0 \quad \text{et} \quad G(\Omega) = 0 \quad (1)$$

où Ω est une forme, U_{ad} le domaine admissible dans lequel Ω est restreint, $J : D \rightarrow \mathbf{R}$ est la fonction objectif à minimiser, $C(\Omega)$ et $G(\Omega)$ sont respectivement les contraintes d'inégalités et d'égalités.

Le processus d'optimisation topologique est un processus itératif où chaque itération on effectue des simulations numériques qui sont souvent très coûteuses en terme de temps de calcul. Mettre en place un processus d'adaptation de maillage permettrait à la fois de réduire le coût de chaque simulation numérique et de minimiser les erreurs de discrétisation. La méthode de raffinement adaptatif utilisée est basée sur les estimateurs d'erreurs a posteriori de type Zienkiewicz-Zhu. Ces estimateurs, construits à partir d'un recouvrement du gradient d'une fonction donnée, nous permettent de construire une métrique qui sera utilisée pour contrôler la taille et l'orientation des éléments du maillage. La fonction utilisée pour construire l'estimateur d'erreur peut être par exemple une fonction analytique, une fonction ligne de niveaux ou bien la solution discrète d'une EDP calculée par la méthode des éléments finis.

J'ai mis en place une approche de raffinement de maillage adaptatif isotrope où la fonction considérée a les mêmes caractéristiques dans les trois directions de l'espace puis le raffinement de maillage anisotrope qui doit tenir compte des variations locales directionnelles de la fonction.

*. Adresse de l'entreprise : 2 Bd Thomas Gobert, 91120 Palaiseau

Kap Code

Étude et Management de la Connaissance Client *

Responsable : Monsieur Pierre FOULQUIÉ

Étudiant : Alexandre RUEDA PAYEN

Sujet du stage : Topic Model appliqué à de courts textes dans le cadre de l'analyse de réseaux sociaux pour la pharmacovigilance

Résumé

Le stage a été réalisé au sein de l'équipe *Data Science* (EMS) de *Kap Code* (EMC²) entreprise appartenant à Kappa Santé. Les activités de Kap Code reposent sur l'analyse des réseaux sociaux appliquée à la **pharmacovigilance**. Cube BPF est l'activité consistant à surveiller les effets signalés d'une solution médicamenteuse ou d'un traitement sur les personnes auxquelles ils ont été administrés, via des algorithmes de **Traitement Automatique du Language**, sur les réseaux sociaux. Les résultats sont 'scrapés' sur les réseaux sociaux par le département informatique et restitués sous la forme de tableaux de données avec, pour chaque post, les informations textuelles ainsi que l'émetteur du post. Ces données seront analysées via des Topic Models (apprentissage statistique non supervisé), puis restituées à une équipe d'experts pharmaciens. Le langage de développement avec lequel Kap Code travaille depuis des années est **R**. L'un des principaux enjeux de ce stage était de trouver une méthode robuste à l'éparpillement des données. C'est en particulier le cas des textes courts, tels que les tweets. Une méthode possible consiste ainsi à introduire des informations extérieures permettant de représenter numériquement le sens sémantique de chaque mot utilisé dans le corpus de messages : on appelle cela des *Word Embeddings*. Deux méthodes ont ainsi été testées au cours de notre mission, afin d'améliorer la méthode BiTerm Model (BTM) initialement utilisée : **La méthode Relational BiTerm Model (R-BTM)** et **La méthode Embedding Topic Model (ETM)**. Toutes ces méthodes maximisent les fonctions de vraisemblance d'un modèle statistique via un Échantillonneur de Gibbs. Pour évaluer les performances de ces algorithmes, les modèles sont relancés plusieurs fois sur les mêmes données avec des paramètres d'initialisation aléatoires différents. Ensuite, ils sont comparés via un algorithme développé au cours du stage, évaluant la distance entre deux modèles comme étant la meilleure combinaison entre les topics d'un modèle à l'autre. Une analyse des experts de l'équipe pharmacologie, permet alors de confirmer que les topics inférés par nos méthodes sont consistants et exploitables. Les tests ont porté sur des données "twitter" concernant les opinions des Français sur la gestion de la pandémie de covid19 (10000 à 100000 tweets) et sur des données de forums concernant les préoccupations des internautes francophones sur leurs troubles intestinaux (10000 à 100000 données). Le modèle Embedding Topic Model (ETM) est conservé, du fait de sa versatilité et de son temps de calcul correspondant mieux aux attentes de l'entreprise. Ses résultats pourront être réutilisés dans le cadre de classifications (exemple : analyse sentimentale), de l'analyse de similitude entre messages ou encore dans le cadre de la détection de concepts médicaux.

*. Adresse de l'entreprise : 28 rue d'Enghein, 75010 Paris

ONERA - Office National d'Études et de Recherches Aérospatiales

Département Traitement de l'Image et Systèmes - Unité de Mathématiques et Applications et Calcul Intensif*

Responsable : Monsieur Éric SAVIN

Étudiante : Céléna LOUIS

Sujet du stage : Mise en forme d'une aile d'avion à propulsion électrique distribuée

Résumé

La propulsion électrique distribuée (PED) est considérée aujourd'hui comme une solution envisageable pour améliorer l'efficacité propulsive des transports aériens. L'ONERA a, par exemple, développé le concept DRAGON de propulsion électrique distribuée hybride où des soufflantes connectées à des moteurs électriques sont placées tout le long de l'aile d'avion de telle façon que le placement donne le maximum d'effet aéro-propulsif.

Par ailleurs, plusieurs modèles d'ailes et plusieurs études ont démontré la potentielle utilité des techniques actives de mise en forme des voilures pour la performance, la stabilité et le contrôle de vol des aéronefs comme, par exemple, le programme Active Aeroelastic Wing (AAW) développé par la NASA dans les années 1990.

En s'appuyant sur des recherches précédentes de la NASA, l'objectif de ce stage est de développer un modèle numérique des effets aéroélastiques d'une aile flexible basé sur la théorie des poutres. Ce modèle doit prendre en compte d'une part, les forces et moments aérodynamiques et de propulsion générés par les moteurs, et d'autre part, la rigidité géométrique et les effets de la force suiveuse.

Cette étude a commencé par l'analyse des modèles aéroélastiques et des concepts de PED existants. S'en est suivi l'adaptation à un modèle de poutre puis la formulation éléments finis du problème aéroélastique. Ensuite, un code a été implémenté en Matlab pour résoudre le modèle numérique espace-temps énoncé précédemment.

La seconde partie du stage consistait à résoudre un problème d'optimisation. Étant donné une position et une torsion spécifiques d'une aile, il s'agit de trouver la poussée des propulseurs permettant d'osciller au cours du temps, autour de la position et de la torsion données.

*. Adresse de l'entreprise : 6 Chemin de la Vauve aux Granges, 91120 Palaiseau

Safran Aircraft Engines

Fleet Data Control & Collect *

Responsable : Monsieur Nicolas DUBUS

Étudiant : Makhatch ABDULVAGABOV

Sujet du stage : Consolidation des données d'utilisations moteurs

Résumé

Le stage a pour objectif de mener une étude sur les heures et cycles des moteurs et de produire un dashboard, permettant de comparer les heures et les cycles issus des différentes bases. Il a eu lieu dans l'équipe du service **Fleet Data Control & Collect** sur le site de Safran Aircraft Engines à Montereau-sur-le-Jard. Le stage est constitué de trois projets principaux :

- L'analyse du délai de mise à jour des données.
- La comparaison et la consolidation des données heures/cycles de différentes bases de données.
- La prévision des cycles.

A chaque vol les informations sur l'avion sont recueillies par les compagnies aériennes. Ces données sont ensuite transmises à Safran par l'intermédiaire d'un représentant de l'entreprise auprès du client (par le FSE : Field Service Engineer). En général les données arrivent régulièrement, mais dans certains cas celles-ci sont mis à jour avec un retard important. La première partie du stage constitue une analyse du délai pour identifier les causes pouvant être à l'origine de ce retard. En plus de l'analyse sur les données, une série d'entretiens individuels avec plusieurs FSE, permet d'identifier d'éventuels facteurs impactant le délai et n'apparaissant pas dans les données numériques (facteurs humains).

Les données d'utilisations, dont les heures et les cycles effectués par mois sont présentes dans différentes bases : la base FDM et la base Data Hub. Ces bases se distinguent surtout par l'origine des données. Il peut y avoir des différences sur les paramètres d'utilisations entre ces bases. Il est donc intéressant de les croiser et d'analyser les écarts. Dans le cadre du stage, un dashboard est réalisé permettant de comparer ces paramètres et de suivre l'écart entre les heures et les cycles des bases FDM et Data Hub.

Une étude univariée (prévision des cycles à partir des cycles renseignés dans le passé) est réalisée sur la série temporelle du cumul des cycles. Un modèle SARIMA (Seasonal Autoregressive Integrated Moving Average) est sélectionné pour faire des prévisions. En comparant les résultats avec les données de validation, on obtient un écart moyen en pourcentage de 1.53% ce qui permet de valider le modèle sélectionné.

*. Adresse de l'entreprise : Chemin de Bassin, 77950 Montereau-sur-le-Jard

SUMMIT Sorbonne Université

Mathématiques et Applications *

Responsables : Monsieur Stéphane LABBE et Mesdames Nora AISSIOUENE et Karine LAURENT

Étudiant : Kevin MECHOUK

Sujet du stage : Modélisation et calage de systèmes hydrauliques

Résumé

Le but de ce stage est de développer des méthodes de calage de modèles hydrauliques à l'aide de certaines grandeurs physiques mesurées à différents endroits d'un réseau de tuyaux. La circulation de l'eau dans les réseaux de tuyaux est difficile à modéliser du fait des nombreux phénomènes physiques qui s'y produisent et de la taille des réseaux urbains. Pour cela, des logiciels libres sont utilisés par les organismes de distribution d'eau potable au niveau urbain. J'ai utilisé l'un d'entre eux, Epanet, afin de pouvoir effectuer des simulations hydrauliques au cours du temps.

J'ai commencé par effectuer une étude bibliographique des différentes équations régissant la circulation de l'eau dans un réseau. Puis j'ai créé une interface en python afin d'utiliser les fonctions d'Epanet et d'avoir un contrôle sur le logiciel. J'ai également fait des études bibliographiques sur des algorithmes évolutionnaires. En particulier, j'ai étudié et implémenté des algorithmes génétiques afin d'optimiser des paramètres à l'aide de différentes données mesurées par des capteurs. J'ai codé ces méthodes en Python, puis j'ai testé et comparé les différents résultats obtenus.

Ce stage m'a permis d'améliorer mes compétences en Python, en analyse mathématique et en synthèse d'articles. J'ai pu également travailler sur ma capacité de présentation et ma gestion du travail.

*. Adresse de l'entreprise : 4, Place Jussieu, 75005 Paris

Les étudiants de l'année 2020-2021

| Nom | Prénom | Entreprise | Option | Page |
|--------------|---------------|-------------------|---------------|-------------|
| ABDULVAGABOV | Makhatch | Safran | ps | 20 |
| BCHIR | Chemsedine | CTIF | ps | 7 |
| COUSINO | Théo | GRTgaz | ps | 16 |
| G | P | Brico-Dépot | app,ps | 6 |
| GOUMRI | Samra | dreamRS | ps | 11 |
| GUILBERT | Eloi | EDF | m | 12 |
| HERVE | Paul | BPCE | ps | 5 |
| KRSTEVSKA | Jovana | Axens | app,ps | 3 |
| LOUIS | Céléna | ONERA | m | 19 |
| MA | Kexin | Fotonower | ps | 14 |
| MAHI | Salah-Eddine | Dassault Systèmes | ps | 8 |
| MECHROUK | Kevin | SUMMIT | ps | 21 |
| MILLEQUANT | Marin | DORVAL AM | ps | 10 |
| OUARAB | Sara | EDF | m | 13 |
| RUEDA PAYEN | Alexandre | Kap Code | ps | 18 |
| SEKOURANE | Amine | IRT System X | ps | 17 |
| TANG | Jijun | Axens | ps | 4 |
| UY | Dany | Framatome | ps | 15 |
| ZEGHIDOUR | Abel | Dassault Systèmes | ps | 9 |

Les résumés sont classés dans la brochure par ordre alphabétique de l'entreprise.

Abbréviations

ps : Ingénierie Mathématique Pour l'Entreprise - analyse numérique, calcul scientifique et probabilités, statistiques

m : Ingénierie Mathématique Pour l'Entreprise - analyse numérique, calcul scientifique et mécanique

app : étudiant-e apprenti-e en alternance