## Ingénierie Mathématique Master mention Mathématiques et applications





Responsables: Cindy Guichard et Marie Postel

Laboratoire Jacques-Louis Lions

Barre 15-25 bureau 313 Téléphone : 01 44 27 54 08

 ${\bf Email: cindy.guichard@sorbonne-universite.fr, marie.postel@sorbonne-universite.fr}$ 

# 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 2021-2022

https://m2ingmath.math.upmc.fr/

Secrétariat : Barre 14-15 bureau 206, 4 place Jussieu, 75005 Paris. Sorbonne Université, BC 187, 4 place Jussieu, 75252 Paris cedex 05

Cette brochure constitue la dix-huitième 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 quatrième 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 neuf stages de fin d'études terminés en 2022.

## Responsables pédagogiques dans les différentes spécialisations en 2021-2022

Probabilités - statistiques : Lokmane Abbas Turki, Vincent Lemaire

Laboratoire de probabilités et modèles aléatoires, Campus Pierre et Marie Curie, couloir 16-26 1<sup>er</sup> étage

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

Autres intervenants: M. Abdallah, J.-P. Baudry, I. Kharroubi, Y. Mazroui, 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<sup>e</sup> étage

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

Autres intervenants: L. Boudin, G. Delay, M. Fernandez (Inria), P. Frey, S. Labbé, M. Postel (SU)

M. Cerf (Ariane Group), P. Havé (haveneer), X. Juvigny

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

TAR)

## 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<sup>e</sup> 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 14-15,  $2^e$  étage, 206 Téléphone : 01 44 27 51 14

## Responsable du Master Ingénierie mathématique

Marie Postel

## Ad's up Consulting

BU Data\*

Responsables: Monsieur Pierre GUIO et Madame Solène VINSOT

**Étudiant : Lionel KOBA** 

Sujet du stage : Modélisation des clics et explication des performances des créatives

#### Résumé

Au cours de mon stage j'ai été amené à travailler en priorité sur deux projets à savoir

- Modélisation des clics
- Explication des performances de "créatives" (annonces publicitaires).

La modélisation des clics nous a permis d'identifier les variables qui ont un impact sur la variation du nombre de clics et également d'évaluer l'impact individuel et collectif de ces variables sur les clics en utilisant la méthode des moindres carrés généralisés et le modèle "random effects". Puis, en nous servant d'un modèle de projection ARIMA nous avons prédit le nombre de clics sur un horizon assez proche(30 jours). L'explication des performances des créatives consistait dans un premier temps à réaliser un clustering d'un certain nombre de créatives sur la base de leurs métriques de visibilité et en utilisant la méthode du mélange gaussien. Suite à cette répartition en clusters, nous avons répondu à la problématique qui consiste à identifier les labels qui participent à la réalisation des bonnes performances. Pour ce faire, nous avons réalisé une régression logistique en vue de calculer un score d'appartenance des créatives aux différents clusters précédemment formés et ainsi faire le lien entre la présence ou non d'un label sur une image et sa performance. J'ai également réalisé d'autres missions de façon secondaire, il s'agit entre autres de :

- Data visualisation avec l'outil "Data studio"
- Prise en main des projets déjà existants notamment le "Product Scoring" et le "Bench- mark concurrentiel".
- Implémentation d'un algorithme d'attribution-contribution basé sur les chaînes de Markov.
- Parallélisation de code Python afin d'améliorer la vitesse d'exécution.
- Création de connecteurs API (une API est un ensemble normalisé de classes, de méthodes, de fonctions et de constantes qui sert de façade par laquelle un logiciel offre des services à d'autres logiciels).
- Prise en main des plateformes de collecte de données et de contrôle des campagnes publicitaires telles que Google Ads et Google Analytic

<sup>\*.</sup> Adresse de l'entreprise : 103 Boulevard de Charonne, 75011 Paris

## Axa

Group Risk Management\*

Maître d'apprentissage : Messieurs Rémi MEYNADIER et Hugo RAKOTOARIMANGA

Tuteur académique : Monsieur Bruno DESPRES

**Étudiante : Anna WEISMAN** 

Mission en alternance : Développement d'un générateur stochastique de précipitation à l'échelle

de l'Europe

## Résumé

Mon alternance s'inscrit dans le cadre d'un projet financé par AXA GIE commencé en Septembre 2021 dont l'objectif est de créer un modèle d'inondation à l'échelle Européenne. Ce projet a pour but de mettre à jour la vision du risque d'inondation en Europe de AXA, en couvrant à la fois les inondations pluviales (ruissellement d'eau) et fluviales (débordement de rivière).

Ce projet est une collaboration avec Fathom, une startup basée à Bristol à la pointe de la modélisation du risque d'inondation, et Deltares, un institut de recherche basé à Delft aux Pays Bas spécialisé dans l'eau et les infrastructures. J'ai été engagée pour travailler dans l'équipe de modélisation des risques liés aux catastrophes naturelles de AXA GIE aux côtés de Rémi Meynadier et Hugo Rakotoarimanga afin de développer le premier module du modèle d'inondation : le générateur stochastique de précipitation. L'objectif du générateur stochastique de précipitation est de simuler une série journalière de 10 000 ans de précipitations. Cette série est ensuite intégrée dans un modèle hydrologique et un modèle hydraulique en sortie desquels est généré un catalogue d'événements d'inondation associés chacun à une période de retour, i.e. le temps statistique entre deux occurrences d'un événement naturel d'une intensité donnée. Ce catalogue permet d'estimer les pertes liées à la réassurance d'événements d'inondation. Nous avons développé un générateur de précipitation basé sur des régimes météorologiques et l'alternance de périodes sèches et humides. Ce modèle prend en compte la persistance de la précipitation et préserve les statistiques observées de distribution des extrêmes de pluie à travers l'Europe. 10000 années de précipitations ont été générées de façon reproductible. Avec cette méthodologie, nous sommes capable de générer de nouvelles séquences de précipitations. Ce modèle pourrait être transposé à d'autres zones géographiques.

Nous avons adapté ce modèle afin de prédire l'évolution des extrêmes de précipitation pour différents scénarios de changement climatique. Nous nous sommes basés sur la relation issue de la thermodynamique de Clausius Claypeyron afin de modifier notre modèle en fonction de différents scénarios d'augmentation globale annuelle moyenne de la température par rapport à la période pré-industrielle basés sur les scénarios de forçage radiatif définis par le Groupe d'experts intergouvernemental sur l'évolution du climat (GIEC). Ce modèle prédit une augmentation globale des extrêmes de précipitation avec l'augmentation de la température. Durant cette alternance j'ai également réalisé un travail important sur l'automatisation et la reproductibilité des résultats du modèle, en utilisant Snakemake, un outil de gestion de flux opérationnel qui permet de paralléliser et d'automatiser des tâches.

<sup>\*.</sup> Adresse de l'entreprise : 6 Place de La Pyramide, 92800 Puteaux

## CY Cergy Paris Université

Laboratoire Analyse Géométrie Modélisation – UMR CNRS 8088 \*

**Responsable: Monsieur Christian DAVEAU** 

**Étudiant: Svevo BANDELIER** 

Sujet du stage : Problème de diffusion d'ondes électromagnétiques avec condition d'impédance d'ordre élevé

## Résumé

Le but de ce stage consiste à modéliser des problèmes de diffusion d'ondes électromagnétiques par un obstacle, éventuellement composé de matériaux ayant des propriétés physiques différentes. Lorsque les champs électromagnétiques émis par une antenne rencontrent un obstacle, ils sont à priori diffusés dans toutes les directions. Pour évaluer l'importance du phénomène de diffusion des champs en présence d'un obstacle, on utilise une grandeur : la surface équivalente radar (SER). Cette quantité traduit la visibilité de l'obstacle par un radar. Pour ne pas être détecté l'obstacle doit avoir la SER la plus faible possible. Pour cela, on étudie la forme du corps qui minimise les reflets, et on applique une couche de diéletrique sur la surface ou une partie de la surface de l'obstacle. L'électromagnétisme est régi par les équations de Maxwell posées dans  $\mathbb{R}^3$ . On a donc à traiter un domaine non-borné, ce qui exclut les méthodes par éléments finis. Dans ce domaine, les champs électromagnétiques sont modélisés par des méthodes intégrales. Cette méthode repose dans le fait que les champs extérieurs à l'obstacle sont exprimés à partir d'intégrales faisant intervenir leurs traces tangentielles sur la frontière de l'obstacle. L'intérieur de l'obstacle peut être décrit par différentes méthodes. La méthode que nous avons choisie est le développement de conditions approchées d'impédance à la frontière de l'obstacle avec l'air, ce qui conduit à ne mailler que la frontière. Une condition d'impédance relie les traces tangentielles des champs électriques et magnétiques en tout point de la frontière. Historiquement, une première condition, dite de Leontovich, reliait les deux traces par un coefficient multiplicateur. Depuis, des techniques plus précises ont été mises au point, comme des conditions d'impédance d'ordre élevé (HOIBC) . Un code HOIBC existe déjà, mais seulement en 3D. J'ai été chargé d'implémenter un code 2D dont l'intérêt est de mettre au point des techniques destinées à des développements ultérieurs. L'objectif de ce code est le calcul de SER. Le passage de HOIBC du 3D au 2D n'est pas trivial car on doit revenir à la théorie des équations de Maxwell. Les fonctions de Green ne sont plus les mêmes, donc il en découle des équations différentes avec de nouvelles intégrales singulières à traiter. De ce fait les méthodes pour les discrétiser et les approcher numériquement conduisent à l'implémentation d'un code 2D à partir de zéro. La validation du code a été effectuée par le traitement de différents exemples dont les résultats étaient connus.

<sup>\*.</sup> Adresse de l'entreprise : 2 Avenue Adolphe Chauvin, 95300 Pontoise

#### EDF R&D

Laboratoire National d'Hydraulique et Environnement \*
Maître d'apprentissage : Monsieur Florent TACCONE

Tuteur académique : Monsieur Stéphane LABBÉ

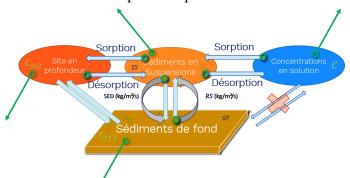
**Étudiant : Zied AMAMA** 

Mission en alternance : Modélisation des interactions entre sédiments et radionucléides dans

**TELEMAC** 

## Résumé

Mon alternance a été réalisée au Laboratoire National d'Hydraulique et Environnement du sous groupe Recherche et Développement de EDF. Ma mission principale était d'améliorer et de développer le module "qualité de l'eau", dans un logiciel de modélisation pour les écoulements à surface libre. Dans le cadre du projet Plateforme Hydro-Environnementale (PHE), des travaux sur la modélisation simplifiée des interactions de sorption et de désorption mettent en forme un modèle couplant le transport sédimentaire (SISYPHE) avec les micropolluants et radionucléides (RN). Un approfondissement de ces méthodes a ensuite été développé dans le sous module MICROPOL (micropolluants) de WAQTEL (module qualité de TELEMAC-MASCARET\*). Le module MICROPOL traite aussi des interactions de sorption désorption par usage d'équations à l'équilibre (Kd), mais le transport sédimentaire est représenté par un flux fictif.



Modèle d'échanges réversibles à deux étapes implémenté dans TELEMAC

C'est dans le coeur de ces actions que s'inscrit mon alternance. Le principal objectif est d'améliorer ces modèles d'interactions entre sédiments et radionucléides en 2D et 3D dans un solveur des équations de Navier-Stokes par la méthode des éléments finis. Ceci en implémentant des termes sources qui tiennent compte de l'aspect cinétique des échanges. Il conviendra alors de comparer les apports de l'implémentation d'une dynamique sédimentaire propre dans la modélisation des concentrations de RN. Lors de cette année en apprentissage, j'ai pu rejoindre une équipe de développeurs, en ayant principalement pour tâche d'implémenter des évolutions dans le logiciel, de les tester sur de nombreux cas théoriques/réels, ainsi que de documenter les nouvelles fonctionnalités pour le meilleur usage des utilisateurs.

<sup>\*.</sup> Adresse de l'entreprise : 6 Quai Watier, 78400 Chatou

<sup>\*.</sup> http://www.opentelemac.org/

## Hôtel des Lys

\*

**Responsable: Monsieur Laurent YANG** 

**Étudiant: Cheng WANG** 

Sujet du stage : Implémentation et évaluation d'un modèle d'apprentissage automatique pour la prédiction des annulations de réservation d'hôtel avec l'apprentissage automatique

#### Résumé

Mots clés : Prévision d'annulation, Data science, Industrie Hôtelière, Gestion des recettes, Apprentissage Statistique

La science des données est une discipline qui a un développement rapide depuis les dernières années grâce à l'essor du Big Data, Et l'industrie hôtelière ne fait pas exception. Diverses problématiques font l'objet de sujets exploratoires afin d'être résolues par des algorithmes de machine learning. Les annulations sont un des points mystérieux de la gestion des revenus hôteliers en raison de leur impact sur les systèmes de gestion de réservations. En fait, les raisons qui poussent les clients à annuler sont toujours floues.

L'objectif de ce mémoire est de prédire les annulations de chambres d'hôtel en utilisant des ensembles de données provenant du système de gestion de revenus et de d'expliquer la variable qui donne le plus d'effet. Le problème de la prédiction des annulations est un problème de classification binaire, car les deux résultats possibles sont l'annulation ou la non-annulation.

L'analyse débute par la préparation d'une base de données sur laquelle seront prédites les annulations de réservations. Ces données proviennent du Medialog1 et des plateformes de vente en ligne (par exemple, Booking.com et Expedia) qui enregistrent les données de réservation. L'acquisition de données pertinentes représente un travail important du fait que le bon score vient de la qualité des données. La construction de notre jeu de données est une tâche principale dans la première phase du stage.

Divers algorithmes ont été implémentés pour modéliser les annulations de réservation afin de faire des prédictions. Ce sont par exemple les algorithmes de régression linéaire (LR), d'arbre de décision , de foret aléatoire et "extreme gradient boosting" (XGBoost). LR modèle est un modèle basique avec lequel nous avons comparé les autres modèles.

Les données ont été fournies par un hôtel spécifique et les algorithmes d'apprentissage automatique utilisés dans cette étude sont "Random Forest", "XGBoost" et "Logit". les deux premiers sont des modèles basés sur des arbres, qui créent des arbres de décision afin de faire des prédictions dans un problème de classification. Ils sont appelés arbres de classification. L'objectif d'un arbre de classification est de déterminer une variable de résultat qualitative en effectuant des divisions binaires par étapes, où les différents résultats sont appelés classes. Le modèle "Logit", ou régression logistique, est une forme de régression binaire qui est utilisée comme modèle de référence dans ce travail. Nos principaux résultats indiquent que Random Forest est le modèle le plus performant sur les données de l'hôtel avec une précision proche de 80%.

<sup>\*.</sup> Adresse de l'entreprise : 16 rue Richaud ,78000 Versailles

## **IFPEN**

Direction Sciences de la Terre/Département Physique numérique des milieux poreux \*

Responsables: Messieurs Daniele COLOMBO et Jérémy FREY

**Étudiant : David JERNIVAL** 

Sujet du stage : Implémentation de la méthode CS-FEM dans le logiciel Code\_Aster

#### Résumé

Ce stage entre dans le cadre d'un projet à caractère prospectif : le stockage de CO<sub>2</sub> dans des réservoirs déplétés et des aquifères. Les risques mécaniques associés à ces stockages, tels que l'intégrité des roches couvertures, et la stabilité des failles potentiellement présentes dans le secteur, doivent impérativement être évaluées. Pour ce faire IFPEN s'appuie sur un prototype de couplage itératif entre ses simulateurs réservoirs pour l'écoulement multiphasique en milieu poreux, et Code\_Aster pour la mécanique. Les maillages utilisés, hérités de la modélisation réservoir dans le domaine pétrolier, sont pensés pour la méthode des volumes finis et non adaptés aux éléments finis (Code\_Aster). La composante mécanique du couplage devient alors problématique, car la présence de mailles dégénérées ou présentant un mauvais rapport de forme dans le maillage, va dégrader la qualité de la solution obtenue, ou rendre le système non inversible.

L'objectif a donc été d'implémenter dans Code\_Aster, un logiciel de mécanique des structures open-source développé par EDF, une méthode prometteuse pour lever ces difficultés : le CS-FEM (Cell-based Smoothed-FEM). Cette méthode, à la différence d'une méthode d'éléments finis classique (FEM), ne s'appuie pas sur la transformation isoparamétrique des éléments, la rendant théoriquement moins sensible à la qualité du maillage, et capable d'utiliser des polyèdres non supportés en FEM.

En somme, le stage a d'abord consisté à se familiariser avec la littérature connue avec la classe de méthodes S-FEM (Smoothed-FEM), puis à prendre en main le logiciel Code\_Aster. Nous avons ensuite implémenté une première approche spécifique aux éléments finis triangulaires en 2D, avant de développer une approche 3D plus générique, permettant de traiter différents éléments connus comme l'hexaèdre, le tétraèdre, la pyramide ou encore le pentaèdre.

L'implémentation, si elle a été un succès, n'a toutefois pas satisfait les promesses attendues pour le cas 3D. Nous avons effectivement réalisé des tests de validation en déplacement à partir d'un cas-test issu de la littérature. Pour le CS-FEM nous avons testé plusieurs sous-découpages par type d'éléments et comparé la solution obtenue à la solution FEM, puis à la solution FS-FEM, une autre méthode S-FEM disponible dans un code prototype développé à IFPEN. Pour les maillages pentaè-driques, tétraèdriques et pyramidaux, les solutions CS-FEM stables (avec un nombre suffisant de sous-domaines) sont équivalentes aux solutions FEM à maillage équivalent. Seul le cas des éléments hexaèdriques montrait une meilleure précision que la solution FEM. Mais sur un cas réel (un modèle simplifié du bassin de Neuquen en Argentine en maillage hexaèdrique), la différence entre les solutions obtenues par les deux méthodes devient négligeable.

<sup>\*.</sup> Adresse de l'entreprise : 1-4 Avenue du Bois-Préau, 92500 Rueil-Malmaison

**ONERA - Office National d'Études et de Recherches Aérospatiales** DTIS-MACI \*

Responsables : Messieurs François ROGIER et Stéphane BRULL

Étudiante: Nathalie BONAMY PARRILLA

Sujet du stage : Schémas LBM et applications aux plasmas

#### Résumé

Ce stage a été réalisé en collaboration avec l'Institut de Mathématiques de Bordeaux. Les applications des plasmas sont de plus en plus nombreuses dans l'industrie. En effet, elles peuvent être utilisées dans divers domaines, par exemple dans l'aéronautique. Des actionneurs plasma peuvent être utilisés sur les ailes d'avions pour recoller l'écoulement de l'air autour de l'aile. Ceci permet de gagner en performance, puisque la portance est améliorée. La modélisation des plasmas devient alors un enjeu intéressant pour ces applications. Des équations fluides peuvent modéliser le comportement des plasmas, cependant une autre approche est possible à travers la modélisation cinétique. Il s'agit d'une approche mésoscopique, qui contient plus d'informations sur la physique à l'oeuvre dans un plasma que l'approche macroscopique. Les schémas numériques sur ce type d'équations peuvent être plus simples à calculer, notamment grâce à des opérateurs de relaxation qui viennent remplacer des opérateurs de collision qua-dratiques plus complexes. Une méthode possible consiste à discrétiser l'espace des vitesses en un nombre fini de propagations, il s'agit de la méthode lattice Boltzmann ou "LBM". Cette méthode permet de paralléliser très facilement le calcul de la solution numérique. Elle est cependant, assez peu utilisée pour les plasmas actuellement. L'objectif de ce stage était donc de simplifier un modèle d'équations cinétiques, modélisant un type de plasma particulier, en vue d'écrire des schémas "LB". Le travail réalisé au cours de ce stage se poursuit en thèse.

<sup>\*.</sup> Adresse de l'entreprise : 2 Av. Edouard Belin, 31000 Toulouse

## Safran

Safran Tech, Centre de Recherche & Technologie \* Responsable : Monsieur Lennart GULIKERS

Étudiant : Saif-Ul-Islam MKOUBOI

Sujet du stage : Inférence des paramètres de santé dans le cadre d'un système sous déterminé

#### Résumé

L'étude et la surveillance de la santé des moteurs d'avions fait partie des sujets de recherche émergents et qui suscitent l'intérêt croissant des grandes industries et entreprises aérospatiales. Le but principal étant d'anticiper et de prédire les défaillances potentielles afin de mener des actions de maintenance préventives et d'éviter les évènements imprévus. Chez Safran Tech, l'analyse des données pour la surveillance de l'état des moteurs d'avions et d'hélicoptères est au cœur des programmes de recherche depuis plusieurs années avec de bons résultats obtenus. Un projet a été mis en route avec comme objectif d'introduire les principales équations physiques éventuellement utiles à l'élaboration d'indicateurs de santé et de fournir des équations reliant les variables décrivant l'état du moteur pour ensuite les utiliser pour déterminer les conditions dans lesquelles l'inférence des rendements du moteur est possible. Cette étude a permis notamment d'identifier les capteurs qui doivent obligatoirement être présents sur le moteur.

Dans ce stage, nous considérons un système avec des paramètres dépendant du temps, plus exactement un moteur d'hélicoptère dégradé, qui est observé à travers des capteurs de sortie. Nous nous intéressons à l'inférence des paramètres dans le cas où le nombre de paramètres dépasse le nombre de capteurs ce qui correspond donc à un cadre sous-déterminé. Sans contraintes supplémentaires, cela conduit à des valeurs multiples (en nombre éventuellement infini) pour les paramètres qui sont cohérents avec les sorties des capteurs. Plusieurs méthodes sont proposées pour surmonter ce problème avec plus ou moins de succès. Nous considérons le réglage des paramètres dépendant du temps à l'échelle des vols. Un objectif d'optimisation où la norme de la différence entre les paramètres de deux pas de temps consécutifs, c'est-à-dire 2 vols consécutifs sont ajoutés avec un poids à l'erreur d'ajustement des données a été proposé.

Nous avons dans un premier temps vérifié la cohérence de ce modèle d'optimisation avec notre problème étudié puis proposé une méthode de résolution numérique nous permettant d'effectuer nos expériences, d'identifier et mettre en exergue les difficultés qui peuvent être rencontrées et empêcher l'inférence de nos paramètres de santé. N'ayant pas accès à de vraies données nous avons travaillé sur les données issues d'un simulateur qui génère des données bruitées. Mais avant cela nous avons mené plusieurs expériences de cas dit "collège" avec des modèles que nous avons définis nous-même. Cette étape nous a permis d'avoir une idée sur le comportement de notre méthode et de confirmer son bon fonctionnement. Ensuite nous avons mené les expériences nous permettant de comparer nos résultats avec celles issues d'une autre méthode de résolution déjà faite auparavant. Et pour finir nous avons proposé et testé une idée alternative d'ajustement de la fonction objectif.

<sup>\*.</sup> Adresse de l'entreprise : Rue des jeunes Bois, 78117 Châteaufort

## **SUMMIT Sorbonne Université**

Mathématiques et Applications \*

Responsables: Monsieur Nicolas CLUZEL et Madame Siyun WANG

**Étudiant : Raydje DENON** 

Sujet du stage : Prédiction de la tendance de l'indicateur Covid-19 Obépine

#### Résumé

Le réseau Obépine est un réseau national français de prélèvements et d'analyses de la concentration du virus du covid-19 dans les eaux usées à travers le territoire. Ces prélèvements sont effectués dans 216 stations d'épuration à travers le territoire environ 2 fois par semaine. Ils permettent ensuite de construire un indicateur de la concentration du covid-19 dans les eaux usées traitées par la station d'épuration, nous avons donc affaire à des séries temporelles. Cependant de part la nature aléatoire des prélèvements et le lisseur de Kalman utilisé, nos séries temporelles sont discontinues, sparses et irrégulières.

Le but de ce stage était de réussir à prédire la tendance pour la semaine suivante de l'in- dicateur covid-19 (station par stations). Le problème de prédiction s'est transformé en un problème de classification, consistant à classer la tendance de la semaine suivante dans trois catégories décrites par les mots clefs : stable, montée et descente. J'ai codé majoritairement en python pour ce stage, tout en ayant besoin d'utiliser R de temps en temps.

Après une analyse de l'autocovariance des séries et un échec de modélisation ARIMA, j'ai implémenté une extraction manuelle de variables sur les courbes. Puis j'ai effectué une sélection de ces variables grâce à une régression logistique et une forêt d'arbres aléatoires. Les résultats n'étant pas assez concluants, principalement à cause de la discontinuité des courbes, j'ai utilisé diverses architectures de réseaux de neurones et enrichi les données afin d'améliorer les prédictions.

Ce stage m'a permis d'améliorer mes compétences en programmation Python, d'augmenter ma maîtrise des bibliothèques liées à l'apprentissage profond, à la construction d'arbres, améliorer mon esprit d'analyse et de déduction et de travailler mes compétences de présentation.

<sup>\*.</sup> Adresse de l'entreprise : 4, Place Jussieu, 75005 Paris

## **SUMMIT Sorbonne Université**

Mathématiques et Applications \*

Responsables: Madame Nora AISSIOUENE et Messieurs Mohammed BOU- ZIANE et Idir

**MEHIDI** 

**Étudiante : Marine DUMON** 

Sujet du stage : Méthodes de calage de paramètre dans un réseau de distribution d'eau po-

table

#### Résumé

La distribution de l'eau potable jusqu'à nos habitations s'effectue grâce à un réseau souterrain composé de conduites et de différentes structures supposées garantir l'approvisionnement en eau potable. Certains paramètres du réseau tels que les hauteurs d'eau dans les noeuds d'approvisionnement ne sont pas connus avec certitude ce qui peut avoir des conséquences importantes pour les usagers comme des fuites non identifiées ou des débits insuffisants.

SUMMIT travaille sur un projet avec l'entreprise Eau de Paris pour créer une méthode de calage de paramètres et ainsi concevoir un système de surveillance en temps réel de son réseau de distribution d'eau potable. C'est dans ce projet que je suis intervenue afin d'adapter en temps réel la méthode de calage déjà implémentée et de trouver des méthodes afin d'améliorer les résultats.

Après une étude bibliographique sur le modèle mathématique et la simulation numérique de la circulation de l'eau dans un réseau de distribution, j'ai effectué plusieurs études afin de comprendre, d'évaluer et de choisir une fonction coût à minimiser lors du calage. En particulier, j'ai travaillé sur une analyse de paysage destinée à analyser et comprendre les difficultés de l'optimisation puis sur une recherche d'indicateurs de fonctions coût afin de proposer une méthode d'évaluation de ces dernières. Plusieurs méthodes de calage en temps réel ont ensuite été testées pour obtenir des améliorations dans les résultats du calage tout d'abord sur des cas simples mais également sur des cas réels.

<sup>\*.</sup> Adresse de l'entreprise : 4, Place Jussieu, 75005 Paris

## Les étudiants de l'année 2021-2022

Nom	Prénom	Entreprise	Option	Page
AMAMA	Zied	EDF R&D	m-app	6
BANDELIER	Svevo	Cy Cergy Paris Université	m	5
BONAMY PARRILLA	Nathalie	ONERA	m	9
DENON	Raydje	SUMMIT	ps	11
DJAOUT	El hacène	SUMMIT	ps	
DUMON	Marine	SUMMIT	ps	12
JERNIVAL	David	IFPEN	m	8
KOBA	Lionel	Ad's Up	ps	3
MKOUBOI	Saif-Ul-Islam.	Safran	ps	10
WANG	Chengyu	Hôtel des Lys	ps	7
WEISMAN	Anna	Axa	m-app	4

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