

Internship in the laboratory of Medical Psychology and Behavioral Neurobiology, Tübingen

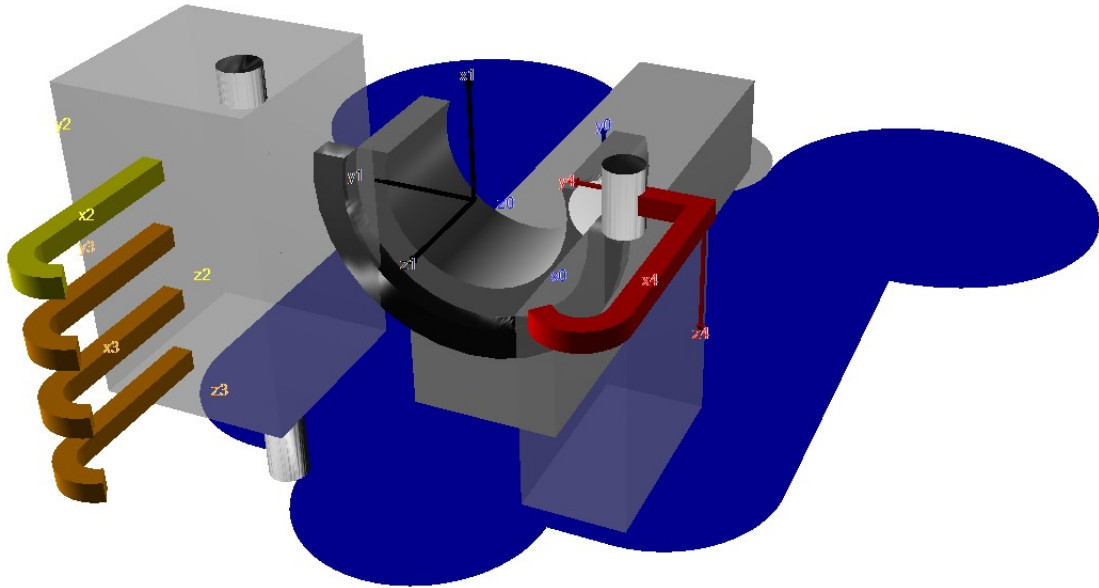


Figure 1: The Exoskeleton

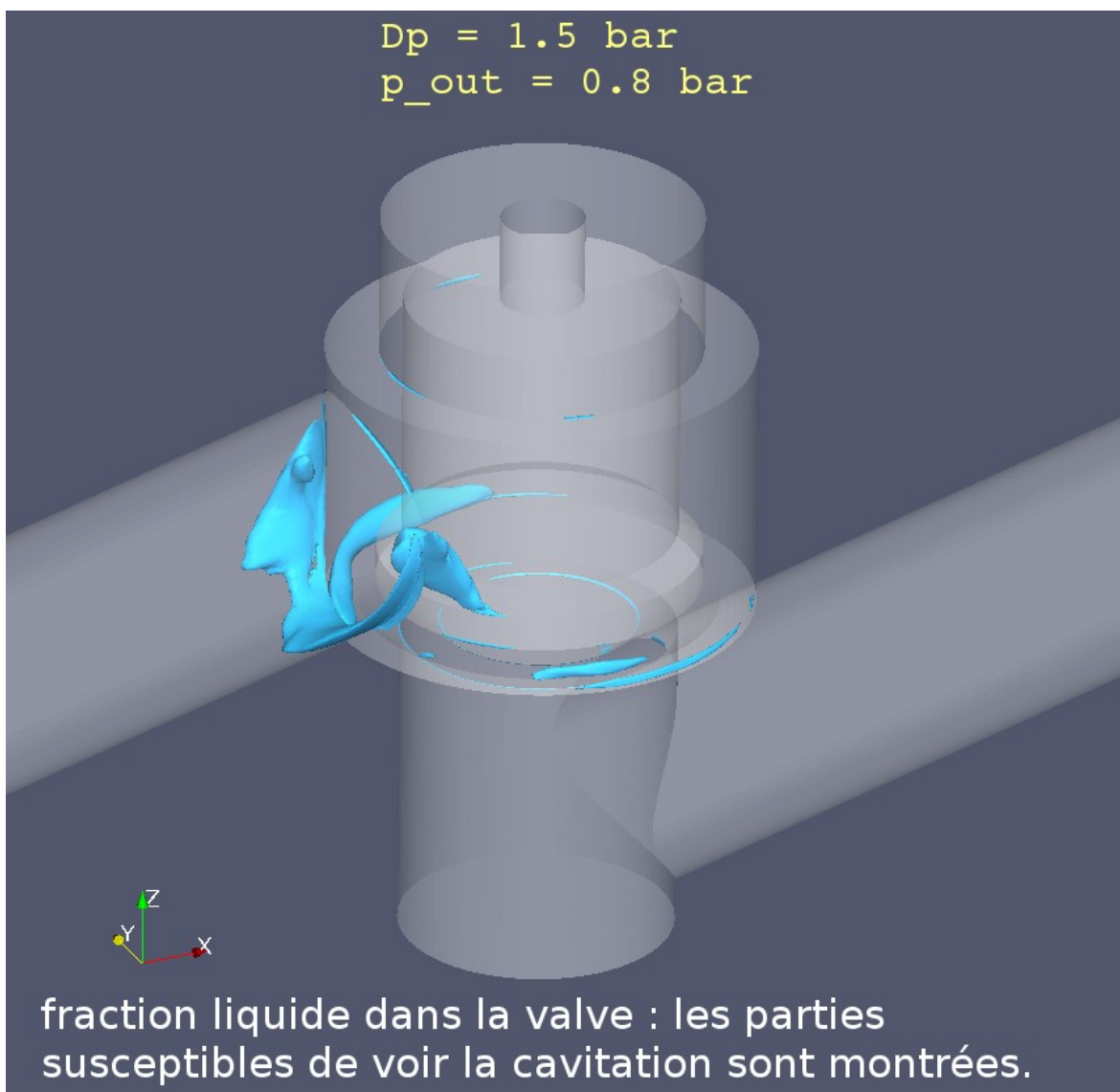
I have worked within a group of researchers investigating on neuro rehabilitation of stroke patients. They are using a robotic exoskeleton of the hand and the forearm in order to restore the normal activity of the arm of the patients. The name of the robot is the Exo. During the experiments, the Exo is sending position values of all its degrees of freedom. In this internship I have developed a software for its visualization in 3D. The software has 2 parts : an online part and an offline one. The online part is dedicated to the view of the robot in real time. This means that while the robot is moving, my virtual model is also moving. The offline part consists in visualizing data of the robot after the experiments, so not in real time.

The software is very useful to my colleagues as they are able to see precisely what the robot is doing from all the angles they wish and they can also view the source of errors of the robot as they can simulate whatever they want.

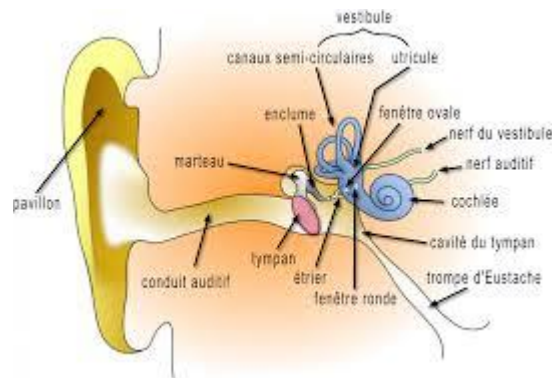
La cavitation est un phénomène qui apparaît quand, à température constante, la pression d'un liquide diminue assez pour être inférieure à la pression de saturation. Dès lors, des bulles de gaz se forment : c'est la cavitation. Ces bulles peuvent alors se déplacer dans la valve / conduite, et imploder peu après. Ce phénomène provoque alors une onde sonore, et peut dégrader les matériaux environnant.

La norme européenne existante concernant la détection de la cavitation dans les valves repose seulement un critère acoustique. Cependant la cavitation peut avoir lieu sans bruit une fois qu'un régime critique est atteint.

Ce stage à l'université de la Corogne (Espagne, Galice), s'inscrit dans l'établissement d'une nouvelle norme sur la détection de cavitation dans les valves de contrôle (qui peuvent notamment jouer un rôle important dans les centrales nucléaires). J'ai donc du modéliser des écoulements avec cavitation dans une valve, afin de voir comment la cavitation affecte le débit en sortie.

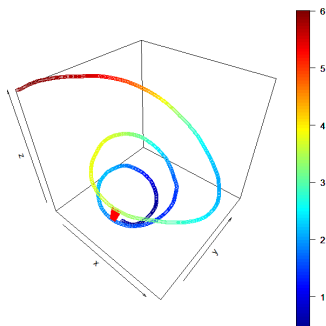


Utilisation d'Analyse Statistique pour l'Etude Anatomique de la Cochlée

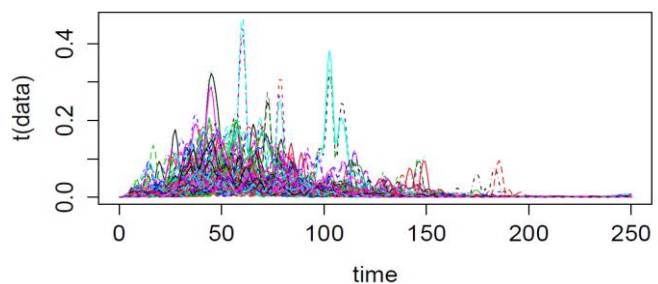


La cochlée représente la partie auditive de l'oreille interne et est localisée dans l'os temporal. Son nom provient du grec qui signifie "coquillage". Cet organe creux ayant un axe osseux communie avec l'oreille moyenne et constitue la dernière étape de l'intégration du son avant le nerf auditif. Pour faire simple la cochlée a pour rôle, grâce à un jeu de vibrations, de transformer les sons en signaux électriques qui vont stimuler le système nerveux et ainsi envoyer l'information jusqu'au cerveau.

Le but du stage a été de présenter aux biologistes et au médecin des éléments concrets qui leur permettent de différencier les cochlées des hommes et des femmes, les taxons *Homo*, *Paranthropus*, *Australopithecus*, *Gorilla* et *Pongo*, et de regarder si certaines caractéristiques sont propres à chaque individu. Statistiquement, le but était de prédire la variable à expliquer $Y=f(X)$ qui était, dans la première partie du stage le sexe, et dans la seconde le taxon, avec X les distances point à point des matrices numérisées en 3D avec 1000 points de mesure dans un premier temps, puis les transformées de Fourier des otoémissions ensuite. Il s'agissait donc de trouver la fonction f qui prédit le mieux à partir de diverses méthodes.



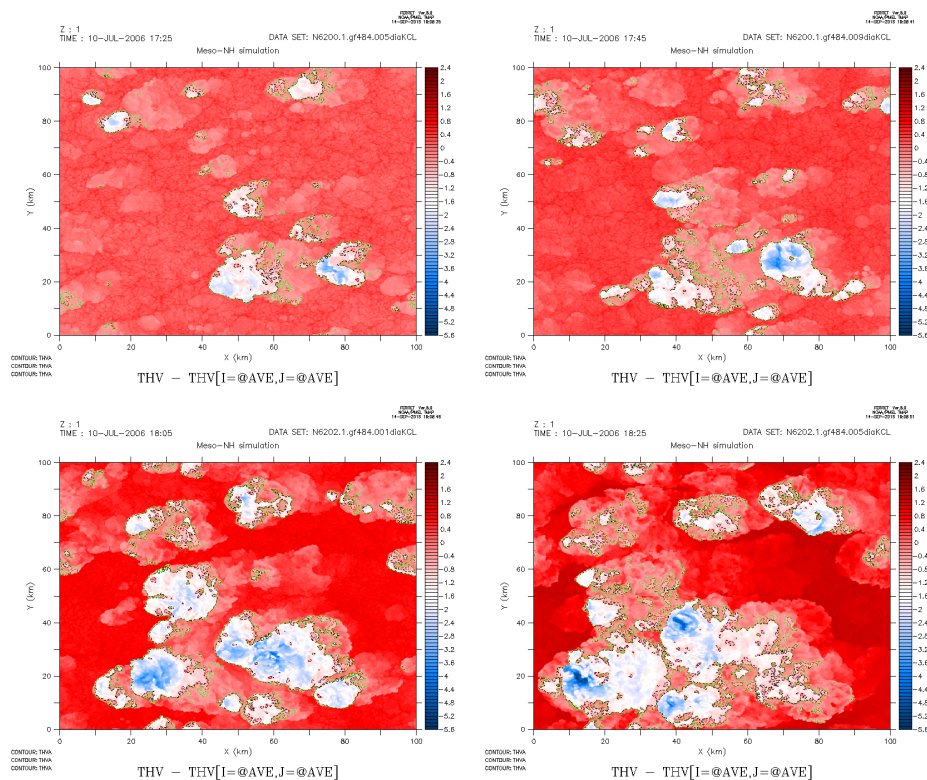
Représentation des distances jugées significativement différentes entre les hommes et les femmes



Transformées de Fourier des courbes d'otoémissions

Dynamique des courants de densité simulés par LES: évaluation et apport pour la paramétrisation

La formation de courants de densité, courants d'air descendant qui se forment lors de l'évaporation des précipitations, est un phénomène fréquemment associé aux épisodes orageux, notamment dans les zones relativement arides comme au Sahel, en Afrique de l'Ouest. Ce phénomène est encore peu représenté dans les modèles de climat. Ces structures prennent la forme de poches froides qui s'étalent à la surface de la terre, et génèrent de fortes rafales de vent, qui peuvent à leur tour engendrer de la convection profonde. La connaissance de ces structures est encore limitée du fait de leur difficile observation. Lors de la campagne de terrain effectuée en 2006 dans le cadre du programme d'Analyses Multidisciplinaires de la Mous-son Africaine, ou AMMA, mené par différents acteurs internationaux et initié par le CNRM, un système convectif (orage) s'est initié à proximité d'une station de mesure. Associé à ce système, un courant de densité a pu être observé. De plus, l'équipe d'accueil a effectué plusieurs simulations à très haute résolution, reproduisant ce cas d'étude. Ces simulations sont de type Large Eddy Simulation (LES) et permettent de résoudre précisément les processus turbulents. Des schémas numérique modélisent les phénomènes microphysiques. Ce stage a pour objectif d'analyser, grâce à des outils statistiques et numériques, les caractéristiques des courants de densité représentés numériquement et d'en comprendre leur cycle de vie. Pour ce faire, nous avons couplé les simulations à un outil de tracking, ce qui nous a permis de suivre l'évolution des caractéristiques morphologiques, thermodynamiques et dynamiques des courants de densité au cours de leur développement. Dans un objectif de généralisation, nous avons également étudié des simulations à plus faible résolution sur un domaine couvrant la majeure partie de l'Afrique de l'ouest.



Stage à Météo France

Mise en oeuvre et évaluation d'une méthode automatisée d'homogénéisation des séries

Une série d'observations météorologiques est dite homogène lorsque les conditions de mesure n'ont pas varié au cours du temps. La plupart du temps, les changements des conditions de mesures provoquent des sauts dans les séries, appelés ruptures (traits verticaux sur la *Figure 1.*). Elles ne sont cependant pas le reflet de la réelle évolution climatique. La détection et la correction de ses ruptures est donc indispensable avant toute étude climatique fiable (*Figure 2.*) : il s'agit de l'homogénéisation. Actuellement, un programme d'homogénéisation appelé HOMER est reconnu et utilisé dans le monde entier. Il nécessite cependant une intervention humaine qui peut être longue, fastidieuse et coûteuse de part la formation nécessaire. Une demande de plus en plus importante est donc faite sur le développement d'une méthode automatisée d'homogénéisation des séries climatiques. Alexis Hannart a essayé de répondre à cette problématique dans son article Hannart *et al.*(2014). Mon travail consistait à implémenter cette méthode et à la tester sur plusieurs jeux de données afin de voir si les résultats étaient aussi satisfaisants que ceux obtenus manuellement. Afin de se rapprocher au plus près des résultats souhaités, le protocole optimal consiste à utiliser la méthode de manière semi-automatique. De cette manière, le protocole fournit des premiers résultats très satisfaisants pour les températures minimales et maximales. Pour les précipitations et les durées d'insolation les performances sont cependant moins bonnes. Bien que d'autres tests soient encore nécessaires afin d'évaluer la méthode de manière plus systématique, cette dernière constitue cependant un gain de temps considérable et une importante aide à la décision pour les personnes en charge de l'homogénéisation.

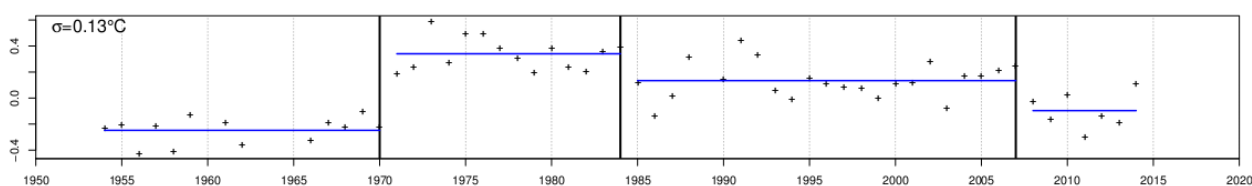


Figure 1. Ruptures présentes sur une série climatique

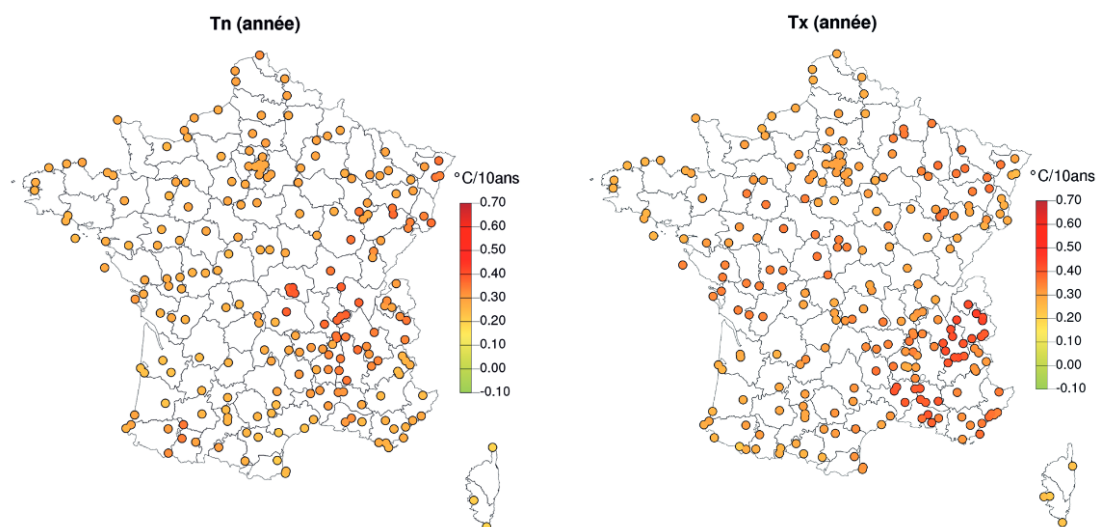


Figure 2. Tendances annuelles des séries homogénéisées de températures minimales (Tn) et maximales (Tx) en °C par décennie, sur la période 1959-2009. Les points cerclés de noir indiquent les tendances significatives au seuil de confiance 95%.

Analyse de Sensibilité des Modèles Agronomiques à l'estimation de la Réserve en eau des Sols

Les sols, en tant que réservoirs d'eau verte, sont essentiels dans les écosystèmes agricoles. Leur fonctionnement hydrique est décrit dans les modèles agronomiques par leur Réserve Utile (RU), i.e. la quantité d'eau disponible pour la croissance de la plante. Cette propriété hydrique peut être estimée selon différentes méthodes, chacune introduisant une incertitude liée aux erreurs de mesure. L'objectif du projet RUEdesSOLS est, entre autres, d'évaluer l'impact du choix de la méthode et de l'incertitude associée sur les sorties des modèles (rendement, qualité de la production, besoin en eau, etc.), afin de proposer une méthodologie pour l'estimation de la RU qui puisse être utilisée dans un large champ de conditions agro-pédo-climatiques.

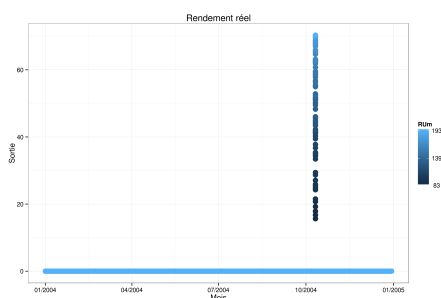


FIGURE 1 – Rendement calculé en sortie du modèle à la date de récolte, dégradé de couleur selon la valeur de la RU en entrée du modèle

Mon stage consiste à analyser d'abord la relation entre l'incertitude sur la RU et l'incertitude des sorties du modèle agronomique AqYield, par exemple le rendement et la satisfaction hydrique. À l'aide du logiciel R, je vais mettre en oeuvre des solutions efficaces par simulation pour la propagation d'incertitudes et développer des stratégies algorithmiques pour trouver des seuils. Ensuite, je pourrai hiérarchiser l'influence de la RU par rapport à d'autres facteurs (le climat, le taux d'argile, etc.) grâce à une analyse de sensibilité.

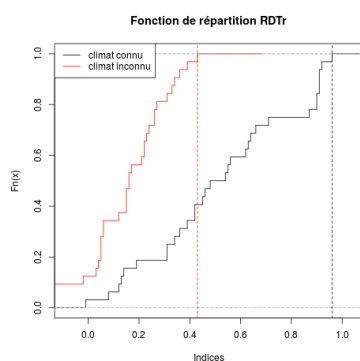
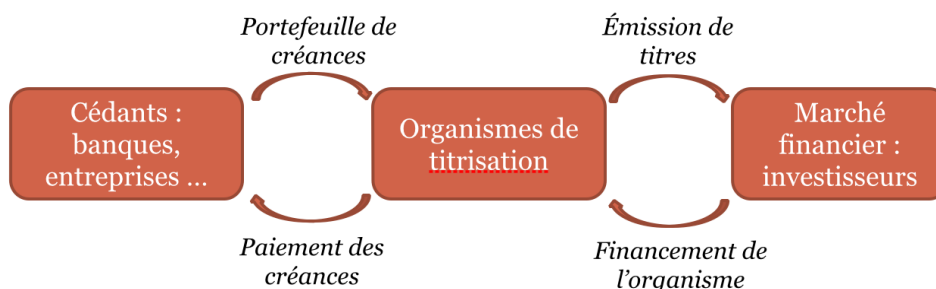


FIGURE 2 – Fonction de répartition de l'indice de sensibilité du premier ordre associé à la RU selon si le climat est fixé (climat connu) ou s'il est considéré comme un facteur (climat inconnu), pour l'observation de la sortie rendement

Résumé de stage

Chargé d'analyse statistique transversale sur la titrisation en France

Le stage s'est déroulé à la Banque de France et plus particulièrement au sein de la Direction Générale des Statistiques (DGS). J'ai rejoint le service de l'Épargne Financière dans le but de travailler sur un projet transversal relatif au processus de titrisation, notamment sur le logiciel statistique SAS.

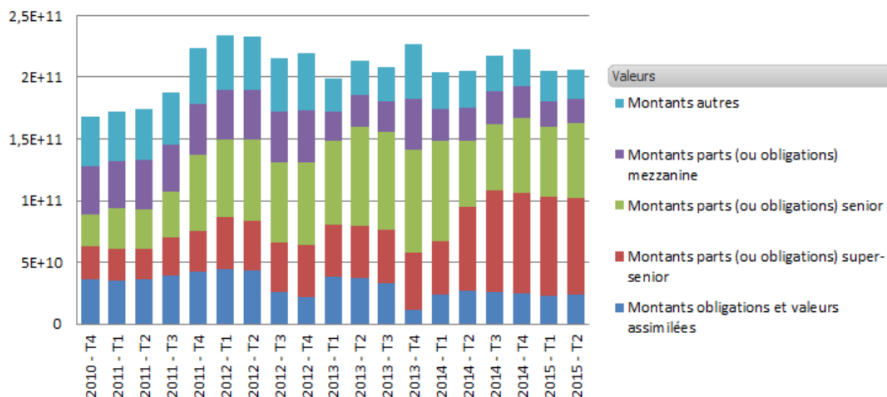


La titrisation se définit comme une technique financière visant à rendre liquide et négociable sur un marché financier un produit initialement non liquide (tel qu'un crédit bancaire), et contribuant ainsi au bon fonctionnement des circuits de financement de l'économie.

Au sein de la DGS, cinq services présentent dans leur activité la collecte d'informations concernant la titrisation en France, mais à des niveaux différents et auprès d'organismes divers. Prises individuellement, ces collectes ne permettent pas d'avoir une vision globale du poids de la titrisation en France, ni de présenter des résultats statistiques généraux.

Ma mission a d'abord consisté à organiser ces différentes informations sous forme d'une base de données commune. C'est la première phase nécessaire à l'analyse statistique : l'exploration des données (étude des données, détections de valeurs atypiques, etc.). L'objectif était, à la suite de cette partie « gestion de données », de procéder à une analyse statistique descriptive de la base obtenue (répartition des secteurs emprunteurs, proportion d'actifs sous-jacents, évolution de flux, etc.). Finalement, les résultats de cette analyse devaient être présentés sous la forme d'un tableau de bord formé à partir d'indicateurs graphiques et nécessitant le moins de mises à jour manuelles possibles.

Montants titres émis



“Détection de déplacements humains”

ALSETT, Toulouse

De plus en plus d'individus ont constamment leur smartphone sur eux lors de leur déplacement. Le smartphone subit donc les mêmes accélérations, décélérations, arrêts, changements de directions et, plus généralement, les mêmes forces que l'utilisateur.

L'intérêt du stage est de déterminer si, à partir des signaux transmis par les capteurs d'un smartphone, nous sommes capable d'identifier l'activité de déplacement de l'individu. Nous voulons être capable de prédire : “étant donné l'allure des signaux transmis par les capteurs de ton smartphone, je sais que tu es en train de courir”.

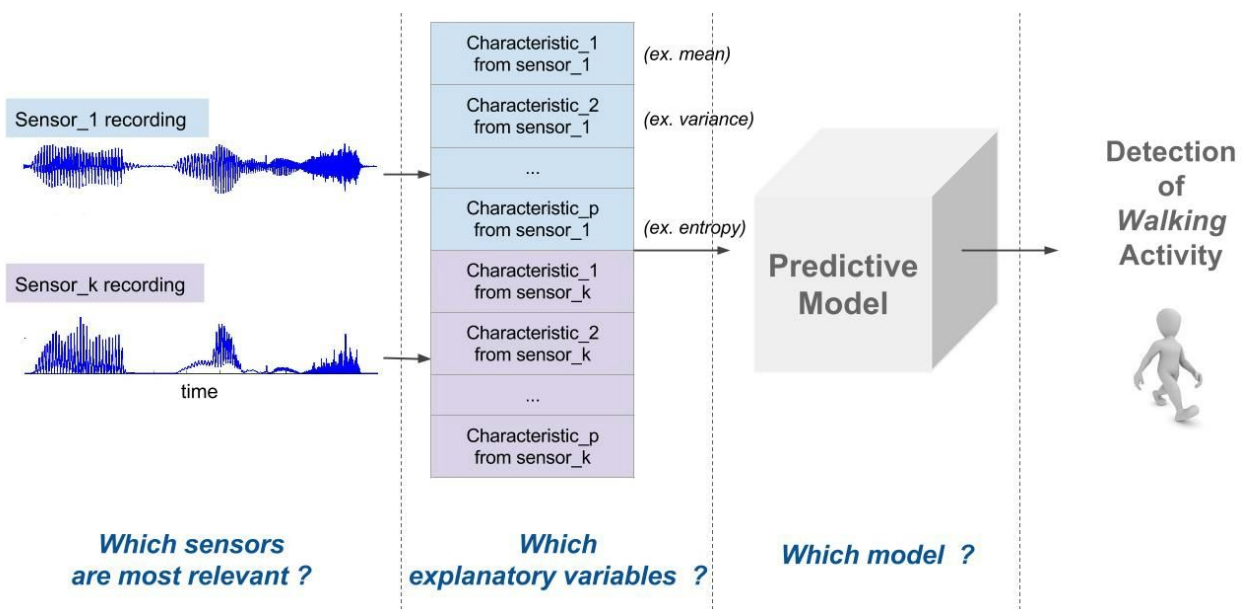
Pour simplifier, nous ne considérons que 6 déplacements élémentaires : la marche, la marche rapide, la course, le piétinement, l'immobilité et les déplacements lents et aléatoires. Ces déplacements constituent donc les modalités de ma variable cible.

Tous les capteurs du smartphone (accélération linéaire, gyroscope, accéléromètre, etc...) sont pris en compte dans l'étude. Ces capteurs et plus précisément certaines caractéristiques sur leurs signaux temporels sont les variables explicatives du modèle.

La première étape est la récolte des données. Elle demande d'imaginer assez de scénarios pour couvrir l'ensemble des déplacements possibles et que ces derniers soient réalisés par des profils de tout genre.

Une fois les données collectées, un travail d'exploration est nécessaire pour déceler une potentielle défaillance de certains capteurs et pour comprendre leurs comportements généraux.

Une sélection des capteurs les plus pertinents amène à l'obtention de la matrice sur laquelle les algorithmes de Machine Learning sont appliqués. Plusieurs approches ont été testées : le Partial Least Square Regression et notamment la version sparse, les agrégations de modèles tels que le Random Forest, le Bagging et le Boosting.



N-SIDE, en Belgique

L'objectif de ce stage était de produire un outil dédié à la régression gaussienne, utilisable et compréhensible par des personnes non statisticiennes. La première partie du stage consistait à faire un état de l'art des pratiques de l'entreprise, ainsi qu'à analyser les besoins en statistiques. Les besoins les plus importants résidaient dans le pré traitement des données, les méthodes de sélection de variables, et les critères de comparaison des modèles. Pour y remédier, des séances de formation hebdomadaires ont été organisées. Lors de ces séances, les outils statistiques d'exploration multidimensionnelle mais également les différents critères applicables aux modèles de régression gaussienne (test de Fisher, Akaïke, erreur quadratique moyenne,...) ont été présentés. Pour faciliter et automatiser le plus possible les études statistiques des consultants, la deuxième partie du stage consistait à créer un package R. Ce package permet de détecter des éventuels outliers dans une base de données, de comparer des modèles de régression gaussienne en mettant en place une séparation entre données apprentissage et données tests, mais également de visualiser les résidus des modèles. Le package a été créé pour les consultants de l'entreprise : un code couleur a été installé afin de les prévenir lorsqu'une situation est à risque, les noms des fonctions ainsi que leurs arguments sont simples à retenir, une documentation très détaillée avec des exemples est disponible. Tous les graphiques sont enregistrés automatiquement dans leur répertoire de travail, ce qui leur permet de garder une trace de l'avancement de leur étude.

