



RAPPORT DE STAGE DE M1 MATHÉMATIQUES APPLIQUÉES POUR L'INGÉNIERIE, L'INDUSTRIE ET L'INNOVATION (MAPI³)

Maintenance prédictive sur vélo

Benjamin BEYRIE
M1 MAPI³

Tuteurs :
Antoine PICOT
Gilles TREDAN

Enseignant responsable :
Sophie JAN

Stage effectué du 24 avril 2017 au 26 juillet 2017 conjointement avec le laboratoire LAPLACE
site ENSEEIHT, et le LAAS

Remerciements

Je voudrais tout d'abord remercier mes deux maîtres de stage, Antoine Picot (LAPLACE) et Gilles Tredan (LAAS), qui m'ont donné leur confiance et m'ont accompagné tout au long de ces 3 mois de stage.

Je remercie également Sophie Jan, qui a accepté l'invitation de devenir ma tutrice référente de l'université Paul Sabatier.

Je tiens aussi à remercier Christophe Bertero, Roberto Pasqua, et toute l'équipe du LAAS travaillant sur le projet CLUE, qui m'ont beaucoup aidé sans jamais faillir.

Je souhaite finalement remercier Marie-Helene Vignal, Luca Amodei ainsi que toute l'équipe pédagogique de la formation MApI³, qui nous soutiennent toute l'année au dépend de leur temps personnel.

Résumé

Ce stage a pour objectif d'explorer les possibilités de maintenance prédictive offertes par un accéléromètre embarqué sur un vélo.

Pour ce faire, nous avons constitué un jeu de données de références contenant des enregistrements d'accélération de vélos à l'état connu (gonflé, dégonflé,...) sur un parcours type. Dès lors, nous avons pu étudier les signaux accélérométriques avec des séries de Fourier ou par des décompositions en ondelettes, plusieurs indicateurs ont été extraits des signaux accélérométriques, à la fois en temporel et en fréquentiel.

Une fois traitées, nous avons utilisé des techniques d'apprentissage supervisé sur les indicateurs extraits des données. Notre objectif était de prédire l'état du vélo à l'aide des seules données accélérométriques.

Les scores obtenus sur la précision des prédictions après l'apprentissage supervisé nous permettent de dire qu'il est possible de déduire d'un signal accélérométrique, l'état du vélo.

Dans l'avenir, un microcontrôleur embarqué gérera des calculs en local, tandis qu'une transmission longue portée de type LORA enverra les rapports ou jeux de données sur un serveur.

Table des matières

1	Introduction	1
1.1	Établissement d'accueil	1
1.2	Description du projet "CLUE" : des vélos connectés	1
1.3	Principe du stage	2
1.4	Accélémètre et protocole de capture de données	2
1.5	Visualisation des données brutes	4
2	Traitemennt des données accélérométiques	5
2.1	Transformée de fourier	5
2.1.1	Transformée de fourier locale (STFT)	5
2.1.2	Représentation temps/fréquence de la STFT	5
2.2	Transformée en ondelettes	6
2.2.1	Transformée en ondelettes discrètes du signal(DWT)	6
2.2.2	Représentation temps/échelle de la DWT	6
2.3	Indicateurs statistiques	7
3	Apprentissage supervisé à 2 états	8
3.1	Principe général	8
3.1.1	Extraire les données	8
3.1.2	Séparation des données	8
3.1.3	Préparation et contrôle	8
3.1.4	Entrainement de l'échantillon d'apprentissage	9
3.1.5	Prédiction de l'échantillon de validation	9
3.2	Fourier ou Ondelettes ?	10
3.2.1	Protocole de comparaison	10
3.2.2	Résultats	10
3.3	Choix des axes	13
3.3.1	Protocole de comparaison	13
3.3.2	Résultats	13
3.4	Bilan	14
4	Apprentissage supervisé à 3 états	15
4.1	Résultats	15
4.1.1	Echantillons de validation :	15
4.1.2	Echantillons de test :	16
4.2	Conclusion sur la partie apprentissage supervisé	16
5	Discussions	17
5.1	Dimension embarquée	17
5.1.1	Principe	17
5.1.2	Coût de calcul	17
5.2	Liste des pannes possible	18
5.3	Autocritique	18
5.3.1	Points négatifs	18
5.3.2	Points positifs	18
6	Conclusion	19

Chapitre 1

Introduction

1.1 Établissement d'accueil

Le Laboratoire Plasma et Conversion d'Energie est une Unité Mixte de Recherche du Centre National de la Recherche Scientifique (CNRS), de l'Institut National Polytechnique de Toulouse (INPT) et de l'Université Paul Sabatier (UPS).

Les recherches menées au LAPLACE, allant de l'étude des mécanismes physiques de conversion d'énergie jusqu'aux systèmes qui les mettent en œuvre, se situent donc au cœur de ces préoccupations. Les grands domaines d'applications englobent la production, le transport, la gestion, la conversion et l'usage de l'électricité.

Le groupe CODIASE (COmmande et DIAGnostique des Systèmes Electriques), groupe du LAPLACE dans lequel j'ai effectué mon stage, voit ses activités se situer au carrefour de l'énergie électrique, de l'automatique, et du traitement du signal . Elles concernent un large domaine d'applications couvrant les dispositifs de conversion de l'énergie électrique comme les machines électriques ou les générateurs électrochimiques et les dispositifs de traitement de l'énergie électrique comme les convertisseurs statiques et autres procédés visant à améliorer la qualité de l'énergie électrique mise à disposition.

Le savoir-faire du groupe repose sur la création d'expertises dans les domaines de la commande et du diagnostic de ces systèmes avec un axe méthodologique fort empreint de modélisation fine et de démarches théoriques adaptées.

1.2 Description du projet "CLUE" : des vélos connectés

L'objectif du projet est d'équiper une partie des vélos évoluant dans Toulouse d'un ensemble de capteurs afin d'étudier les déplacements des usagers, mais aussi de profiter du réseau de capteurs mobiles ainsi déployé pour collecter des informations sur l'environnement urbain.

La Maison du Vélo, partenaire du projet, est une association qui propose notamment la location de vélos à la semaine, au mois ou à l'année. Une partie de ces vélos à déjà été instrumentée. La rotation rapide des locations (un mois en majorité) permettra d'assurer la maintenance des dispositifs. Aussi, un apport en connaissance des techniciens présents est un plus.

Chaque dispositif est constitué d'un ensemble de capteurs, d'une petite réserve d'énergie alimentée par la dynamo des vélos, d'une solution de stockage, et d'une interface réseau permettant la collecte des données.

La collecte des données s'effectue principalement à l'aide d'un réseau Lora¹ déployé sur l'agglomération. La portée espérée est de l'ordre de quelques kilomètres, et des stations de collecte seront déployées chez les partenaires et soutiens du projet. Les données seront de plus stockées localement sur chaque plateforme et récupérées en fin de location.

1. Transmition sans fil longue portée à bas débit

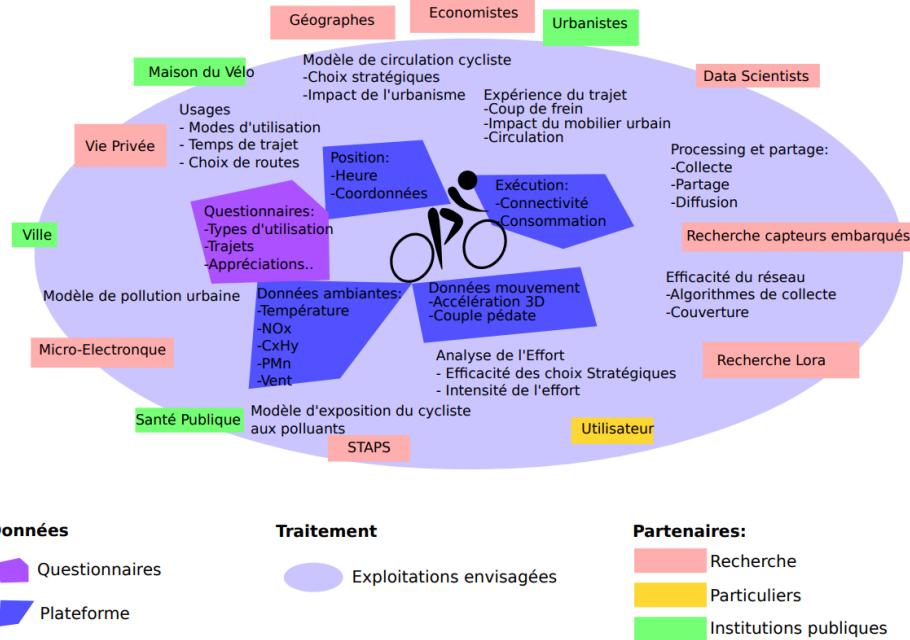


Illustration 1: Schéma d'exploitation des données

1.3 Principe du stage

La maintenance prédictive des vélos est un enjeu important pour éviter des coûts non nécessaires sur l'usure des vélos.

La maintenance des vélos de La Maison du Vélo, par exemple, a un coût matériel mais aussi humain (un pneu dégonflé va peut-être entraîner une crevaison ou une jante déformée si elle est vue trop tard).

La maintenance prédictives des vélos à l'aide de données accélémétriques permettra de voir au plus vite un défaut à corriger, ce qui permettra le maintien de la flotte.

1.4 Accéléromètre et protocole de capture de données

L'accéléromètre LIS2DH12 (fig 1.1) est de type 3 axes (X, Y et Z).

Il peut mesurer des accélérations dans les bornes de 2g/4g/8g/16g et est capable de capturer les informations à une fréquence allant de 1 Hz à 5,3 kHz.

Il a une résolution de 8 bits en sortie des données.

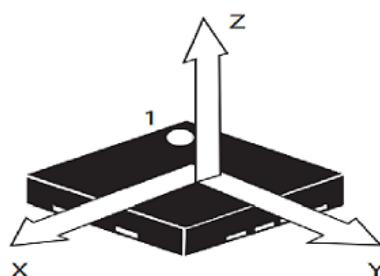


FIGURE 1.1 – Accéléromètre

L'accéléromètre est soudé sur un circuit imprimé (fig 1.2), avec entre autre un capteur GPS.



FIGURE 1.2 – Circuit imprimé

Protocole de capture des données d'apprentissage :

Nous avons paramétré l'accéléromètre sur la plage -2g/+2g et 32Hz pour les acquisitions car les accélérations à vélo ne dépassent pas les 1g en théorie, et pour avoir un domaine fréquentiel lors du traitement du signal allant de 0Hz à 16Hz.

Le "golden run" est réalisé sur un parcours fixé (fig 1.3, Autour du LAAS). L'idée est de découpler le problème de la route du problème de la pression des pneus (Le golden run est un cas facile où l'on relache les contraintes).

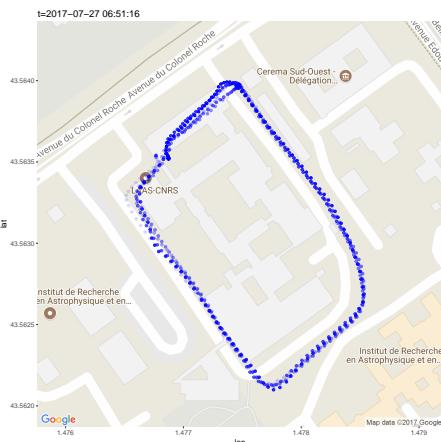


FIGURE 1.3 – Golden run

Il se compose de 5 tours par état du vélo, effectués à une vitesse sensiblement constante. Un tour dure environ 2 minutes.

Ce terrain comporte une partie route, une partie pavée, du plat, des trottoirs, nous l'avons donc choisi car il est complet.

1.5 Visualisation des données brutes

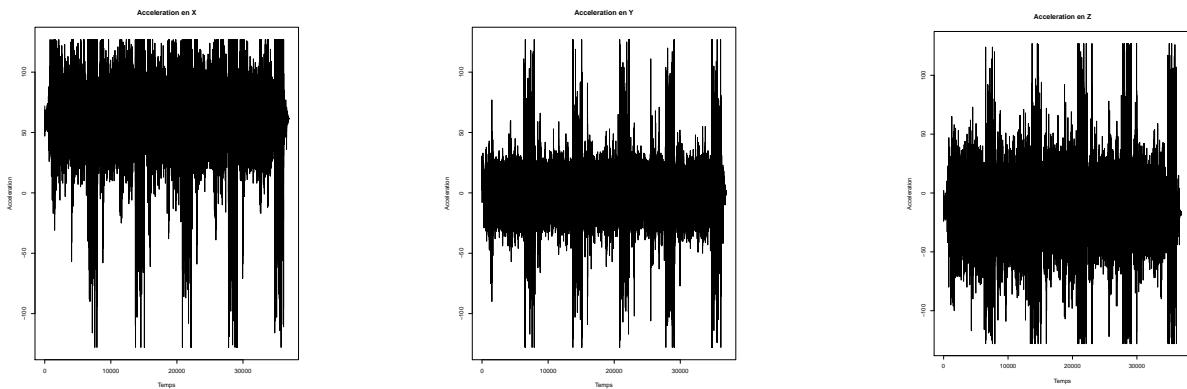


FIGURE 1.4 – Accélérations en X, Y et Z

L'axe X représente la verticale au sol, il prend donc naturellement les 1g de la gravité terrestre en plus, l'axe Y est dirigé sur le côté, quand l'axe Z est dirigé vers l'avant. On constate qu'un pattern se reproduit 5 fois, ce qui est cohérent avec les 5 tours réalisés.

De plus nous pouvons constater que les données sont bruitées, c'est pour cela que nous avons besoin d'effectuer du pré-traitement.

Chapitre 2

Traitement des données accélérométriques

Nous avons employées deux méthodes pour traiter et débruiter ces données.

1. **La transformée de fourier** : Représentation temps/fréquence, l'intervale des fréquence étant de 0Hz à 16Hz.
2. **La transformée en ondelettes** : Représentatiion temps/échelle, avec une décomposition en ondelettes jusqu'au niveau 8.

Ces méthodes sont intéressantes d'une part pour leur tolérance au bruit, d'autre part pour leur coût de calcul, et enfin pour leur disponibilité.

2.1 Transformée de fourier

Rappel de la définition mathématique de la transformée de fourier.

$$S(k) = \sum_{n=0}^{N-1} s(n) \cdot e^{-2i\pi k \frac{n}{N}}$$

avec s le signal, N la taille de l'échantillon et $0 \leq k \leq N$

2.1.1 Transformée de fourier locale (STFT)

Le principe est le même que la transformée de fourier, sauf qu'elle est réduite à une fenêtre glissante le long du signal.

$$STFT\{s(n)\}(m, \omega) \equiv S(m, \omega) = \sum_{n=0}^{N-1} s(n)w(n-m) \cdot e^{i\omega n}$$

avec w la fenêtre glissante, et m le facteur de translation.

2.1.2 Représentation temps/fréquence de la STFT

Paramètres de la STFT :

Taille de la fenêtre : 32 ; cela équivaut à un intervalle d'une seconde car on capture à 32Hz.

Distance de décalage de la fenêtre : 32 ; cela équivaut à décaler la fenêtre d'une seconde à l'autre. Il est possible de représenter les résultats de cette transformée de fourier (fig 2.1).

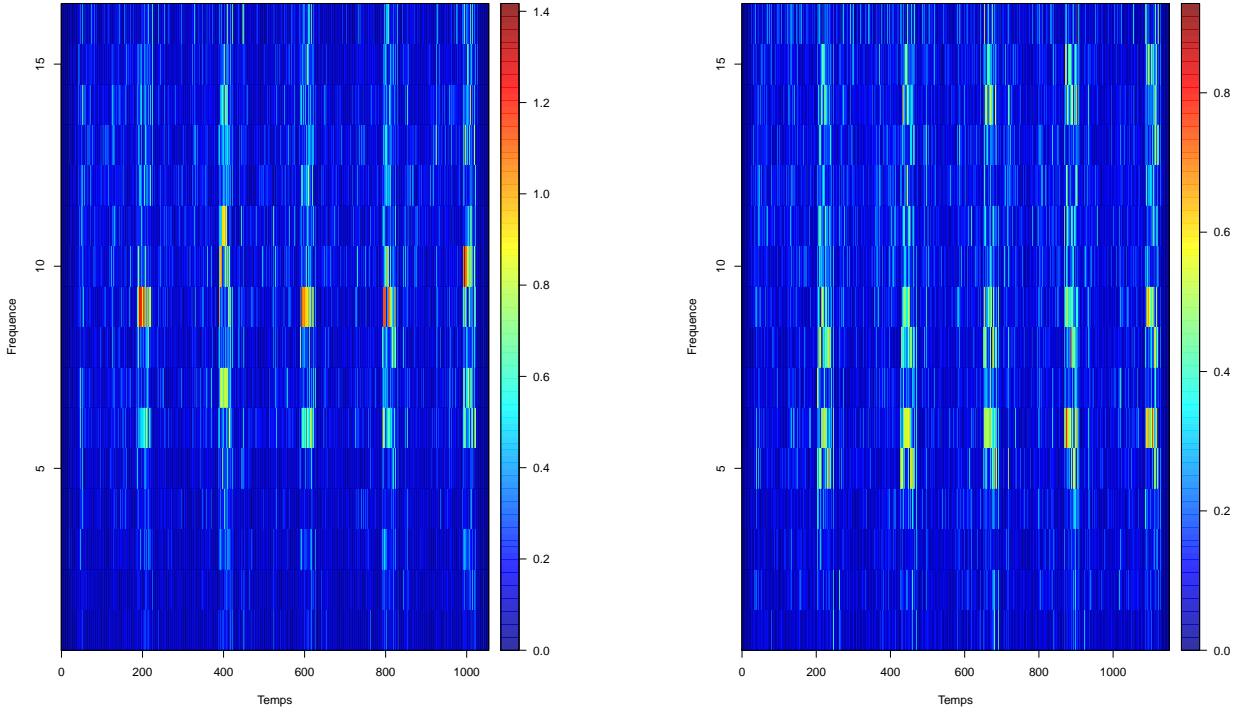


FIGURE 2.1 – Représentation temps/fréquence de la transformée de fourier, à gauche vélo gonflé, à droite vélo dégonflé

On constate encore un patern revenant 5 fois, dans les hautes fréquences autour de 10Hz, correspondant à un fort bruit dans le signal. En réalité, c'est le moment où l'on passe sur la zone pavée.
 On constate aussi que pour un vélo gonflé, le signal est plus fort et plus concentré vers les hautes fréquences, tandis que pour le vélo dégonflé, le signal à tendance à se répartir sur toute la bande de fréquence mais avec une force moindre.

2.2 Transformée en ondelettes

Rappel de la définition mathématique de la transformée en ondelettes.

Soit Ψ une ondelette mère. On définit ainsi une famille $\Psi_{s,\tau}$ (où $\{s,\tau\} \in \mathbb{R}^{+*} \times \mathbb{R}$) d'ondelettes à partir de l'ondelette mère :

$$\forall t \in \mathbb{R}, \Psi_{s,\tau}(t) = \frac{1}{\sqrt{s}} \cdot \Psi\left(\frac{t-\tau}{s}\right)$$

avec τ le facteur de translation et s le facteur d'échelle

2.2.1 Transformée en ondelettes discrètes du signal(DWT)

La transformée en ondelettes discrète translate et dilate l'ondelette selon des valeurs discrètes. Ces coefficients s et τ seront discrétisés de la manière suivante : $s = s_0^j$ et $\tau = k\tau_0 s_0^j$ avec $s_0 > 1$ et $\tau > 0$.

$$DWT\{x(t)\}(\tau_0, s_0) = \frac{1}{\sqrt{s_0^j}} \int_{-\infty}^{\infty} x(t) \Psi\left(\frac{t - k\tau_0 s_0^j}{s_0^j}\right) dt$$

avec s_0^j le facteur d'échelle, τ_0 le facteur de translation, k et j des entiers.

La transformée en ondelettes discrète peut s'interpréter comme un filtre passe bande dont la largeur est déterminée par le facteur s .

2.2.2 Représentation temps/échelle de la DWT

Comme pour la transformée de fourier locale, nous allons prendre des intervalles d'une secondes pour la transformée en ondelettes continues. Il est possible de représenter les résultats de cette transformée en ondelettes (fig 2.2).

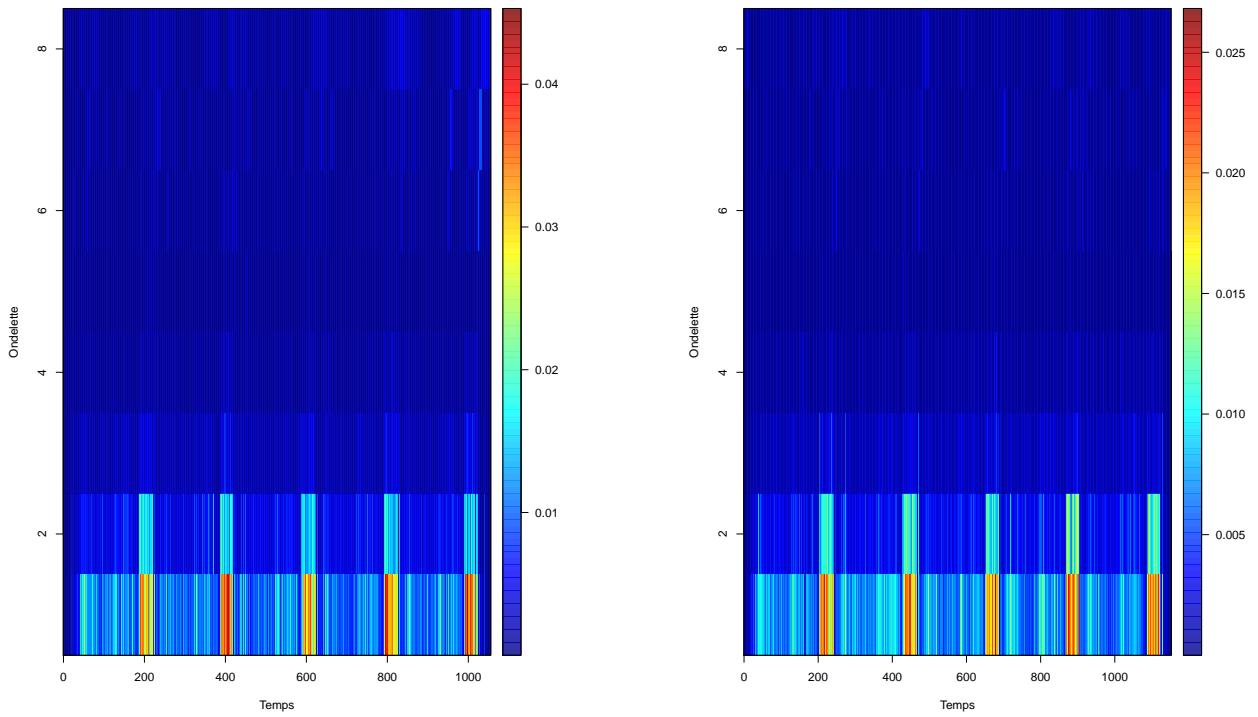


FIGURE 2.2 – Représentation temps/échelle de la décomposition en ondelettes, à gauche vélo gonflé, à droite vélo dégonflé

On constate encore un patern revenant 5 fois, surtout dans les deux premiers niveaux de décompositions en ondelettes. En réalité, c'est le moment où l'on passe sur la zone pavée.
On constate aussi que pour un vélo gonflé, le signal est plus fort que pour le vélo dégonflé (0.04 contre 0.025).

2.3 Indicateurs statistiques

Pour compléter ces deux méthodes et rajouter de l'information utile pour notre futur apprentissage supervisé, nous allons rajouter 8 indicateurs statistiques à chaque seconde.

On construit ces indicateurs pour chaque secondes, ils représentent donc à un instant t , les indicateurs des fréquences allant de 0Hz à 16Hz pour la transformée de fourier, ou bien les indicateurs des échelles allant de 1 à 8 pour les décompositions en ondelettes.

1. Le minimum
2. Le maximum
3. La valeur crête à crête
4. La moyenne
5. La médiane
6. La variance
7. L'asymétrie
8. Le kurtosis

Au final, on a en sortie une matrice avec les lignes en 2 parties.

La première partie des lignes représente soit les fréquences allant de 1Hz à 16Hz de la transformée de fourier du signal accélérométrique, soit les 8 premières décompositions en ondelettes du signal accélérométrique.
La deuxième partie des lignes représente les indicateurs statistiques correspondants.

Chapitre 3

Apprentissage supervisé à 2 états

Maintenant que les signaux sont pré-traités, nous allons essayer de les exploiter pour prédire le niveau de gonflage du pneu.

L'apprentissage supervisé se fait en plusieurs étapes :

1. 1ère étape : Extraire un jeu de données
2. 2ème étape : Séparer ce jeu de données en un échantillon d'apprentissage et un échantillon de validation
3. 3ème étape : Préparer et contrôler les données
4. 4ème étape : Entrainement sur l'échantillon d'apprentissage
5. (4bis étape) : Validation des entraînements sur un échantillon de validation
6. 5ème étape : Prédiction sur l'échantillon test

3.1 Principe général

On suppose $y \in \{\text{gonflé}, \text{dégonflé}\}$, on cherche f tel que $y = f(x)$.
On entraîne donc en faisant $\underset{f}{\operatorname{argmin}} \| f(x) - y \|$.

3.1.1 Extraire les données

Les données se présentent sous forme de dataframes, c'est à dire un grand jeu de données au format matriciel. Une partie de la matrice contient des valeurs (différentes valeurs de fréquences/echelles, identificateurs statistiques).

Une autre partie de la matrice contient des labels correspondant à l'état du vélo lors de la prise (gonflé ou dégonflé).

Cette matrice est de l'ordre de 1000×20 .

3.1.2 Séparation des données

On partitionne notre jeu de données en deux.
Dans notre cas, 50% ira à l'échantillon d'apprentissage, et 50% ira à l'échantillon de validation (échantillon qui sert à valider si l'apprentissage s'est bien passé).

3.1.3 Préparation et contrôle

La préparation des données est vivement recommandée pour améliorer la qualité des résultats. Il existe toute sorte de fonctions pour préparer les données.

Après plusieurs essais, nous choisirons deux méthodes :

1. Centrer et réduire les données : $\frac{W - Esperence}{EcartType}$, avec W un point.

2. Projeter sur un cercle les données : $\frac{W}{\|W\|}$, avec W un point.

On rajoutera ensuite un paramètre de validation croisée pour s'assurer une meilleure fiabilité et réduire la variance.

3.1.4 Entrainement de l'échantillon d'apprentissage

Il existe beaucoup de méthodes permettant d'entraîner un échantillon d'apprentissage.
Après plusieurs essais, nous avons gardé trois modèles robustes :

1. Les réseaux de neurones (nnet) : Un réseau neuronal est l'association, en un graphe plus ou moins complexe, d'objets élémentaires, les neurones formels. Les principaux réseaux se distinguent par l'organisation du graphe (en couches, complets...), c'est-à-dire leur architecture, son niveau de complexité (le nombre de neurones, présence ou non de boucles de rétroaction dans le réseau), par le type des neurones (leurs fonctions de transition ou d'activation) et enfin par l'objectif visé : apprentissage supervisé.
2. Random forest (rf) : Random forest est une technique d'apprentissage supervisée qui combine une technique d'agrégation, le "Bagging", et une technique particulière d'induction d'arbres de décision. Lors de la construction de l'arbre, pour initier la segmentation d'un noeud, la méthode effectue dans un premier temps une sélection aléatoire parmi les variables candidates, sélection à partir de laquelle, dans un deuxième temps, elle cherche la variable de segmentation.
3. Le boosting (boost) : Le principe du boosting est issu de la combinaison de classificateurs (appelés également hypothèses). Par itérations successives, la connaissance d'un classifieur faible (weak classifier) est ajoutée au classifieur final (strong classifier). On appelle apprenant faible un algorithme qui fournit des classificateurs faibles, capables de reconnaître deux classes au moins aussi bien que le hasard ne le ferait (c'est-à-dire qu'il ne se trompe pas plus d'une fois sur deux en moyenne, si la distribution des classes est équilibrée). Le classifieur fourni est pondéré par la qualité de sa classification : mieux il classe, plus il sera important. Les exemples mal classés sont boostés pour qu'ils aient davantage d'importance vis-à-vis de l'apprenant faible au prochain tour, afin qu'il pallie le manque.

3.1.5 Prédiction de l'échantillon de validation

A partir des modèles entraînés, et d'un échantillon de test (échantillon quelconque extérieur à l'échantillon d'apprentissage et de validation), essayer de trouver avec le plus de précision possible l'état du vélo de cet échantillon test.

Principe d'évaluation des classificateurs :

$$\sum_{x,y \in validation} \|\hat{y} - y\| = \sum \|\widehat{f(x)} - y\|$$

avec f le classifieur entraîné, x l'échantillon de test et y la cible.

Si tout se passe bien, la prédiction devrait être éloignée du taux de 50% de points bien classés.

3.2 Fourier ou Ondelettes ?

Nous allons comparer ces deux méthodes pour savoir laquelle sélectionner dans notre objectif de réaliser de la maintenance prédictive sur des vélos à partir de signaux accélérométriques.

3.2.1 Protocole de comparaison

Nous observons sur 3 méthodes d'apprentissage supervisé, quelle type de traitement du signal nous permet d'avoir le meilleur score au classement sur un échantillon de test quelconque.

Nous gardons les meilleurs scores obtenus pour chaque méthode.

Nous utilisons en données les axes X, Y et Z (Nous aurions tout aussi pu prendre indépendamment les axes de l'accéléromètre). Nous verrons plus tard le choix des axes. Pour finir, nous comparons aussi les deux méthodes de préparation des données, la méthode "centrée réduite" (scaling and centering que l'on notera sc) et la méthode "projection sur un cercle" (spatialSign que l'on notera ss).

3.2.2 Résultats

Méthode	Score Fourier	Score Ondelettes	Préparation
nnet	Précision : 0.759 Kappa ¹ : 0.516	Précision : 0.787 Kappa : 0.571	sc
rf	Précision : 0.798 Kappa : 0.593	Précision : 0.862 Kappa : 0.724	sc
boost	Précision : 0.795 Kappa : 0.589	Précision : 0.870 Kappa : 0.738	sc
nnet	Précision : 0.780 Kappa : 0.558	Précision : 0.813 Kappa : 0.625	ss
rf	Précision : 0.791 Kappa : 0.579	Précision : 0.832 Kappa : 0.663	ss
boost	Précision : 0.800 Kappa : 0.598	Précision : 0.837 Kappa : 0.673	ss

FIGURE 3.1 – Tableau comparatif entre fourier et ondelettes, échantillon de validation

Méthode	Score Fourier	Score Ondelettes	Préparation
nnet	Précision : 0.697	Précision : 0.576	sc
rf	Précision : 0.539	Précision : 0.654	sc
boost	Précision : 0.548	Précision : 0.673	sc
nnet	Précision : 0.641	Précision : 0.568	ss
rf	Précision : 0.648	Précision : 0.616	ss
boost	Précision : 0.637	Précision : 0.653	ss

FIGURE 3.2 – Tableau comparatif entre fourier et ondelettes, échantillon de test quelconque

Le Kappa de Cohen (Kappa) se mesure comme suivant :

$$K = \frac{P - Pe}{1 - Pe}$$

avec P l'accord relatif entre deux observateurs et Pe la probabilité d'un accord aléatoire. Ici nos scores sont tous supérieurs à 0.5 donc l'apprentissage est bon.

Même si les ondelettes ont de meilleurs résultats sur les échantillons de validations, elles ne sont pas aussi performantes sur des échantillons qui sortent du cadre de l'entraînement. De ce fait, on préférera prendre la méthode transformée de fourier avec pour préparation la projection sur un cercle, qui nous semble plus robuste au changement de terrain avec des précisions de même ordre. De plus, fourier est bien moins gourmand en ressources que les ondelettes, de ce fait, il sera beaucoup facile d'implémenter fourier sur un microcontrôleur.

1. Kappa de Cohen ; mesure l'accord entre deux observateurs ($k < 0$ Désaccord, $0 < k < 0.5$ Accord faible, $0.5 < k < 1$ Accord fort)

Pour rajouter de la fiabilité aux méthodes, nous traçons les "courbes ROC" (receiver operating characteristic) (fig 3.3) liées aux méthodes de classification.

Une "courbe ROC" donne le taux de vrais positifs (fraction des positifs qui sont effectivement détectés) en fonction du taux de faux positifs (fraction des négatifs qui sont détectés (incorrectement)).

Une classification aléatoire nous donnerait une courbe linéaire allant de (0,0) à (1,1). Plus l'aire sous la courbe est grande, meilleure est la classification.

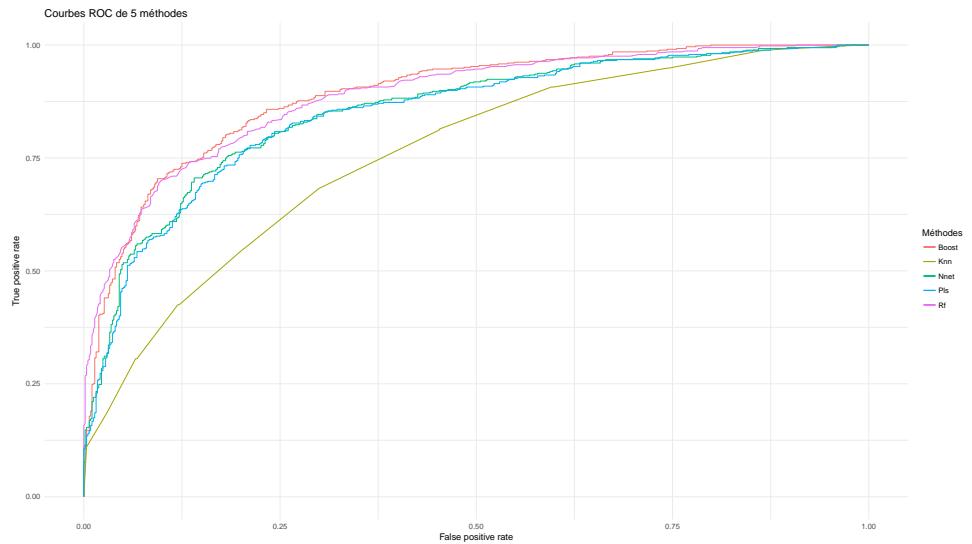


FIGURE 3.3 – Courbes ROC pour 5 classificateurs

On constate que les classificateurs nnet, rf et boost se comportent comme nous le souhaitons (aire sous la courbe assez grande), ce qui signifie que les performances de ces méthodes sont très bonnes. Représentons ensuite les prédictions en fonction du lieu d'enregistrement afin de comprendre l'impact de la route sur les performances (fig 3.4).



FIGURE 3.4 – Représentation spatiale d'une des classifications
Fourier à gauche, Ondelettes à droite

Les représentations (fig 3.4) sont sensiblement les mêmes, on voit que les classificateurs se trompent presque aux mêmes endroits, il est donc naturel de se poser la question si cela est corrélé avec la vitesse (par exemple un arrêt à un feu rouge, un stop, etc...).

Affichons la même courbe que celle pour fourier à gauche, en y mettant un seuil pour n'afficher que les points dont la vitesse est supérieure à 12km/h (allure normale à vélo), puis affichons à droite la densité des prédictions gonflé/dégonflé par vitesse (fig 3.5) :

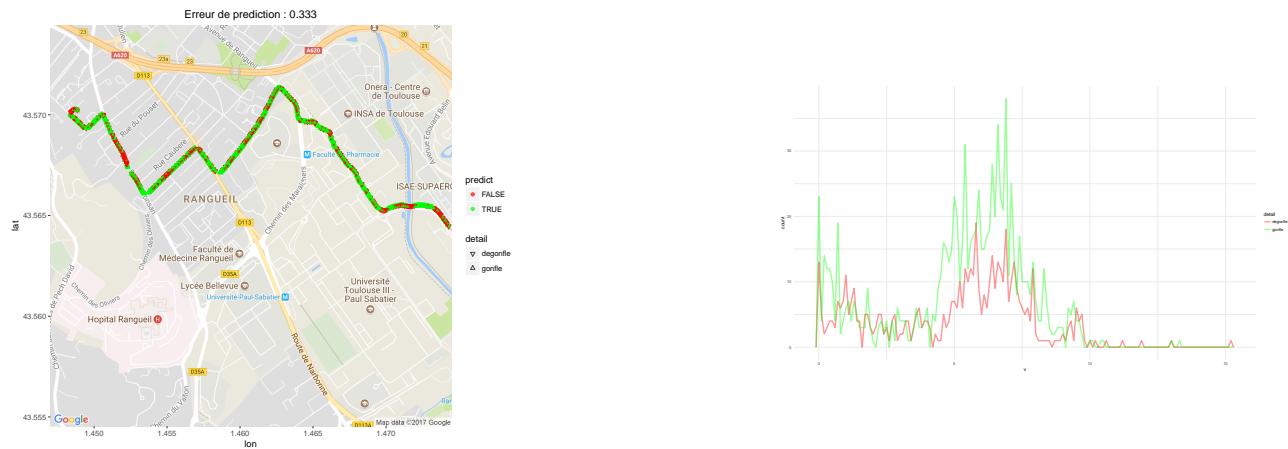


FIGURE 3.5 – Comparaison du taux d'erreur en fonction de la vitesse, à gauche le parcours, à droite le taux d'erreur en fonction de la vitesse

L'erreur de prédiction est réduite, ce qui laisse à croire que les vitesses trop basses sont difficiles à classer, et donc induisent des incertitudes.

On voit que les densités des courbes avant 10km/h (ou environ 3m/s) se superposent (donc mêmes densités), tandis qu'après, la courbe verte (les biens classés) est au dessus de la courbe rouge (les mals classés). On a donc un classificateur qui n'arrive pas à gérer des vitesses trop peu importantes (il classe à 50/50), il est donc bon de fixer un seuil en deçà duquel il sera inutile de traiter les données, ce qui induira de meilleures performances de prédiction sur notre classifieur.

3.3 Choix des axes

Il est bon de se poser la question : Quel axe prendre pour détecter au mieux l'état d'un vélo ? Garder seulement l'axe X semble raisonnable pour un pneu dégonflé, sachant que celui-ci jouera sur l'amortissement vertical du vélo. Prendre les axes Y et Z peut rajouter de l'information, et nous serviraient peut-être pour d'autres défauts que les pneus dégonflés.

(Rappel : X est dirigé vers le bas du vélo, Y vers l'avant du vélo et Z sur le côté gauche du vélo.)

3.3.1 Protocole de comparaison

Nous allons comparer le score de 4 combinaisons différentes, pour un signal traité avec transformée de Fourier, et un "preProcess" de projection sur un cercle :

1. L'axe X
2. L'axe Y
3. L'axe Z
4. Les axes X+Y+Z

3.3.2 Résultats

Méthode	Score X	Score Y	Score Z	Score X+Y+Z
nnet	Précision : 0.525	Précision : 0.607	Précision : 0.656	Précision : 0.641
rf	Précision : 0.694	Précision : 0.599	Précision : 0.535	Précision : 0.638
boost	Précision : 0.628	Précision : 0.630	Précision : 0.564	Précision : 0.637

FIGURE 3.6 – Tableau comparatif entre les axes X, Y, Z et X+Y+Z

Le choix se porte donc vers la combinaison X+Y+Z, d'une part parce qu'elle obtient des résultats satisfaisants, et d'autre part, pour avoir une meilleure robustesse avec plus de données. Représentons ensuite les prédictions en fonction de l'axe pour la méthode de Boosting, où nous constatons que le choix des trois axes est plus régulier qu'un seul axe (fig 3.7) :



FIGURE 3.7 – Représentation spatiale de la classification
X à gauche, X+Y+Z à droite

3.4 Bilan

On a choisi de traiter le signal par des transformées de fourier au dépend des transformées en ondelettes, la raison est que fourier est plus précis sur un terrain quelconque qui sort du terrain d'apprentissage, alors que les ondelettes fonctionnent mieux sur le terrain d'apprentissage, mais ce n'est pas ce que l'on souhaite (sur-apprentissage).

On a aussi vu que la préparation de données de type projection sur un cercle aide à avoir des meilleurs résultats de prédictions.

La vitesse est à négliger si elle est trop basse. Enfin, on prendra l'ensemble des données X+Y+Z pour avoir les résultats les plus fiables.

Par curiosité, voici les 16 variables les plus importantes pour random forest (voir fig 3.8), avec un vélo gonflé et un vélo dégonflé (on voit que la moyenne, la variance ou la fréquence à 8Hz reviennent souvent) :

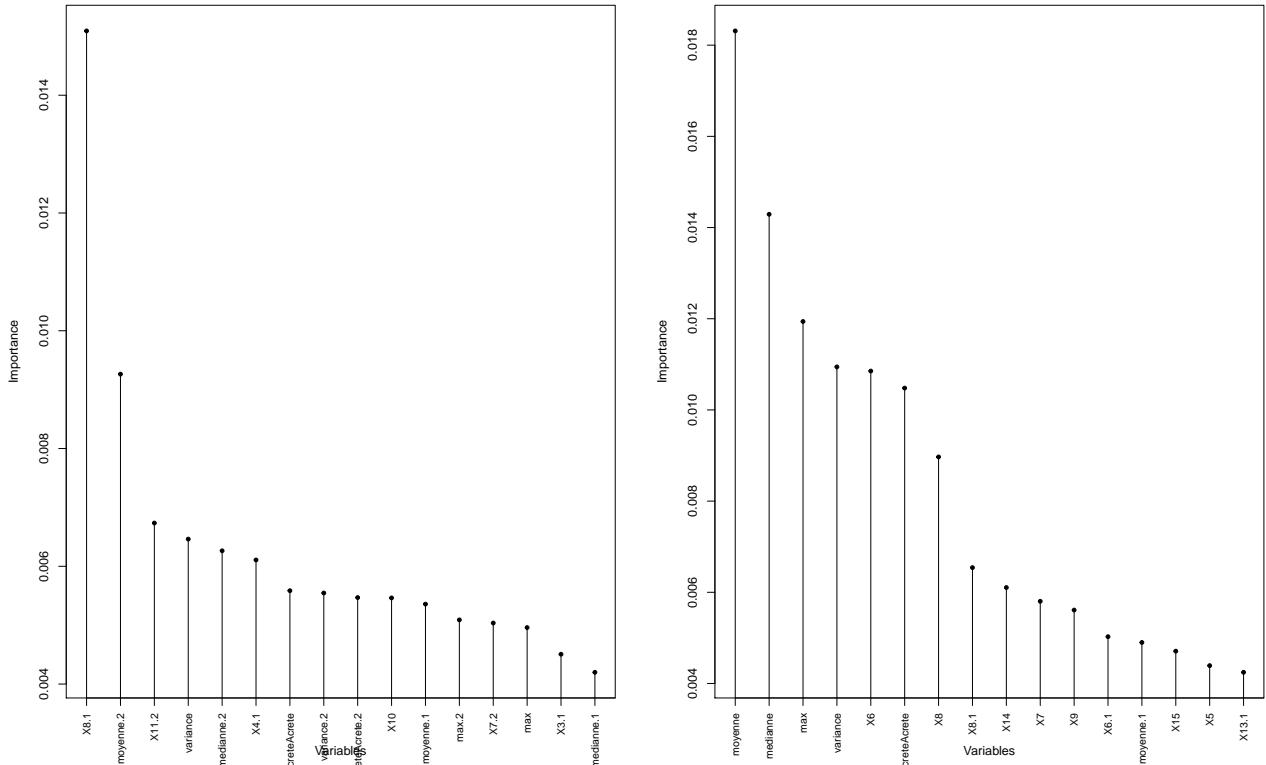


FIGURE 3.8 – 16 plus importantes variables, vélo gonflé :: : dégonflé

Chapitre 4

Apprentissage supervisé à 3 états

La classification binaire est intéressante, cependant, la liste des défauts sur un vélo est bien plus grande, nous allons voir si le fait de rajouter un état (pneu surgonflé) à la classification nous permet toujours de pouvoir prédire le défaut rencontré par le vélo. En théorie, une classification aléatoire à 3 états nous donnerait en répartition 1/3 à chacuns des états (surgonflé, gonflé, dégonflé).

4.1 Résultats

4.1.1 Echantillons de validation :

Il est bon de montrer en premier lieu les résultats de la classification sur les données de validation (autour du LAAS).



FIGURE 4.1 – Méthode boosting : vélo surgonflé :: :gonflé :: :dégonflé

Nous obtenons de bons résultats, avec comme taux d'erreur le plus fort à 13.9% pour le vélo surgonflé. Nous allons maintenant effectuer la classification sur échantillon de test.

4.1.2 Echantillons de test :

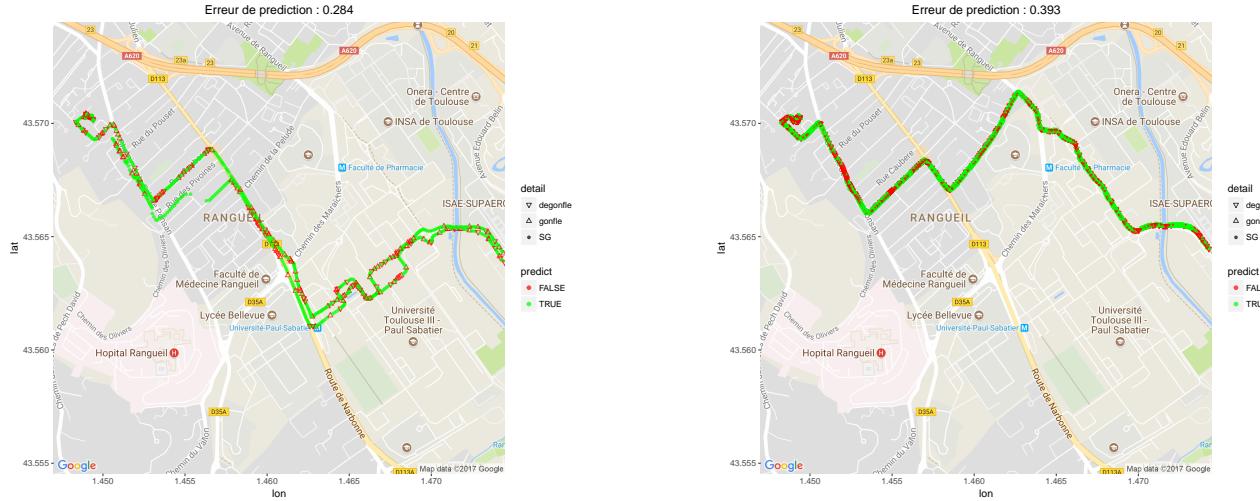


FIGURE 4.2 – Méthode boosting : vélo surgonflé : : gonflé

Nous pouvons constater que les résultats de la classification (voir fig 4.2) sont vraiment bons, de l'ordre de 70% de bien classés.

4.2 Conclusion sur la partie apprentissage supervisé

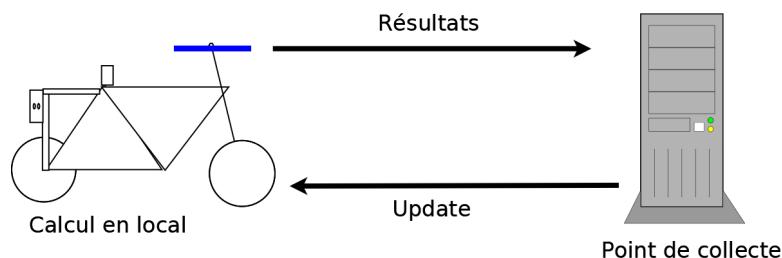
Le but était d'étudier la prédictibilité d'une panne sur un vélo à partir de données accéléromètres. Nous avions peur que l'algorithme de classification ai de bons résultats uniquement sur des données de validation. Finalement, nous avons réussi à montrer qu'il est possible, sur un trajet quelconque, de prédire l'état du vélo.

Chapitre 5

Discussions

5.1 Dimension embarquée

5.1.1 Principe



- Un modèle de traitement du signal est préparé sur un serveur puis est implémenté dans le microcontrôleur embarqué.
- L'accéléromètre sur le vélo récolte des données, puis le microcontrôleur fait du pré-traitement pour envoyer les données sur un point de collecte.
- Les informations reçues par le point de collecte sont prédites pour savoir quel est l'état du vélo.
- Une mise à jour est envoyée au microcontrôleur, pour lui dire l'état du vélo. Si il a un défaut, on peut imaginer un "bip" sonore pour prévenir le cycliste.

Actuellement, nous déchargeons les données issues de la carte SD à la main, mais notre objectif est d'obtenir une prédiction en direct depuis le vélo.

5.1.2 Coût de calcul

Le microcontrôleur n'a pas la même puissance de calcul que les ordinateurs de nos jours. Il faut alors se demander, quelle partie est à calculer sur le microcontrôleur, et quelle partie sur un serveur. De plus, il faut réfléchir à la fréquence d'échantillonage, car le volume de données monte vite, et l'envoi en LORA est limité ($32\text{Hz} \times 38\text{ bits} \times \text{nombre de secondes}$, environ 1200bits/s juste pour l'accéléromètre). Toutes ces questions sur le coût sont encore à l'étude.

5.2 Liste des pannes possible

Nous sommes aussi allés à La Maison du Vélo, située dans le centre ville de Toulouse, pour parler à des mécaniciens du vélo, et connaître les pannes les plus récurrentes. Voici un tableau récapitulatif de ce qui serait intéressant de tester, sachant que nous avons traité seulement le cas pneu surgonflé/gonflé/dégonflé :

Panne	Fréquence(1=fréquent, 3=rarement)	Description sur le signal accéléromètre
roue dégonflée	1	Amortissement sur l'axe vertical
roue voilée	1	Vibration régulière sur les 3 axes, surtout l'axe dirigé vers l'avant du vélo
roue plate	1.5	Vibration régulière sur la verticale du vélo
nexus mal réglé	2	Aléatoirement un acoups dans l'axe dirigé vers l'avant (le nexus passe mal une vitesse de temps en temps),
axe du vélo	2	Signal sensiblement le même, peut-être un léger décalage sur la calibration de l'accéléromètre (l'axe vertical ne prendra pas 1g de la gravité mais un peu moins, et les autres axes prendront le reste)
chaîne détendue	2	Acoups de temps en temps sur l'axe dirigé vers l'avant du vélo
freins	3	Très dur à détecter sur le signal, le cycliste produira un peu plus d'effort pour avancer à la même vitesse

FIGURE 5.1 – Tableau comparatif entre différentes pannes

Les combinaisons de plusieurs pannes pourraient aussi être très intéressantes à étudier.

5.3 Autocritique

5.3.1 Points négatifs

- Un départ très lent, beaucoup d'obstacles ont ralenti les début du stage, notamment la difficulté d'extraire un jeu de données de la carte SD et donc en extraire de l'information.
- Des lacunes sur les autres domaines que les mathématiques. Ce stage a été très multidisciplinaire, j'ai vu (avec les yeux) de l'électronique, de l'informatique, des mathématiques, des notions de physiques.
- Mon manque d'initiative, je suis habitué à travailler sur des sujets de TP bien encadrés avec des consignes claires et précises, du coup j'attendais que l'on me donne des ordres. Ici on est dans le monde de la recherche, on attendait de moi plus d'inspiration.

5.3.2 Points positifs

- L'environnement sous windows et le fait de coder sous le logiciel R, que j'ai déjà beaucoup utilisé lors de mon cursus de M1 MApI³, m'a permis d'effectuer rapidement des tâches qu'on me donnait, et de pouvoir tester rapidement plusieurs cas.
- Je suis toujours à l'écoute de ce que l'on me dit, et essaye de faire au mieux, et si jamais je bloque, je n'hésite pas à contacter mes tuteurs.
- Le côté multidisciplinaire a aussi son avantage, on touche à tout et on ne s'ennuie pas, j'ai toujours eu du travail à réaliser.

Chapitre 6

Conclusion

L'objectif premier de ce stage, était de savoir si il était possible de prédire, à partir de données accéléromètres, un défaut sur un vélo. Le cas échéant, dans un deuxième temps embarquer le système réalisé pour pouvoir prédire en temps réel l'état du vélo.

A l'aide de traitement du signal et d'apprentissage supervisé, nous avons vu qu'il est possible, seulement à partir de données accéléromètres, de prédire l'état du vélo sur un trajet quelconque.

Le deuxième objectif du stage, la partie avec le système embarqué, n'a pas pu être réalisé faute de temps. Le sujet étant exploratoire, nous ne pouvions pas fixer de planning précis sur quand la première partie allait être finie.

Pour finir, ce stage a été pour moi très enrichissant, j'ai touché à beaucoup de domaines scientifiques, et j'ai pu apporter ma pierre à l'édifice d'un projet grandissant.